



Social and dimensional comparison effects in general and domain-specific test anxiety: a nested factor modeling approach

Irma Talić¹ · Jörn R. Sparfeldt² · Jens Möller³ · Karl-Heinz Renner¹ · Samuel Greiff⁴ · Christoph Niepel⁴

Accepted: 4 July 2023 / Published online: 30 August 2023
© The Author(s) 2023

Abstract

The generalized internal/external frame of reference (GI/E) model assumes social and dimensional achievement comparisons to form self-perceptions. These domain-specific comparisons have been shown to shape two facets of test anxiety (i.e., worry and emotionality) both directly and indirectly through academic self-concepts. However, examinations of such domain-specific relations have rarely integrated general components, although the hierarchical nature of both test anxiety and academic self-concept is well-known. Thus, the present study implemented a nested factor modeling approach. We examined social and dimensional comparison effects on worry and emotionality as well as mediation effects of academic self-concepts in the math and verbal domains while controlling for general components. We contrasted this approach with the conventionally used first-order factor model where general components were not considered. Data from $N=348$ German secondary school students ($M_{\text{age}}=15.3$ years, Grades 9–10) were analyzed using structural equation models. Direct negative within-domain and positive cross-domain achievement-anxiety relations emerged, yet, the pattern of cross-domain relations changed across modeling approaches. Only the nested factor model showed indirect cross-domain mediation relations. Our findings suggest the importance of structural representations of hierarchical constructs. The nested factor model approach enhanced predictions within the GI/E model, particularly those related to dimensional comparisons.

Keywords Test anxiety · Worry · Emotionality · Structural equation modeling · Nested factor model · Generalized internal/external frame of reference model

Introduction

Test anxiety (TA) comprises a set of detrimental reactions to potential failure in evaluative situations (Zeidner, 1998). Such reactions can ultimately lead to severe educational disadvantages (Zeidner, 2020). Major research efforts have been devoted to predicting TA with the goal of preventing this experience and creating effective interventions

(von der Embse et al., 2013). To predict TA, domain-specific achievement-based comparison processes can be drawn on (Arens et al., 2017; Marsh, 1988; Schilling et al., 2005; Streblov, 2004). Using the internal/external frame of reference (I/E) model, social (e.g., “How good am I in math compared with my classmates?”) and dimensional (e.g., “How good am I in math compared with German?”) comparison processes were postulated to shape the formation of the domain-specific academic self-concept (ASC; Marsh, 1986). ASC is typically defined as students’ self-perceptions of their own competence (Marsh & Craven, 2006). In extending the I/E model to the generalized I/E (GI/E) model (Möller et al., 2016), TA was included as an outcome variable. Generalizing results from ASC research to TA research is facilitated on the basis of a link between the two constructs that comprises causality (i.e., ASC as determinant of TA; Marsh, 1988; see also Schilling et al., 2005) and structural similarities (i.e., hierarchical structure with general component at the apex; Gogol et al., 2016).

Irma Talić
irma.talic@unibw.de

¹ Institute of Psychology, University of the Bundeswehr Munich, Munich, Germany

² Department of Educational Science, Saarland University, Saarbrücken, Germany

³ Institute for Psychology of Learning and Instruction, Kiel University, Kiel, Germany

⁴ Present Address: Department of Behavioral and Cognitive Sciences, University of Luxembourg, Esch-Sur-Alzette, Luxembourg

Domain-specificity plays a pivotal role in these social and dimensional comparison processes within the GI/E model. Regarding TA, however, general (e.g., Cassady & Johnson, 2002) and domain-specific (e.g., Sparfeldt et al., 2005) approaches seem to coexist in parallel. These are rarely combined, even though the consideration of general TA can alter domain-specific achievement-anxiety relations (Devine et al., 2012). Methodologically, the consideration of a hierarchical general factor can be achieved with a nested factor (NF) model (Gogol et al., 2017; Gogol et al., 2016, for ASC see also Arens et al., 2021; Brunner et al., 2009; Brunner et al., 2010). In this model, general variance is distinguished from domain-specific variance, which seems suitable with regard to the domain-specific processes that the GI/E model examines.

The overarching objective of the present study is therefore the application of the NF model within the GI/E framework. Specifically, domain-specific social and dimensional comparison processes regarding the TA facets (i.e., worry and emotionality) in the math and verbal domains are investigated, while controlling for general TA. In doing so, we examined both direct achievement-anxiety paths and indirect mediation paths through the ASC while controlling for general ASC. To illustrate differences between this model and the conventional choice of modeling strategy, we contrasted our NF model against a first-order factor (FOF) model in which general components were not considered and general variance was subsumed in domain-specific components. To achieve this, we first give an overview on the structure of TA and reiterate theoretical and empirical considerations for the inclusion of TA (facets) as outcome in the GI/E model, before introducing the NF model and proposing its application in investigating social and dimensional comparison effects in the GI/E model.

The Structure of Test Anxiety (TA)

Exam- or test-related concerns about failure or negative consequences can be seen in a set of detrimental phenomenological, physiological, and behavioral responses denoted as TA (Zeidner, 1998). TA is of high interest to researchers and practitioners due to its detrimental associations with academic achievement (Barroso et al., 2021; Hembree, 1988; von der Embse et al., 2018), and subjective well-being (Steinmayr et al., 2016). In the light of these relations, efforts have been directed towards investigating TA to enhance the understanding of its structure and antecedents.

Multidimensionality

In general, multidimensionality can be conceived with regard to different aspects, for instance, multidimensionality in terms of domain-specificity and multidimensionality in terms of different construct components or facets (Arens et al., 2011). Multidimensionality in terms of domain-specificity is discussed in detail in the paragraph on TA's hierarchy (see Section *Hierarchy: Domain-Specificity and Generality*), contrasting domain-specific and general approaches to TA. Here, we discuss multidimensionality in terms of different construct facets. Liebert and Morris (1967) identified two fundamental facets of TA, worry and emotionality. To date, there is a wide agreement on these two facets (e.g., Zeidner, 2020). Worry as the cognitive facet encompasses negative self-talk and failure-focused expectations or cognitions and generally shows stronger negative relations with academic achievement (e.g., Hembree, 1988; Steinmayr et al., 2016; von der Embse et al., 2018). Emotionality refers to perceived autonomic hyperarousal (e.g., rapid heartbeat or sweating) or feelings of nervousness (Morris et al., 1981). Worry and emotionality tend to be moderately correlated with each other but are conceptualized as two separate TA dimensions that respond to different stimuli in evaluative situations (Morris et al., 1981). This two-dimensional facet distinction has received consistent empirical support (e.g., Gogol et al., 2017; Hembree, 1988; Sparfeldt et al., 2013).

Hierarchy: Domain-Specificity and Generality

Multidimensionality in terms of domain-specificity has been investigated in TA, where both general approaches, assessing TA with regard to test situations *in general* (e.g., Cassady & Johnson, 2002) and domain-specific approaches, assessing TA with regard to specific domains or subjects (e.g., Sparfeldt et al., 2005) are present in the literature. Without discounting the general nature of TA, the importance of considering different anxiety contexts has been emphasized throughout, visible in the conceptualization of TA as a *situation-specific* or *contextualized* personality trait (see Zeidner, 2020). Accordingly, distinct domain (i.e., school subject) factors need to be considered in both TA facets, worry and emotionality (Sparfeldt et al., 2005). Yet, employing distinct domain-specific factors only (i.e., not considering general manifestations) cannot adequately represent the hierarchical structure of TA. Pointing out the importance of considering general *and* domain-specific manifestations simultaneously, Devine et al. (2012) reported changes in the pattern of results in achievement-anxiety relations when controlling for general anxiety levels.

The Link between TA and ASC

Some of these construct characteristics (i.e., multidimensionality, domain-specificity, generality, hierarchy) apply not only to TA (i.e., phenomenological, physiological, and behavioral reactions associated with possible failure in tests or other evaluative situations; Zeidner, 1998), but also ASC (i.e., students' self-perceptions of their own competence; Marsh & Craven, 2006)—a key determinant of student achievement (Möller et al., 2020). The link between TA and ASC is discussed with regard to (a) structural similarities between TA and ASC, (b) causal relations between TA and ASC, and (c) the transfer of ASC to TA research. Similar to TA, (a) the structure of ASC has been subject to investigation (e.g., Brunner et al., 2010, for an overview of different structural models and their implications see Arens et al., 2021). Some recent works have placed emphasis on structural similarities of TA and ASC: Specifically, both constructs were paralleled with regard to their domain-specific structure with a general component at the apex of the hierarchy (Gogol et al., 2017; Gogol et al., 2016).¹ Further, (b) a causal link between TA and ASC has been discussed with ASC as predictor of TA (i.e., low [high] ASC leading to higher [lower] TA in the same domain; Marsh, 1988). Empirical evidence for negative relations between ASC and TA has repeatedly been reported (Ahmed et al., 2012; Gogol et al., 2017; von der Embse et al., 2018). Despite presumed reciprocal relations between ASC and TA, the effect of ASC on TA seems to be more crucial (Schilling et al., 2005). Correspondingly, achievement-TA relations have been shown to be mediated through ASC (Arens et al., 2017; Schilling et al., 2005). Finally, (c) on the basis of a causal relationship between TA and ASC, Marsh (1988) first suggested a possible application of ASC research-derived results to TA: "If self-concept is a causal determinant of anxiety, then processes affecting self-concept should also affect anxiety" (p. 139). In particular, this has been done with regard to the GI/E model.

