Same same but different: the role of subjective domain similarity in the longitudinal interplay among achievement and self-concept in multiple academic domains

Learning and Individual Differences 102 (2023) 102270

Contents lists available at ScienceDirect

Learning and Individual Differences

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

Same same but different: The role of subjective domain similarity in the longitudinal interplay among achievement and self-concept in multiple academic domains

Fabio Sticca [a,*], Thomas Goetz [b], Jens Möller [c], Franz Eberle [d], Kou Murayma [e], Richard Shavelson [f]

[a] Institute for Educational Support for Behaviour, Social-Emotional, and Psychomotor Development, University of Teacher Education in Special Needs, Zurich, Switzerland
[b] Department of Developmental and Educational Psychology, Faculty of Psychology, University of Vienna, Vienna, Austria
[c] Institute for Psychology of Learning and Instruction, Kiel University, Kiel, Germany
[d] Institute of Education, University of Zurich, Zurich, Switzerland
[e] Hector Research Institute of Education Sciences and Psychology, University of Tübingen, Tübingen, Germany
[f] Graduate School of Education, Stanford University, CA, USA

ARTICLE INFO

Keywords:
Academic self-concept
Academic achievement
Subjective domain similarity

ABSTRACT

The present study examined the associations between grades and self-concept within and between four academic domains from an intraindividual perspective. Further, we explored whether students' subjective domain similarity moderated intraindividual between-domain effects of achievement on self-concept and vice versa. A sample of 756 Swiss high-school students reported on their academic self-concept in mathematics, German (native), English, and French on three measurement occasions across high school. Students reported on the subjective domain similarity. School administrators reported students' grades. Achievement in one domain had a positive effect on self-concept within the same domain and a negative effect on cuncurrent and later self-concept in other domains. Conversely, self-concept in one domain had a positive effect on achievement in the same domain and a negative effect on cuncurrent and later achievement in other domains. Further, subjective domain similarity attenuated the negative effect of achievement in one domain on self-concepts in another domain on the same measurement occasions. However, subjective domain similarity was not found to moderate the effect of achievement in one domain on change in self-concepts in another domain or vice versa.

Public significance statement: Academic achievement leads to a better academic self-concept within a given domain (e.g., mathematics) and vice-versa. However, higher achievement in one domain (e.g., mathematics) can lead to worse self-concept in another domain (e.g., first language). The present study shows that between domain effects of achievement on self-concept might be more pronounced if the student that is evaluating her/his self-concept in the two domains perceives the two domains at hand as different rather than similar. Accordingly, teachers might want to make the similarities and links between different academic domains more visible to students to avoid negative effects of positive feedback in one domain on self-concept in other domains.

In many school systems, students regularly receive feedback on their academic achievement through grades. As such, grades are positively linked to students' academic self-concept[1] (defined as an individual's self-perception of her or his academic achievement in a given academic domain; Shavelson et al., 1976). However, while students who received higher grades in math have higher self-concept in math, they also tend to have a lower self-concept in a verbal domain (controlling for their achievement in the verbal domain), and vice versa (Marsh, 1986). As a result, achievement in a mathematical domain is positively correlated with achievement in a verbal domain, but the respective self-concepts are uncorrelated (for a review, see Möller et al., 2009). Different theories and models have been formulated and combined to explain these

* Corresponding author at: University of Teacher Education in Special Needs, Schaffhauserstrasse 239, 8050 Zurich, Switzerland.
E-mail address: fabio.sticca@hfh.ch (F. Sticca).

[1] Academic achievement and academic self-concept will be abbreviated with achievement and self-concept in the following. Exceptions will be clearly indicated.

https://doi.org/10.1016/j.lindif.2023.102270
Received 14 July 2022; Received in revised form 5 January 2023; Accepted 10 February 2023
Available online 27 February 2023

findings, as negative effects[2] on students' self-concept can have an undesirable effect on students' academic and professional development (Möller et al., 2020). Furthermore, a better understanding of the psychological mechanism at hand would be instrumental for interventions aimed at supporting students' self-concept across multiple domains. These theories and models as well as respective empirical evidence are shortly outlined in the following to contextualize the aims of the present study: To examine the phenomenon from an intraindividual perspective and to explore the intraindividual role of subjective domain similarity (SDS) in the reciprocal longitudinal associations between achievement and self-concept.

The two basic models are the Internal/External frame of reference Model (I/EM; Marsh, 1986) and the Reciprocal Effects Model (REM; Marsh & Craven, 2006). A one hand, the I/EM is a cross-sectional model that assumes that students use an *external* frame of reference to compare their own achievement to their peers' achievement within a given domain (i.e., social comparison; Festinger, 1954), while they use an *internal* frame of reference to compare their achievement in one domain to their achievement in another domain (i.e., dimensional comparisons; Möller & Köller, 2001). For comparisons between a mathematical and a verbal domain, the effect of achievement on self-concept was found to be positive *within* domains (i.e., skill development; Möller et al., 2009) and negative *between* domains (i.e., contrast effects; Möller et al., 2009). On the other hand, the REM assumes that there are several positive longitudinal effects between achievement and self-concept within a domain (Calsyn & Kenny, 1977), namely the effects of achievement on self-concept (i.e., skill development; Möller et al., 2009) as well as of self-concept on achievement (i.e., self-enhancement effect; Valentine et al., 2004), achievement on achievement (i.e., autoregressive effect of achievement) and self-concept on self-concept (i.e., autoregressive effect of self-concept).

Focusing on the typical comparison of a mathematical and a verbal domain, Möller et al. (2009) performed a meta-analysis of 69 studies. The results of the study fully supported the existence of contrast effects. Concerning the methodology, Möller and Marsh (2013) reviewed evidence from introspective, correlational, and experimental studies, which strengthened the empirical underpinning of contrast effects. In a recent update of this review, Möller et al. (2020) included as many as 121 studies. The authors found support for the assumption of contrast effects between two subjects when one is from the math and one from the verbal domain. Contrast effects were found to be weaker between pairs of subjects from the verbal domain. Moreover, contrast effects were close to zero between two domains from the math/science domain. Finally, the authors found that there were no significant assimilation effects in any domain combination.

The I/EM and the REM were combined into the *reciprocal* I/EM (RI/EM) by Möller et al. (2011), resulting in a comprehensive model about the longitudinal interplay among achievement and self-concept in multiple domains. In addition to the effects stemming from the I/EM and the REM, the RI/EM added the following *between-domain* longitudinal effects: (1) reciprocal between-domain effects between achievements, (2) reciprocal between-domain effects between self-concepts, (3) between-domain effects of self-concept on achievement. Since different terminologies have been used within the RI/EM (Möller et al., 2011; Niepel et al., 2014), such as *dimensional comparison* effects vs. *internal frame of reference* effects, the term *dimensional comparison* effects will be adopted for the present study. Further, the following central distinctions will be made to facilitate the distinction between the effects: The term *dimensional achievement comparison* effect will be used to indicate a between-domain effect of achievement on self-concept. Further, *achievement contrast* effect will be used to indicate a negative effect, and *achievement assimilation* effect will be used to indicate a positive dimensional achievement comparison effect. Similarly, the term *dimensional self-concept comparison* effects will be used to indicate between-domain effects of self-concept on achievement, and *self-concept contrast* effect will be used to indicate a negative effect, while *self-concept assimilation* effect will be used to indicate a positive dimensional achievement comparison effect.

