Arens, A. Katrin; Morin, Alexandre J. S.; Watermann, Rainer

Relations between classroom disciplinary problems and student motivation. Achievement as a potential mediator?

formal und inhaltlich überarbeitete Version der Originalveröffentlichung in:
formally and content revised edition of the original source in:
Learning and instruction 39 (2015), S. 184-193

Relations between classroom disciplinary problems and student motivation: Achievement as a potential mediator?

A. Katrin Arens*, German Institute for International Educational Research, Germany

Alexandre J. S. Morin*, Institute for Positive Psychology and Education, Australian Catholic University, Australia

Rainer Watermann, Freie Universität Berlin, Germany

* Authors (AKA and AJSM) contributed equally to this paper and should both be considered first author.

Acknowledgements

Involvement of the second author in the preparation of this paper was made possible by grants from the Australian Research Council (DP130102713; DP140101559).

This is the prepublication version of the following manuscript:

Arens, A.K., Morin, A.J.S., & Watermann, R. (2015). Relations between classroom disciplinary problems and student motivation: Achievement as a potential mediator? *Learning and Instruction, 39*, 184–193. http://doi.org/10.1016/j.learninstruc.2015.07.001

This paper is not the copy of record and may not exactly replicate the authoritative document published in *Learning and Instruction.*

Abstract

This study examined the relation between classroom disciplinary problems in language classes, student achievement, and three facets of student motivation: competence self-perceptions, test anxiety, and engagement. The analyses were conducted with the German sample from the Progress in International Reading Literacy Study (PIRLS) 2006 ($N = 7,899$). The results demonstrated that discipline problems are directly and negatively related to achievement and to all motivation constructs considered. In most cases, the relation between classroom disciplinary problems and motivation constructs was mediated by verbal achievement. Boys were found to report more frequent discipline problems in classrooms than girls. This study contributes to research by assessing the impact of classroom disciplinary problems using doubly latent multilevel structural equation models in order to properly disaggregate effects occurring at the student, versus classroom level.

Keywords: classroom management; motivation; achievement; mediation; gender

Classroom management encompasses actions taken by the teacher to maintain order and maximize on-task time (Evertson & Weinstein, 2006; Kunter, Baumert, & Köller, 2007). Research supports the role of effective classroom management as a key determinant of learning and achievement (Hattie, 2009; Seidel & Shavelson, 2007). Fewer studies have looked at the relations between classroom management and student motivation, yet they confirm the beneficent effects of effective classroom management. For example, Piwowar, Thiel, and Ophardt (2013) evaluated the effectiveness of classroom management training for secondary school teachers. Students of participating teachers showed an increase in their class engagement compared to students taught by control teachers.

Classroom management encompasses different facets that potentially share differential relations with students' outcomes (e.g., Seidel & Shavelson, 2007). For example, the TARGET framework describes six instructional strategies (Task, Authority, Recognition, Grouping, Evaluation, and Time) that have been shown to facilitate the adoption of a mastery goal structure in classrooms, and to help improve student motivation and achievement (Ames, 1992; Bergsmann, Lüftenegger, Jöstl, Schober, & Spiel, 2013; Urdan, 2004). In this study, we focus on classroom disciplinary problems as an indicator of inadequate classroom management. In Seidel and Shavelson' (2007) recent meta-analysis of teacher effects on learning, classroom discipline belonged to an "organization of learning" component of teaching, which was demonstrated to have a substantial impact on student achievement. The related (opposite) construct of classroom chaos has also been shown to have a negative effect on students' achievement in Marsh et al.'s (2012) study. Empirical results suggest that classroom disciplinary problems might also negatively impact student motivation. For instance, rule clarity and teacher monitoring (indicating low levels of classroom disciplinary problems) were found to enhance students' interest in math (Kunter et al., 2007).

In this study, we explore the relations between classroom disciplinary problems and three motivational outcomes (i.e., students' self-perceptions of competence, anxiety, and engagement). Given the consistently found relations between effective classroom management (including classroom discipline) and achievement on the one hand (Seidel & Shavelson, 2007), and between achievement and motivation on the other hand (Hancock, 2001; Marsh & Craven, 2006), we additionally test whether student achievement might mediate the association between classroom disciplinary problems and motivational outcomes. To ensure conceptual clarity in the identification of these relations, we rely on doubly latent multilevel structural equation models (Lüdtke et al., 2008; Lüdtke, Marsh, Robitzsch, & Trautwein, 2011; Marsh et al., 2009) allowing us to locate the effects occurring at the classroom level versus the individual student level.

1. Contextual and Climate Effects

In research on classroom characteristics, it is important to distinguish between contextual and climate effects (Marsh et al., 2012; Morin, Marsh, Nagengast, & Scalas, 2014). Contextual effects are built from meaningful individual characteristics that are aggregated at the classroom level where they take a different meaning. A simple example is the gender composition of classroom. As both individual and classroom components of variables involved in contextual effects are meaningful in their own right, contextual effects need to be controlled for corresponding individual effects (Marsh et al., 2012; Morin et al., 2014). A well-documented contextual effect of direct relevance here is the big-fish-little-pond effect (BFLPE; Marsh, 1987, 2007; Marsh et al., 2008) according to which students' academic self-concept is positively related to students' individual levels of achievement, but negatively related to class-average achievement when controlling for individual achievement. The BFLPE emerges from social comparison processes involved in the construction of students' self-concept (Möller, Pohlmann, Köller, & Marsh, 2009): Students evaluate their own relative standing in the classroom by comparing their own level of achievement with that of their classmates. Realizing that one's own achievement falls short of the average class achievement yields negative effects on students' self-concept, and these negative effects have been shown to be shared among all students composing the classroom.

Climate effects result from the direct assessment of classroom constructs, i.e., when students are directly asked to rate classroom characteristics. Thus, instead of rating their own characteristics (such as their own discipline in the classroom), students are directly asked to evaluate their classroom (such as their perceptions of disciplinary problems occurring in the classroom) and are thus theoretically interchangeable. Climate effects therefore depict students' shared perceptions of their classroom environment. Given that all students are asked to rate the same objective environment rather than to rate themselves in this environment, residual inter-individual differences (occurring at the student level once shared classroom perceptions are controlled) in ratings of classroom climate are a form of measurement error (related to inter-rater agreement in relation to ratings of classroom characteristics) that needs to be controlled in the model. More precisely, we refer to the student-level component of these climate ratings as "residuals" because, in multilevel models, this component reflects inter-individual deviations from the average rating provided by all students forming the classroom. These student-level residuals of classroom climate ratings may still play a substantive role in the interpretation of the results, yet it is critical for the effects of such residual ratings to be interpreted while keeping in mind their nature (i.e., residualized inter-individual differences in perceptions) (Marsh et al., 2012; Morin et al., 2014).

Of direct relevance to this study, when considered at the classroom level, the effects of disciplinary problems represent climate effects, while those of academic achievement represent contextual effects. The above discussion makes it clear that the effects of classroom disciplinary problems should first and foremost be studied at the classroom level and properly represented as climate effects. Nonetheless, some studies have investigated perceptions of the classroom environment and student outcomes at the individual level only (e.g., Greene, Miller, Crowson, Duke, & Akey, 2004; Patrick et al., 2007). Other studies have relied on a more proper multilevel approach. For instance, Frenzel, Pekrun, and Goetz (2007) demonstrated positive relations between students' inter-individual deviations in their perceptions of teaching quality and their enjoyment of mathematics lessons, but negative relations between class-average evaluations of teaching quality and class-average levels of students' enjoyment. Kunter et al. (2007) demonstrated that students' inter-individual deviations in their perceptions of their teachers' rule clarity and monitoring were positively related to changes in individual levels of students' interest in math, while the class-average perceptions of rule clarity and monitoring were unrelated to changes in class-average levels of interest in math. Marsh et al. (2012) examined classroom level relations between social comparison focus and classroom chaos on the one hand, and math achievement and math self-concept on the other hand. Their results demonstrated a negative effect of classroom chaos on math achievement. In turn, classrooms characterized by a higher social comparison focus were found to be characterized by higher levels of students' achievement and self-concept. Finally, Morin et al. (2014) documented the direct effect of a composite factor of classroom climate on math achievement as well as its mediated relation through math self-efficacy.

Taken together, these studies demonstrate the importance of relying on models allowing for a proper disaggregation of the individual, versus classroom, components of these relations. So far, these studies have mainly focused on secondary school students' perceptions related to math classrooms. Therefore, it remains an open question whether similar associations between dimensions of classroom management and student outcomes also exist in language classes and for younger students. Furthermore, although some of these studies have focused on motivation constructs besides achievement, they commonly consider only a single component of motivation at a time. Hence, there is a need to extend these studies to examine a broader range of motivational constructs simultaneously.

2. A Multidimensional Approach to Motivation

By examining the relations between classroom disciplinary problems and three motivation outcomes (self-perceptions of competence, test anxiety and engagement), our study is anchored in current conceptions of motivation as a multidimensional construct (e.g., Murphy & Alexander, 2000). For instance, Martin (2007) differentiates between behavioral and cognitive dimensions of motivation, which can manifest themselves in adaptive or maladaptive forms. The constructs considered in this study fit within this framework with self-perceptions of competence describing an adaptive cognition, test anxiety reflecting a maladaptive cognition, and engagement representing an adaptive behavior.

