Relations between teachers' emotional exhaustion and students' educational outcomes

A. Katrin Arens*, German Institute for International Educational Research
Alexandre J. S. Morin*, Institute for Positive Psychology and Education, Australian Catholic University

* 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).

Correspondence regarding this investigation should be sent to A. Katrin Arens, German Institute for International Educational Research, Centre for Research on Education and Human Development and Centre for Research on Individual Development and Adaptive Education of Children at Risk (IDeA), Schloßstr. 29, D-60486 Frankfurt am Main, Germany, Email: arens@dipf.de

This is the prepublication version of the following manuscript:

Arens, A.K., & Morin, A.J.S. (2016). Relations between teachers' emotional exhaustion and students' educational outcomes. *Journal of Educational Psychology, 108*, 800–813. http://doi.org/10.1037/edu0000105

This paper is not the copy of record and may not exactly replicate the authoritative document published in *Journal of Educational Psychology*.

Abstract

Studies investigating the effects of emotional exhaustion among teachers have primarily focused on its relations with teacher-related outcome variables but little research has been done for examining its relations with student outcomes. Therefore, this study examines the relations between teachers' emotional exhaustion and educational outcomes among students. Students' educational outcomes considered here cover a wide range of cognitive (i.e., achievement in terms of school grades and standardized achievement test scores) and non-cognitive (competence self-perceptions, school satisfaction, and perceptions of teacher support) outcomes. The analyses are based on the PIRLS 2006 German data including 380 teachers and 7899 4th grade students. The results demonstrated direct negative relations between teachers' emotional exhaustion and the class average of students' school grades, standardized achievement test scores, school satisfaction, and perceptions of teacher support, but not competence self-perceptions. At the individual student level, the results showed significant relations between non-cognitive outcomes and academic achievement.

Keywords. Teacher emotional exhaustion; student achievement; academic self-concept; school satisfaction; perceptions of teacher support

Because of its high prevalence and important practical implications, teacher burnout has been extensively studied within the field of educational psychology. This research has demonstrated clear associations between teacher burnout and the emergence of multiple negative consequences for the teachers themselves, such as lower levels of job satisfaction (e.g., Skaalvik & Skaalvik, 2010), poorer self-reported physical health (Hakanen, Bakker, & Schaufeli, 2006), lower self-efficacy (Brouwers & Tomic, 2000; Evers, Brouwers, & Tomic, 2002; Schwarzer & Hallum, 2008; Skaalvik & Skaalvik, 2007, 2010), and higher levels of intentions to leave the teaching profession (Leung & Lee, 2006). In contrast, the possible relation between teacher burnout and student outcomes has yet to be systematically investigated. For this purpose, the present study aims to investigate the relation between teachers' levels of emotional exhaustion – a core component of the burnout syndrome – and students' educational outcomes. In this context, this study adopts a broad perspective by considering both cognitive (i.e., achievement) and non-cognitive (i.e., competence self-perceptions, perceptions of teacher support, and school satisfaction) educational outcomes.

Teachers' Emotional Exhaustion

Burnout is defined as a syndrome consisting of emotional exhaustion, depersonalization, and reduced personal accomplishment (Maslach & Jackson, 1981; Maslach, Jackson, & Leiter, 1996; Maslach, Schaufeli, & Leiter, 2001). Emotional exhaustion is defined as feelings of emotional overstrain and reduced emotional resources. Depersonalization refers to feelings of cognitive distance, indifference, or cynicism toward one's job service recipients. Reduced personal accomplishment involves feelings of inefficacy to successfully complete work demands and a missing sense of personal accomplishment. Although any employee might develop burnout, it was originally and is still mainly studied in professions characterized by frequent, and highly demanding, social interactions, such as employees working in the health care, social services, or educational systems (Maslach et al., 2001). Previous research showed that these professional groups, in particular teachers which are the focus of the current study (e.g., Friedman, 2000; Hakanen et al., 2006; Vandenberghe & Huberman, 1999), presented a particularly high risk of burnout development.