Evidence for the RI/EM is comparably scarce, as the number of studies examining this complex model is modest. Möller et al. (2011) as well as Möller et al. (2014) examined the RI/EM using Mathematics and German (native) as academic domains of interest, while Chen et al. (2013) contrasted the domains of mathematics and Chinese (native), and Niepel et al. (2014) examined the RI/EM using mathematics, German (native), and English (foreign). Taken together, support for both the skill development and the self-enhancement hypotheses from the REM was consistent across studies. Inconsistent results, however, were gathered regarding dimensional comparison effects. In this regard, Möller et al. (2011) found mixed results and results reported by Chen et al. (2013) were predominantly not in line with theoretical predictions. Möller et al. (2014) found evidence for dimensional achievement comparison effects but not for dimensional self-concept effects, independently of the use of grades or test scores. Finally, the study by Niepel et al. (2014) represents a *generalized* extension of the RI/EM because it included (1) a comparison between math and German as well as math and English, but also added (2) the longitudinal comparison between German (native) and English (foreign). Herein, dimensional comparisons were most evident in the comparison between math and German as well as between math and English. The comparison of German and English, however, yielded no significant dimensional comparison effects. In sum, the predictions about dimensional comparisons could only partially be confirmed, with the important note that no assimilation effects were found for German and English.

The external frame of reference has been adequately described by the theory of social comparison, while the corresponding theoretical elaboration of the internal frame of reference has been missing for a long time. This was done by the Dimensional Comparison Theory formulated by Möller and Marsh (2013) and the *generalized* I/EM (gI/EM) developed by Möller et al. (2015, 2016). Whereas previous discussions of differences in dimensional achievement comparison effects were based on the idea of an objective "mathematical to verbal" continuum[3] (Möller et al., 2020), dimensional comparison theory highlights the *subjective domain similarity* (SDS) between two dimensions. Subjective domain similarity can be considered with respect to many different aspects of the two domains (Goetz et al., 2014; Haag & Götz, 2012). For instance, students could think about similarities in importance of grades, in difficulty, in pace of instruction, in levels of achievement, in their self-concept, or in how strongly the topics are related to everyday life. However, these different facets can also be thought of on a more general level, as has been proposed in emotion research, where different facets of an emotion were pooled together (Gogol et al., 2014). In fact, a recent study showed that different aspects of similarity can be attributed to an underling and internally consistent (i.e., reliable) latent variable that can be understood as an overall SDS (Wolff et al., 2021). The role that dimensional comparison theory attributes to SDS is that of a moderator of dimensional comparisons (Möller et al., 2015, 2016): If two domains are perceived as being very dissimilar or far (e.g., mathematics and native language), an achievement contrast effect would be predicted. Conversely, if the two domains are perceived as very similar or near (e.

[2] In the following, the term "effect" is understood as a link between two variables that does not necessarily imply causality.

[3] The mathematical end of this continuum contains domains in with mathematics and logical operations are central, such as mathematics or physics. In contrast, the verbal end of the continuum contains domains in which language, text comprehension, reading, and writing are more central, such as native and foreign languages. This idea is not based on specific types of similarity but rather on an overall difference in terms of their position on the mathematical to verbal continuum (Möller et al., 2020).

g., mathematics and physics), no effect or even an achievement assimilation effect would be expected (Marsh et al., 2015). Hence, the effect of obtaining good grades in one domain on the self-concept in another domain depends on how the student orders things in his internal frame of reference. For example, if a specific student considers German and English to be very similar, they are represented in the same dimension within the student's internal frame of reference. Thus, positive feedback (e.g., grades) in either one of the two domains might generalize on student's self-concept that fall within the same dimension. In contrast, if the students thinks that two domains as very different, they will be represented in two different dimensions within her/his internal frame of reference. As a result, the effect of the improved grade in German will have a positive effect on self-concept in German but a negative effect on the self-concept in English. This is most likely to happen, if the students thinks that it is not possible to be good in both domains.

To date, examinations about potential moderators of dimensional comparison effects are rare. Results of the meta-analysis by Möller et al. (2009) suggested that the I/EM was confirmed irrespectively of participants' age, their gender, and country, which suggests that these variables do not moderate dimensional achievement comparison effects. Individual moderators that were studied so far are (1) students' subjective similarity of achievement in the two domains of interest (Skaalvik & Rankin, 1992; Wolff et al., 2021), (2) students' intelligence (Steinmayr & Spinath, 2015), (3) teachers' diagnostic competence (rated by their students; Wolff et al., 2021; Zimmermann et al., 2017), (4) students' belief in a negative association between mathematical and verbal abilities (Möller et al., 2006; Wolff et al., 2021), and (4) overall domain (diss-)similarity (Helm et al., 2016; Wolff et al., 2021). Helm et al. (2016) explored the role of experimentally manipulated perceived domain similarities on differences in self-concepts in two domains. The authors reported that lower perceived domain similarities in math and German led to stronger differences in self-concepts between math and German. The same was reported for the comparisons between math and physics, as well as between English and German. Wolff et al. (2021) included five of these potential moderators into their replication and extension study and concluded that only the (a) belief in the negative association between two domains and (b) the overall domain similarity moderated dimensional achievement comparison effects. As for the domain similarity, the authors found that the more similar a mathematical and a verbal domain were perceived, the less pronounced were the dimensional achievement comparison effects.

Previous studies have examined the various models and the role of SDS in terms of differences between students. In the present study, we propose that the moderating effect of SDS might not be limited to comparisons *between* various students with differing levels of SDS for a specific pair of domains (*interindividual* moderation). Instead, the moderation might also be present *within* a given individual that makes multiple comparisons across more than two domains (*intraindividual* moderation): Within the same student, dimensional comparison effects might be more negative when the student compares two domains that she/he evaluates as being very different (e.g., math and German). Conversely, dimensional comparison effects might be less negative when the same student compares subjectively similar domains (e.g., French and English). While these studies outlined above enhanced our understanding of the psychological mechanisms behind Dimensional Comparison Theory, there has yet to be a study that examines these mechanisms from an intraindividual perspective with a focus on the longitudinal role of SDS across multiple domains.

1. Aims and hypotheses

The aim of the present study was to replicate results from the gI/EM and the gRI/EM and to extend them by examining the moderating role of SDS from an intraindividual perspective. As for the gI/EM, we hypothesized that (H1a) previous results about the core predictions of the gI/EM could be replicated and that (H1b) dimensional achievement comparison effects would be less strongly negative with increasing SDS (i.e., positive moderation of a negative main effect). Regarding the gRI/EM, knowledge about the gRI/EM needs to be enriched by expanding the evidence and by including the SDS as a moderator of intraindividual dimensional comparison effects. Herein, we assumed that (H2a) positive skill development and self-enhancement effects, and dimensional achievement comparison effects as well as dimensional self-concept comparison effects could be supported, and (H2a) that the two forms of dimentional comparison effects would be less strongly negative with increasing SDS.

Besides the central aim of expanding our knowledge on the role of SDS both in the gI/EM and the gRI/EM, we shifted our attention away from specific combinations of domains and focused on the intraindividual role of SDS across all six pairs of domains (see below), which represents an intraindividual approach that is in line with the generalized extensions of the I/EM and the RI/EM.

2. Method

2.1. Sample and procedure

Three assessments of students' academic self-concepts covering the three academic years of upper-track school (i.e., Grades 9 to 11) in the German-speaking part of Switzerland (i.e., Gymnasium) were carried out in the spring of 2012 (T1), 2013 (T2), and 2014 (T3). Fig. 1 shows the timing of the assessment of the various constructs. From all German-speaking upper-track schools in Switzerland where the four academic domains of mathematics, German, English, and French were taught in Grades 9 to 11, eight schools out of 21 were randomly selected for participation in the present study. Three assessments with a time lag of one year were carried out in the classrooms during a single, 45-minute lesson using a paper and pencil questionnaire. All students in the 45 Grade 9 classrooms from these eight schools were eligible to participate.