Engagement describes students' observable behaviors in the classroom, including their active participation. Engagement has been shown to be positively related to student achievement (Alexander, Entwisle, & Dauber, 1993; Fredricks, Blumenfeld, & Paris, 2004) and has also been used as a valuable outcome in its own right (Skinner, Wellborn, & Connell, 1990). Importantly, students' level of engagement has previously been found to be influenced by various facets of classroom management. For instance, Skinner and Belmont (1993) showed that students' perceptions of classroom structure predicted behavioral engagement. Likewise, Patrick, Ryan, and Kaplan (2007) showed that students' perceptions of the classroom social environment (teacher support, promotion of respect and task-related interaction, support) were related to their level of engagement.

Self-perceptions of competence, widely explored in self-concept research (Marsh, 2007; Marsh & Craven, 2006), relate to students' self-evaluations of their own abilities in specific domains. Numerous studies have demonstrated substantial relations between competence self-perceptions and achievement, and showed these relations to be domain-specific (e.g., higher relations between math competence self-perceptions and math achievement than between math competence self-perceptions and verbal achievement) and reciprocal (achievement impacts self-perceptions and vice versa) (Marsh, 2007; Marsh & Craven, 2006). By integrating competence self-perceptions, this study is in line with previous research (Marsh et al., 2012; Morin et al., 2014) investigating the relations between facets of classroom management, achievement, and students' self-perceptions of competence (self-efficacy in

Morin et al., 2014; self-concept in Marsh et al., 2012) as a motivational construct.

Although defined as a maladaptive cognition in Martin's (2007) framework, test anxiety has been investigated from different perspectives (Zeidner, 1998). Liebert and Morris (1967) proposed that test anxiety encompasses worry (a cognitive component including concerns and rumination), and emotion (physiological affective reactions, e.g., sweating or rapid heartbeat). Test anxiety also includes a behavioral component characterized by avoidance behaviors (Elliot & McGregor, 1999). These components have been found to display negative relations with educational outcomes including achievement, learning strategies, and effort (Frenzel et al., 2007; Zeidner, 1998). Classroom characteristics have also been found to contribute to students' test anxiety. For instance, Helmke (1988) noted that classrooms lacking structured instruction practices reinforced the negative effect of test anxiety on achievement. Pintrich, Roeser, and De Groot (1994) showed that students reported higher levels of test anxiety when they perceived fewer opportunities for cooperative learning.

3. Mediation through Achievement

In sum, the motivation constructs considered here all fit within Martin's (2007) multidimensional framework of student motivation, have been demonstrated to vary as a function of classroom management facets (Helmke, 1988; Marsh et al., 2012; Patrick et al., 2007), and have been found to be significantly related to students' achievement (Fredricks et al., 2004; Marsh & Craven, 2006). Thus, the presumed negative association between classroom disciplinary problems and student motivation is likely to reflect, at least in part, the detrimental effects of disciplinary problems on achievement. Indeed, research also supports a negative association between classroom disciplinary problems and students' achievement (Cameron Ponitz, Rimm-Kaufman, Brock, & Nathanson, 2009; Marsh et al., 2012). Many possible mechanisms may explain this relation. For example, teachers' lack of assertiveness in the classroom might lead to difficulties in effectively transferring knowledge and skills to the students, thus impacting achievement. Furthermore, disciplinary problems might reduce the time students spend on learning, also leading to lower levels of student achievement. Lower levels of achievement might then result in lower levels of students' perceptions of their own competence and willingness to engage at school, as well as to higher levels of test anxiety.

The present study thus examines the associations between classroom disciplinary problems and motivation, and tests whether this association is direct or mediated through achievement. The possibility that these associations could occur at both the individual student and classroom levels raises interesting perspectives deserving further investigation (Frenzel et al., 2007; Kunter et al., 2007; Marsh et al., 2012; Morin et al., 2014). Importantly, this possibility reinforces the need to rely on multilevel models to properly disaggregate the portion of the effects occurring at the student versus classroom level.

Students' achievement can be measured by school grades or standardized achievement tests. School grades might be more strongly related to motivation compared to test scores, as they are more obvious to students and entail information about their relative standing in the class (Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2005). Thus, school grades are likely to have a determining impact on motivation. In contrast, due to the lack of corresponding explicit feedback, students are less aware of their relative standing on standardized tests, suggesting that test scores may have a more limited impact on motivational outcomes.

Conversely, school grades are commonly allocated by the teacher using the class as a frame of reference (i.e., "grading on a curve"; for an extensive discussion and demonstration, see Marsh, et al., 2014). Thus, school grades tend to be distributed more similarly across classes than standardized achievement test scores, with the highest grade assigned to the relatively best students within the class and the lowest grade to the poorest achieving students within the class. This tendency makes school grades harder to compare across classes and schools. Standardized achievement tests therefore tend to provide a more reliable indicator of students' absolute levels of achievement. As such, standardized achievement test scores are likely to be more strongly affected than school grades by classroom disciplinary problems that affect the quality and duration of classroom learning experiences. Given the assumed differential relations of the two achievement indicators (school grades vs. standardized achievement test scores) to motivational outcomes (higher relations with school grades) and discipline problems (higher relations with standardized achievement test scores), it is worthwhile to test the presumed mediated relations across both achievement indicators.

4. The Present Study

This study examines the relations between classroom disciplinary problems, achievement, and motivation. In doing so, we consider direct as well as indirect relations mediated through achievement. All of these relations are estimated both at the classroom and student level to properly disentangle the individual versus classroom components of these relations. This study tests the following hypotheses:

Hypothesis 1 (H1). Classroom levels of disciplinary problems will negatively predict classroom levels of achievement, competence self-perceptions and engagement, and positively predict classroom levels of test anxiety.

Hypothesis 2 (H2). The effects of classroom disciplinary problems on classroom levels of achievement will be more pronounced for standardized test scores than for school grades.

Hypothesis 3 (H3). In line with the BFLPE (Marsh, 1987, 2007), classroom levels of achievement will negatively predict classroom levels of competence self-perceptions and engagement, and positively predict classroom levels of test anxiety.

Hypothesis 4 (H4). Individual levels of achievement will positively predict individual levels of competence self-perceptions and engagement, and negatively predict individual levels of test anxiety.

Hypothesis 5 (H5). At both levels, the effects of achievement on motivational outcomes will be more pronounced for school grades than for standardized achievement test scores.

Hypothesis 6 (H6). At the classroom level, the effects of classroom disciplinary problems on motivational outcomes will be significantly mediated by achievement.

We leave as open research question whether inter-individual residualized ratings of classroom disciplinary problems will significantly relate to achievement and motivation. Finally, we also explore whether boys and girls differ in their perceptions of classroom disciplinary problems as an additional Research Question (RQ1). Little research has been conducted on gender differences in the perception of classroom characteristics although there is some indication that girls and boys differ in their classroom experiences (e.g., Gentry, Gable, & Rizza, 2002). Therefore, more research seems to be needed in this area.

5. Method

5.1 Sample

This study uses the German data from the Progress in International Reading Literacy Study (PIRLS) 2006 (for further information, see Bos et al., 2007; Mullis, Martin, Kennedy, & Foy, 2007), made available by the Research Data Centre at the Institute for Educational Quality Improvement (Berlin, Germany). This data set includes 7,899 German elementary school students (4,051 boys; 51.3%) from 405 different fourth grade classes (M_{age} = 10.46 years; SD = 0.51).

5.2 Instruments

5.2.1 Classroom Disciplinary Problems. Five items from the student survey were used to assess disciplinary problems. Students had to report how often the teachers from their language (German) classes had difficulties maintaining discipline and structure in the classroom (α = .78; e.g., "Our teacher has to wait a long time until the class is quiet") on a 4-point scale (1=every lesson, 2=in most of the lessons, 3=in a few lessons, 4=never). The scale was reversed-coded so that higher values indicated more frequent disciplinary problems.

5.2.2 Academic Achievement. Verbal achievement (the same domain for which the students reported on the occurrence of classroom disciplinary problems) was first measured by the PIRLS standardized reading achievement test (composite reliability coefficient for Germany: KR-20 = .86; Mullis et al., 2007). PIRLS relied on a multi-matrix design so that each individual student received only a subset of texts (out of a pool of ten). Students were asked to read the texts silently and to answer multiple-choice (four response options) and open-ended questions about the texts. PIRLS 2006 used an item response theory approach to obtain comparable test scores for each student, and provides a set of five plausible values of reading achievement for each student. These values are randomly selected from a distribution of achievement scores that approximates the student's true ability (Mullis et al., 2007). All analyses including reading achievement were conducted separately for each of the five plausible values and properly aggregated afterwards (Little & Rubin, 2002).

Verbal achievement was also assessed using teacher-assigned school grades in German. In Germany, school grades range from 1 as the best grade to 6 as the lowest grade. Grades 5 and 6, which represent deficient and inadequate accomplishments, are seldom used in elementary school. Thus, the German PIRLS 2006 data set combines grades 4, 5, and 6 to one category (4 and lower), and thus contains only four categories of school grades: "1" (excellent), "2" (good), "3" (satisfactory), and "4 and lower" (poor). All analyses were conducted with reversed-coded grades so that higher values represent higher achievement.

5.2.3 Competence Self-Perceptions. In PIRLS 2006, competence self-perceptions related to school in general are assessed by 17 items (e.g., "I often fail in tests"), rated on a 4-point scale (1=true to 4=not true). To establish a parsimonious and valid measurement model for competence self-perceptions which could be subsequently incorporated into the complex multilevel model, we retained the five items with the highest factor loadings (α = .82). Higher scores on this scale indicate higher levels of competence self-perceptions.