The Generalized Internal/External Frame of Reference (GI/E) Model

The observation that math and verbal ASCs were nearly uncorrelated despite the substantial correlations of math and verbal achievement indicators led to the development of the I/E model that aimed to explain the formation of ASCs through achievement-based social and dimensional

¹ In addition and analogously to TA, one could also distinguish an affective and a cognitive facet in ASC as another similarity between the two constructs (see Arens et al., 2011). In the present study, however, we defined ASC as self-perceptions of competence, emphasizing its cognitive nature.

comparison processes (Marsh, 1986). The I/E model was later extended to the so-called the GI/E model (Möller et al., 2016) such that constructs other than ASC (i.e., in our case TA) could be considered as outcome variables of these comparisons. Within the GI/E model, where multiple domains are considered, the social and dimensional comparisons show in the direct within and cross-domain paths.

Direct Paths: Social and Dimensional Comparisons

When engaging in social comparisons, students draw on an external frame of reference and compare their achievement in one domain with relevant others' achievements in the same domain (Möller et al., 2016). Social comparison effects thereby appear as negative within-domain achievement-anxiety relations (e.g., higher math achievement is related to lower math TA; Arens et al., 2017). Dimensional comparisons require an internal frame of reference as students compare their own achievements across domains (Möller & Marsh, 2013; Möller et al., 2016). Dimensional comparison effects appear as positive cross-domain achievement-anxiety relations (e.g., higher math achievement is related to higher German TA; Arens et al., 2017).

With regard to the ASC, empirical support for GI/E-hypothesized comparison processes is ample (see meta-analysis by Möller et al., 2020). With regard to TA, evidence for GI/E-hypothesized social comparison effects on worry and emotionality has been consistent, whereas evidence for dimensional comparison effects has been tied to math but not verbal TA (Arens et al., 2017; Schilling et al., 2005). In another study, dimensional comparison effects were observed in the verbal domain, yet applied to a measure of TA that did not differentiate between the worry and emotionality components (Marsh, 1988). Another study identified social comparison effects in math and two verbal domains, as well as dimensional comparison effects between one verbal domain (i.e., French) but not another verbal domain (i.e., German) and math anxiety, and between the two verbal domains in a multilingual context (van der Westhuizen et al., 2022). These findings were derived using a unidimensional measure of TA that was not differentiating different facets. Thus, there is some evidence for the relevance of social and dimensional comparison processes in the formation of domain-specific TA. To gain further insight into the mechanism of these processes, mediation analyses have been conducted using the ASC as mediator.

Indirect Paths: Mediation via Academic Self-Concept (ASC)

Based on theoretical assumptions (i.e., school grades as source for self-perceptions of academic competence that in

turn influence further socio-affective variables such as TA) and empirical relations between achievement and ASC on the one hand (Möller et al., 2020), and between ASC and TA on the other hand (see Section *The Link between TA and ASC*), the possible mediation of achievement-anxiety relations through ASC seems straightforward. Accordingly, some empirical support has been reported (Arens et al., 2017; Schilling et al., 2005). In the GI/E model, the mediation of social comparisons is indicated by substantial indirect within-domain paths (e.g., between achievement in math and TA in math through the math ASC) and the algebraic sign of the indirect path would be negative (i.e., multiplying a positive within-domain achievement-ASC relation with a negative within-domain ASC-TA relation). Inversely, the mediation of dimensional comparisons is indicated by substantial indirect cross-domain paths (e.g., between achievement in math and TA in German through the German ASC) and the sign of this indirect path would be positive (i.e., multiplying a negative cross-domain achievement-ASC relation with a negative within-domain ASC-TA relation).

Arens et al. (2017) reported significant indirect paths for both social and dimensional comparisons in both the math and verbal domains. Schilling et al. (2005) reported significantly lower direct achievement-anxiety relations after including ASCs, that did not reach statistical significance in the verbal domain, suggesting differential mediation effects in both domains. Hence, evidence for a mediation of GI/E-based achievement-anxiety relations through ASCs is provided, yet worth replicating. In addition, even though all hypothesized relations within the GI/E model relate to domain-specific construct manifestations, the structural representation of TA and ASC implemented in these studies (Arens et al., 2017; Marsh, 1988; Schilling et al., 2005; van der Westhuizen et al., 2022) did not allow for a precise disentanglement of different domain-specific construct components (e.g., a clear differentiation of math anxiety from verbal anxiety or general anxiety). To prevent a conglomeration of different domain-specific and general construct components that blur the examination of strict within- and cross-domain relations within the GI/E model, we argue for adapting the constructs' structural representations.

Structural Representation within Nested Factors

To best represent the multidimensional and hierarchical construct structure of both TA and ASC, we chose a nested factor (NF) model (Gustafsson & Balke, 1993). This model can be used to decompose a given manifestation (e.g., math worry) into its general component, its domain-specific component, and measurement error (Eid et al., 2017). A general factor is specified to influence all (general and

domain-specific) items, whereas domain-specific factors are specified to additionally influence their respective domain-specific items. The general factor serves as the reference domain, that is, the general items do not form their own domain-specific factor but load directly on the general factor along with all the other domain-specific items. Domain-specific factors are interpreted as residual factors (i.e., the part of the domain-specific manifestation that is not accounted for by the general component) and are thus uncorrelated with the general factor. Different domain-specific factors can correlate. This model is also referred to as the bifactor (S-1)-model because it has one domain-specific factor less than the number of domains that are included (Eid et al., 2017).

The NF model has been validated with regard to the ASC (i.e., Nested Marsh/Shavelson Model; Arens et al., 2021; Brunner et al., 2010; Brunner et al., 2009). Given the structural similarities of ASC and TA, it has been applied to TA as well (Gogol et al., 2017; Gogol et al., 2016). The NF model takes account of general manifestations operating at the apex of domain-specific manifestations (Brunner et al., 2010). In contrast to a higher-order factor model, where all domain-specific latent factors load on a higher-order latent general factor, the NF model shows superior model fit when correlations between the domain-specific factors are low (Arens et al., 2021). The NF model thus offers flexibility in representing relations between domain-specific factors as positive, negative, or zero, while retaining the meaning of the general factor due to its defined reference domain irrespective of the number and scope of the domains that are included (Eid et al., 2017).

Yet, within the GI/E model, the predominant modeling strategy is the first-order factor (FOF) model in which domain-specific items load on their respective domain-specific factors only. Hierarchical structures cannot be represented in this model (Arens et al., 2021). The domain-specific factors in the FOF model represent a mixture of general and domain-specific variances, which can distort relations with correlates (Brunner et al., 2009). In contrast to this mixture, in the NF model, domain-specific factors have a clear meaning and operate independently from general levels. This separation seems fruitful particularly in the GI/E framework in which the domain-specific relations are of utmost interest.

The Present Study

In the present study, we therefore demonstrated the application of the NF modeling approach in investigating relations in the GI/E model in contrast to the widely used FOF models. Specifically, we examined the role of social and dimensional comparisons in the formation of the TA facets worry and emotionality while disentangling general and domain-specific

components and hence purifying domain-specific relations—the core of the GI/E model. Contrasting the FOF to the NF models within the GI/E framework allowed us to examine the difference in result patterns concerning (mediated) social and dimensional comparison effects on the two TA facets in dependence on the modeling strategy. In other words, we controlled for the influence of general TA and ASC levels on domain-specific TA and ASC manifestations and with this, potentially draw a more complete picture of social and dimensional comparisons—comparisons that students naturally engage in and thus hold important implications both on theoretical (e.g., investigating the strength of dimensional comparisons when general levels are controlled for) and practical grounds (e.g., adjusting psychoeducation in TA interventions).

A careful synthesis of the current literature indicated that our study is the first one to examine direct and ASC-mediated social and dimensional comparison effects on the TA facets worry and emotionality in the math and verbal domains in the GI/E model (Arens et al., 2017; Schilling et al., 2005), that controlled for general TA and ASC using an NF modeling strategy (Gogol et al., 2017; Gogol et al., 2016).

In our first research question (RQ), we aimed to replicate findings reported in previous studies (Arens et al., 2017; Schilling et al., 2005) that used FOF models to observe social comparison effects of grades as achievement indicators on facets of TA in both math and German and dimensional comparison effects on facets of TA in math but not in German in German secondary school students. Further, in Arens et al. (2017), all direct achievement-TA paths were fully mediated by ASC in both domains, whereas Schilling et al. (2005) reported full mediations in German and partial mediations in math.

RQ1: Replication. Can prior work (Arens et al., 2017; Schilling et al., 2005) be replicated with regard to (a) FOF GI/E model relations and (b) FOF mediated GI/E model relations?

Second, we addressed our focal RQ, which aimed at investigating social and dimensional comparison effects on facets of TA when controlling for general manifestations in the NF model. In doing so, we examined direct paths between grades and facets of TA as well as indirect, ASC-mediated paths. Hereby, general TA and ASC were controlled for within nested factors. Hence, this RQ addressed the question of whether domain-specific social and dimensional comparisons influence facets of TA irrespective of general TA and whether these relations are mediated by domain-specific ASCs when controlling for general ASC.

RQ2: Extension. Can (a) GI/E model relations and (b) mediated GI/E model relations be detected when employing the NF modeling approach? How will (c) statistical predictions differ across the NF versus FOF models?

Third, the NF models include additional paths that are not formalized in the original GI/E model. Transferring domain-specific processes to general processes, we examined relations between achievement and general facets of TA, as well as their mediation by general ASC.

RQ3: Ancillary. How do additional paths between grades and general worry and emotionality show in the NF GI/E model, and are these paths mediated by general ASC in the NF mediated GI/E model?

Method

Procedure and Participants

The present work is part of the larger “Dynamics of Academic Self-Concept in Everyday Life” (DynASCEL) project (Niepel et al., 2022) on students’ perceptions of academic competence and learning environments, where an intensive longitudinal experience sampling design was embedded in a paper-and-pen pre- and post-assessment. In the present study, only selected data from the pre-assessment were relevant for our research questions.² We recruited a convenience sample of $N=348$ German secondary school students (43.1% of whom were male students based on $n=340$ students with available gender information) attending the ninth ($n=288$) and 10th ($n=60$) grades of the highest ability track (i.e., the German *Gymnasium*). Students were nested within 18 classrooms from six schools located in four different German federal states (i.e., Baden-Württemberg, Mecklenburg-Vorpommern, Nordrhein-Westfalen, Rheinland-Pfalz). Participants reported a mean age of 15.3 years ($SD=0.66$, range = 13.3 to 17.4 years; based on $n=335$). Student clusters within classrooms were stable across school subjects and across school grades, such that students were asked to refer to the same math and German test situations.