While a total of $N = 996$ students participated in the present study in at least one of the three waves, we decided to focus on students that spoke German at home to avoid complications due to different mother languages combined with the importance of verbal academic domains in the present study. A total of $n = 112$ students reported speaking another language at home and $n = 128$ students did not report this information. Accordingly, a total of 756 students that participated in the first assessment (43.5 % female; mean age 15.6 years, $SD = 0.60$), 554 that participated in the second assessment (44.2 % female; mean age 16.6 years, $SD = 0.57$), and 474 that participated in the third assessment (43.9 % female; mean age 17.6 years, $SD = 0.58$) were included in the present study. Attrition was mainly due to one school dropping out of the study after the first assessment ($n = 129$), to students leaving the school they were initially assessed at, and to students being absent during data collection. To avoid a substantial drop in statistical power due to the reduction in sample size, 42 students were additionally recruited at T2, and 38 students were additionally recruited at T3. A subsample of 465 (61.5 %) students participated in all three assessments, while 98 (13.0 %) participated in two assessments, and 193 (25.5 %) participated in only one assessment.

Regarding participants' nationality, 92.8 % of the participants were born in Switzerland (as compared to 90.7 % for the total sample), while 5.7 % were born in other European countries and 1.6 % were born outside of Europe. As for participants' parents' nationality, the respective percentages were 76.3 % and 16.0 % for mothers and 78.5 % and 15.4 % for fathers. Turning to parents' education, 32.4 % of the participants' mothers and 47.8 % of their fathers held a university or college degree. Of those parents without a university degree, 37.4 % of mothers and 48.1 % of fathers held a vocational college degree, and 18.3 % of mothers and 9.8 % of fathers had a high-school diploma. Only 0.4 % of the participants' parents had not completed high school.

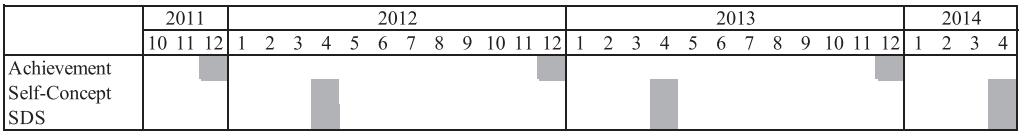

Fig. 1. Timing of the assessments of academic achievement and self-concept.
Note. Grey slots indicate the timing of the assessments. SDS = Subjective Domain Similarity.

2.2. Ethics statement

The present study was conducted in compliance with ethical standards expressed in the WMA Declaration of Helsinki. Furthermore, the study has been approved and all study procedures have been deemed appropriate by the Institutional Review Board of the University of Konstanz. Prior to each assessment, participants were informed that participation was voluntary and that they could discontinue their involvement at any time. All parents and caregivers were informed of the study aims and procedures, with all protocols approved by schools' principals and teachers. The heads of schools and the teachers approved the study protocol. After the data were collected and entered, all identifiers linking participants to their data were deleted. Thus, analyses were conducted on depersonalized data. After each assessment, participants were compensated with a small gift, such as chocolate and entry into a prize draw.

2.3. Study measures

2.3.1. Academic achievement

Each student's midyear grades (i.e., grades obtained in December (see Fig. 1) of the previous year, roughly four months before the assessments at T1, T2, and T3) in mathematics, German, English, and French were provided by the respective school administrators at each assessment and were linked to the individual data using anonymous identification codes. In Switzerland, grades range from 1 (insufficient) to 6 (excellent) with 4 being the threshold for a sufficient grade. Half grades (e.g., 4.5) are also common in Switzerland. Grades are generally determined by the results that students obtain in their exams across a term.

2.3.2. Academic self-concept

The Self-Description Questionnaire (Marsh & O'Neill, 1984) was used to assess self-concept in mathematics, German, English, and French at all measurement occasions (see Fig. 1). The scale encompassed a total of three items: (1) I get good marks in [DOMAIN]; (2) [DOMAIN] is one of my best domains; and (3) I have always done well in [DOMAIN]. Response options consisted of a 5-point Likert scale from 1 = strongly disagree to 5 = strongly agree. The internal consistencies (Cronbach's alpha) were found to range from 0.84 to 0.91 across domains and assessments (Fig. 2).

2.3.3. Subjective domain similarity

At each measurement occasion (see Fig. 1), students' self-reported academic domain similarity was assessed for each of the six pairs of academic domains. The wording of the respective items was *How similar are [DOMAIN A] and [DOMAIN B] to you?* Response options consisted of a 5-point Likert scale from 1 = *not similar at all* to 5 = *very similar*.

Appendix A shows the mean scores and standard deviations of academic achievement, academic self-concept (latent means and standard

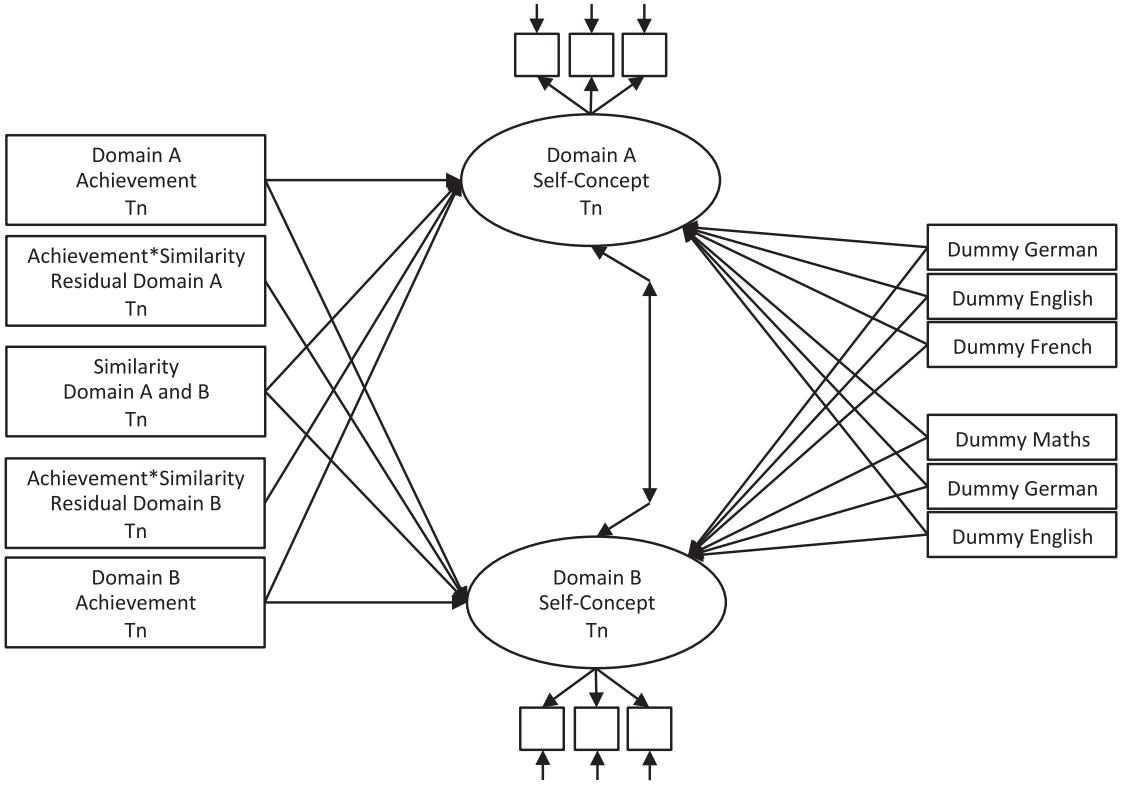

Fig. 2. Conceptual model for the analyses for the I/EM (cross-sectional).
Note. Correlations among manifest variables are not displayed.

deviations), and self-reported domain similarities at each assessment and for each academic domain as well as correlations to each other and to all other study variables.