5.2.4 Test Anxiety. Three items (α = .81; e.g., "When the teacher announces a test, I feel nervous") were used to assess test anxiety. These items were rated on a 4-point scale (1=true to 4=not true), and reversed-coded so that higher values represented higher levels of anxiety.

5.2.5 Engagement. Five items ($\alpha = .80$; e.g., "I often do not feel like playing an active part in German lessons") were used to assess students' active engagement in language (German) lessons. These items were rated on a 4-point scale (1=true to 4=not true), and reversed-coded so that higher values represented higher levels of engagement.

5.3 Analyses

Analyses were conducted using Mplus 7.11 robust maximum likelihood (MLR) estimator (Muthén & Muthén, 2012). Full Information Maximum Likelihood estimation (FIML; Enders, 2010) was used to handle the missing data present at the item level: 15.78% for classroom disciplinary problems, 10.91% for competence self-perceptions, 11.27% for test anxiety, 11.18 % for engagement, and 3.76% for school grades. There were no missing values on the plausible values for the achievement test.

All models are doubly latent multilevel structural equation models (ML-SEM, also labeled MSEM; for a technical presentation, see Lüdtke, et al., 2008, 2011; Marsh et al., 2009; for a practical introduction, see Marsh et al., 2012; Morin et al., 2014). These models are called *doubly latent* because they provide a control for sampling error (based on the inter-rater agreement between students forming a classroom) when forming classroom-level aggregates, and for measurement error in estimating constructs at the item-level.

Classroom disciplinary problems were modeled at the classroom level (L2), while inter-individual residualized differences in perceptions of classroom disciplinary problems were modeled at the student level (L1). Achievement, competence self-perceptions, engagement, and test anxiety were modeled at both levels. L2 effects of classroom disciplinary problems on achievement and motivation were modeled as climate effects, whereas L2 effects of achievement on motivation constructs were modeled as contextual effects. As noted by Marsh et al. (2009, 2012), the proper representation of contextual effects involve a grand-mean centering of the variables (versus group-mean centering for climate effects). However, the Mplus implementation of doubly latent ML-SEM model relies on an implicit group-mean centering at L1 of variables included at both levels to achieve of proper disaggregation of L1 and L2 effects. To obtain proper estimates of contextual effects, the group-mean centered results were converted to their grand-mean centered-equivalent following the procedures presented in Marsh et al. (2012) and Morin et al. (2014). We provide a more extensive discussion of centering issues in the online supplements. Finally, gender (0=female; 1=male) was integrated as a L1 predictor. The full ML-SEM model tested in this study is presented in Figure 1. Given the complexity of the models, separate models were estimated including either standardized achievement test scores or school grades.

Although key relations were modelled at both levels, we assume that the effects of classroom disciplinary problems will be mainly located at L2, whereas the effects of achievement will be mainly located at L1. Higher levels of disciplinary problems at L2 are expected to decrease L2 levels of achievement, which are in turn expected to decrease L2 levels of motivation. This assumption is based on the previously described BFLPE (Marsh, 2007), according to which negative relations are expected between class-average levels of achievement and students' self-concept (competence self-perceptions). In this study, we expect BFLPE to extend to all additional motivation constructs considered. In turn, higher L1 levels of achievement are expected to increase L1 levels of motivation.

We first estimated a multilevel confirmatory factor analytic (ML-CFA) model where all constructs were assessed from their items, at both levels, using latent aggregation to create L2 variables. To facilitate interpretations and limit non-essential multicollinearity, all variables were standardized prior to estimation (Marsh et al., 2012; Morin et al., 2014). The ML-CFA were first estimated freely at both levels and re-estimated by constraining factor loadings to be invariant across levels, which helps increase the stability and accuracy of ML-SEM models and ensures that the constructs are comparable across levels (e.g., Lüdtke et al., 2011; Morin et al., 2014). Starting from this ML-CFA model of metric invariance, we then estimated the a priori ML-SEM models illustrated in Figure 1.

Goodness of fit was assessed with the robust $\chi 2$ test statistic, Root Mean Square Error of Approximation (RMSEA), the Tucker-Lewis Index (TLI), and the Comparative Fit Index (CFI). Typical cut-off scores taken to respectively reflect excellent and adequate fit to the data were used: (i) CFI and TLI $\geq .95$ and $\geq .90$; (ii) RMSEA $\leq .06$ and $\leq .08$ (Hu, & Bentler, 1999; Marsh, Hau, & Wen, 2004; Morin et al., 2014). We report unstandardized and standardized regression coefficients, as well as effect size indicators, which are interpreted as in multiple regression or SEM (for details on how to obtain proper standardized and effects sizes estimates, see Marsh et al., 2012; Morin et al., 2014). The relative magnitude of properly standardized coefficients can be directly compared across levels. Indirect effects between classroom disciplinary problems and motivation constructs as mediated by achievement were also calculated as the product of the two components paths[1].

6. Results

6.1 Preliminary Analyses

A critical assumption of ML-SEM is the presence of variability at L2, which is assessed with the intraclass correlation coefficient (ICC1). ICC1 should ideally be close to or higher than .1, but are seldom larger then .3 (Lüdtke et al., 2008, 2011). In this study, ICC1 values are satisfactory for classroom disciplinary problems (.15), competence self-perceptions (.13), school grades (.11), and achievement test scores (.21), but low for test anxiety (.05) and engagement (.03). Lower levels of L2 variability for the motivation constructs are consistent with our expectation that the relations between achievement and motivation are likely to occur more substantially at L1. Interestingly, ICC1 is lower for school grades than achievement scores, consistent with the "grading-on-a-curve" phenomenon (Marsh et al., 2014).

In ML-SEM, it is also important to assess the agreement among students in their ratings of

[1] Tests of statistical significance for indirect effects are better estimated through bootstrap confidence intervals through alternative methods (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002). Unfortunately, these tests are not available with doubly latent ML-SEM so we report significance tests calculated using Mplus MODEL CONSTRAINT command.

the L2 construct (i.e., inter-reliability). This is typically assessed with the ICC2 indicator, which is interpreted in line with other reliability measures (Lüdtke, Robitzsch, Trautwein, & Kunter, 2009; Marsh et al., 2012). Low ICC2 values are not a problem per se in doubly latent ML-SEM model as this source of measurement error is explicitly controlled through the latent aggregation process (Marsh et al., 2012; Morin et al., 2014). ICC2 values are acceptable for classroom disciplinary problems (.77), competence self-perceptions (.75), school grades (.71), and achievement test scores (.84), but lower for test anxiety (.49) and engagement (.41) – reinforcing the importance of relying on latent aggregation in the estimation of the models.

Finally, it is important to verify the adequacy of the a priori ML-CFA measurement models, as well the composite reliability of the latent constructs (e.g., Morin et al., 2014). The results from the preliminary ML-CFA are reported in Table 1. The results showed that these models provide an adequate fit to the data, and that factor loadings are invariant across levels. The specific results for these models can be consulted in the online supplements. From these results, we calculated composite reliability coefficients at L1 and L2 using McDonald's (1970) omega (ω) coefficient. Although these coefficients are all satisfactory, they still show imperfect reliability, reinforcing the need to rely on doubly latent models providing a control for this form of measurement error: classroom disciplinary problems, ω_{L1}= .75, ω_{L2}= .97; competence self-perceptions, ω_{L1}= .82, ω_{L2}= .95; test anxiety, ω_{L1}= .81, ω_{L2}= .95; and engagement, ω_{L1}= .80, ω_{L2}= .89.

6.2 Final Models

6.2.1 Standardized Achievement Test Scores. The goodness of fit indices and parameter estimates for the complete ML-SEM model based on standardized achievement test scores are respectively reported in Tables 1 and 2. These results indicate an adequate fit to the data. Starting at the classroom level (L2), and consistent with H1, the results show that classrooms with higher levels of disciplinary problems tend to present lower levels of achievement and motivation (i.e., lower levels of engagement and competence self-perceptions, and higher levels of test anxiety). The effects of disciplinary problems appear more pronounced for achievement than for motivational constructs. Finally, higher levels of classroom achievement are also related to lower levels of engagement and competence self-perceptions, consistent with the BFLPE and H3, but not to higher levels of test anxiety. Finally, consistent with H6, the L2 indirect effects of classroom disciplinary problems on motivational constructs, as mediated by achievement, are significant and substantial for engagement (.10, $p < .01$) and competence self-perceptions (.20, $p < .01$), but not test anxiety (-.03, *ns*).

At the individual (L1) level, the results show that students' residualized perceptions of classroom disciplinary problems present negative associations with achievement and motivation constructs. In addition, supporting H4, students with higher levels of achievement tend to present higher levels of engagement and competence self-perceptions, and lower levels of test anxiety. Standardized coefficients (properly standardized in relation to the total variance) show that the relations between achievement and students' outcomes are more pronounced at L1 than L2, in accordance with our expectations. Furthermore, the indirect effects of residualized perceptions of classroom disciplinary problems on motivational constructs, as mediated by achievement, are significant for engagement ($\beta = -.04$, $p < .01$), competence self-perceptions ($\beta = -.13$, $p < .01$), and test anxiety ($\beta = -.09$, $p < .01$). Finally, when compared to girls, boys report higher levels of residualized perceptions of their classroom disciplinary problems and competence self-perceptions, but lower levels of achievement, test anxiety, and engagement.