The APA Ethics Code (American Psychological Association, 2020) was considered in all stages of the research process to ensure scientific accuracy whilst protecting the rights and welfare of the minor participants. Specifically, student participation was voluntary, participants could withdraw from the study at any time without stating any reasons and without facing any negative consequences, and written parental consent was obtained for all participating students. Students, parents, and schools were exhaustively informed

² Data from the larger research project, including data used in the present study, have been and will be used in other manuscripts, yet addressing different research questions (e.g., Dörendahl et al., 2021; Franzen et al., 2022; Hausen et al., 2022). Data on the present study’s core variables (i.e., domain-specific and general worry and emotionality) have not been reported elsewhere.

on the study's purposes and subsequent data processing. All measures and procedures were approved by the local ethics review panel of the University of Luxembourg and by all involved education authorities in the respective four German federal states.

Measures

Test Anxiety (TA)

TA was assessed with general (i.e., school in general) and domain-specific (i.e., math and verbal domains) adaptations of worry and emotionality items based on the German Test Anxiety Inventory (TAI-G; Hodapp, 1991; Hodapp et al., 2011). Following the introduction “*In evaluative situations (e.g., tests, written or oral examinations)*,” students responded to the five parallel-worded item stems for worry (e.g., “*I worry about my results*”) and emotionality, each (e.g., “*I feel anxious*”). The items were presented in a grid format as first introduced by Rost and Sparfeldt (2002), where the item stems were presented in rows with a placeholder “*...*” (for the target domain), and the target domains (i.e., school, math, German) were presented in columns. The students related the items stems from the rows to the target domain in the column and responded using a 6-point Likert scale in the cells of the grid, ranging from 1 (*almost never*) to 6 (*almost always*) such that higher scores represented higher TA (see also Sparfeldt et al., 2013; Sparfeldt et al., 2005). Domain-specific worry and emotionality ratings presented in this format have been shown to be reliable with ω coefficients ≥ 0.91 and measurement invariant across school subjects (Schneider et al., 2022).

Report Card Grades

Students reported their math and German grades from their last report card, which we used as academic achievement indicators. Self-reported and actual grades tend to be highly correlated in German school student samples, indicating the reliability and validity of self-reported grades ($r \geq 0.91$, Sparfeldt et al., 2008; see $r \geq 0.76$ for grades 9 and 10 in a German-speaking Swiss sample reported by Sticca et al., 2017 and $r = 0.88$ for grades 7 and 8 across three German school tracks reported by Dickhäuser & Plenter, 2005). School grades in Germany are measured on a 6-point Likert scale, which we recoded so that higher values corresponded with higher achievements, ranging from 1 (*insufficient*) to 6 (*excellent*).

Academic Self-Concept (ASC)

General (i.e., school in general) and domain-specific (i.e., math and verbal domains) ASCs were assessed using six parallel-worded items each, which were based on the

well-validated and reliable Self-Description Questionnaire (SDQ; Marsh et al., 1983) and the short scale by Gogol et al. (2014). An example item is “*I am good at [most school subjects] / [math] / [German]*.” Gogol and colleagues (2014) reported reliability coefficients of $\omega \geq 0.75$ for their three-item short scales. Items were rated on a 6-point Likert scale ranging from 0 (*does not apply at all*) to 5 (*completely applies*) such that higher values indicated higher ASCs.

Statistical Analyses

Statistical analyses were performed within the structural equation modeling (SEM) framework using the software package Mplus 8.3 (Muthén & Muthén, 1998–2017). To adjust standard errors for the nonindependence of observations because students were clustered in classrooms, we used the “TYPE = COMPLEX” option. Correlated uniqueness was considered by allowing for correlated residual variances between parallel-worded items. We used the MLR estimator to obtain robust standard errors and deal with missing data (Kaplan, 2009; Muthén & Muthén, 1998–2017). The percentages of missing values ranged from 1.7% to 3.2% for worry, 2.6% to 4.9% for emotionality, and 2.3% to 3.7% for ASC across domains. 4.3% and 5.2% were missing in grades in math and German, respectively.

To reduce the complexity and support the power of the model, we (a) reduced the number of indicators per factor (Hoyle & Gottfredson, 2015), selecting three items for each TA facet and ASC out of the larger item sets based on the size of the factor loadings (see also Marsh et al., 2006).³ Such short scales have been shown to measure TA and ASC reliably (Gogol et al., 2014). Further, we (b) adopted a two-step approach (Anderson & Gerbing, 1988), in which we, first, conducted confirmatory factor analyses based on which we extracted values for factor loadings, item residual variances, and exogenous factor variances. Second, we fixed these parameters to these values when specifying the structural models. To enter school grades as latent single-item factors, we followed the procedure illustrated by Kline (2016), fixing factor loadings to 1 and fixing residual variances to constant values that were derived from the indicators' empirical sample variance and the reliability estimate reported in previous work (Sparfeldt et al., 2008).

To estimate the replicability of result patterns in the FOF model (RQ1), we specified the FOF GI/E model with domain-specific grades as predictors, domain-specific worry and

³ To test the robustness of our results, we additionally ran all models with the respective full item sets. Descriptively, the model fits were lower compared with the models using three-item scales (i.e., CFI $\geq .939$, RMSEA $\leq .061$, SRMR $\leq .062$). The pattern of significant within- and cross-domain paths was identical across models using three items versus models using the full item sets.

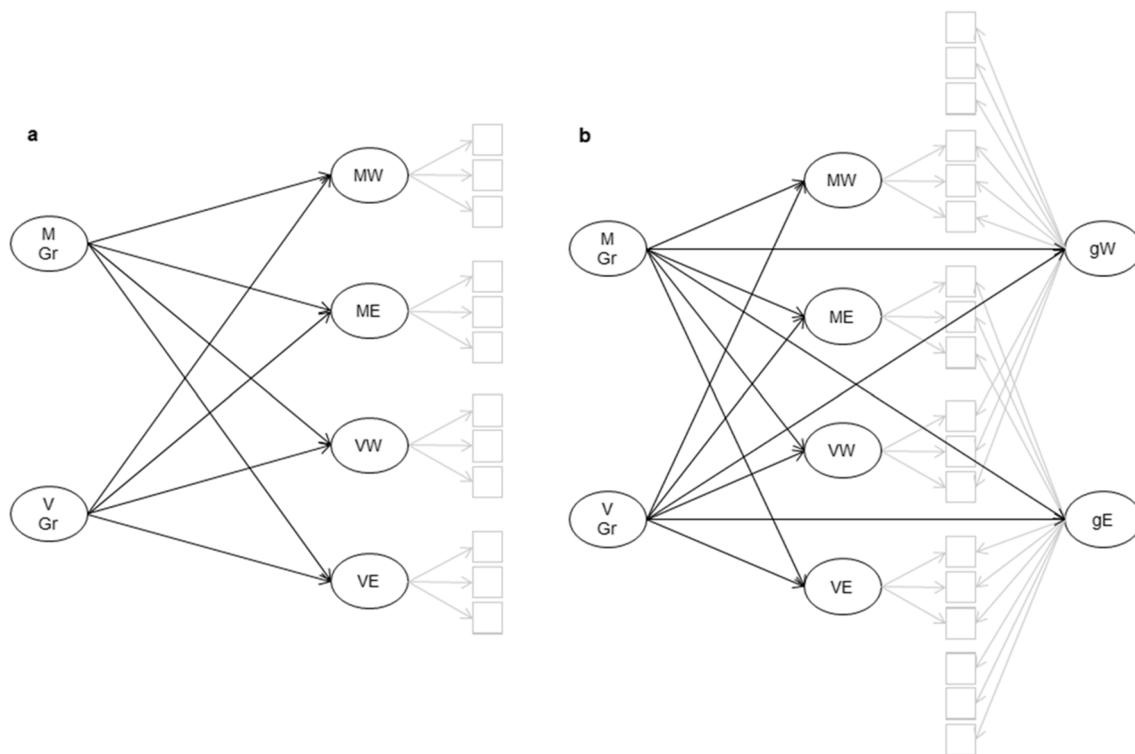


Fig. 1 Different modeling approaches in testing GI/E relations (a) the first-order factor (FOF) GI/E model, where facets of test anxiety (TA; i.e., worry and emotionality) are predicted by grades via within- and cross-domain paths in two domains (Model 1b), and (b) the nested factor (NF) GI/E model, where domain-specific and general facets of TA are predicted by grades via within- and cross-domain

paths in two domains (Model 3b). M Gr = Math grade; V Gr = German grade; MW = Math worry; ME = Math emotionality; VW = German worry; VE = German emotionality; gW = general worry; gE = general emotionality. For better clarity of presentation, measurement models are displayed in grey, and item residual variances, factor variances, and correlational paths have been omitted

emotionality as criteria, and within- or cross-domain regression paths between each predictor and each criterion. Importantly, domain-specific factors influenced their corresponding domain-specific indicators only (see Fig. 1a). In the FOF mediated GI/E model, domain-specific ASCs were included as mediator variables. Indirect relations were requested using the “MODEL INDIRECT” option in Mplus. Analogous to Arens et al. (2017), we focused on the paths for which the ASC and TA facets belonged to the same domain. To estimate the sizes of the indirect effects, we calculated squared standardized indirect path coefficients as measures of explained variance in accordance with Lachowicz et al.’s (2018) recommendations. Thus, cut-off criteria for proportions of explained variance were applied (i.e., small = 2%, medium = 15%, large = 25%; Cohen, 1988).