2.4. Analysis strategy

Analyses were performed using Mplus 7.4. Missing data was addressed using the Full Information Maximum Likelihood (FIML) procedure.

The main aim of the present study was to shed light on the intra-individual moderating role of SDS in both the gI/E and the gRI/EM. Most importantly, it was assumed that the moderation of SDS would be independent of the specific pair of academic domains. In other words, the academic domains themselves and their combination were assumed not be the cause of the various between-domain effects. Instead, it was assumed that the SDS moderated these between-domain effects (Möller et al., 2020). Accordingly, the moderating effect was assumed to be a mechanism at work when a given person compares two random academic domains. To test these assumptions, data regarding multiple domains needs to be restructured. The statistical rationale for this restructured data is described at length in Appendix B. In short, twelve rows of data were created for each participant: Six rows represented all possible combinations of the four domains (i.e., 4*(4-1)/2), and another six rows represented the reversed combinations. This resulted in a sample size of $N = 9072$ (i.e., 756*12). This type of restructuring is common when using paired data (G. Chen et al., 2017; Kenny et al., 2006).

Preliminary analyses showed that the intraclass correlation (ICC; level 1 = individual, level 2 = classroom) of self-concept ranged from 0.023 to 0.086, while those for achievement were ranged from 0.049 to 0.136, and those for SDS ranged from 0.026 to 0.125. These results revealed that a relevant portion of the variance in the three variables lay at the classroom level. However, the classroom level could *not* be taken into account, as the Huber-White sandwich estimator (i.e., TYPE = COMPLEX option in Mplus; Freedman, 2006) had to be used to correct for the nesting of the data within individuals as a result of the restructuration of the data (see below).

2.4.1. Examination of the gI/EM: cross-sectional perspective

To examine the gI/EM (cross-sectional model), three structural equation models were constructed, one for each assessment. All models were run on the restructured data with twelve rows per participant. Achievement of both domain A and B were modeled as manifest predictors of the latent scores of self-concept of both domain A and B. The two exogenous achievement scores were allowed to correlate, as were the residuals of the two endogenous latent self-concept scores. The identification of the latent variables in this and all following models was achieved using the effect-coding method (Little, 2013). Herein, instead of using the reference indicator method (i.e., fixing the loading of a single indicator at 1 and its intercept at 0), the mean of the loadings is fixed at 1 while the mean of the intercepts is fixed to 0. Thus, the effect coding method ensures a more balanced representation of all indicators of a given latent variable. The SDS was then added to the model as a moderator. Specifically, orthogonalized residual based interaction terms (Little, Bovaird, & Widaman, 2006) were added to the model: Scores of achievement in domain A were multiplied with SDS of domain A and B and the results of this multiplication was regressed on the two variables that composed it. The residuals of these two regressions were saved and added to the model as manifest predictors of the latent self-concept in the respective other domain. The main effects of the SDS on the self-concept in domain A and B were also added to the model. Finally, in order to take into account the clustering within students as well as to correct for the multiplied data set, we used the Huber-White sandwich estimator (i.e., TYPE = COMPLEX option in Mplus; Freedman, 2006), and also added dummy variables representing domain A (three dummies with mathematics as reference group) and domain B (three dummies with French as a reference group) in order to account for potential random-intercept effects of the combination of domains. The inclusion of the dummy variables is essentially equivalent to fixed-effects models to deal with nested data (McNeish & Stapleton, 2016). Fig. 1 illustrates the resulting model.

2.4.2. Examination of the gRI/EM: longitudinal perspective

The analysis strategy of the gRI/EM was organized into four main steps. First, latent-state models were used to evaluate the measurement invariance of the self-concept measure. Appendix C shows that scalar invariance could be established for all domains.

In a second step, a multivariate cross-lagged model was built using the manifest scores of achievement in domains A and B as well as the latent scores of self-concept in domains A and B. The structure if the gRI/EM deviated from the usual structure of a cross-lagged model because scores of achievement were represented by mid-term grades obtained in December of a given year, while self-concept was assessed in the spring of the following year without a specific time of reference. This peculiarity was considered by modeling achievement at T1 (i.e., December 2011) in domain A as a predictor of self-concept at T1 (i.e., spring 2012) in domains A and B as well as of achievement at T2 in domains A and B. Self-concept at T1 in domain A was then modeled as a predictor of achievement at T2 in domains A and B as well as self-concept at T2 in domains A and B. The remaining effects that referred to later time points were modeled using the same logic. The effects from achievement/self-concept in domain B on achievement/self-concept in domains A and B were mirrored to those presented above. Fig. 3 shows a simplified statistical model for the analyses for the gRI/EM.

In the third step, the main effects of SDS, the interaction effects between achievement and SDS as well as self-concept and SDS were added to the model. SDS at T1 was modeled as a predictor of self-concept at T1 in domains A and B as well as achievement at T2 in domains A and B. Further, SDS at T1 was modeled as (1) a moderator of the effect of achievement at T1 in domain A on the self-concept at T1 in domain B, and vice versa (i.e., moderation of dimensional achievement comparison effects), and (2) as a moderator of the effect of self-concept at T1 in domain A on achievement at T2 in domains B, and vice versa (i.e., moderation of dimensional self-concept comparison effects). The remaining effects that referred to later time points were again modeled using the same logic. The interaction between the manifest achievement and the manifest SDS variables was modeled as described above for the gI/EM. In contrast, the interaction between the latent self-concept and the manifest SDS was modeled using the latent orthogonalization technique[4] (Little, Bovaird, & Widaman, 2006). This new interaction term was then modeled as a predictor of the respective outcomes of interest. Appendix D shows a conceptual model for the model with interactions. Again, we used the Huber-White sandwich estimator (i.e., TYPE = COMPLEX option in Mplus; Freedman, 2006) to take into account the clustering within students and to correct for the multiplied data set. Dummy variables representing domain A and domain B were used take random-intercept effects of the pairs of domains into account.

In the fourth and last step, the stability of the within and between domain associations across time was examined. This step was undertaken to test for the stability of the effects over time. Technical details about this procedure are outlined in Appendix E.

[4] In the case of a latent variable interacting with a manifest variable this technique consists in computing the multiplication of each item of the latent variable (i.e., three indicators of self-concept) with the manifest variable (i.e., SDS). The multiplied variables are then all regressed onto the three indicators and the manifest variable in separate regression models and the residuals of these regressions are saved. The results of this procedure are three variables that are then included in the structural equation model as manifest indicators of a latent orthogonalized interaction term.