6.2.2 School Grades. The goodness of fit indices and parameter estimates for the complete ML-SEM model based on school grades are respectively reported in Tables 1 and 3. With few exceptions, these results replicate those obtained for standardized achievement test scores, and similarly support H1, H3, H4 and H6[2]. For this reason, we mainly focus on areas where results substantively differ across models. First, and supporting H2, the L2 effects of classroom disciplinary problems on achievement appear more pronounced for standardized achievement test scores (β = -.35) than school grades (β = -.09). Second, and supporting H5, the effects of achievement on motivational outcomes appear more pronounced for school grades (L2: for test anxiety, engagement, and competence self-perceptions, respectively β = .09, -.16, -.18; L1: β = -.25, .34, .45) than standardized achievement test scores (L2: β = .01, -.08, -.11; L1: β = -.20, .24, .33). It is noteworthy that the L2 relation between school grades and test anxiety is significant, whereas the same relation is not significant for standardized achievement tests scores.

7. Discussion

This study examined the relations between disciplinary problems in language classes and three dimensions of students' motivation (i.e., competence self-perceptions, test anxiety, and engagement). In addition to the examination of direct relations between these constructs, we also tested whether these relations were mediated through achievement. In doing so, we contrasted two indicators of achievement (school grades and standardized achievement test scores).

To properly distinguish effects located at the individual and classroom levels, while ensuring proper control for measurement errors in the assessment of the constructs and in the aggregation of individual ratings into L2 constructs, this study relied on doubly latent ML-SEM (Marsh et al., 2012; Morin et al., 2014). Therefore, the effects of L2 classroom disciplinary problems were properly modeled as climate effects, reflecting the effects of students' shared perceptions of their classroom experiences. In contrast, the effects of classroom levels of achievement were modeled as contextual effects and properly controlled

[2] Supporting H6, all indirect effects are significant at both L2 (engagement: .08; competence self-perceptions: .13; test anxiety: -.08; all $p < .01$) and L1 (engagement: -.05; competence self-perceptions: -.14; test anxiety: -.09 all $p < .01$).

for corresponding effects of individual levels of achievement. The results revealed direct relations between classroom levels of disciplinary problems, and classroom levels of achievement and motivation. Thus, in classes characterized by a higher level of disciplinary problems, students tended to present lower levels of achievement, competence self-perceptions and engagement, and higher levels of test anxiety. Similar relations were observed at the individual level, showing that students perceiving higher levels of disciplinary problems tended to present lower levels of achievement and motivation.

Most of the L1 and L2 relations between classroom disciplinary problems (or residualized inter-individual perceptions of these problems) and motivational outcomes proved to be mediated through students' achievement. The only non-significant indirect relation was between classroom levels of disciplinary problems, standardized achievement test scores, and test anxiety. However, this indirect relation was significant when school grades were modeled as the mediator. Our analyses therefore revealed that classroom disciplinary problems are both directly and indirectly associated with students' motivational outcomes, and that this relation occurs both at the student and classroom levels. This conclusion supports results from previous studies which also demonstrated that various facets of (in)efficient classroom management was significantly related to students' achievement (e.g., Cameron Ponitz et al., 2009; Marsh et al., 2012), and motivation (e.g., Frenzel et al., 2007; Kunter et al., 2007; Patrick et al., 2007; Skinner & Belmont, 1993).

However, whereas Frenzel et al. (2007) report opposite relations between indicators of teaching quality and enjoyment of mathematics at the individual (positive) or classroom (negative) levels, the current study shows that perceptions of classroom disciplinary problems are negatively associated with student motivation at both levels. Similarly, the only previous study of similar relations relying on doubly latent ML-SEM models failed to report significant relations between classroom chaos and students' self-concept (Marsh et al., 2012). In contrast, the current study found significant negative associations between classroom disciplinary problems and students' competence self-perceptions. Possible explanations for these discrepant results may potentially lie in the differences between these studies in the school subject considered (math in these other studies; verbal in this study), age of the sample (elementary in this study vs. secondary in the other studies) and measures of classroom characteristics (classroom chaos, teaching quality, classroom disciplinary problems) or students' outcomes (self-concept, math enjoyment, competence self-perceptions). Clearly, further studies are needed to better document the observed relations by examining how they change when integrating different facets of classroom management including positive (e.g., rule clarity: Kunter et al., 2007) and negative (e.g., disciplinary problems as in this study) manifestations and when considering alternative motivational constructs (e.g., interest: Kunter et al., 2007; goal orientations: Lüftenegger, Van de Schoot, Schober, Finsterwald, & Spiel, 2014). At this stage, it is important to note that even though the three motivational constructs considered here are well-supported by theory and research (Martin, 2007), they do not cover the full spectrum of students' motivation. Given that this study relies on an analysis of a representative sample of the German student population taken from the PIRLS 2006, our selection of constructs had to be partly determined by what was available in this data set.

Although significant indirect relations were observed at both the classroom and individual level, it is important to note that that these relations were of opposite directions due to the change in the direction of the relations between achievement and motivational outcomes at both levels. More precisely, while the relation between classroom levels of academic achievement and motivational outcomes was negative, the relation between individual levels of achievement and motivational outcomes was positive. This finding is fully in line with the BFLPE (Marsh, 1987, 2007). The BFLPE is one of the most extensively replicated effects in educational psychology (Marsh, 2007; Marsh et al., 2008) and originally refers to the differential effect of students' individual versus classroom levels of achievement on students' academic self-concept. However, the current study extends evidence in favor of the BFLPE to test anxiety and engagement (Zeidner & Schleyer, 1999).

Other interesting implications of the current results stem from the comparison of models based on school grades or standardized achievement test scores. Although both series of analyses converged on highly similar results, the magnitude of some relations differed as a function of the retained achievement indicator. First, the relations between classroom disciplinary problems and achievement were more pronounced for standardized achievement test scores than for school grades. This result is in line with the idea that standardized achievement test scores provide a more reliable indicator of students' learning and thus appear to be more sensitive to disruptions of the learning process that might occur in classrooms as a result of disciplinary problems (e.g., Marsh et al., 2014). In contrast, the relations between achievement and motivational outcomes appeared to be more pronounced for school grades than for standardized achievement test scores. This result supports the idea that school grades, due to their greater salience for students in regards to their relative standing within their classes, are likely to have a greater motivational impact (Marsh et al., 2005). Although the ICC1 associated with school grades was lower than the ICC1 associated with standardized achievement test scores, consistent with the grading-on-a-curve phenomenon (Marsh et al., 2014), it is noteworthy that classroom disciplinary problems were still found to be negatively related to average classroom-levels of school grades. This finding shows that teachers' grading practices might not solely depend on the performance of students within single classes but might also be influenced by more general achievement standards.

Finally, our results revealed that boys tended to perceive higher levels of classroom disciplinary problems in classrooms than girls. While one might expect girls to be more sensitive to discipline problems in classrooms, this findings may reflect the fact that boys are more likely to be the source of classroom disciplinary problems (Kaplan, Gheen, & Midgley, 2002; Lahey et al., 2000). Furthermore, our results also replicate previous research results showing that girls tend to present higher levels of test anxiety (Cassady & Johnson, 2002) and engagement (Furrer & Skinner, 2003; Marks, 2000), but lower levels of competence self-perceptions (De Fraine, Van Damme, & Onghena, 2007).

In sum, this study enhances our knowledge on the association between classroom disciplinary problems (as one facet of classroom management), achievement, and motivation. Based on its findings, researchers and practitioners should be aware that disciplinary problems present significant and substantial relations with student achievement and motivation, and that these effects occur at the individual student and classroom levels. Thus, the development of interventions designed to help teachers to rely on efficient, and equitable (to ensure that individual students benefit from it), strategies of classroom management appears to be a valuable avenue for future research. Despite these strengths, the present study has also some shortcomings. First, our study is cross-sectional, suggesting that longitudinal studies are needed to more clearly investigate the direction of the observed relations, and experimental studies are needed to establish causality. Whereas achievement was assumed to mediate the effects of classroom disciplinary problems on student motivation, it is equally possible to conceptualize student motivation as a mediator of the relation between classroom disciplinary problems and student achievement (Bronstein, Ginsburg, & Herrera, 2005; Morin et al., 2014). Other variables beyond achievement might also serve as potential mediators of the relations between classroom disciplinary problems (and classroom management more generally) and motivation and should be considered more thoroughly in future studies. For example, Kunter et al. (2007) demonstrated that the relations between students' perceptions of rule clarity and teacher monitoring and students' interest in math were mediated through students' experience of intrinsic need satisfaction. Similarly, Patrick et al. (2007) showed that students' mastery goal orientations acted as a mediator of the relation between classroom environment and students' engagement. In addition, moderation might also operate in the associations between classroom management, student achievement, and student motivation. For example, Hancock (2001) demonstrated that students diagnosed with high levels of test anxiety displayed lower levels of motivation and achievement when they were exposed to classrooms with a strong evaluation focus but did not differ from students with low levels of test anxiety in less evaluative classrooms. Therefore, a wide array of students' characteristics, which could not be taken into account in the present study, might participate in the way classroom characteristics influence student achievement and motivation. Finally, multiple methods should be applied to assess classroom characteristics so that student reports on classroom management (students' perceptions of disciplinary problems in this study) should be preferably combined with objective observational data (Urdan, 2004).