To investigate GI/E-hypothesized relations while controlling for general components (RQ2), we next implemented the NF modeling approach by adding general factors (the S-1 specification according to Eid et al., 2017). To this end, domain-specific (i.e., math and German) worry and emotionality items were specified to load on their respective domain-specific latent factors. In addition, we specified

general factors, which influenced all worry or emotionality domain-specific and general items. Correlations between the general and its domain-specific factors (e.g., general worry and math worry) were fixed to zero. Relations among domain-specific factors were allowed. In the NF GI/E model, domain-specific grades were entered as predictors, and domain-specific and general worry and emotionality were entered as criteria (see Fig. 1b). In the NF mediated GI/E model, we added domain-specific and general ASCs as mediator variables. Here, ASC was represented analogously within nested factors.

Clearly, the NF modeling approach yielded relations that had not been formalized in the original GI/E model (i.e., paths between grades and general TA and their mediation via general ASC), which we additionally addressed in RQ3.

For model evaluation, we followed the recommended cut-off criteria in the absolute goodness-of-fit indices CFI, RMSEA, and SRMR, where values of $CFI \geq 0.95$, $RMSEA \leq 0.06$, and $SRMR \leq 0.08$ are considered to indicate a good fit to the data (Hu & Bentler, 1999).

Table 1 Goodness-of-fit Indices for Tested Models

Model	MLR χ^2 (<i>df</i>)	CFI	RMSEA [90% CI]	SRMR
Preliminary confirmatory factor analyses				
1a First-order factor measurement model with six factors (i.e., worry, emotionality, and grades in math and German)	72.915 (58)	0.995	0.027 [0.000; 0.045]	0.024
2a First-order factor measurement model with eight factors (i.e., worry, emotionality, academic self-concepts, and grades in math and German)	294.289 (135)	0.970	0.058 [0.049; 0.067]	0.032
3a Nested factor measurement model with eight factors (i.e., worry and emotionality in math, German and general, and grades in math and German)	151.966 (118)	0.993	0.029 [0.012; 0.041]	0.025
4a Nested factor measurement model with 11 factors (i.e., worry, emotionality, and academic self-concept in math, German and general, and grades in math and German)	490.707 (285)	0.974	0.046 [0.039; 0.052]	0.033
First-order factor models				
1b First-order factor GI/E model with six factors (i.e., worry, emotionality, and grades in math and German)	66.227 (80)	1.000	0.000 [0.000; 0.017]	0.024
2b First-order mediated GI/E model with eight factors (i.e., worry, emotionality, academic self-concept, and grades in math and German)	278.925 (168)	0.979	0.044 [0.034; 0.052]	0.034
Nested factor models				
3b Nested factor GI/E model with eight factors (i.e., worry and emotionality in math, German and general, and grades in math and German)	140.231 (162)	1.000	0.000 [0.000; 0.012]	0.028
4b Nested factor mediated GI/E model with 11 factors (i.e., worry, emotionality, and academic self-concept in math, German and general, and grades in math and German)	493.847 (359)	0.983	0.033 [0.025; 0.040]	0.041

MLR = maximum likelihood estimation with robust standard errors, CFI = comparative fit index, RMSEA = root mean square error of approximation, CI = confidence interval, SRMR = standardized root mean square residual

Results

Preliminary Analyses

All preliminary confirmatory factor analyses showed a good fit to the data for both the FOF and NF modeling approaches (see Models 1a, 2a, 3a, and 4a in Table 1).⁴ Table 2 presents the standardized factor loadings as well as the corresponding McDonald's ω reliability coefficients.⁵ All factor loadings differed significantly from zero and were moderate to large in both modeling approaches. The reliability coefficients of the three-item scales were $\omega \geq 0.88$ in the FOF model and $\omega \geq 0.70$ in the NF model across factors and domains (see Table 2).

The latent factor correlations across the modeling approaches can be found in Table 3. Significantly negative within-domain relations between TA and grades and ASCs were observed in all domain-specific TA facets except for German worry, where only one relation to German ASC differed significantly from

zero in the NF model. Relations within a facet changed considerably across modeling approaches (i.e., $\rho = .72$, $p < .001$ [$\rho = .33$, $p = .016$] for math and German worry in the FOF [NF] model, and $\rho = .59$, $p < .001$ [$\rho = -.11$, $p = .408$] for math and German emotionality in the FOF [NF] model). Across domains and modeling approaches, domain-specific worry and emotionality were moderately correlated with each other (i.e., $\rho = .45$ to $\rho = .58$). In the NF model, general worry and emotionality were positively correlated at $\rho = .55$ ($p < .001$), and each was negatively related to the math grade (i.e., general worry: $\rho = -.15$, $p = .044$; general emotionality: $\rho = -.21$, $p = .001$).

The First-Order Factor (FOF) GI/E Model

First, we addressed RQ1 to replicate prior findings with (a) the FOF GI/E and (b) the FOF mediated GI/E model. The (a) FOF GI/E model showed an excellent fit to the data (Model 1b in Table 1). Table 4 presents the standardized path coefficients and standard errors. Negative within-domain paths between grades and facets of TA, indicating social comparison effects, differed significantly from zero for worry and emotionality in math (math grade \rightarrow math worry, $\beta = -.38$ and math grade \rightarrow math emotionality, $\beta = -.47$, $ps < .001$) and emotionality in German (German grade \rightarrow German emotionality, $\beta = -.21$, $p = .001$). Statistically significant positive cross-domain paths between math [German] grades and facets of TA in German [math], indicating dimensional contrast effects, were only observed between the German grade

⁴ To ensure the applicability of the NF models compared with higher-order factor models in which general factors load on domain-specific latent factors, we also computed higher-order factor models. These models showed inadequate model fits when we considered the combination of the aforementioned fit indices.

⁵ Note that in Table 2, only Models 2a and 4a are reported because these include all examined constructs. Standardized factor loadings differed between Model 1a versus 2a and between Model 3a versus 4a only negligibly with $\Delta\lambda \leq .006$ in all cases.

Table 2 Standardized Factor Loadings and McDonald’s ω Reliability Coefficient

Factor	First-Order Factor measurement model (Model 2a)				Nested Factor measurement model (Model 4a)			
	Item 1	Item 2	Item 3	ω	Item 1	Item 2	Item 3	ω
Worry _{math}	.807	.892	.880	.897	.401	.498	.458	.702
Worry _{German}	.798	.867	.867	.883	.560	.638	.597	.792
Emotionality _{math}	.858	.931	.892	.923	.440	.472	.485	.763
Emotionality _{German}	.860	.888	.853	.900	.600	.592	.577	.807
ASC _{math}	.909	.952	.916	.947	.688	.815	.741	.924
ASC _{German}	.889	.936	.922	.940	.733	.850	.791	.924
	Ranges Item 1 through Item 9			ω	Ranges Item 1 through Item 9			ω
Worry _{general}	NA			NA	$.570 \leq \lambda \leq .875$.939
Emotionality _{general}	NA			NA	$.616 \leq \lambda \leq .888$.954
ASC _{general}	NA			NA	$.407 \leq \lambda \leq .856$.944

ASC = academic self-concept, NA = not applicable in first-order factor modeling

All reported standardized factor loadings were significant at $p < .001$

and worry in math (German grade \rightarrow math worry, $\beta = .19$, $p = .005$). Hence, prior work could be replicated regarding social comparison effects on all domain-specific TA facets except for worry in German. Dimensional comparison effects on both TA facets in math (Arens et al., 2017; Schilling et al., 2005) were only replicated with regard to worry in math.

To examine (b) the FOF mediated G/E model, we added math and German ASCs as mediator variables. The model showed a good fit to the data (Model 2b in Table 1). Table 5 presents the standardized direct path coefficients along with standard errors. Only one direct path between grades and facets of TA, both within and across domains, was statistically

significantly different from zero (i.e., math grade \rightarrow math worry, $\beta = -.28$, $p = .035$). The direct relations between grades and ASC were all significantly positive within matching domains (math grade \rightarrow math ASC, $\beta = .82$ and German grade \rightarrow German ASC, $\beta = .80$, $ps < .001$) and negative across non-matching domains (i.e., math grade \rightarrow German ASC, $\beta = -.34$, $p < .001$, and German grade \rightarrow math ASC, $\beta = -.18$, $p = .008$), replicating the original I/E pattern. Direct paths between ASC and facets of TA reached statistical significance in a few cases, and if so, they were negative within domains (i.e., math ASC \rightarrow math emotionality, $\beta = -.46$, $p < .001$, and German ASC \rightarrow German emotionality, $\beta = -.27$, $p = .006$) and positive across

Table 3 Latent Factor Correlations in the First-Order Factor and Nested Factor Modeling Approach

Factor Correlations	1	2	3	4	5	6	7	8	9	10	11
1. Worry _{math}	–	.723***	.575***	.303***	–.293***	.007	–.269***	.158***	NA	NA	NA
2. Worry _{German}	.334*	–	.349***	.471***	–.111	.034	–.087	.063	NA	NA	NA
3. Emotionality _{math}	.454***	–.055	–	.586***	–.423***	–.138	–.515***	.086	NA	NA	NA
4. Emotionality _{German}	–.023	.448***	–.107	–	–.085	–.196*	–.116*	–.273***	NA	NA	NA
5. Grade _{math}	–.306***	–.004	–.449***	.115*	–	.483***	.736***	.028	NA	NA	NA
6. Grade _{German}	–.058	.006	–.064	–.156*	.482***	–	.207***	.634***	NA	NA	NA
7. ASC _{math}	–.242*	–.041	–.427***	.212***	.453***	–.202***	–	–.028	NA	NA	NA
8. ASC _{German}	.123	–.140*	.342***	–.430***	–.327***	.352***	–.436***	–	NA	NA	NA
9. Worry _{general}	.000 ^a	.000 ^a	.098	–.057	–.149*	.045	–.161*	.200***	–	NA	NA
10. Emotionality _{general}	.047	.029	.000 ^a	.000 ^a	–.212*	–.105	–.253***	.096	.545**	–	NA
11. ASC _{general}	.087	.126*	–.199***	–.009	.628***	.651***	.000 ^a	.000 ^a	–.061	–.199*	–