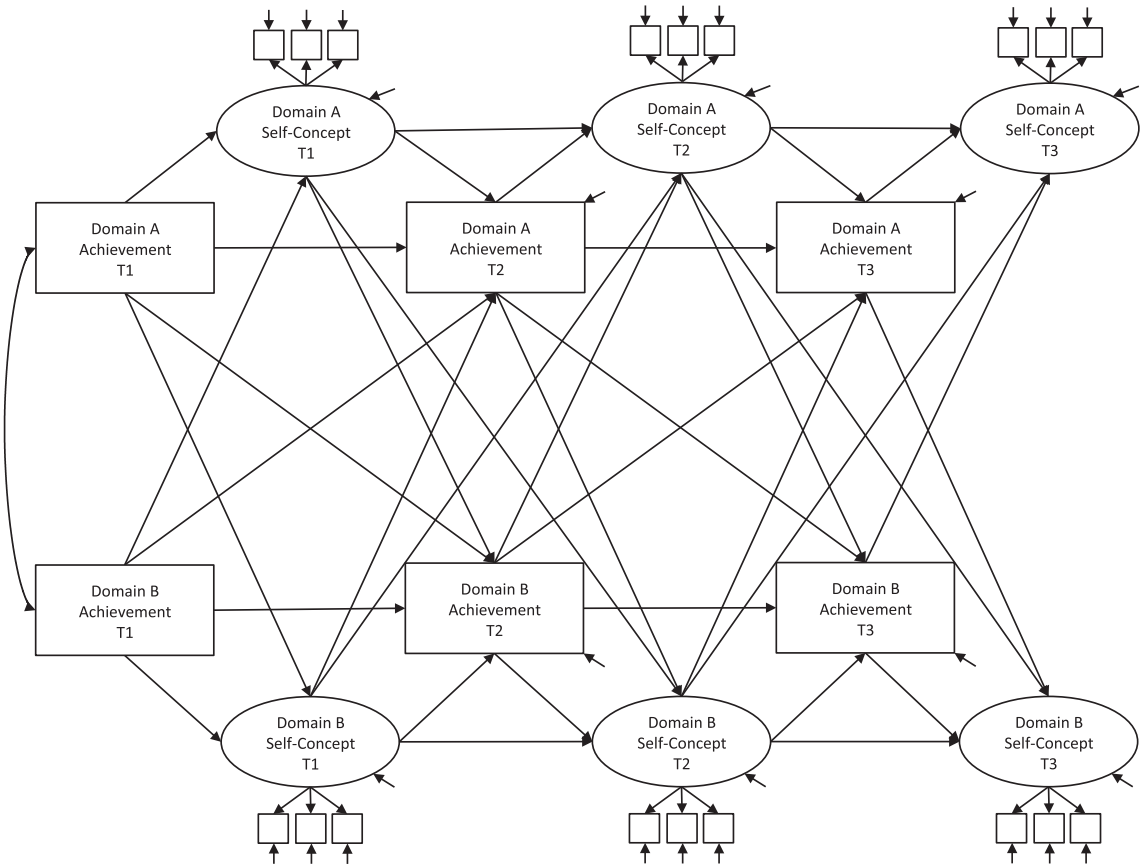

Fig. 3. Simplified statistical model for the analyses for the RI/EM (Longitudinal).
Note. Correlations between pairs of residuals are not displayed (e.g., between T2 achievement in domain A and T2 achievement in domain B). Correlated uniqueness is also not displayed.

3. Results

3.1. Descriptive results and correlations about the subjective domain similarity

To have a better understanding of the nature of the items that were used to measure SDS, descriptive results will be shortly outlined here. Descriptive results can be found in Appendix A. The highest mean similarity was observed for the English-French combination (3.47 to 3.56), followed by German-English (2.72 to 3.03), German-French (2.76 to 2.82), mathematics-German (1.71 to 1.83), mathematics-English (1.68 to 1.75), and mathematics-French (1.33 to 1.37). Standard deviations were almost all in the range of 0.90 to 1.00, with a somewhat lower variability in the mathematics-French combination (0.68 to 0.84). Descriptive statistics for the SDS obtained after restructuring the data set (i.e., within-person data with 12 rows per participant) were $M = 2.31$ and $SD = 1.19$ at T1, $M = 2.38$ and $SD = 1.22$ at T2, $M = 2.34$ and $SD = 1.22$ at T3. These results suggest that the average similarity across all domains was in the middle of the scale (i.e., 1 = *not similar at all* to 5 = *very similar*) and that the standard deviation of this averaged within-person SDS was higher than the standard deviation within the respective pairs of domains (see above).

Correlations of SDS over time were found to be moderate across all domains (0.30 to 0.43) and to be lower than those of self-concept (0.50 to 0.73) and achievement (0.67 to 0.89). Further, correlations between SDS and the respective self-concepts scores were found to be generally below $r = 0.10$. Moreover, associations between SDS and achievement were found to be almost exclusively under $r = 0.10$. Finally, correlations between SDS were generally moderate among combinations that contain mathematics (0.26 to 0.47) as well as among those combinations that do not contain mathematics (0.19 to 0.47). In contrast, correlations of SDS among combinations that contained and those that did not contain mathematics were generally lower (−0.07 to 0.23).

3.2. Results for the gI/EM

The results of the examination of the role of SDS in the I/EM are reported in Table 1. All models were found to exhibit a good fit to the data. Notably, the pattern of results appeared to be quite similar across time: Skill development effects were found to be positive and large in

Table 1
Standardized results of the three cross-sectional models for the examination of the I/EM not based on specific domain comparisons ($N = 9072$).

	T 1	T 2	T 3
ACH => SC within-domain	0.82***	0.79***	0.74***
ACH => SC between-domain	−0.13***	−0.11***	−0.12***
SDS => SC	0.02	0.07***	0.05**
ACH * SDS => SC	0.04***	0.05**	0.04**
χ^2/df	1180.15	510.43	359.72
df	56	56	56
CFI	0.95*	0.97	0.98
RMSEA	0.05	0.03	0.03
SRMR	0.02	0.02	0.02

Note. ACH = Achievement; SC = Self-Concept; SDS = Subjective SDS; χ^2 = Chi-Square; df = Degrees of Freedom; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; SRMR; Standardized Root Mean Square Residual.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

magnitude. Conversely, dimensional achievement comparison effects were found to be negative and rather small. Effects of SDS were found to be positive and very small. Finally, the effect of the interaction between between-domain achievement and SDS on self-concept was found to be positive and small. Nevertheless, we note that the size of the interaction effect was found to be around 1/2 to 1/3 of the size of the dimensional achievement comparison main effect. This proportion is relevant because it shows that dimensional achievement comparison effects would tend towards zero for comparisons between domains with scores of SDS that are three standard deviations above the mean, which is generally the case for comparisons between English and French (i.e., domains reported as being rather similar). In contrast, dimensional achievement comparison effects would tend towards $\beta = 0.20$ for comparisons between domains with scores of SDS that are three standard deviations below the mean, which is generally the case for comparisons between math and German (i.e., domains reported as being rather different).

To better understand these results, we computed a series of separate between-person models for all six combinations of domains, direction of comparison (e.g., math on German plus German on math), and three time points, which resulted in 6 * 2 * 3 = 36 models. The results of these models are reported in the Supplemental Material Table 1. The results from these 36 between-person models show that dimensional achievement comparison effects vary between $\beta = -0.10$ and -0.26 for the math vs. German (and vice versa; i.e., domains judged as very different) combination and are all highly significant. The pattern for math vs. English and math vs. French (i.e., domains judged as less different) is somewhat less pronounced, while the pattern for the comparisons of German vs. English as well as German vs. French (i.e., domains judged as more similar) show first cases of non-significant dimensional achievement comparison effects. Finally, the English vs. French (i.e., domains judged as rather similar) comparisons show the highest number of non-significant dimensional achievement comparison effects. As for the effect of SDS in these between-person models, a very inconsistent pattern ranging from significantly negative to significantly positive effects were found. In sum, these separate models show that the largest factor to determine the size of dimensional achievement comparison effects, is the combination itself, while the variation around this main factor is less relevant.

3.3. Results for the gRI/EM without homogeneity of effects constraints

The model with scalar invariance constraints that was built to examine the moderating role of SDS in the context of the gRI/EM was found to fit the data well (see Appendix C). The model including directional effects and dummy variables for the pairs of domains also fitted the data well ($\chi^2 = 8553.27$; $df = 845$; CFI = 0.93, RMSEA = 0.03; SRMR = 0.03). Table 2 shows that for the cross-sectional part of the model, results were found to be virtually identical to those reported above in the examination of the gI/EM: Achievement in one domain had a positive and very strong effect on self-concept in the corresponding domain, while its effect on self-concept in the non-corresponding domain was negative and small in magnitude. Furthermore, SDS had a negligible effect on self-concept and, most importantly, was found to positively moderate the dimensional achievement comparison effect, with a small effect size that was nonetheless close to 1/3 of the dimensional achievement comparison main effect.