8. References

Alexander, K.L., Entwisle, D.R., & Dauber, S.L. (1993). First-grade classroom behavior: Its short and long-term consequences for school performance. *Child Development, 64,* 801–814. doi: 10.1111/j.1467-8624.1993.tb02944.x

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84*, 261-271. doi:10.1037/0022-0663.84.3.261

Bergsmann, E., Lüftenegger, M., Jöstl, G., Schober, B., & Spiel, C. (2013). The role of classroom structure in fostering students' school functioning: A comprehensive and application-oriented approach. *Learning and Individual Differences, 26*, 131-138. doi:10.1016/j.lindif.2013.05.005

Bronstein, P., Ginsburg, G.S., & Herrera, I. S. (2005). Parental predictors of motivational orientation in early adolescence: A longitudinal study. *Journal of Youth and Adolescence, 34,* 559-575. doi:10.1007/s10964-005-8946-0

Bos, W., Hornberg, S., Arnold, K.-H., Faust, G., Fried, L., Lankes, E.-M., ... Valtin, R. (2010): Internationale Grundschul-Lese-Untersuchung 2006 (IGLU/PIRLS 2006). Version: 1. IQB – Institut zur Qualitätsentwicklung im Bildungswesen. Datensatz. http://doi.org/10.5159/IQB_IGLU_2006_v1

Cameron Ponitz, C., Rimm-Kaufman, S.E., Brock, L.L., & Nathanson, L. (2009). Early adjustment, gender differences, and classroom organizational climate in first grade. *The Elementary School Journal, 110*, 142-162. doi:10.1086/605470

Cassady, J.C., & Johnson, R.E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology, 27*, 270-295. doi:10.1006/ceps.2001.1094

De Fraine, B., Van Damme, J., & Onghena, P. (2007). A longitudinal analysis of gender differences in academic self-concept and language achievement: A multivariate multilevel latent growth approach. *Contemporary Educational Psychology, 32*, 132-150. doi:10.1016/j.cedpsych.2006.10.005

Elliot, A.J., & McGregor, H.A. (1999). Test anxiety and the hierarchical model of approach and avoidance achievement motivation. J*ournal of Personality and Social Psychology, 76*, 628-644. doi:10.1037/0022-3514.76.4.628

Enders, C.K. (2010). *Applied missing data analysis*. New York: Guilford.

Evertson, C.M., & Weinstein, C.S. (Eds,), (2006). *Handbook of classroom management. Research, practice and contemporary issues*. Mahwah, NJ: Erlbaum.

Fredricks, J.A., Blumenfeld, P.C., & Paris, A.H. (2004). School engagement. Potential of the concept, state of evidence. *Review of Educational Research, 74*, 59-109. doi:10.3102/00346543074001059

Frenzel, A.C., Pekrun, R., & Goetz, T. (2007). Perceived learning environment and students' emotional experiences: A multilevel analysis of mathematics classrooms. *Learning and Instruction, 17*, 478-493. doi:10.1016/j.learninstruc.2007.09.001

Furrer, C., & Skinner, E. (2003). Sense of relatedness as a factor in children's academic engagement and performance. *Journal of Educational Psychology, 95,* 148-162. doi:10.1037/0022-0663.95.1.148

Gentry, M., Gable, R.K., & Rizza, M. G. (2002). Students' perceptions of classroom activities: Are there grade-level and gender differences? *Journal of Educational Psychology, 94*, 539-544. doi:10.1037//0022-0663.94.3.539

Greene, B.A., Miller, R.B., Crowson, H.M., Duke, B.L., & Akey, K.L. (2004). Predicting high school students' cognitive engagement and achievement: Contributions of classroom perceptions and motivation. *Contemporary Educational Psychology, 29*, 462-482. doi:10.1016/j.cedpsych.2004.01.006

Hancock, D.R. (2001). Effects of test anxiety and evaluative threat on students' achievement and motivation. *The Journal of Educational Research, 94*, 284-290. doi:10.1080/00220670109598764

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* London: Routledge.

Helmke, A. (1988). The role of classroom context factors for the achievement-impairing effect of test anxiety. *Anxiety Research, 1*, 37-52. doi:10.1080/10615808808248219

Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55. doi:10.1080/10705519909540118

Kaplan, A., Gheen, M., & Midgley, C. (2002). Classroom goal structure and student disruptive behaviour. *British Journal of Educational Psychology, 72*, 191-211. doi:10.1348/000709902158847

Kunter, M., Baumert, J., & Köller, O. (2007). Effective classroom management and the development of subject-related interest. *Learning and Instruction, 17,* 494-509. doi:10.1016/j.learninstruc.2007.09.002

Lahey, B.B., Schwab-Stone, M., Goodman, S.H., Waldman, I.D., Canino, G., Rathouz, P.J., Miller, T.L., Dennis, K.D., Bird, H.R., & Jensen, P.S. (2000). Age and gender differences in oppositional behavior and conduct problems: A cross-sectional household study of middle childhood and adolescence. *Journal of Abnormal Psychology, 109,* 488-503. doi:10.1037//002I-843X. 109.3.481

Liebert, R.M., & Morris, L.W. (1967). Cognitive and emotional components of test anxiety: A distinction and some initial data. *Psychological Reports, 20*, 975-978.

Little, R., & Rubin, D. (2002). *Statistical analysis with missing data*. New York: John Wiley.

Lüdtke, O., Marsh, H.W, Robitzsch, A., & Trautwein, U. (2011). A 2×2 taxonomy of multilevel latent contextual models: Accuracy-bias tradeoffs in full and partial error-correction models. *Psychological Methods, 16*, 444-467. doi:10.1037/a0024376

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, *13*, 203–229. doi:10.1037/a0012869

Lüdtke, O., Robitzsch, A., Trautwein, U., & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings of classroom and school characteristics in multilevel modeling. *Contemporary Educational Psychology, 34*, 120-131. doi:10.1016/j.cedpsych.2008.12.001

Lüftenegger, M., Van de Schoot, R., Schober, B., Finsterwald, M., & Spiel, C. (2014). Promotion of students' mastery goal orientations: Does TARGET work?. *Educational Psychology, 34*, 451-469. doi:10.1080/01443410.2013.814189

MacKinnon, D.P., Lockwood, C.M., Hoffman, J.M., West, S.G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7*, 83-104. doi:10.1037/1082-989X.7.1.83

Marks, H.M. (2000). Student engagement in instructional activity: Patterns in the elementary, middle, and high school years. *American Educational Research Journal, 37*, 153-184. doi:10.3102/00028312037001153

McDonald, R. P. (1970). The theoretical foundations of princi¬pal factor analysis, canonical factor analysis, and alpha fac¬tor analysis. *British Journal of Mathematical and Statistical Psychology, 23*, 1–21. doi: 10.1111/j.2044-8317.1970.tb00432.x

Marsh, H.W. (1987). The Big-Fish-Little-Pond Effect on academic self-concept. *Journal of*

Educational Psychology, 79, 280-295. doi:10.1037/0022-0663.79.3.280

Marsh, H.W. (2007). *Self-concept theory, measurement and research: The role of self-concept in educational psychology.* Leicester, UK: British Psychological Society.

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. doi:10.1111/j.1745-6916.2006.00010.x

Marsh, H.W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to cutoff values for fit indexes and dangers in overgeneralizing Hu & Bentler's (1999). *Structural Equation Modeling, 11,* 320-341. doi:10.1207/s15328007sem1103_2

Marsh, H.W., Kuyper, H., Seaton, M., Parker, P.D., Morin, A.J.S., Möller, J., & Abduljabbar, A.S. (2014). Dimensional comparison theory: An extension of the internal/external frame of reference effect on academic self-concept formation, *Contemporary Educational Psychology, 39*, 326-341. doi:10.1016/j.cedpsych.2014.08.003

Marsh, H.W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A.J.S., Adbuljabbar, A., & Köller, O. (2012). Classroom climate and contextual effects. Methodological issues in the evaluation of group-level effects. *Educational Psychologist, 47*, 106-124. doi:10.1080/00461520.2012.670488

Marsh, H.W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B.O., & Nagengast, B. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research, 44*, 764-802. doi:10.1080/00273170903333665

Marsh, H.W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K.T., O'Mara, A. J., &Craven, R.G. (2008). The big-fish-little-pond-effect stands up to critical scrutiny: Implications for theory, methodology, and future research. *Educational Psychology Review*, *20*, 319–350. doi:10.1007/s10648-008-9075-6

Marsh, H.W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades and standardized test scores: Reciprocal effects models of causal ordering. *Child Development, 76*, 397-416. doi:10.1111/j.1467-8624.2005.00853.x

Martin, A.J. (2007). Examining a multidimensional model of student motivation and engagement using a construct validation approach. *British Journal of Educational Psychology, 77*, 413-440. doi:10.1348/000709906X118036

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

Morin, A.J.S., Marsh, H.W., & Nagengast, B., & Scalas, L.F. (2014). Doubly latent multilevel analyses of classroom climate: An illustration. *Journal of Experimental Education, 82,* 143-167. doi:10.1080/00220973.2013.769412

Mullis, I.V., Martin, M.O., Kennedy, A.M., & Foy, P. (2007). *PIRLS 2006 international report. IEA's progress in international reading literacy study in primary schools in 40 countries.* Chestnut Hill, MA: International Study Center.

Murphy, P.K., & Alexander, P.A. (2000). A motivated exploration of motivation terminology. *Contemporary Educational Psychology, 25*, 3-53. doi:10.1006/ceps.1999.1019

Muthén, L.K., & Muthén, B.O. (2012). *Mplus User's Guide*. Seventh Edition. Los Angeles, CA: Muthén & Muthén.