Correlations below the diagonal represent correlations within the nested factor model containing all examined factors (i.e., Model 4a), whereas correlations above the diagonal represent correlations within the first-order factor model containing all examined factors (i.e., Model 2a). ASC = academic self-concept, NA = not applicable in first-order factor modeling

^aFixed to zero due to nested factor modeling

* $p < .05$, *** $p < .001$

Table 4 Standardized Path Coefficients for the First-Order Factor and Nested Factor GI/E Model

	First-Order Factor GI/E Model (Model 1b)		Nested Factor GI/E Model (Model 3b)	
	β	SE	β	SE
Within-domain paths from grades to test anxiety				
Grade _{math} → Worry _{math}	-.381***	0.071	-.372**	0.108
Grade _{math} → Emotionality _{math}	-.466***	0.049	-.560***	0.082
Grade _{German} → Worry _{German}	.105	0.086	.001	0.111
Grade _{German} → Emotionality _{German}	-.209**	0.065	-.286***	0.051
Cross-domain paths from grades to test anxiety				
Grade _{math} → Worry _{German}	-.164	0.095	-.015	0.120
Grade _{math} → Emotionality _{German}	.014	0.056	.246***	0.058
Grade _{German} → Worry _{math}	.186**	0.067	.116	0.101
Grade _{German} → Emotionality _{math}	.082	0.062	.201*	0.081
Additional paths				
Grade _{math} → Worry _{general}	NA	NA	-.204**	0.077
Grade _{math} → Emotionality _{general}	NA	NA	-.205**	0.074
Grade _{German} → Worry _{general}	NA	NA	.140*	0.067
Grade _{German} → Emotionality _{general}	NA	NA	-.021	0.073

SE = standard error, NA = not applicable in first-order factor modeling

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

domains (i.e., German ASC → math worry, $\beta = .11, p = .042$, and German ASC → math emotionality, $\beta = .15, p = .029$). Finally, indirect paths (see Table 6) were significantly negative in two out of four cases within matching domains (i.e., math grade → math ASC → math emotionality, $\beta = -.38, p = .001$ and German grade → German ASC → German emotionality, $\beta = -.22, p = .006$). None of the four indirect paths across non-matching domains were significantly different from zero. Thus, in contrast to prior work, we found within-domain mediations that were related to emotionality only.

The Nested Factor (NF) GI/E model

Applying the Nested Factor (NF) approach to the GI/E model

Subsequently, we addressed RQ2, aimed at applying the NF modeling approach to (a) the GI/E model and (b) the mediated GI/E model and (c) contrasting statistical predictions in the NF versus FOF models. The (a) NF GI/E model showed an excellent fit to the data (Model 3b in Table 1). Statistically significant negative within-domain paths were found in three out of four cases (i.e., math grade → math worry, $\beta = -.37, p = .001$, math grade → math emotionality, $\beta = -.56, p < .001$ and German grade → German emotionality, $\beta = -.29, p < .001$), indicating social comparison effects. Statistically significant positive cross-domain paths were found only for emotionality (i.e., math grade → German emotionality, $\beta = .25, p < .001$, and German grade → math

emotionality, $\beta = .20, p = .013$), indicating dimensional comparison effects (see Table 4).

The (b) NF mediated GI/E model, including general and domain-specific ASCs as mediator variables, showed a good fit to the data (see Model 4b in Table 1). Standardized direct path coefficients (Table 5) indicated one significantly negative within-domain path (i.e., math grade → math worry, $\beta = -.35, p = .002$) and no significant direct cross-domain paths between grades and facets of TA. Concerning indirect (Table 6) within-domain paths, all except for the path between math grade and math worry via math ASC were significantly negative (i.e., math grade → math ASC → math emotionality, $\beta = -.24$, German grade → German ASC → German worry, $\beta = -.16$, and German grade → German ASC → German emotionality, $\beta = -.30$, all $ps < .01$). Indirect cross-domain paths were significantly positive for all except for the path between German grade and math worry via math ASC (i.e., math grade → German ASC → German worry, $\beta = .15$, math grade → German ASC → German emotionality, $\beta = .30$, and German grade → math ASC → math emotionality, $\beta = .19$, all $ps < .01$). Thus, a mediation of social and dimensional comparison effects on facets of TA via ASCs was supported for math emotionality and German worry and emotionality.

Finally, we (c) contrasted the statistical predictions made by the FOF versus NF models. In the nonmediated set of models (Model 1b vs. Model 3b; Table 4), the pattern of significant within-domain relations, indicative of social comparison effects, did not change, but the

Table 5 Standardized Direct Path Coefficients and Standard Errors for the First-Order Factor and the Nested Factor GI/E Mediation Model

	First-Order Factor GI/E Mediation Model (Model 2b)		Nested Factor GI/E Mediation Model (Model 4b)	
	β	<i>SE</i>	β	<i>SE</i>
Direct within-domain paths from grades to test anxiety				
Grade _{math} → Worry _{math}	-.280*	0.133	-.349**	0.112
Grade _{math} → Emotionality _{math}	-.027	0.136	-.208	0.111
Grade _{German} → Worry _{German}	.124	0.126	.170	0.168
Grade _{German} → Emotionality _{German}	-.012	0.117	.056	0.107
Direct cross-domain paths from grades to test anxiety				
Grade _{math} → Worry _{German}	-.190	0.138	-.150	0.182
Grade _{math} → Emotionality _{German}	.041	.128	-.083	0.100
Grade _{German} → Worry _{math}	.086	.092	.117	0.110
Grade _{German} → Emotionality _{math}	-.123	0.089	-.088	0.101
Additional paths				
Grade _{math} → Worry _{general}	NA	NA	-.224**	0.078
Grade _{math} → Emotionality _{general}	NA	NA	-.162	0.088
Grade _{German} → Worry _{general}	NA	NA	.186*	0.076
Grade _{German} → Emotionality _{general}	NA	NA	.138	0.090
Direct paths from grades to academic self-concept (mediator)				
Grade _{math} → ASC _{math}	.817***	0.050	.738***	0.050
Grade _{math} → ASC _{German}	-.342***	0.080	-.655***	0.077
Grade _{math} → ASC _{general}	NA	NA	.396***	0.066
Grade _{German} → ASC _{math}	-.175**	.067	-.579***	0.064
Grade _{German} → ASC _{German}	.798***	0.039	.665***	0.062
Grade _{German} → ASC _{general}	NA	NA	.477***	0.052
Direct paths from academic self-concept (mediator) to test anxiety				
ASC _{math} → Worry _{math}	-.078	0.111	-.053	0.078
ASC _{math} → Emotionality _{math}	-.461***	0.119	-.325***	0.088
ASC _{math} → Worry _{German}	.026	0.099	-.014	0.118
ASC _{math} → Emotionality _{German}	-.151	0.108	.044	0.083
ASC _{German} → Worry _{math}	.112*	0.055	-.036	0.066
ASC _{German} → Emotionality _{math}	.149*	0.068	.149	0.090
ASC _{German} → Worry _{German}	-.006	0.065	-.234**	0.071
ASC _{German} → Emotionality _{German}	-.273**	0.100	-.458***	0.089
ASC _{general} → Worry _{general}	NA	NA	-.034	0.104
ASC _{general} → Emotionality _{general}	NA	NA	-.215	0.126

ASC = academic self-concept, SE = standard error, NA = not applicable in first-order factor modeling

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

strength of grade-emotionality associations was descriptively higher in the NF model. However, the pattern of results regarding positive cross-domain paths, indicative of dimensional comparison effects, changed in the FOF model (i.e., one path to math worry) versus the NF model (i.e., two paths to emotionality in both domains). In the mediated set of models (Model 2b vs. 4b; Tables 5 and 6), the resulting pattern of significant and nonsignificant direct paths was virtually the same. Concerning the indirect paths, a pronounced change was visible, with one additional significant indirect within-domain and

three additional indirect cross-domain paths in the NF mediated GI/E model as opposed to the FOF mediated GI/E model. Thus, particularly cross-domain paths (both direct and indirect) changed with respect to the modeling strategy.

General paths

The NF models yielded additional general paths, which we addressed in our ancillary RQ3. In the NF GI/E model (i.e., Model 3b, see Table 4), the math grade was significantly

Table 6 Standardized Indirect Path Coefficients, Standard Errors and Effect Sizes for the First-Order Factor and the Nested Factor GI/E Mediation Model

	First-Order Factor GI/E Mediation Model (Model 2b)			Nested Factor GI/E Mediation Model (Model 4b)		
	β_{ind}	<i>SE</i>	β_{ind}^2	β_{ind}	<i>SE</i>	β_{ind}^2
Indirect within-domain paths						
Grade _{math} → ASC _{math} → Worry _{math}	-.064	0.093	.004	-.039	0.057	.001
Grade _{math} → ASC _{math} → Emotionality _{math}	-.377**	0.112	.142	-.240***	0.068	.058
Grade _{German} → ASC _{German} → Worry _{German}	-.005	0.052	.000	-.156**	0.046	.024
Grade _{German} → ASC _{German} → Emotionality _{German}	-.217**	0.080	.047	-.304***	0.068	.092
Indirect cross-domain paths						
Grade _{math} → ASC _{German} → Worry _{German}	.002	0.023	.000	.153**	0.050	.023
Grade _{math} → ASC _{German} → Emotionality _{German}	.093	0.048	.009	.300***	0.070	.090
Grade _{German} → ASC _{math} → Worry _{math}	.014	0.022	.000	.031	0.046	.001
Grade _{German} → ASC _{math} → Emotionality _{math}	.081	0.042	.007	.188***	0.049	.035
Additional indirect paths						
Grade _{math} → ASC _{general} → Worry _{general}	NA	NA	NA	-.014	0.040	.000
Grade _{math} → ASC _{general} → Emotionality _{general}	NA	NA	NA	-.085	0.049	.007
Grade _{German} → ASC _{general} → Worry _{general}	NA	NA	NA	-.016	0.050	.000
Grade _{German} → ASC _{general} → Emotionality _{general}	NA	NA	NA	-.103	0.059	.010

ASC = academic self-concept, SE = standard error, NA = not applicable in first-order factor modeling

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

negatively related to general worry and emotionality (math grade → general worry, $\beta = -.20, p = .008$ and math grade → general emotionality, $\beta = -.21, p = .006$), whereas the German grade was positively related to general worry (German grade → general worry, $\beta = .14, p = .037$). In the NF mediated GI/E model (i.e., Model 4b, see Table 5), two of these relations were significantly different from zero (i.e., math grade → general worry, $\beta = -.22, p = .004$, and German grade → general worry, $\beta = .19, p = .015$; see Table 5). With regard to the general ASC, significant positive relations with both grades were found (math grade → general ASC, $\beta = .40$, and German grade → general ASC, $\beta = .48, ps < .001$). However, there were neither significant direct paths between general ASC and general worry or emotionality nor significant indirect paths between grades and general facets of TA via general ASC (see Table 6). In conclusion, direct paths to general TA were found with regard to worry, but they were not mediated by the general ASC.