Among the three components of burnout, emotional exhaustion appears to be the most critical. It has often been positioned to emerge first in the development of burnout, thus in turn leading to higher levels of depersonalization and reduced feelings of personal accomplishment (e.g., Byrne, 1994; Maslach et al., 1996, 2001). Further substantiating the central role of emotional exhaustion among the three components of burnout, emotional exhaustion has been found to present the most consistent relations with outcomes out of all burnout components (Cordes & Dougherty, 1993; Demerouti, Bakker, Nachreiner, & Schaufeli, 2001; Green, Walkey, & Taylor, 1991; Halbesleben & Bowler, 2007). In particular, emotional exhaustion appears to be the burnout component that best predicts (in terms of explaining the most variance) decreases in work performance (e.g., Wright, & Bonett, 1993). Accordingly, Wright and Cropanzano (1998) demonstrated that emotional exhaustion was related to job performance and voluntary employee turnover when controlling for positive and negative affectivity. Moreover, even alternative conceptualizations of burnout include a dimension representing psychological/emotional exhaustion (e.g., the job demands-resources model of burnout proposed by Demerouti et al., 20011) while other authors (e.g., Moore, 2000) note the importance of the emotional exhaustion construct in and of itself.

The proposed central role of emotional exhaustion in the burnout syndrome seems to translate to teachers' burnout. For instance, in Chan's (2006) study of Hong Kong secondary school teachers, emotional exhaustion emerged as a significant predictor of depersonalization, which in turn predicted lower levels of personal accomplishment. Furthermore, ratings of emotional exhaustion more strongly contribute to a total score of teachers' burnout than other burnout components (Grayson & Alvarez, 2008). Consequently, many studies on teacher burnout have focused solely on emotional exhaustion (Dicke et al., 2015; Klusmann, Kunter, Trautwein, Lüdtke, & Baumert, 2008a, 2008b).

In terms of consequences for the teachers themselves, burnout has been found to be significantly associated with lower levels of motivation toward teaching (Hakanen et al., 2006; Schaufeli & Salanova, 2007) and greater levels of doubts regarding the ability to teach effectively (Dicke et al., 2015; Evers et al., 2002; Schwarzer & Hallum, 2008; Skaalvik & Skaalvik, 2007, 2010). However, little research has been conducted with respect to the possible consequences of teacher burnout on student outcomes. The present study focuses on this understudied area through a systematic investigation of the relation between teachers' levels of emotional exhaustion (as the core component of the burnout syndrome) and student outcomes. Given the known negative relations between teacher burnout and their own levels of work performance (e.g., Cropanzano, Rupp, & Byrne, 2003), motivation (e.g., Hakanen et al., 2006), and self-efficacy (e.g., Dicke et al., 2015), it seems reasonable to assume that emotionally exhausted teachers may lack the necessary resources to provide an adequate and supportive learning environment to students, in turn potentially leading to lower educational outcomes among students. In this regard, this study examines the relations between teachers' levels of emotional exhaustion and students' cognitive outcomes in terms of achievement and between teachers' levels of emotional exhaustion and three non-cognitive outcomes, namely students' competence self-perceptions, perceptions of teacher support, and school satisfaction.

Teachers' Emotional Exhaustion and Student Outcomes

Relations with Student Achievement

In any study focusing on the impact of schooling on students' educational outcomes, achievement represents a key outcome to consider. As noted above, teacher self-efficacy has been previously shown to be negatively impacted by teacher burnout (Dicke et al., 2015; Evers et al., 2002; Schwarzer & Hallum, 2008; Skaalvik & Skaalvik, 2007, 2010). Teacher self-efficacy itself has been found to be positively related to students' levels of achievement (Goddard, Hoy, & Woolfolk Hoy, 2000; Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998). Moreover, research has shown that teachers suffering from high levels of emotional exhaustion rely on less efficient classroom management strategies and report more difficulties in coping with students' disruptive classroom behaviors (e.g., Chang, 2009; Evers, Tomic, & Brouwers, 2004). Furthermore, Klusmann et al. (2008b) revealed that teachers characterized by low levels of occupational engagement and resilience (referred to as resigned [R-type] teachers; Schaarschmidt, Kieschke, & Fischer, 1999) tend to present higher levels of emotional exhaustion and to be less favorably rated by their students regarding their instructional tempo, levels of cognitive activation, and provision of personal support relative to teachers characterized by relatively high levels of engagement and resilience (referred to as healthy-ambitious [H-type] teachers). Altogether, these findings suggest that emotionally exhausted teachers may tend to provide a less favorable instructional context to their students, which in turn could lead to lower levels of achievement among students.