Patrick, H., Ryan, A.M., & Kaplan, A. (2007). Early adolescents' perceptions of the classroom social environment, motivational beliefs, and engagement. *Journal of Educational Psychology, 99,* 83-98. doi:10.1037/0022-0663.99.1.83

Pintrich, P.R., Roeser, R.W., & de Groot, E.A.M. (1994). Classroom and individual differences in early adolescents' motivation and self-regulated learning. *Journal of Early Adolescence, 14,* 139-161. doi:10.1177/027243169401400204

Piwowar, V., Thiel, F., & Ophardt, D. (2013). Training inservice teachers' competencies in classroom management. A quasi-experimental study with teachers of secondary schools. *Teaching and Teacher Education, 30*, 1-12. doi:10.1111/j.1745-6916.2006.00010.x

Urdan, T. (2004). Using multiple methods to assess students' perceptions of classroom goal structures. *European Psychologist, 9,* 222-231. doi:10.1027/1016-9040.9.4.222

Seidel, T., & Shavelson, R.J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research, 77*, 454-499. doi:10.3102/0034654307310317

Skinner, E.A., & Belmont, M.J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology, 85*, 571-581. doi:10.1037/0022-0663.85.4.571

Skinner, E.A., Wellborn, J.G., & Connell, J.P. (1990). What it takes to do well in school and whether I've got it: A process model of perceived control and children's engagement and achievement in school. *Journal of Educational Psychology, 82*, 22-32. doi:10.1037/0022-0663.82.1.22

Zeidner, M. (1998). *Test Anxiety: State of the Art*. New York: Plenum.

Zeidner, M., & Schleyer, E. (1999). The big-fish-little-pond effect for academic self-concept, test anxiety, and school grades in gifted children. *Contemporary Educational Psychology, 24*, 305-329. doi:10.1006/ceps.1998.098

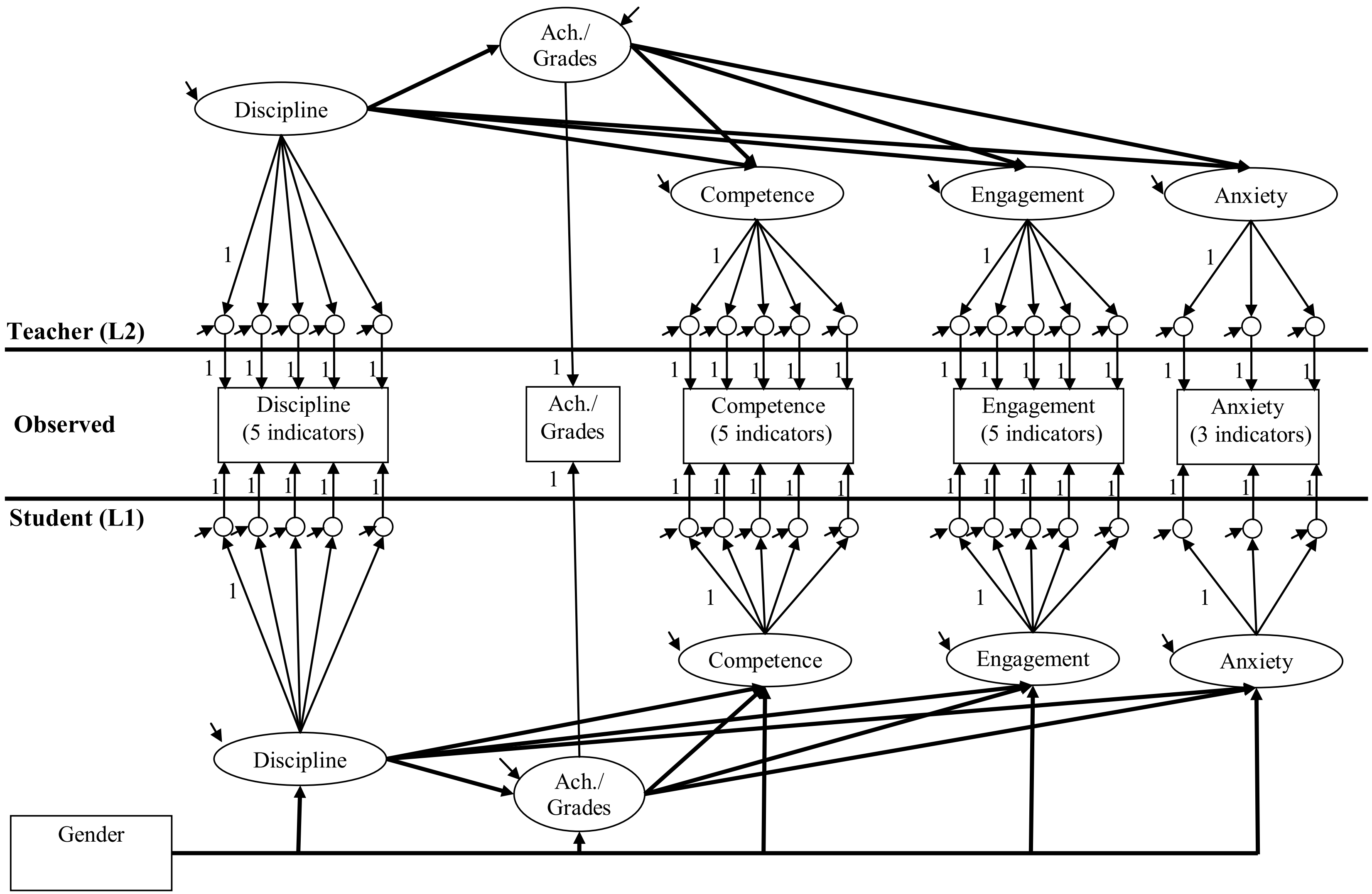

Figure 1. Doubly latent multilevel structural equation (ML-SEM) model tested in the current study. *Note.* For parsimony, factor correlations are not shown in the figure.

Table 1

Goodness-of-fit Indices

	χ^2	df	CFI	TLI	RMSEA	
Standardized Achievement Test Scores						
1	1676.625	300	.970	.964	.024	CFA, free factor loadings
2	1695.027	314	.970	.965	.024	CFA, invariant factor loadings across levels
3	1694.787	314	.970	.965	.024	SEM, predictive model
School Grades						
4	1719.062	300	.970	.963	.024	CFA, free factor loadings
5	1749.474	314	.969	.964	.024	CFA, invariant factor loadings across levels
6	1749.263	314	.969	.964	.024	SEM, predictive model

Note. df = degrees of freedom; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation. SEM = Structural Equation Modeling.

Table 2

Effects from the Multilevel Predictive Models using Standardized Achievement Test Scores as an Achievement Indicator

	Est. (S.E.)	Std. (S.E.)	ES (S.E).
L2 (Classroom level)			
Disciplinary problems → test anxiety	0.153 (0.070)*	0.054 (0.025)*	0.056 (0.026)*
Disciplinary problems → engagement	-0.167 (0.034)**	-0.109 (0.022)**	-0.111 (0.023)**
Disciplinary problems → competence self-perceptions	-0.133 (0.052)*	-0.058 (0.022)*	-0.059 (0.023)*
Disciplinary problems → achievement	-1.140 (0.139)**	-0.346 (0.042)**	-0.414 (0.050)**
Achievement → test anxiety (contextual)	0.029 (0.048)	0.014 (0.024)	0.015 (0.024)
Achievement → engagement (contextual)	-0.091 (0.024)**	-0.082 (0.022)**	-0.084 (0.022)**
Achievement → competence self-perceptions (contextual)	-0.175 (0.037)**	-0.105 (0.023)**	-0.108 (0.024)**
L1 (Student level)			
Disciplinary problems (residualized) → test anxiety	0.389 (0.026)**	0.264 (0.017)**	0.273 (0.018)**
Disciplinary problems (residualized) → engagement	-0.296 (0.016)**	-0.369 (0.020)**	-0.376 (0.021)**
Disciplinary problems (residualized) → competence self-perceptions	-0.427 (0.022)**	-0.353 (0.017)**	-0.362 (0.017)**
Disciplinary problems (residualized) → achievement	-0.478 (0.024)**	-0.277 (0.013)**	-0.331 (0.015)**
Achievement → test anxiety	-0.218 (0.016)**	-0.201 (0.015)**	-0.208 (0.015)**
Achievement → engagement	0.141 (0.009)**	0.239 (0.015)**	0.244 (0.016)**
Achievement → competence self-perceptions	0.295 (0.013)**	0.332 (0.015)**	0.340 (0.015)**
Gender → disciplinary problems	0.104 (0.017)**	0.077 (0.013)**	0.087 (0.015)**
Gender → test anxiety	-0.293 (0.022)**	-0.167 (0.012)**	-0.172 (0.013)**
Gender → engagement	-0.024 (0.011)*	-0.025 (0.011)*	-0.025 (0.012)*
Gender → competence self-perceptions	0.052 (0.016)**	0.036 (0.011)**	0.037 (0.011)**
Gender → achievement	-0.063 (0.023)**	-0.031 (0.011)**	-0.037 (0.013)**

Note. Est. = unstandardized parameter estimate; Est. = unstandardized parameter estimates; S.E.= standard error of the estimate; Std. = standardized parameter estimate; ES = effect size; Gender was coded as a dichotomous variable with 0 = female, 1 = male.
* $p < .05$; ** $p < .01$.