Discussion

The present study combined a general with a domain-specific approach using the NF model to examine the role of social and dimensional comparison effects on the formation of two facets of TA (i.e., worry and emotionality)

in two different domains (i.e., math and verbal) within the GI/E model. The overarching aim was to apply an NF modeling strategy to the GI/E model to control for general proportions of TA within domain-specific TA, ultimately purifying domain-specific relations to academic achievement and self-concept. We investigated these relations in NF models and also contrasted them against relations identified with conventional modeling strategies that do not consider hierarchical construct structures (i.e., FOF models). In doing so, we examined domain-specific achievement-based relations—the core of the GI/E model—while controlling for general components. To the best of our knowledge, this is the first study to do so with regard to social and dimensional comparison effects on facets of TA.

First-Order Factor (FOF) Models: A Replication

First, when employing the FOF modeling strategy, we were able to replicate prior findings to a large extent (RQ1). Specifically, previous studies had reported social comparison effects on both facets in the math and verbal domains and dimensional comparison effects on both facets in math only (Arens et al., 2017; Schilling et al., 2005; see Marsh, 1988, who also found dimensional comparison effects on an English TA measure that did not differentiate worry and emotionality, and van der Westhuizen et al., 2022, who found dimensional

comparison effects of French achievement on math TA, and of French [German] achievement on German [French] TA, again using TA measures that did not differentiate worry and emotionality). In the present study, we found social comparison effects on both facets in math and on emotionality in German. In addition, we replicated the dimensional comparison effects on worry in math. The effect on math emotionality did not reach statistical significance in our case but was almost identical in its effect size (Arens et al., 2017) such that the lack of significance was likely a matter of statistical power. TA in German (which was the native language for most students) did not seem to be subject to dimensional comparisons in either study, possibly due to the fact that self-perceptions in the verbal domain are not as restricted to the school context as in math domains. Accordingly, students have more sources of self-evaluation (other than academic math achievement) that impact their verbal-domain TA (Arens et al., 2017).

Concerning the role of domain-specific math and German ASCs as mediators of the achievement-anxiety relations, in the FOF model, we were only able to replicate a considerably smaller proportion of relations as compared with Arens et al. (2017), who reported full mediations on both facets in both domains, and Schilling et al. (2005), who reported full mediations on both facets in German and partial mediations on both facets in math. In the present study, we observed only within-domain mediations on emotionality in both domains. With regard to worry, we did not find significant indirect paths at all, even though the large overlap between the cognitive facet of worry and ASC has been discussed (Arens et al., 2017). However, our observed effect sizes were in the expected direction.

Nested Factor (NF) Models: An Extension

Second, we successfully applied the NF framework to the GI/E model, clearly showing GI/E-hypothesized relations in both direct and mediated NF models (RQ2). We demonstrated social comparison effects on all facets except for worry in German (analogous to social comparison effects found in the FOF model). However, in contrast to the FOF model, we found dimensional comparison effects on emotionality in both domains, suggesting that the emotionality component (as opposed to worry) is more susceptible to dimensional comparisons when general TA is controlled for. In all comparisons between the FOF and NF model results, it is crucial to keep in mind that the interpretation of the domain-specific factors varies according to the modeling strategy, with domain-specific factors in the NF model representing residual variance that is not explained by the general factor (Arens et al., 2021). When we examined the factor correlations, it became clear that the domain-specific worry factors were more strongly related to each other in the FOF model than the domain-specific emotionality factors were. Indeed, when we employed the NF model, the correlation

between the domain-specific worry factors remained significantly different from zero (after the general worry factor was included), whereas the correlation between the domain-specific emotionality factors did not differ from zero (after the general emotionality factor was included). General factors in the NF model are defined by their unique items, that is, their reference domain (Eid et al., 2017). The significantly positive relation between the domain-specific worry factors thus indicated substantial common variance even after general school-specific worry was controlled for. The remaining relation between the math and German worry factors that is not tied to the school context, suggests lower school-specificity for general worry as opposed to general emotionality. On the one hand, this still significant correlation between the math and German worry factors might describe worry cognitions that are tied specifically to math and German (and that are independent to school in general) as two core school subjects (e.g., high emphasis placed on these subjects or higher amount of weekly lessons). On the other hand, it might also be possible that worry entails more general components not restricted to evaluative situations within the school context but rather school-unrelated other life domains (e.g., generalized worry). In line with this reasoning, Hock et al. (2022) showed that the stable, trait manifestation of worry was prevalent in situations that are generally perceived as threatening and aversive even when a state scale is administered that assesses worry with the instruction “*right now, at this very moment*” using latent state-trait analyses. Emotionality could be more tied to evaluative situations in the school context also due to its temporal proximity to the evaluative situation, whereas worry might also occur days or weeks prior to the situation during exam or test preparations (e.g., Schilling et al., 2005), thus possibly mixing with other worries unrelated to the school context. Furthermore, worry that is associated with future-directed “What if”-type of questions (e.g., “What if I fail in this exam?”) may collapse with “Why”-type of questions typical for rumination and directed to the past (e.g., “Why did I fail in past exams?”, “Why am I such a failure?”) in the course of processing an upcoming exam, highlighting the time-overarching character of worry compared to emotionality (Renner et al., 2018). To conclude, further research is needed to clarify the psychological meaning of this significant correlation.

With regard to the NF mediated model, considerable changes were evident compared with the FOF mediated model—particularly concerning indirect cross-domain paths (i.e., dimensional comparison effects). Three out of four indirect within-domain and three out of four indirect cross-domain paths reached statistical significance (in contrast to a total of two out of eight paths in the FOF model). Specifically, both TA facets in German were mediated by German ASC, and emotionality in math was mediated by math ASC, both within and across domains. To understand

why dimensional comparisons in particular are affected by the modeling strategy, one has to keep in mind that dimensional comparisons are internal comparisons across domains (i.e., students comparing their own abilities across domains). By applying the NF model, domain-specific manifestations are purified, ultimately meaning that they truly reflect domain-specific manifestations (i.e., students perceiving different levels of TA across domains) and not a mixture of domain-specific and general components (i.e., students perceiving themselves to be generally more or less anxious than others across school domains). In other words, the removal of confounded (i.e., domain-specific and general) variance enabled the detection of intraindividual (dimensional) relations.

Third and finally, we found significant relations in the NF models with regard to general TA (RQ3). Both facets were negatively predicted by the math grade, whereas general worry was positively predicted by the German grade. One reason for this finding might be that the portion of worry previously attributed to math worry is more dominant in the general worry factor than the portion previously attributed to German worry. In other words, the salience of worry attributed to math for general worry might be higher than that of worry attributed to German. This difference would explain negative [positive] relations with math [German] grades. Indeed, factor loadings of items loading on their domain-specific worry factor were descriptively lower for math than for German after the general factor was included.

No indirect path was visible when general ASC was considered as a mediator (i.e., no paths between math and German grades and general worry and emotionality mediated by general ASC). The concept of domain-specific mediations of achievement-anxiety relations does not seem to translate to general relations. It might be the case that the (respectively negative and positive) effects of math and German grades on general worry through general ASC (which captures the variance shared between math and German ASCs, which are, in turn, positively related to math and German grades, respectively) was canceled out due to relations that went in opposite directions. Such an occurrence emphasizes the caution researchers need to exercise when interpreting the general factor and its relations. Yet, one advantage of the NF model (e.g., in contrast to the higher-order factor model) includes the invariance of the meaning of the general factor due to its ties to the reference domain (Eid et al., 2017). Also, neither the FOF nor the NF model considers item cross-loadings. Given these rather strict requirements, the good fit to the data is all the more convincing. The NF modeling strategy and its implications have been examined in contrast to other modeling strategies with regard to a number of psychological constructs outside of TA and ASC, for instance in the individual clinical assessment of depressive symptoms (Heinrich et al., 2020) and attention

deficit hyperactivity disorder (Eid, 2020), highlighting its importance in more applied settings.

Limitations and Future Research

Our study has some limitations. First, we employed cross-sectional data and thereby cannot draw conclusions about causality. Yet, we chose to refer to social and dimensional comparison *effects* to remain in line with prior research on the GI/E model. Longitudinal and experimental studies that were designed to infer causality have supported GI/E-based assumptions for the ASC (see Niepel et al., 2014; Wolff et al., 2021). Further, our sample was limited to ninth- and 10th-grade students from the highest ability track in Germany. In order to improve the generalizability of our results, further research is needed across different age groups, school tracks, and countries. School-track specific differences in achievement-anxiety relations have previously been identified when comparing the highest ability track to other school tracks (Penk et al., 2014). Another study found that achievement-anxiety relations differed across grade levels with the strongest negative relations in the middle (sixth through eighth) grades and the lowest negative relations in the higher (ninth through 12th) grades (von der Embse et al., 2018), where our sample was located. A recent meta-analysis identified small to moderate, statistically significant negative achievement-anxiety relations in math across 747 effect sizes, also identifying grade level as one moderator of the strength of these relations (Barroso et al., 2021).