Student achievement is commonly measured either by standardized achievement test scores or school grades. School grades are allocated by the teacher, and have been shown to be influenced by a "grading-on-a-curve" phenomenon whereby the class itself acts as a frame of reference for the distribution of grades (Marsh et al., 2014). For this reason, school grades tend to be distributed similarly from one class to the other, with the highest grades assigned to the relatively best performing students and the lowest grades to the poorest achieving students within the class. This tendency makes school grades harder to compare across classes and schools whereas standardized achievement test scores provide a more accurate indicator of students' absolute levels of achievement which can be more readily compared across classes, schools, and even nations. For this reason, major international studies of educational outcomes, such as the Progress in International Reading Literacy Study (PIRLS;

timssandpirls.bc.edu/), the Trends in International Mathematics and Science Study (TIMSS; timssandpirls.bc.edu/), and the Programme for International Student Assessment (PISA; www.oecd.org/pisa/) invest considerable resources to provide standardized achievement test scores that are comparable across countries, schools, and classes.

In the present study, it is likely that teachers' level of emotional exhaustion may present differential associations with both achievement indicators. Teachers' emotional exhaustion might lead to lower levels of their self-efficacy for teaching and instructional quality (Chang, 2009; Klusmann et al., 2008b) in turn diminishing students' mastery of course content and pre-specified achievement standards as assessed in standardized achievement tests. The allocation of school grades, in contrast, has been found to be subject to the "grading-on-a-curve" phenomenon leading to similar distributions of grades across classes irrespective of students' absolute levels of attainments of pre-specified curriculum standards. Moreover, school grades do not only reflect students' absolute levels of learning, but are also influenced by their behavior and effort in class (Brookhart, 1993, 1994; McMillan, Myran, & Workman, 2002; Zimmermann, Schütte, Taskinen, & Köller, 2013). Finally, the allocation of school grades is also subject to teachers' grading and evaluation practices, such as the leniency, strictness, or arbitrariness. Given these multiple possible sources of influence on school grades, it is harder to clearly anticipate the nature of the effects of teachers' emotional exhaustion on school grades. Thus, this study systematically assesses the extent to which the findings regarding the relation between teachers' emotional exhaustion and student achievement would replicate across both achievement indicators.

Relations with Non-cognitive Outcomes

Competence self-perceptions. Klusmann et al. (2008b) showed that students taught by teachers presenting higher levels of emotional exhaustion significantly differed from students taught by teachers presenting lower levels of emotional exhaustion in their experience of autonomy and self-perceived competence. The latter construct, namely students' competence self-perceptions (also referred to as academic self-concept) has often been reported to be positively associated with a wide variety of educational outcomes including effort, coursework selection, positive affect, and academic achievement (Marsh, 2007; Marsh & Craven, 2006). Competence self-perceptions are known to emerge from a combination of social (i.e., contrasting their own level of achievement with that of others) and dimensional (i.e., contrasting their own levels of achievement across subject areas) comparison processes (Möller, Pohlmann, Köller, & Marsh, 2009). As noted above, emotionally exhausted teachers may provide students with less favorable classroom environments in terms of instructional quality and classroom management, leading students to struggle more to achieve mastery of the course content. This might trigger a dimensional comparison process leading to a decrease of their competence self-perceptions in the specific area taught by the exhausted teacher compared to other subject areas for which students may experience a more optimal classroom environment provided by the teachers. Competence self-perceptions are also known to depend upon the availability of proper feedback processes (Craven, Marsh, & Debus, 1991). As with other instructional practices (Chang, 2009; Kunter et al., 2008b), the provision of adequate feedback is likely to be affected by teachers' levels of emotional exhaustion, which may contribute to low levels of students' competence self-perceptions.

Perceptions of teacher support. Teachers suffering from emotional exhaustion may also lack the adequate emotional resources to establish a positive relationship with their students and to provide adequate levels of support. Teacher support encompasses a broad array of instructional strategies aiming to facilitate students' academic progress through the provision of personalized monitoring, advice, and encouragement covering not only academic objectives but also socio-emotional needs (e.g., Klem & Connel, 2004; Patrick & Ryan, 2005; Reddy, Rhodes, & Mulhall, 2003; Ryan & Patrick, 2001) and has been shown to be positively related with student achievement (e.g., Furrer & Skinner, 2003; Hamre & Pianta, 2001, 2005; Hughes, Luo, Kwok, & Loyd, 2008; Spilt, Hughes, Wu, & Kwok, 2012). Research shows that emotionally exhausted teachers tend to withdraw from positive student–teacher relations and to provide inadequate support to their students (Chang, 2009). Accordingly, Hoglund, Kingle, and Hosan (2015) showed teacher burnout to be a negative predictor of growth in quality of student–teacher relations. Similarly, Hamre, Pianta, Downer, and Mashburn (2008) reported a positive relation between teacher depression and student–teacher conflicts. Yoon (2002) showed an association between teachers' stress level and the frequency of negative student–teacher relations. Finally, focusing on the opposite of teachers' burnout, Jennings and Greenberg (2009) noted the role of teachers' wellbeing in the prediction of healthy student–teacher relations, which themselves are known to predict student outcomes including achievement and motivation.