As for the longitudinal part of the model, self-concept was found to be highly stable, whilst the relative stability of achievement was found to be moderate. Both the skill development and the self-enhancement effects were found to be positive and of moderate magnitude. Regarding the between-domain within-construct effects, the grade-on-grade effects were both found to be positive and small to medium in size, while the self-concept on self-concept effects were positive, and very small. Moreover, regarding the between-domain cross-construct effects, both the dimensional achievement comparison and the dimensional self-concept comparison effects were consistently found to be negative and small. Furthermore, SDS was partially found to have a positive yet very small effect on self-concept but had no effect on achievement. Finally, SDS was not found to longitudinally moderate any of the effects, with effect sizes being very close to zero.

Table 2
Standardized results of the longitudinal models for the examination of the RI/EM ($n = 8976$).

Model without homogeneity of effects constraints			
	Cross-sectional	Longitudinal 1	Longitudinal 2
Effects on self-concept (SC)			
ACH => SC within-domain	0.81***	0.47***	0.41***
ACH => SC between-domain	−0.13***	−0.11***	−0.10***
SDS => SC	0.02	0.05**	0.03*
ACH * SDS => SC	0.04***	0.01	0.01
SC => SC within-domain		0.54***	0.58***
SC => SC between-domain		0.05***	0.03+
Effects on achievement (ACH)			
ACH => ACH within-domain		0.36***	0.36***
ACH => ACH between-domain		0.19***	0.19***
SDS => ACH		−0.02	−0.02
SC * SDS => ACH		0.01	0.01
SC => ACH within-domain		0.26***	0.39***
SC => ACH between-domain		−0.09***	−0.09**

Model with homogeneity of effects constraints		
	Cross-sectional	Longitudinal
Effects on self-concept (SC)		
ACH => SC within-domain	0.81***	0.44***
ACH => SC between-domain	−0.12***	−0.11***
SDS => SC	0.02	0.03***
ACH * SDS => SC	0.04***	0.01+
SC => SC within-domain		0.56***
SC => SC between-domain		0.04***
Effects on achievement (ACH)		
ACH => ACH within-domain		0.35***
ACH => ACH between-domain		0.18***
SDS => ACH		0.01
SC * SDS => ACH		0.01
SC => ACH within-domain		0.32***
SC => ACH between-domain		−0.08***

Note. ACH = Achievement; SC = Self-Concept; SDS = Subjective SDS.
+ $p < .10$.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

3.4. Results for the gRI/EM with homogeneity of effects constraints

The simultaneous equality constraints on all longitudinal effects in the model did not lead to a meaningful deterioration in model fit ($\Delta\chi^2 = 96.36$; $\Delta df = 50$; ΔCFI = −0.002, ΔRMSEA = −0.001; ΔSRMR = 0.001). Thus, the homogeneity of the various longitudinal effects could be supported, which suggests that the psychological mechanisms at hand were stable over time. Results of the model with homogeneity of effects constraints are displayed in Table 2. The cross-sectional component of the model was in line with results reported in the cross-sectional examination of the gI/EM: Achievement in one domain had a positive and very strong effect on self-concept in the corresponding domain, while its effect on self-concept in the non-corresponding domain was negative and small in magnitude. Furthermore, SDS had no effect on self-concept and was found to positively moderate the dimensional achievement comparison effect, with an effect size that was as high as 1/3 of the dimensional achievement comparison main effect.

Concerning the longitudinal part of the model, self-concept was

confirmed to be highly stable, whilst the relative stability of achievement was found to be moderate. Further, the skill development as well as self-enhancement effects were supported. The between domain grade on grade effects were found to be significant and small to medium in size. Moreover, the between domain self-concept on self-concept effects were significant and very small sized. Further, both the dimensional achievement comparison and the dimensional self-concept comparison effects were found to be significant and small in magnitude. Furthermore, SDS was found to have a positive main effect on self-concept, but not on achievement. Finally, SDS was not found to moderate neither the dimensional achievement comparison nor the dimensional self-concept comparison effects with effect sizes being again very close to zero.

4. Discussion

The present study pursued two main aims, namely, to examine the role of the subjective domain similarity (SDS) as a potential moderator in the gI/EM and to expand our knowledge about the gRI/EM by both replicating and expanding previous results and by including the SDS as a moderator of longitudinal between-domain effects. Additionally, an overall approach was taken to examine the hypotheses of interest for a total of six possible comparisons among four academic domains, namely mathematics, German, English, and French within each student. Thus, this was the first study to elucidate the role of SDS as a moderator of both cross-sectional and longitudinal between-domain effects in the context of the gRI/EM. The following discussion will begin with a brief section about the nature of the SDS variables, followed by results for to the gI/EM and to the gRI/EM. Finally, implications for practice will be discussed.

4.1. Discussion of subjective domain similarity measures

Descriptive results for SDS revealed that a very similar pattern of mean scores was found for the various combinations of domains across the three measurement occasions. Additionally, the apparent difference in SDS mean scores between all combinations that contain mathematics and those that do not contain mathematics supports the notion of the a similarity continuum form mathematical to verbal domains (Möller et al., 2020). Our results suggest that the sequence on this continuum might consist of mathematics on one end and French on the other end (most different pair), while German as a native language and English as a foreign language seem to be equally distant from mathematics and German but still to be more apart from each other than each one is to French. This pattern might indicate that SDS is a multidimensional concept and that students consider multiple dimensions of the domains when they are asked to rate the overall similarity between two domains. Moreover, the present results also suggest, that there is a relevant amount of intraindividual variability within the same pair of domains, which suggests that similarity is a subjective aspect to some degree (Wolff et al., 2021). Further, correlational results suggest that SDS are not strongly related to self-concept and/or achievement measures, while there seems to be a systematic pattern of associations between SDS from different pairs of domains. In line with the notion of the math to verbal continuum, pairs of domains that contain mathematics were moderately correlated, as were pairs that did not contain mathematics. In contrast, pairs that contained mathematics were weakly correlated to pairs that did not contain mathematics. However, it must be noted that the domains that were considered in the present study were limited to mathematics and different language domains. Future studies might expand these comparisons to other domains such as history, sports, and arts. In conclusion, the single items that were used to assess SDS might not just be mere reflections of self-concepts and/or achievement and are useful indicators students' complex perceptions of how similar two domains are to each other all things considered.

4.2. Discussion of results about the gI/EM

Regarding the main predictions of the gI/EM (hypothesis H1a), cross-sectional results from all three waves of assessment were very consistent: Self-enhancement effects were found to be positive and strong, while dimensional achievement comparison effects were found to be negative and comparably weak. These results are in line with both theoretical assumptions and previous studies (e.g., Lösch et al., 2017; Möller et al., 2009; Schmidt et al., 2017), although the magnitude of the dimensional achievement comparison effects was found to be somewhat lower and that of self-enhancement was found to be higher, which might be due to the more homogeneous sample of students in a high-achieving track (i.e. Gymnasium). Considering that our results stemmed from a domain-unspecific approach and accordingly restructured data, it is noteworthy that they match those from previous studies.