Finally, the present findings are restricted to the math and one verbal (i.e., German as language of instruction) domain, such that further research incorporating multiple other domains is warranted. If researchers are particularly interested in (cross-domain) dimensional comparison effects, the inclusion of other domains is recommendable. When assuming a continuum with math and verbal subjects as contrary endpoints, the perceived subject similarity is thought to moderate dimensional comparisons, such that they can even be found to work in the opposite direction (i.e., so-called assimilation effects as opposed to contrast effects; Möller & Marsh, 2013). Yet, a recent meta-analysis on the GI/E model with ASC as the outcome variable did not find such assimilation effects across 505 data sets (Möller et al., 2020). Employing an NF model to purify domain-specific relations considerably advances insights into dimensional comparisons. In addition, if multiple domains from the math-verbal continuum are included, the NF model enables closer examinations of contrast and assimilation effects and their occurrences when extracting the variance that is shared across domain-specific manifestations. For instance, nonsignificant relations between math grades and German TA might be revealed in the NF model as a result of positive relations between math grades and pure German TA in

combination with negative relations between math grades and the portion of German TA that is positively correlated with math TA (i.e., general TA).

Implications and Conclusion

In their everyday school lives, students encounter various peers whose abilities in domain-specific domains serve as references (i.e., social comparisons) as well as various different domains in which a student's own abilities serve as a reference (i.e., dimensional comparisons) that shape students' socio-affective experiences and perceptions (e.g., TA, ASC). In the present study, we used NF models to consider (a) general worry and emotionality levels to identify social and dimensional comparison effects on purely domain-specific worry and emotionality and (b) general ASC levels to examine mediation effects of social and dimensional comparison effects through purely domain-specific ASCs.

Our approach facilitated the detection of dimensional comparison effects and differential worry and emotionality characteristics when controlling for general components in NF models—both directly and indirectly via domain-specific ASCs. Our findings thus have several implications. This study combines conceptual considerations (i.e., TA and ASC as domain-specific and hierarchical constructs on the one hand and the interest in domain-specific relations in the GI/E model on the other hand) with methodological considerations (i.e., the NF model as adequate representation of hierarchical constructs). In this study, we therefore argue for matching methodological approaches to conceptual ideas, and thus provide new directions for future research within the GI/E model. One example for the potential incremental value of using NF models in testing GI/E relations is the controversy on assimilation effects in dimensional comparisons (Möller et al., 2020) by suggesting modeling strategy as potential moderator of dimensional comparisons. In this article, we present both the result patterns yielded by the FOF and the NF models such that the effect of modeling strategy on content-related relations is highlighted. Further, the implementation of the NF model offers new insights on the proportions of general and domain-specific components within constructs (e.g., by comparing correlations among domain-specific components before and after the inclusion of a general, overarching factor).

Practically, examining the interplay of social and dimensional comparisons on the formation of the two TA facets worry and emotionality is of interest given TA's undesirability. Achievement feedback to students (e.g., in the form of school grades) is a common occurrence in daily school life. Thus, achievement-based comparison processes within- and across domains could be of interest in applied educational contexts where raising student, teacher or parent awareness for dimensional comparisons might buffer their detrimental impact (i.e., students performing lower in subject A than in another subject B might develop higher TA in subject A

if they make dimensional comparisons). Wolff and Möller (2021) demonstrated that minimal interventions may lower such negative influences of dimensional comparison effects. One advantage of combining the NF modeling approach with the GI/E framework is the opportunity to evaluate achievement-TA associations more precisely with regard to general or domain-specific manifestations. Similarly, in an applied setting, TA interventions could be evaluated with regard to their effectiveness concerning general or domain-specific TA.

To conclude, the application of NF models to GI/E models matches conceptual and methodological considerations and offers novel insights on the impact of general TA and ASC levels on domain-specific social and dimensional comparisons.

Funding Open Access funding enabled and organized by Projekt DEAL. This research was funded by a grant from the Luxembourg National Research Fund (FNR) to Christoph Niepel (C16/SC/11333571). All statements expressed in this article are the authors' and do not reflect the official opinions or policies of the authors' host affiliations or any of the supporting institutions.

Data availability Data cannot be made available because of data protection concerns. Readers interested in the data can contact the first author by email (irma.talic@unibw.de).

Declarations

Ethical Approval All procedures were approved by the local ethics review panel of the University of Luxembourg and by all involved education authorities.

Ethical Standards and Informed Consent All procedures in the current study were performed in compliance with ethical standards. Participation was voluntary and written parental informed consent was obtained for all participating students.

Conflict of Interest The authors declare no conflicts of interest.

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

- Ahmed, W., Minnaert, A., Kuyper, H., & van der Werf, G. (2012). Reciprocal relationships between math self-concept and math anxiety. *Learning and Individual Differences*, 22(3), 385–389. <https://doi.org/10.1016/j.lindif.2011.12.004>

- American Psychological Association. (2020). *Publication manual of the American Psychological Association 2020: the official guide to APA style (7th ed.)*. American Psychological Association.
- Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*(3), 411–423. <https://doi.org/10.1037/0033-2909.103.3.411>
- Arens, A. K., Yeung, A. S., Craven, R. G., & Hasselhorn, M. (2011). The twofold multidimensionality of academic self-concept: Domain specificity and separation between competence and affect components. *Journal of Educational Psychology*, *103*(4), 970–981. <https://doi.org/10.1037/a0025047>
- Arens, A. K., Becker, M., & Möller, J. (2017). Social and dimensional comparisons in math and verbal test anxiety: Within- and cross-domain relations with achievement and the mediating role of academic self-concept. *Contemporary Educational Psychology*, *51*, 240–252. <https://doi.org/10.1016/j.cedpsych.2017.08.005>
- Arens, A. K., Jansen, M., Preckel, F., Schmidt, I., & Brunner, M. (2021). The structure of academic self-concept: A methodological review and empirical illustration of central models. *Review of Educational Research*, *91*(1), 34–72. <https://doi.org/10.3102/0034654320972186>
- Barroso, C., Ganley, C. M., McGraw, A. L., Geer, E. A., Hart, S. A., & Daucourt, M. C. (2021). A meta-analysis of the relation between math anxiety and math achievement. *Psychological Bulletin*, *147*(2), 134–168. <https://doi.org/10.1037/bul0000307>
- Brunner, M., Keller, U., Hornung, C., Reichert, M., & Martin, R. (2009). The cross-cultural generalizability of a new structural model of academic self-concepts. *Learning and Individual Differences*, *19*(4), 387–403. <https://doi.org/10.1016/j.lindif.2008.11.008>
- Brunner, M., Keller, U., Dierendonck, C., Reichert, M., Ugen, S., Fischbach, A., & Martin, R. (2010). The structure of academic self-concepts revisited: The nested Marsh/Shavelson model. *Journal of Educational Psychology*, *102*(4), 964–981. <https://doi.org/10.1037/a0019644>
- Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology*, *27*(2), 270–295. <https://doi.org/10.1006/ceps.2001.1094>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Erlbaum.
- Devine, A., Fawcett, K., Szücs, D., & Dowker, A. (2012). Gender differences in mathematics anxiety and the relation to mathematics performance while controlling for test anxiety. *Behavioral and Brain Functions*, *8*(33), 1–9. <https://doi.org/10.1186/1744-9081-8-33>
- Dickhäuser, O., & Plenter, I. (2005). „Letztes Halbjahr stand ich zwei“: Zur Akkuratheit selbst berichteter Noten [On the accuracy of self-reported school marks]. *Zeitschrift Für Pädagogische Psychologie [german Journal of Educational Psychology]*, *19*(4), 219–224. <https://doi.org/10.1024/1010-0652.19.4.219>
- Dörendahl, J., Scherer, R., Greiff, S., Martin, R., & Niepel, C. (2021). Dimensional comparisons in the formation of domain specific achievement goals. *Motivation Science*, *7*(3), 306–318. <https://doi.org/10.1037/mot0000203>
- Eid, M. (2020). Multi-faceted constructs in abnormal psychology: Implications of the bifactor S-1 model for individual clinical assessment. *Journal of Abnormal Child Psychology*, *48*(7), 895–900. <https://doi.org/10.1007/s10802-020-00624-9>
- Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychological Methods*, *22*(3), 541–562. <https://doi.org/10.1037/met0000083>
- Franzen, P., Arens, A. K., Greiff, S., & Niepel, C. (2022). Student profiles of self-concept and interest in four domains: A latent transition analysis. *Learning and Individual Differences*, *95*(5), 102–139. <https://doi.org/10.1016/j.lindif.2022.102139>
- 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>
- Gogol, K., Brunner, M., Preckel, F., Goetz, T., & Martin, R. (2016). Developmental dynamics of general and school-subject-domain-specific components of academic self-concept, academic interest, and academic anxiety. *Frontiers in Psychology*, *7*, 356. <https://doi.org/10.3389/fpsyg.2016.00356>
- Gogol, K., Brunner, M., Martin, R., Preckel, F., & Goetz, T. (2017). Affect and motivation within and between school subjects: Development and validation of an integrative structural model of academic self-concept, interest, and anxiety. *Contemporary Educational Psychology*, *49*, 46–65. <https://doi.org/10.1016/j.cedpsych.2016.11.003>
- Gustafsson, J. E., & Balke, G. (1993). General and domain-specific abilities as predictors of school achievement. *Multivariate Behavioral Research*, *28*(4), 407–434. https://doi.org/10.1207/s15327906mbr2804_2
- Hausen, J. E., Möller, J., Greiff, S., & Niepel, C. (2022). Students’ personality and state academic self-concept: Predicting differences in mean level and within-person variability in everyday school life. *Journal of Educational Psychology*, *114*(6), 1394–10411. <https://doi.org/10.1037/edu0000760>
- Heinrich, M., Zagorscak, P., Eid, M., & Knaevelsrud, C. (2020). Giving G a meaning: An application of the bifactor-(S-1) approach to realize a more symptom-oriented modeling of the Beck depression inventory–II. *Assessment*, *27*(7), 1429–1447. <https://doi.org/10.1177/1073191118803738>
- Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research*, *58*(1), 47–77. <https://doi.org/10.3102/00346543058001047>
- Hodapp, V. (1991). Das Prüfungsängstlichkeitsinventar TAI-G: Eine erweiterte und modifizierte Version mit vier Komponenten [The Test Anxiety Inventory TAI-G: An expanded and modified version with four components]. *Zeitschrift Für Pädagogische Psychologie [german Journal of Educational Psychology]*, *5*(2), 121–130.
- Hodapp, V., Rohrmann, S., & Ringeisen, T. (2011). *Prüfungsangstfragebogen: PAF. [Test Anxiety Questionnaire]*. Hogrefe.
- Hock, M., Renner, K. H., Bergner-Köther, R., & Laux, L. (2022). Separating states and traits in anxiety and depression. [Manuscript submitted for publication]
- Hoyle, R. H., & Gottfredson, N. C. (2015). Sample size considerations in prevention research applications of multilevel modeling and structural equation modeling. *Prevention Science*, *16*(7), 987–996. <https://doi.org/10.1007/s11121-014-0489-8>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Kaplan, D. (2009). *Structural equation modeling: Foundations and extensions* (2nd ed.). Sage.
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). The Guilford Press.
- Lachowicz, M. J., Preacher, K. J., & Kelley, K. (2018). A novel measure of effect size for mediation analysis. *Psychological Methods*, *23*(2), 244–261. <https://doi.org/10.1037/met0000165>
- Liebert, R. M., & Morris, L. W. (1967). Cognitive and emotional components of test anxiety: A distinction and some initial data. *Psychological Reports*, *20*(3), 975–978. <https://doi.org/10.2466/pr0.1967.20.3.975>
- Marsh, H. W. (1988). The content domain-specificity of math and English anxieties: The high school and beyond study. *Anxiety Research*, *1*(2), 137–149. <https://doi.org/10.1080/10615808808248226>

- Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal*, 23(1), 129–149. <https://doi.org/10.3102/00028312023001129>
- Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, 1(2), 133–163. <https://doi.org/10.1111/j.1745-6916.2006.00010.x>
- Marsh, H. W., Relich, J. D., & Smith, I. D. (1983). Self-concept: The construct validity of interpretations based upon the SDQ. *Journal of Personality and Social Psychology*, 45(1), 173–187. <https://doi.org/10.1037/0022-3514.45.1.173>
- Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2006). Integration of multidimensional self-concept and core personality constructs: Construct validation and relations to well-being and achievement. *Journal of Personality*, 74(2), 403–456. <https://doi.org/10.1111/j.1467-6494.2005.00380.x>
- Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* (8th ed.). Muthén & Muthén.
- Möller, J., & Marsh, H. W. (2013). Dimensional comparison theory. *Psychological Review*, 120(3), 544–560. <https://doi.org/10.1037/a0032459>
- Möller, J., Müller-Kalthoff, H., Helm, F., Nagy, N., & Marsh, H. W. (2016). The generalized internal/external frame of reference model: An extension to dimensional comparison theory. *Frontline Learning Research*, 4(2), 1–11. <https://doi.org/10.14786/flr.v4i2.169>
- 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>
- Morris, L. W., Davis, M. A., & Hutchings, C. H. (1981). Cognitive and emotional components of anxiety: Literature review and a revised worry–emotionality scale. *Journal of Educational Psychology*, 73(4), 541–555. <https://doi.org/10.1037/0022-0663.73.4.541>
- Niepel, C., Marsh, H. W., Guo, J., Pekrun, R., & Möller, J. (2022). Revealing dynamic relations between mathematics self-concept and perceived achievement from lesson to lesson: An experience-sampling study. *Journal of Educational Psychology*, 114(6), 1380–1393. <https://doi.org/10.1037/edu0000716>
- Niepel, C., Brunner, M., & Preckel, F. (2014). The longitudinal interplay of students' academic self-concepts and achievements within and across domains: Replicating and extending the reciprocal internal/external frame of reference model. *Journal of Educational Psychology*, 106(4), 1170–1191. <https://doi.org/10.1037/a0036307>
- Penk, C., Pöhlmann, C., & Roppelt, A. (2014). The role of test-taking motivation for students' performance in low-stakes assessments: An investigation of school-track-specific differences. *Large-Scale Assessments in Education*, 2(1), 1–17. <https://doi.org/10.1186/s40536-014-0005-4>
- Renner, K. H., Hock, M., Bergner-Köther, R., & Laux, L. (2018). Differentiating anxiety and depression: the State-Trait Anxiety-Depression Inventory. *Cognition & Emotion*, 32(7), 1409–1423. <https://doi.org/10.1080/02699931.2016.1266306>
- Rost, D. H., & Sparfeldt, J. R. (2002). Facetten des schulischen Selbstkonzepts: Ein Verfahren zur Messung des differentiellen Selbstkonzepts schulischer Leistungen und Fähigkeiten (DISK-Gitter) [Facets of academic self-concept: Development of a self-concept grid: Psychometric properties and some validity data]. *Diagnostica*, 48(3), 130–140. <https://doi.org/10.1026/0012-1924.48.3.130>
- Schilling, S.R., Sparfeldt, J.R., & John, M. (2005). Besser in Mathe, besorgter in Deutsch? – Beziehungen zwischen Schulleistungen, Selbstkonzepten und Prüfungssängsten im Rahmen des I/E-Modells. In S.R. Schilling, J.R. Sparfeldt & C. Pruisken (Hrsg.), *Aktuelle Aspekte pädagogisch-psychologischer Forschung*. Detlef H. Rost zum 60. Geburtstag (S. 159–178). Waxmann.
- Schneider, R., Sparfeldt, J. R., Niepel, C., Buch, S. R., & Rost, D. H. (2022). Measurement invariance of test anxiety across four school subjects. *European Journal of Psychological Assessment*, 38(5), 356–364. <https://doi.org/10.1027/1015-5759/a000676>
- Sparfeldt, J. R., Schilling, S. R., Rost, D. H., Stelzl, I., & Peipert, D. (2005). Leistungsängstlichkeit: Facetten, Fächer, Fachfacetten? Zur Trennbarkeit nach Angstfacette und Inhaltsbereich [Test anxiety: The relevance of anxiety facets as well as school subjects]. *Zeitschrift für Pädagogische Psychologie [German Journal of Educational Psychology]*, 19(4), 225–236. <https://doi.org/10.1024/1010-0652.19.4.225>
- Sparfeldt, J. R., Buch, S. R., Rost, D. H., & Lehmann, G. (2008). Akkuratess selbstberichteter Zensuren [The accuracy of self-reported grades in school]. *Psychologie in Erziehung und Unterricht*, 55, 68–75.
- Sparfeldt, J. R., Rost, D. H., Baumeister, U. M., & Christ, O. (2013). Test anxiety in written and oral examinations. *Learning and Individual Differences*, 24, 198–203. <https://doi.org/10.1016/j.lindif.2012.12.010>
- Sticca, F., Goetz, T., Bieg, M., Hall, N. C., Eberle, F., & Haag, L. (2017). Examining the accuracy of students' self-reported academic grades from a correlational and a discrepancy perspective: Evidence from a longitudinal study. *PLoS One*, 12(11), e0187367. <https://doi.org/10.1371/journal.pone.0187367>
- Streblo, L. (2004). *Bezugsrahmen und Selbstkonzeptgenese [Frame of reference and genesis of self-concept]*. Waxmann.
- Steinmayr, R., Crede, J., McElvany, N., & Wirthwein, L. (2016). Subjective well-being, test anxiety, academic achievement: Testing for reciprocal effects. *Frontiers in Psychology*, 6, 1994. <https://doi.org/10.3389/fpsyg.2015.01994>
- van der Westhuizen, L., Arens, A. K., Greiff, S., Fischbach, A., & Niepel, C. (2022). The generalized internal/external frame of reference model with academic self-concepts, interests, and anxieties in students from different language backgrounds. *Contemporary Educational Psychology*, 68, 102037. <https://doi.org/10.1016/j.cedpsych.2021.102037>
- von der Embse, N., Jester, D., Roy, D., & Post, J. (2018). Test anxiety effects, predictors, and correlates: A 30-year meta-analytic review. *Journal of Affective Disorders*, 227, 483–493. <https://doi.org/10.1016/j.jad.2017.11.048>
- von der Embse, N., Barterian, J., & Segool, N. (2013). Test anxiety interventions for children and adolescents: A systematic review of treatment studies from 2000–2010. *Psychology in the Schools*, 50(1), 57–71. <https://doi.org/10.1002/pits.21660>
- 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., & Möller, J. (2021). Dimensional comparison theory: Minimal intervention affects strength of dimensional comparison effects. *Journal of Experimental Education*, 89(4), 625–642. <https://doi.org/10.1080/00220973.2020.1843128>
- Zeidner, M. (1998). *Test anxiety: The state of the art*. Springer.
- Zeidner, M. (2020). Test anxiety. In B. J. Carducci, C. S. Nave, A. Di Fabio, D. H. Saklofske, & C. Stough (Eds.), *The Wiley encyclopedia of personality and individual differences: Personality processes and individual differences* (Vol. III, pp. 445–449). John Wiley & Sons.

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