School satisfaction. Students' school satisfaction can be defined as students' positive affect and attitude toward school (Baker, Dilly, Aupperlee, & Patil, 2003; Huebner & McCullough, 2000) and has been found to be associated with a wide array of desirable outcomes including achievement, engagement, learning behavior, and lower levels of school dropout (Elmore & Huebner, 2010; Huebner & Gilman 2006; Ladd, Buhs, & Seid, 2000). Previous research has shown that students' school satisfaction is facilitated by students' exposure to a favorable instructional climate, positive student–teacher relations, and teacher support (DeSantis-King, Huebner, Suldo, & Valois, 2006; Huebner & Gilman, 2006; Ito & Smith, 2006; Suldo & Shaffer, 2008; Zullig, Huebner, & Patton, 2011). All of these variables related to students' school satisfaction have also been found to be significantly associated with teachers' levels of emotional exhaustion (e.g., Chang, 2009; Grayson & Alvarez, 2008; Jennings & Greenberg, 2009; Klusmann et al., 2008b; Yoon, 2002). This suggests that teachers' emotional exhaustion might also be related to students' levels of school satisfaction.

The Present Study

This study examines the relations between teachers' levels of emotional exhaustion and a wide range of students' cognitive (i.e., achievement) and non-cognitive (i.e., competence self-perceptions, perceptions of teacher support, and school satisfaction) outcomes. As these non-cognitive outcomes have all been found to be highly related to students' levels of achievement (Gregory & Weinstein, 2004; Hamre & Pianta, 2001, 2005; Huebner & Gilman, 2006; Marsh & Craven, 2006; Martin & Downson, 2009), this study also evaluates their direct role in the prediction of students' academic achievement. All of these relations are estimated in multilevel models in order to properly disentangle the individual versus classroom components of these relations. In essence, the effects of teachers' emotional exhaustion on the various educational outcomes can only be located at the classroom (i.e., teacher) level. In contrast, the relations between students' competence self-perceptions, perceptions of teacher support, and school satisfaction, and their resulting levels of achievement can be expected to occur mainly at the individual student level. We discuss multilevel issues more extensively later in the statistical analyses section. In sum, this study tests the following hypotheses, while systematically contrasting the results obtained for models including school grades and models including standardized achievement test scores as achievement indicators:

> Hypothesis 1 (H1). Teachers' levels of emotional exhaustion will be negatively related to classroom levels of students' competence self-perceptions, perceptions of teacher support, school satisfaction, and achievement.
>
> Hypothesis 2 (H2). Individual levels of students' competence self-perceptions, perceptions of teacher support, and school satisfaction will be positively related to individual levels of achievement.

Method

Sample and Procedure

The present study relies on the German data from the PIRLS 2006 study (timssandpirls.bc.edu/pirls2006/; for further information see Bos et al., 2010; Martin, Mullis, & Kennedy, 2007; Mullis, Martin, Kennedy, & Foy, 2007) made available by the Research Data Centre (Forschungsdatenzentrum, FDZ) at the Institute for Educational Quality Improvement (Institut zur Qualitätsentwicklung im Bildungswesen, IQB) located in Berlin (Germany). PIRLS 2006 is a large-scale study aiming to compare students' reading achievement across 40 participating countries (or 45 participating educational systems). Consistent with other comparative studies of student achievement such as PISA or TIMMS, PIRLS aims to provide information about a nation's relative standing regarding students' achievement in certain core subjects to infer important implications for educational policy and instructional practices. Conducted every five years with the first assessment in 2001, PIRLS measures a wide range of variables all related to the subject domain of reading with 4^{th} grade students. Although the assessment of students' reading comprehension for the sake of literary and informational reading purposes represents the core of the study, PIRLS 2006 also includes questionnaires for students, parents, teachers, and school principals to gain detailed information on students' reading motivation, family background, reading instructions, school environment, and home experiences as presumptive determinants of students' reading achievement. PIRLS 2006 thus provided us with a rich data set to test this study's hypotheses while considering both the individual and classroom levels of analyses. Indeed, very few other data sets, and none as rich as PIRLS, include both teachers' self-reports of their own levels of emotional exhaustion, students' self-reports of multiple non-cognitive educational outcomes, and objective indicators of academic achievement.