Table 3

Effects from the Multilevel Predictive Models using School Grades as an Achievement Indicator

	Est. (S.E.)	Std. (S.E.)	ES (S.E).
L2 (Classroom level)			
Disciplinary problems → test anxiety	0.290 (0.063)**	0.094 (0.020)**	0.097 (0.021)**
Disciplinary problems → engagement	-0.211 (0.031)**	-0.126 (0.018)**	-0.128 (0.019)**
Disciplinary problems → competence self-perceptions	-0.222 (0.046)**	-0.087 (0.018)**	-0.089 (0.019)**
Disciplinary problems → achievement	-0.328 (0.079)**	-0.093 (0.022)**	-0.099 (0.024)**
Achievement → test anxiety (contextual)	0.241 (0.061)**	0.090 (0.022)**	0.092 (0.023)**
Achievement → engagement (contextual)	-0.228 (0.026)**	-0.156 (0.016)**	-0.158 (0.017)**
Achievement → competence self-perceptions (contextual)	-0.406 (0.046)**	-0.183 (0.020)**	-0.187 (0.020)**
L1 (Student level)			
Disciplinary problems (residualized) → test anxiety	0.368 (0.025)**	0.251 (0.017)**	0.259 (0.018)**
Disciplinary problems (residualized) → engagement	-0.274 (0.016)**	-0.343 (0.019)**	-0.348 (0.020)**
Disciplinary problems (residualized) → competence self-perceptions	-0.389 (0.022)**	-0.322 (0.017)**	-0.328 (0.017)**
Disciplinary problems (residualized) → achievement	-0.488 (0.023)**	-0.291 (0.013)**	-0.310 (0.014)**
Achievement → test anxiety	-0.251 (0.013)**	-0.254 (0.013)**	-0.262 (0.014)**
Achievement → engagement	0.181 (0.008)**	0.335 (0.015)**	0.340 (0.015)**
Achievement → competence self-perceptions	0.368 (0.012)**	0.450 (0.015)**	0.459 (0.015)**
Gender → disciplinary problems	0.102 (0.017)**	0.077 (0.013)**	0.085 (0.015)**
Gender → test anxiety	-0.343 (0.022)**	-0.196 (0.013)**	-0.202 (0.013)**
Gender → engagement	0.014 (0.011)	0.014 (0.011)	0.014 (0.011)
Gender → competence self-perceptions	0.126 (0.016)**	0.087 (0.011)**	0.089 (0.011)**
Gender → achievement	-0.248 (0.023)**	-0.124 (0.012)**	-0.131 (0.012)**

Note. Est. = unstandardized parameter estimate; Est. = unstandardized parameter estimates; S.E.= standard error of the estimate; Std. = standardized parameter estimate; ES = effect size; Gender was coded as a dichotomous variable with 0 = female 1 = male.
* $p < .05$; ** $p < .01$.

Online Supplements for:

Relations between classroom disciplinary problems and student motivation: Achievement as a potential mediator?

Authors' note:

These online technical appendices are to be posted on the journal website and hot-linked to the manuscript. If the journal does not offer this possibility, these materials can alternatively be made available upon request from the first author (we will adjust the in-text reference upon acceptance).

We developed these materials mostly to provide additional technical information and to keep the main manuscript from becoming needlessly long.

Centering and Related Statistical Considerations in the Estimation of Level 2 effects

In multilevel models, it is typical to differentiate between group-mean centering and grand-mean centering (Enders & Tofighi, 2007; Lüdtke, et al., 2008, 2011). The mathematical implementation of doubly latent ML-SEM models in Mplus relies on a default group-mean centering at the student level (L1) and grand-mean centering at the classroom level (L2) for all variables that are included at both levels. In other words, L1 ratings of all variables included at both levels in the models are directly expressed as deviation from the classroom mean. Given this default parametrisation, the effects of the L2 variable are removed from the corresponding L1 variable, but the effect of the L2 variable is not controlled for the L1-effect—although both effects are estimated as independent from one another (Enders & Tofighi, 2007; Marsh et al., 2009, 2012). Although this procedure is appropriate for climate effects as it results is properly disaggregated L1 ratings that simply reflect deviations from class-average ratings, it creates interpretation problems for contextual effects where both L1 and L2 ratings retain meaning in and of themselves (i.e., not in residualized form) (Marsh et al., 2012; Morin et al., 2014). Indeed, for climate effects, students within the same class are asked to rate common L2 constructs, so that the main construct being assessed is naturally located at L2 and independent from the L1 counterpart. More precisely, students' ratings of classroom climate reflect two components: The shared agreement that represents the climate effect (the L2 construct), and the residual L1 deviation from this class average. With group-mean centering, these two components are estimated as independent from one another, which corresponds to the appropriate interpretation of climate effects as reflecting the effect of the L2 variable, not the L1 ratings by individual students (Marsh et al., 2012; Morin et al., 2014). In contrast, for contextual effects, both the L1 and L2 components remain meaningful in and of themselves, so that group-mean centered results are not fully appropriate. Fortunately, it is possible to convert group-mean centered estimates to grand-mean centered estimates of L2 effects by subtracting the L1-effect from the L2-effect. More precisely, an additional parameter representing the difference between the L2 and L1 coefficients can be calculated using the Mplus model constraint function which provides a directly interpretable estimate of contextual effects and tests of statistical significance (Enders & Tofighi, 2007; Marsh et al., 2009, 2012; Morin et al., 2014). This additional parameter provides an estimate of the L2 effect equivalent to that obtained using a grand-mean centering procedure (i.e., resulting in the estimation of partial regression slopes at L2 that are controlled for the influence of the L1 variable). Therefore, this new parameter provides a direct test of whether the L2 contextual variable (representing the class average) really adds something to the effects of the main L1 construct.

Enders, C.K. & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12,* 121-138.doi: 10.1037/1082-989X.12.2.121

Lüdtke, O., Marsh, H.W, Robitzsch, A., & Trautwein, U. (2011). A 2×2 taxonomy of multilevel latent contextual models: Accuracy-bias tradeoffs in full and partial error-correction models. *Psychological Methods, 16*, 444-467. doi:10.1037/a0024376

Lüdtke, O., Marsh, H.W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, *13*, 203-229. doi:10.1037/a0012869

Marsh, H.W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A.J.S., Adbuljabbar, A., & Köller, O. (2012). Classroom climate and contextual effects. Methodological issues in the evaluation of group-level effects. *Educational Psychologist, 47*, 106-124. doi:10.1080/00461520.2012.670488

Marsh, H.W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B.O., & Nagengast, B. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research, 44*, 764-802. doi:10.1080/00273170903333665

Morin, A.J.S., Marsh, H.W., & Nagengast, B., & Scalas, L.F. (2014). Doubly latent multilevel analyses of classroom climate: An illustration. *Journal of Experimental Education, 82,* 143-167. doi:10.1080/00220973.2013.769412

Table S1

Factor Loadings and Item Uniquenesses of the Invariant Multilevel Confirmatory Factor Analytic Model (see Model 2, see Table 1)

	Disciplinary Problems				Test Anxiety				Engagement				Competence Self-perceptions			
	Loadings		Uniquenesses		Loadings		Uniquenesses		Loadings		Uniquenesses		Loadings		Uniquenesses	
	Est. (S.E.)	Std. (S.E.)	Est. (S.E.)	Std. (S.E.)	Est. (S.E.)	Std. (S.E.)	Est. (S.E.)	Std. (S.E.)	Est. (S.E.)	Std. (S.E.)	Est. (S.E.)	Std. (S.E.)	Est. (S.E.)	Std. (S.E.)	Est. (S.E.)	Std. (S.E.)
Level 1																
Indicator 1	1.000	0.635	0.534	0.597	1.000	0.871	0.231	0.242	1.000	0.477	0.750	0.772	1.000	0.713	0.481	0.492
	(0.000)	(0.011)	(0.015)	(0.014)	(0.000)	(0.007)	(0.012)	(0.012)	(0.000)	(0.011)	(0.015)	(0.011)	(0.000)	(0.009)	(0.013)	(0.013)
Indicator 2	0.862	0.534	0.670	0.715	0.755	0.653	0.555	0.573	1.033	0.491	0.743	0.759	0.918	0.653	0.563	0.574
	(0.023)	(0.012)	(0.015)	(0.013)	(0.014)	(0.010)	(0.013)	(0.013)	(0.033)	(0.012)	(0.016)	(0.012)	(0.016)	(0.009)	(0.013)	(0.012)
Indicator 3	0.849	0.525	0.682	0.724	0.887	0.768	0.396	0.410	1.749	0.829	0.307	0.312	1.039	0.741	0.441	0.451
	(0.026)	(0.013)	(0.017)	(0.014)	(0.014)	(0.009)	(0.014)	(0.014)	(0.047)	(0.007)	(0.011)	(0.011)	(0.019)	(0.008)	(0.012)	(0.011)
Indicator 4	1.189	0.761	0.369	0.421					1.414	0.671	0.539	0.550	0.910	0.652	0.557	0.575
	(0.028)	(0.009)	(0.013)	(0.014)					(0.038)	(0.009)	(0.013)	(0.012)	(0.019)	(0.010)	(0.014)	(0.013)
Indicator 5	0.951	0.598	0.584	0.643					1.773	0.841	0.287	0.292	0.982	0.699	0.499	0.511
	(0.023)	(0.011)	(0.015)	(0.013)					(0.047)	(0.007)	(0.012)	(0.012)	(0.019)	(0.008)	(0.012)	(0.012)
Level 2																
Indicator 1	1.000	0.895	0.024	0.200	1.000	0.982	0.002	0.036	1.000	0.531	0.022	0.718	1.000	0.903	0.006	0.184
	(0.000)	(0.024)	(0.005)	(0.043)	(0.000)	(0.020)	(0.002)	(0.040)	(0.000)	(0.062)	(0.005)	(0.066)	(0.000)	(0.044)	(0.003)	(0.080)
Indicator 2	0.862	0.963	0.006	0.072	0.755	0.876	0.009	0.233	1.033	0.638	0.014	0.593	0.918	0.966	0.001	0.067
	(0.023)	(0.024)	(0.004)	(0.046)	(0.014)	(0.040)	(0.003)	(0.069)	(0.033)	(0.076)	(0.004)	(0.096)	(0.016)	(0.039)	(0.002)	(0.076)
Indicator 3	0.849	0.967	0.005	0.065	0.887	0.963	0.003	0.072	1.749	0.998	0.000	0.004	1.039	0.978	0.001	0.043
	(0.026)	(0.023)	(0.004)	(0.045)	(0.014)	(0.029)	(0.003)	(0.055)	(0.047)	(0.002)	(0.000)	(0.003)	(0.019)	(0.030)	(0.002)	(0.059)
Indicator 4	1.189	0.957	0.013	0.084					1.414	0.855	0.006	0.270	0.910	0.752	0.016	0.434
	(0.028)	(0.015)	(0.004)	(0.028)					(0.038)	(0.057)	(0.003)	(0.097)	(0.019)	(0.049)	(0.004)	(0.073)
Indicator 5	0.951	0.888	0.024	0.212					1.773	0.973	0.002	0.053	0.982	0.902	0.005	0.187
	(0.023)	(0.021)	(0.004)	(0.037)					(0.047)	(0.025)	(0.001)	(0.049)	(0.019)	(0.049)	(0.003)	(0.088)