As for the role of SDS within the gI/EM, our results suggested that the dimensional achievement comparison effect between two random domains was slightly less pronounced when students perceived the two domains as more similar. This result was obtained in all three waves of assessment, thus confirming our Hypothesis H1b. The effect size of the moderation was around 1/2 to 1/3 of the main effect, suggesting that the effect seemed to have some relevance and that between domain effects were close to zero for comparisons of domains that were perceived as maximally similar (i.e., the interaction effect would cancel out the main effect of the grade, resulting in a slope close to zero). These results are in line with previous studies using the same approach (Wolff et al., 2021) although the effect sizes that were found in the three waves of assessment were smaller. This might be due to the use of a single item for the SDS, which results in attenuated effects, as well as to the relatively homogeneous sample with respect to the school track. Further, the present results maps onto those from studies that took a similar theoretical approach but used other constructs (i.e., belief in a negative interdependence of abilities; Möller et al., 2006) or methods (i.e., cross-sectional and experimental approach; Helm et al., 2016; Möller et al., 2006). As such, the present study contributes to and expands the literature on the gRI/EM and Dimensional Comparison Theory, as it provides further evidence for the assumption about the SDS as a moderator of dimensional achievement comparison effects can be supported when taking an interindividual approach such as the one presented here (Wolff et al., 2021). We were able to find support for the notion that the factor that is linked to dimensional achievement comparison effects is the SDS as opposed to an a priori classification of which domains are to be considered as more and which as less similar in terms of a relative position of a mathematical to verbal continuum.

4.3. Discussion of results about the gRI/EM

Turning to the hypotheses H2a concerning the main assumptions of the gRI/EM, the additional longitudinal portion showed that longitudinal self-enhancement effects from one year to the other were still positive and of considerable magnitude in both lags between waves, which indicates the presence of a positive component in the longitudinal cycle between achievement and self-concept. Further, dimensional achievement comparison effects were found to be negative and of the same magnitude as its cross-sectional counterpart. The reverse effects of self-concept on achievement showed a very similar pattern: On the one hand, skill development effects were positive and moderate. On the other hand, dimensional self-concept comparison effects were negative and comparably small. These results are in line with theoretical assumptions (Möller & Marsh, 2013) and in particular the within domain results match those of previous studies, while those about between-domain effects were only partly in line with previous studies, as these found mixed results (e.g., Chen et al., 2013; Möller et al., 2011; Niepel et al., 2014). The difference between dimensional achievement comparison and dimensional self-concept comparison effects appeared to be small in the present study, which is in line with results reported by

Niepel et al. (2014) but less so with those reported by Möller et al. (2011). This balance between negative between-domain effects suggests, that interventions aimed to optimize achievement in one domain might impact self-concept in other domains and interventions aimed to optimize self-concept in one domain might impact achievement in other domains. Indeed, previous studies have shown that the positive effects on achievement and self-concept within a domain seem to have the upper hand on the negative effects between domains, which implies that the final outcome of these social and dimensional comparisons is positive (Möller et al., 2011).

On top of the cross-construct results presented above, the results about within construct effects showed that the relative stability of self-concept was positive and moderate to high, while the between domain effects of self-concept on self-concept were very close to zero. The relative stability of achievement was also found to be positive and moderate, while the between domain effect of grades on grades was positive but comparably small. These results are also in line with theoretical assumptions and also generally in line with previous studies about the gRI/EM (e.g., S.-K. Chen et al., 2013; Möller et al., 2011; Niepel et al., 2014). The finding that these effects are generally positive in direction strengthens the notion that the negative component of the entire cycle is to be looked for in the between-domain cross-construct associations, which confirms the domain specificity of self-concepts (Möller et al., 2011). Taken together, these findings give further credit to the notion that the final outcome of these comparisons processes is desirable (Möller et al., 2011).

Finally, relating to the moderating role of SDS and the respective hypothesis H2b, the longitudinal part of the model did not yield any significant nor meaningfully sized effects. This result is in contrast with our hypotheses and with theoretical assumptions (Möller & Marsh, 2013). Given that this is the first study to examine this overall moderation of SDS on the bidirectional association between achievement and self-concept, this result needs to be interpreted with caution before it is replicated in further studies. In particular, the question about the reason for the presence of a cross-sectional effect coupled with the absence of a longitudinal effect arises as far as dimensional comparison effects are concerned, while the absence of evidence for moderations of longitudinal between-domain contrast effects is discussed below.

From a theoretical point of view, it might be argued that the moderating effect of SDS on dimensional achievement comparison effects is quite weak, which might explain its disappearance in the context of prolonged time frames. Nevertheless, the “cross-sectional” effect in the present study represents a short-term longitudinal effect, as self-concepts were assessed in spring of the respective year, while the actual grades referred were obtained in December and pertained to the previous semester. Although no change was modeled, we might still argue that this is neither a clearly cross-sectional, nor a clearly longitudinal model. This observation leads to the question about the theoretical assumptions of the timing and duration of dimensional comparison effects as well as of the moderating effects of SDS in Dimensional Comparison Theory. No a-priori assumptions exist in Dimensional Comparison Theory in this regard and future studies might examine if the assumed moderating effect of SDS can be found from one typical grading period to the next (e.g., semester or trimester).

Another potential issue is the stability of perceived similarity across time. Since topics within subjects might change across time, it must be assumed that the perceived similarity is also subjected to fluctuation. This also has implications from a methodological perspective, as multiple options of assessment timing and statistical modeling can be chosen. For instance, if the similarity is assessed at every wave (as opposed to between waves for instance), one might use the assessment of similarity at T1 as a moderator of the effect of achievement in domain A at T1 on self-concept in domain B at T2. Alternatively, one might also use the assessment of similarity at T2 as a moderator of the effect of achievement in domain A at T1 on self-concept in domain B at T2 (which is close to what was done in the present study). In sum, one would need to have theoretical guidelines about the timing of these effects (see paragraph above) and to assess/model data accordingly to rule out the possibility of a potentially better suited strategy.

4.4. Implications for practice

An important question that arises from the present results is whether the negative between-domain effects of achievement on self-concept should be minimized through educational interventions. On the one hand, positive achievement in one domain would ideally not interfere with the self-concept in other domains. In other words, contrast effects should be minimized. In line with previous studies, the present results suggest that contrast effects can be prevented by increasing students' perceived similarity between academic domains (Helm et al., 2016; Wolff et al., 2021). Thus, an important implication for teachers and parents is that similarities and connections among a variety of academic domains should be made visible to students. For instance, teachers might help students in understanding how mathematical concepts can be applied to other domains such as sports, geography, or economics and they could also show how an overarching theme (e.g., the school garden) could be addressed from different academic domain perspectives (e.g., physics, mathematics, biology, chemistry, aesthetics) so that commonalities among subjects are identified, e.g., in terms of design goals. Teachers could also repeatedly point out in class that similar issues are also addressed in other academic domains (e.g., grammar in the mother tongue and foreign language) and are sometimes dealt with in a similar way. Such interventions would thus soften the boundaries between school subjects.

On the other hand, it is argued that negative between domain effects and the resulting differential motivational patterns might be important for professional specialization, which is a key component of western economic systems and its link to economic growth is a core assumption in economics (Ethier, 1982; Romer, 1987). Although this is a plausible line of reasoning, we think it is rather problematic to achieve domain specialization by negative effects of positive performance in one domain on self-concept in another domain. Self-concepts are of great importance to humans, for example, in terms of future performance and self-worth. Therefore, any mechanism that weakens the self-concept should be avoided from a humanistic point of view. Specialization can also be promoted by strengthening interests in a particular domain without degrading other domains.

Another implication is that educators and parents need to be aware of the complexity of the reciprocal associations of achievement and self-concept (Möller et al., 2011). A key contribution of our study is that educators and parents should be aware that students' thinking about subject domain similarities plays an important role in this regard. Specifically, students that perceive academic domains as very different might tend to choose an academic or professional specialization based on the self-concept that they developed in the domain in which they achieved better grades over their academic career. But that also means they invest fewer resources in other domains, even if they are interested in those domains. Raising teacher and parent awareness of such processes can reduce the risk of a premature and maladaptive investment of students into a specific domain (Möller & Marsh, 2013). For example, parents should try to avoid statements that reinforce differences between subject domains, such as statements based on false subjective beliefs, like “I also had good grades in English, I'm such a language guy, so I wasn't so good at math.”