The German sample from PIRLS 2006 analyzed in the present study includes 380 teachers who taught 7899 4^{th} grade elementary school students. Teachers participating in PIRLS 2006 are generally those teachers who give reading instructions to the students wherefore the teacher sample considered here are students' teachers of German language. The participating teachers had a mean age of 47.10 years (*SD* = 9.36; range from 26 to 65 years), and an average of 21.73 years (*SD* = 11.45; range from 1 to 4 years) of teaching experience. As common for German elementary schools, the majority of teachers were female (*N* = 338). The sample of students included 3848 girls and 4051 boys, had a mean age of 10.46 years (*SD* = 0.51), and were mainly born in Germany (84.7%). These students attended 414 classes with an average class size of 18.24 students. The student sample includes students from all federal states of Germany and was drawn according to guidelines ensuring representativeness for the German 4^{th} grade student population (for further details see Hornberg, Bos, Buddeberg, Potthoff, & Stubbe, 2007). Socio-economic status (SES) was assessed based on the professional occupation of students' mother and father (retaining the highest of both), which were coded into 6 categories based on the Erikson, Goldthorpe, and Portocarero (1979) classification scheme: Category I (high-grade professionals): 17.3%; Category II (lower-grade professionals): 15.0% ; Category III (routine non-manual employees): 6.7%; Category IV (small business proprietors with our without employees): 7.1%; Category V (skilled manual employees and technicians): 19.1%; Category VI (semi-skilled, unskilled, and agricultural employees): 18.0%; Missing values: 16.8%.

Although the exact time of data collection differed slightly across schools, all schools completed the full PIRLS 2006 assessment over a relatively short period of time near the end of the 2005-2006 school year (between March and June 2006). These measures included a questionnaire for the teachers, which included the assessment of emotional exhaustion. This questionnaire was distributed to the teachers before the start of the testing sessions conducted with the students, but in the same period (March-June 2006). Student measures were administered within a single testing session covering two days. The student survey included measures of students' competence self-perceptions, perceptions of teacher support, and school satisfaction and was completed by the students after the standardized reading achievement test, within the same testing session. School grades in German were obtained during the same period from the teachers, and reflect students' accomplishments in reading and writing so far in the course of the current school year.

Measures

Emotional exhaustion. Teachers provided ratings of their own perceived level of emotional exhaustion in relation to their profession. Five items, rated on 4-point Likert scale [true (1), almost true (2), hardly true (3), or not true (4)], were used for this purpose (e.g., "I often feel exhausted at school"; α = .68). For the ease of interpretation, all items were coded so that higher values reflected higher levels of emotional exhaustion.

Competence self-perceptions. Students' competence self-perceptions regarding the academic domain were assessed with a 10-item scale (e.g., "I often fail at class exams", "I simply cannot remember things", "Compared to others, I am simply not so good at school"; α = .88), all rated on a 4-point Likert scale [true (1), almost true (2), hardly true (3), or not true (4)] and coded so that higher values represented higher levels of competence self-perceptions.

Teacher support. Students' perceptions of the degree to which their German teacher provided support to students in the classroom were assessed using 11 items (e.g., "Our German teacher gives advice to individual students how they could improve their learning"; "Our German teacher realizes when we are bored during the lesson"; α = .88). These items were all rated on a 4-point Likert scale [true (1), almost true (2), hardly true (3), or not true (4)] and coded so that higher values represented higher levels of perceived teacher support.

School satisfaction. Students' level of school satisfaction was measured through 6 items (e.g., "I feel safe at school"; "I like to go to school"; α = .76), all rated on a 4-point Likert scale [true (1), almost true (2), hardly true (3), or not true (4)] and coded so that higher values represented higher levels of school satisfaction.