Note. All loadings and uniquenesses are significant ($p < .05$); Est. = unstandardized parameter estimates; S.E.= standard error of the estimate; Std. = standardized parameter estimate.

Table S2

Factor Covariances (below the diagonal), Factor Variances (in the diagonal), and Factor Correlations (above the diagonal) of the Invariant Multilevel Confirmatory Factor Analytic Model (see Model 2, see Table 1)

	Disciplinary Problems	Test Anxiety	Engagement	Competence Self-perceptions	Achievement	Gender
Level 1						
Disciplinary Problems	0.360*	0.333*	-0.467*	-0.481*	-0.335*	0.087*
Test Anxiety	0.170*	0.724*	-0.535*	-0.588*	-0.302*	-0.134*
Engagement	-0.132*	-0.214*	0.221*	0.732*	0.387*	-0.075*
Competence Self-perceptions	-0.203*	-0.353*	0.243*	0.497*	0.481*	-0.019
Achievement	-0.174*	-0.222*	0.157*	0.293*	0.745*	-0.065*
Gender	0.026 *	-0.057*	-0.018*	-0.007	-0.028*	0.250*
Level 2						
Disciplinary Problems	0.098*	0.507*	-0.749*	-0.540*	-0.633*	
Test Anxiety	0.036*	0.052*	-0.594	-0.767*	-0.600*	
Engagement	-0.022*	-0.013*	0.009*	0.887*	0.656*	
Competence Self-perceptions	-0.027*	-0.027*	0.013*	0.025*	0.602*	
Achievement	-0.112*	-0.077*	0.035*	0.053*	0.318*	

Note. * $p < .05$.

Table S3

Factor Loadings and Item Uniquenesses of the Invariant Multilevel Confirmatory Factor Analytic Model (see Model 5, see Table 1)

	Disciplinary Problems				Test Anxiety				Engagement				Competence Self-perceptions			
	Loadings		Uniquenesses		Loadings		Uniquenesses		Loadings		Uniquenesses		Loadings		Uniquenesses	
	Est. (S.E.)	Std. (S.E.)	Est. (S.E.)	Std. (S.E.)	Est. (S.E.)	Std. (S.E.)	Est. (S.E.)	Std. (S.E.)	Est. (S.E.)	Std. (S.E.)	Est. (S.E.)	Std. (S.E.)	Est. (S.E.)	Std. (S.E.)	Est. (S.E.)	Std. (S.E.)
Level 1																
Indicator 1	1.000	0.635	0.533	0.597	1.000	0.870	0.232	0.244	1.000	0.477	0.750	0.773	1.000	0.717	0.474	0.486
	(0.000)	(0.011)	(0.015)	(0.014)	(0.000)	(0.007)	(0.011)	(0.012)	(0.000)	(0.011)	(0.015)	(0.011)	(0.000)	(0.009)	(0.013)	(0.013)
Indicator 2	0.859	0.532	0.672	0.717	0.756	0.653	0.555	0.574	1.032	0.490	0.743	0.760	0.909	0.650	0.566	0.577
	(0.023)	(0.012)	(0.015)	(0.013)	(0.014)	(0.010)	(0.013)	(0.013)	(0.033)	(0.012)	(0.016)	(0.012)	(0.016)	(0.009)	(0.013)	(0.012)
Indicator 3	0.845	0.523	0.684	0.727	0.889	0.769	0.394	0.409	1.751	0.830	0.306	0.312	1.039	0.745	0.435	0.446
	(0.026)	(0.013)	(0.017)	(0.014)	(0.014)	(0.009)	(0.014)	(0.014)	(0.047)	(0.007)	(0.011)	(0.011)	(0.019)	(0.008)	(0.012)	(0.011)
Indicator 4	1.189	0.762	0.368	0.420					1.412	0.670	0.541	0.551	0.898	0.647	0.564	0.582
	(0.028)	(0.009)	(0.013)	(0.014)					(0.038)	(0.009)	(0.013)	(0.012)	(0.019)	(0.010)	(0.014)	(0.012)
Indicator 5	0.950	0.598	0.584	0.643					1.774	0.842	0.286	0.292	0.973	0.697	0.503	0.515
	(0.023)	(0.011)	(0.015)	(0.013)					(0.047)	(0.007)	(0.012)	(0.012)	(0.019)	(0.008)	(0.012)	(0.012)
Level 2																
Indicator 1	1.000	0.877	0.024	0.231	1.000	0.977	0.002	0.046	1.000	0.482	0.022	0.768	1.000	0.874	0.006	0.237
	(0.000)	(0.027)	(0.005)	(0.047)	(0.000)	(0.024)	(0.002)	(0.048)	(0.000)	(0.061)	(0.005)	(0.059)	(0.000)	(0.051)	(0.003)	(0.090)
Indicator 2	0.859	0.953	0.006	0.092	0.756	0.858	0.009	0.264	1.032	0.585	0.014	0.658	0.909	0.950	0.002	0.098
	(0.023)	(0.028)	(0.004)	(0.053)	(0.014)	(0.044)	(0.003)	(0.075)	(0.033)	(0.078)	(0.004)	(0.091)	(0.016)	(0.052)	(0.002)	(0.098)
Indicator 3	0.845	0.955	0.006	0.087	0.889	0.957	0.003	0.084	1.751	0.998	0.000	0.004	1.039	0.973	0.001	0.053
	(0.026)	(0.028)	(0.004)	(0.053)	(0.014)	(0.034)	(0.003)	(0.065)	(0.047)	(0.000)	(0.000)	(0.001)	(0.019)	(0.037)	(0.002)	(0.072)
Indicator 4	1.189	0.953	0.012	0.092					1.412	0.821	0.006	0.325	0.898	0.708	0.016	0.499
	(0.028)	(0.016)	(0.004)	(0.031)					(0.038)	(0.066)	(0.003)	(0.109)	(0.019)	(0.052)	(0.004)	(0.074)
Indicator 5	0.950	0.871	0.023	0.241					1.774	0.962	0.002	0.074	0.973	0.885	0.005	0.217
	(0.023)	(0.023)	(0.004)	(0.040)					(0.047)	(0.033)	(0.002)	(0.064)	(0.019)	(0.059)	(0.003)	(0.104)

Note. All loadings and uniquenesses are significant ($p < .05$); Est. = unstandardized parameter estimates; S.E.= standard error of the estimate; Std. = standardized parameter estimate.

Table S4

Factor Covariances (below the diagonal), Factor Variances (in the diagonal), and Factor Correlations (above the diagonal) of the Invariant Multilevel Confirmatory Factor Analytic Model (see Model 5, see Table 1)

	Disciplinary Problems	Test Anxiety	Engagement	Competence Self-perceptions	Achievement	Gender
Level 1						
Disciplinary Problems	0.360*	0.333*	-0.465*	-0.480*	-0.322*	0.085*
Test Anxiety	0.170*	0.722*	-0.535*	-0.588*	-0.331*	-0.136*
Engagement	-0.131*	-0.213*	0.221*	0.731*	0.473*	-0.073*
Competence Self-perceptions	-0.204*	-0.354*	0.243*	0.502*	0.581*	-0.016
Achievement	-0.182*	-0.265*	0.210*	0.389*	0.890*	-0.158*
Gender	0.026*	-0.058*	-0.017*	-0.006	-0.074*	0.250*
Level 2						
Disciplinary Problems	0.081*	0.401*	-0.682*	-0.428*	-0.276*	
Test Anxiety	0.024*	0.044*	-0.503*	-0.725*	-0.126	
Engagement	-0.016*	-0.009*	0.007*	0.853*	0.006	
Competence Self-perceptions	-0.017*	-0.021*	0.010*	0.019*	0.032	
Achievement	-0.027*	-0.009	0.000	0.002	0.115*	

Note. * $p < .05$.