4.5. Limitations

The present study is not without limitations. The sample that participated in the present study was not representative for Switzerland, as not all Cantons were represented. Further, although as many as four domains could be considered in the study, many important domains such as those in the middle of the numerical-verbal continuum (e.g.,

history or economics) as well as domains such as physical education remain unstudied. Another issue that must be considered is that grades were assessed a few months before the respective self-concepts within each measurement occasion. This time lag might have led to a weaker association between achievement and self-concept as compared to an assessment of both constructs at the same time. This limitation might also have weakened the potential moderation effect of SDS. Finally, this study was correlational in nature. Although a longitudinal approach over three measurement occasions was taken, causality cannot directly be implied with this study design. In this regard, it is important to note that our results are in line with theoretical expectations and with experimental studies on the role of perceived domain similarity (Helm et al., 2016).

4.6. Conclusion

Decades of research have shown that motivational profiles are domain specific. The present study showed that SDS plays a subtle but consistent role in the process differential motivational development in mathematical and language domains: For pairs of domains that are subjectively perceived as more different, dimensional achievement comparison effects can be expected to be stronger. However, only weak dimensional achievement comparison effects can be expected if the two domains are perceived as very similar. Importantly, the concept of similarity is subjective, which means that students do have differentiated and flexible mental representations of the similarity between various academic domains. Teachers and parents need to be more aware of their attitudes and about their communication about differences between domains, as supporting the perception of similarity between domains might reduce the likelihood of negative between-domain effects and, consequently, of very differentiated self-concepts and resulting differential engagement in different domains among students (Helm et al., 2016). For instance, parents might want to avoid suggesting their children, that they were good at math and thus not so good at German, as such an attitude might foster students' perception of the dissimilarity between math and German.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.lindif.2023.102270.

Declaration of competing interest

The authors declare that there are no conflicts of interest of any kind.

Acknowledgements

This research was supported by a grant from the Swiss National Science Foundation (SNSF) to the second author (100014_131713/1).

References

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

Chen, G., Taylor, P. A., Shin, Y.-W., Reynolds, R. C., & Cox, R. W. (2017). Untangling the relatedness among correlations, part II: Inter-subject correlation group analysis through linear mixed-effects modeling. *NeuroImage, 147*, 825–840. https://doi.org/10.1016/j.neuroimage.2016.08.029

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

Ethier, W. J. (1982). National and international returns to scale in the modern theory of international trade. *The American Economic Review, 72*(3), 389–405. http://www.jstor.org/stable/1831539 http://www.jstor.org/stable/1831539.

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

Freedman, D. A. (2006). On the so-called "Huber Sandwich Estimator" and "Robust Standard Errors". *The American Statistician, 60*(4), 299–302. https://doi.org/10.1198/000313006X152207

Goetz, T., Haag, L., Lipnevich, A. A., Keller, M. M., Frenzel, A. C., & Collier, A. P. M. (2014). Between-domain relations of students' academic emotions and their judgments of school domain similarity. *Frontiers in Psychology, 5*, 1153. https://doi.org/10.3389/fpsyg.2014.01153

Gogol, K., Brunner, M., Goetz, T., Martin, R., Ugen, S., Keller, U., Fischbach, A., & Preckel, F. (2014). "My Questionnaire is Too Long!" The assessments of motivational-affective constructs with three-item and single-item measures. *Contemporary Educational Psychology, 39*(3), 188–205. https://doi.org/10.1016/j.cedpsych.2014.04.002

Haag, L., & Götz, T. (2012). Mathe ist schwierig und Deutsch aktuell: Vergleichende Studie zur Charakterisierung von Schulfächern aus Schülersicht. *Psychologie in Erziehung und Unterricht, 59*(1), 32–46. https://doi.org/10.2378/peu2012.art03d

Helm, F., Mueller-Kalthoff, H., Nagy, N., & Möller, J. (2016). Dimensional comparison theory: Perceived subject similarity impacts on students' self-concepts. *AERA Open, 2*(2), 1–9. https://doi.org/10.1177/2332858416650624

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. Guilford Press.

Little, T. D. (2013). *Longitudinal structural equation modeling*. Guilford Press.

Little, T. D., Bovaird, J. A., & Widaman, K. F. (2006). On the merits of orthogonalizing powered and product terms: Implications for modeling interactions among latent variables. *Structural Equation Modeling: A Multidisciplinary Journal, 13*(4), 497–519. https://doi.org/10.1207/s15328007sem1304_1

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. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal, 23*(1), 129–149. https://doi.org/10.2307/1163048

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science, 1*, 133–163. https://doi.org/10.1111/j.1745-6916.2006.00010.x

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Abduljabbar, A. S., Abdelfattah, F., & Jansen, M. (2015). Dimensional comparison theory: Paradoxical relations between self-beliefs and achievements in multiple domains. *Learning and Instruction, 35*, 16–32. https://doi.org/10.1016/j.learninstruc.2014.08.005

Marsh, H. W., & O'Neill, R. (1984). Self Description Questionnaire III: The construct validity of multidimensional self-concept ratings by late adolescents. *Journal of Educational Measurement, 21*(2), 153–174. https://doi.org/10.1111/j.1745-3984.1984.tb00227.x

McNeish, D., & Stapleton, L. M. (2016). Modeling clustered data with very few clusters. *Multivariate Behavioral Research, 51*(4), 495–518. https://doi.org/10.1080/00273171.2016.1167008

Möller, J., Helm, F., Müller-Kalthoff, H., Nagy, N., & Marsh, H. W. (2015). Dimensional comparisons and their consequences for self-concept, motivation, and emotion. In D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (pp. 430–436). Elsevier.

Möller, J., Helm, F., Müller-Kalthoff, H., Nagy, N., & Marsh, H. W. (2016). The generalized internal/external frame of reference model: An extension to dimensional comparison theory. *Frontline Learning Research, 4*, 1–11. https://doi.org/10.14786/flr.v4i2.169

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

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

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

Möller, J., Streblow, L., & Pohlmann, B. (2006). The belief in a negative interdependence of math and verbal abilities as determinant of academic self-concepts. *The British Journal of Educational Psychology, 76*(Pt 1), 57–70. https://doi.org/10.1348/000709905X37451

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

Möller, J., Zitzmann, S., Helm, F., Machts, N., & Wolff, F. (2020). A meta-analysis of relations between achievement and self-concept. *Review of Educational Research, 90*(3), 376–419. https://doi.org/10.3102/0034654320919354

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

Romer, P. M. (1987). Growth based on increasing returns due to specialization. *The American Economic Review, 77*, 56–62.

Schmidt, I., Brunner, M., & Preckel, F. (2017). Effects of achievement differences for internal/external frame of reference model investigations: A test of robustness of findings over diverse student samples. *British Journal of Educational Psychology*, 1–16. https://doi.org/10.1111/bjep.12198

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research, 46*, 407–441. https://doi.org/10.2307/1170010

Skaalvik, E. M., & Rankin, R. J. (1992). Math and verbal achievement and self-concepts: Testing the internal/external frame of reference model. *The Journal of Early Adolescence, 12*(3), 267–279. https://doi.org/10.1177/0272431692012003003

Steinmayr, R., & Spinath, B. (2015). Intelligence as a potential moderator in the internal/external frame of reference model. An exploratory analysis. *Journal for Educational Research Online/Journal Für Bildungsforschung Online, 7*(1), 198–218.

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

Wolff, F., Zitzmann, S., & Möller, J. (2021). Moderators of dimensional comparison effects: A comprehensive replication study putting prior findings on five moderators to the test and going beyond. *Journal of Educational Psychology, 113*(3), 621–640. https://doi.org/10.1037/edu0000505

Zimmermann, F., Möller, J., & Köller, O. (2017). When students doubt their teachers' diagnostic competence: Moderation in the internal/external frame of reference model. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000196