Academic achievement. Student achievement was first assessed using teacher-assigned school grades in German which summarize students' accomplishments in reading and writing. School grades are typically assigned as a function of students' progress in meeting curriculum standards based on teachers' observations and evaluations of students' performance in a series of oral and written tasks. In PIRLS, these school grades are reported on a 4-point scale (1: excellent; 2: good; 3: satisfactory; 4: poor), and were recoded before the analyses so that higher values represent higher achievement. Student achievement was also 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 (Shoemaker, 1973) 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 (for a more extensive description of this procedure, see Mullis et al., 2007). All analyses including standardized reading achievement were conducted separately for each of the five plausible values and properly aggregated afterwards using multiple imputation procedures (Little & Rubin, 2002) implemented in Mplus (Muthén, & Muthén, 2014).

Statistical Analyses

All analyses were conducted using the Robust Maximum Likelihood (MLR) estimator implemented in Mplus 7.2 (Muthén, & Muthén, 2014) in conjunction with Full Information Maximum Likelihood estimation (FIML; Enders, 2010; Graham, 2012) to handle the relatively low amounts of missing data present at the item level. FIML estimation has been shown to provide unbiased parameter estimates under even very high levels of missing data (e.g., 50%) under Missing At Random (MAR) assumptions, and even in some cases under violations of this assumption (e.g., Enders, 2010; Enders & Bandalos, 2001; Larsen, 2011; Shin, Davidson, & Long, 2009). Moreover, FIML has been found to perform generally better than more computationally intensive multiple imputation procedures in the context of multilevel analyses (e.g., Larsen, 2011). The amounts of missing values ranged from 7.4% to 14.4% per item for students' ratings (competence self-perceptions, perceptions of teacher support, school satisfaction) and school grades, and between 8.6% and 9.7% for teachers' ratings of emotional exhaustion. There were no missing values on the plausible values for standardized achievement test scores.

Students' nesting within classrooms was explicitly taken into account in the estimated multilevel models, with students specified at Level 1 (L1), and teachers (i.e., classroom) at Level 2 (L2). There was no need to control for students' nesting within schools as only one class per school was selected for participation in PIRLS 2006. All models estimated here are doubly latent multilevel structural equation models (ML-SEM). We refer readers to Lüdtke and colleagues (Lüdtke et al., 2008; Lüdtke, Marsh, Robitzsch, & Trautwein, 2011; Marsh et al., 2009) for a technical presentation of these models, and to Marsh et al. (2012) and Morin, Marsh, Nagengast, and Scalas (2014) for a practical introduction to these models. These models are called *doubly latent* because they provide a control for sampling error (based on the inter-rater agreement between students forming a single classroom) when forming L2 aggregates, and for measurement error in estimating L1/L2 constructs starting directly at the item level (Marsh et al., 2012; Morin et al., 2014). In these models, the multiple items for student ratings were used to define a priori latent factors at L1 and L2 reflecting students' competence self-perceptions, perceptions of teacher support, and school satisfaction. At L2, students' ratings represent class aggregates of these constructs, whereas L1 components represent deviations from the average classroom ratings due to the implicit group-mean centering of constructs inherent in the Mplus implementation of doubly latent ML-SEM models (Marsh et al., 2012; Morin et al., 2014). For example, students' perceptions of teacher support reflect a more "objective" indication of teachers' support provided to all students in a class (based on the agreement between students) at L2, but inter-individual differences in these perceptions at L1. Achievement (i.e., school grades or standardized achievement test scores) was assessed by a single indicator at L1 and aggregated to L2 by a latent aggregation process to provide an indicator of class average achievement. Finally, the multiple items completed by the teachers to report their levels of emotional exhaustion were used to define one a priori latent factor located at L2. Given the complexity of the models, separate models were estimated including either standardized achievement test scores (Models 1 to 3) or school grades (Models 4 to 6) as achievement indicators.

We started the analyses with the estimations of simple multilevel confirmatory factor analytic (ML-CFA) measurement models where all constructs were directly estimated based on their items. To facilitate interpretation and limit non-essential multicollinearity, all variables were standardized ($M = 0$, $SD = 1$) prior to the estimation (see Marsh et al., 2012; Morin et al., 2014). These ML-CFA models were first estimated with factor loadings freely estimated at both levels (Models 1 and 4) and re-estimated by constraining factor loadings to invariance across levels (Models 2 and 5). Methodologically, invariant factor loadings help to increase the stability and accuracy of ML-SEM parameter estimates and models by making them more parsimonious (e.g., Lüdtke et al., 2011; Morin et al., 2014). Substantially, invariant factor loadings equate the factor metric across levels helping to ensure that the constructs are comparable at both levels (e.g., Lüdtke et al., 2011; Morin et al., 2014). Based on the ML-CFA models with invariant factor loadings across levels, the two ML-SEM models (Models 3 and 6) illustrated in Figure 1 were estimated to test the research questions. These models included the relations between teachers' levels of emotional exhaustion and student outcomes (achievement, competence self-perceptions, perceptions of teacher support, and school satisfaction) which were assumed to be located at L2, whereas the relations between students' non-cognitive outcomes (competence self-perceptions, perceptions of teacher support, and school satisfaction) and achievement were assumed to be mainly located at L1.

Goodness of fit of the models was assessed with the robust $\chi 2$ test statistic, the comparative fit index (CFI), the Tucker-Lewis Index (TLI), and the root mean square error of approximation (RMSEA) as operationalized in Mplus for MLR estimation (Muthén & Muthén, 2014). Typical cut-off scores taken to respectively reflect excellent and adequate fit to the data were used: (i) CFI and TLI ≥ .95 and ≥ .90; (ii) RMSEA ≤ .06 and ≤ .08 (Hu, & Bentler, 1999; Marsh, Hau, & Wen, 2004; Morin et al., 2014). The relative fit of nested models was compared using typical cut-off scores for ΔCFI and ΔTLI ≤ .01 and ΔRMSEA ≤ .015 taken to indicate equivalent model fit (Chen, 2007, Cheung & Rensvold, 2002). For all models, unstandardized and standardized regression coefficients are reported, and can be interpreted as in multiple regression or SEM. Effect size indicators are also reported, and can be interpreted according to Cohen's guidelines (Cohen, 1988) with values over .10, .30, and .50 reflecting small, moderate, and large effect sizes (for details on how to obtain proper estimates of the standardized predictive coefficients and effects sizes in the context of doubly latent ML-SEM models, see Marsh et al., 2012; Morin et al., 2014). It should be noted that all of these interpretation guidelines only remain rough guidelines which should be used to complement a more detailed examination of parameter estimates, statistical conformity, and theoretical adequacy of the models (Marsh et al., 2004; Morin et al., 2014).

Results

Preliminary Analyses

A critical assumption of any multilevel model is the presence of sufficient variability at L2 to justify analyses conducted at this level. The amount of variability located at L2 is typically assessed with the intraclass correlation coefficient (ICC1), which is calculated as $\frac{\tau_x^2}{\tau_x^2+\sigma_x^2}$ where τ_x^2 is the L2 variance and σ_x^2 is the L1 variance. ICC1 values should be close to or higher than .1, but are seldom larger then .30 (Hedges & Hedberg, 2007; Lüdtke et al., 2008, 2011). In this study, ICC1 values were all satisfactory (competence self-perceptions: ICC1 = .07; perceived teacher support: ICC1 = .12; school satisfaction: ICC1 = .17; standardized achievement test scores: ICC1 = .32; school grades: ICC1 = .11). Not surprisingly, the lowest ICC1 coefficient was associated with students' competence self-perceptions, which is naturally an individual construct, but this value remained in line with estimates from previous studies (e.g., Marsh, Trautwein, Lüdtke, Baumert, & Köller, 2007; Morin et al., 2014; Trautwein, Lüdtke, Marsh, Köller, & Baumert, 2006). Moreover, the ICC1 coefficient was lower for school grades than for standardized achievement test scores indicating that lower variability in school grades occurred at the between-class level than at the within-class level. This suggests that the distribution of school grades was more similar from one class to the other than the distribution of standardized achievement test scores, which is consistent with the "grading-on-a-curve" phenomenon (Marsh et al., 2007, 2014). Finally, the results for the ICC1 suggested that 83% to 93% of the variability in ratings of students' competence self-perceptions, perceived teacher support, school satisfaction, and school grades seemed to occur at L1, allowing us to anticipate that most of the relations between these construct were located at L1.

Another assumption to doubly latent ML-SEM model is that there needs to be sufficient level of agreement (i.e., inter-rater reliability) among students in the ratings of L2 constructs. This is typically assessed with the ICC2 indicator, which directly reflects the reliability of class aggregates and is calculated as $\frac{\tau_x^2}{\tau_x^2 + (\sigma_x^2 / n_j)}$ where n_j is the average size of the classes (Bliese, 2000; Lüdtke et al., 2008, 2011; Raudenbush & Bryk, 2002). ICC2 values are interpreted in line with other reliability measures (e.g., Marsh et al., 2012), and were in the acceptable to satisfactory range in this study: competence self-perceptions: ICC1 = .59; perceived teacher support: ICC2 = .71; school satisfaction: ICC2 = .79; standardized achievement test scores: ICC2 = .90; school grades: ICC2 = .69. Not surprisingly, these results were parallel to those from the ICC1, showing the lowest ICC2 to be associated with competence self-perceptions, and a lower ICC2 for school grades than for standardized achievement test scores. We note however that due to the latent aggregation process used to create the L2 constructs, doubly latent ML-SEM models explicitly control for the level of (dis)agreement among students in their ratings of L2 constructs.

Finally, ML-CFA models were estimated to verify that the a priori measurement model was adequate and to test whether the factor loadings were invariant across levels. The results from these models are reported in Table 1 (Models 1-2 with standardized achievement tests scores as an achievement indicator; Models 4-5 with school grades as an achievement indicator). The results showed that the freely estimated ML-CFA models (Models 1 and 4) provided an adequate fit to the data. The model fit increased when invariance constraints were imposed across levels for the factor loadings (Models 2 and 5). This result supported the invariance of the factor loadings across levels, and that the constructs assessed at both levels can be considered to be fully comparable. The specific results for the final invariant ML-CFA models (i.e., Models 2 and 5), which are reported in the Online Supplements, show that all factor loadings were satisfactory, further confirming the adequacy of the measurement models. From these results, composite reliability coefficients were calculated at L1 and L2. As recommended by Morin et al. (2014; also see Sijtsma, 2008), these coefficients were calculated using McDonald's (1970) omega ($\omega = (\Sigma|\lambda i|)^2 / ([\Sigma|\lambda i|]^2 + \Sigma\delta ii)$ where λi are the factor loadings and δii, the error variances) coefficient for this purpose. These coefficients were all satisfactory: competence self-perceptions: ω_{L1} = .87, ω_{L2} = .98; perceived teacher support: ω_{L1} = .88, ω_{L2} = .98; school satisfaction: ω_{L1} = .74, ω_{L2} = .93; teachers' emotional exhaustion: ω_{L2} = .80. The fact that some of these reliability coefficients are closer to the lower bound of acceptability reinforces the importance of relying on latent model providing a control for measurement errors.

Final ML-SEM Models

Starting from the measurement models assuming the invariance of the factor loadings across levels, the a priori predictive ML-SEM models illustrated in Figure 1 were estimated (Models 3 and 6; Table 1). In these more parsimonious models, rather than freely estimate all possible correlations among constructs, only the a priori predictive paths (i.e., regressions) were estimated. Both of these models provided a satisfactory level of fit to the data, and only resulted in negligible declines (i.e. remaining under the recommended guidelines) in model fit relative to the less restrictive ML-CFA models, thus supporting the adequacy of these models. The results from the model including standardized achievement test scores are reported in the top section of Table 2, whereas those from the model including school grades are reported in the bottom section of Table 2.

When looking at L2, both models converged on highly similar results and provide partial support to Hypothesis 1. More precisely, the results showed negative relations between teachers' levels of emotional exhaustion and the class averages of students' perceptions of teacher support (Model 3: standardized β = -.09, p < .001, Model 6: standardized β = -.09, p < .001) and school satisfaction (Model 3: standardized β = -.08, p < .05, Model 6: standardized β = -.08, p < .05). Teachers' emotional exhaustion was also negatively related with standardized achievement test scores (Model 3: standardized β = -.11, $p < .001$) and school grades (Model 6: standardized β = -.05, $p < .001$). The negative relation between teachers' levels of emotional exhaustion and achievement appeared to be more pronounced for standardized achievement test scores than for school grades. Teachers' emotional exhaustion was not significantly related to the class average of students' competence self-perceptions.

None of the L2 relations between the class averages of students' competence self-perceptions, perceptions of teacher support, school satisfaction and achievement were significant. Rather, and as expected, these relations were located at L1, thus providing support for Hypothesis 2. More precisely, individual levels of students' competence self-perceptions and perceptions of teacher support were both positively related to students' individual levels of standardized achievement test scores (standardized β = .24, $p < .001$, resp. standardized β = .07, $p < .001$) and school grades (standardized β = .34, $p < .001$, resp. standardized β = .07, $p < .001$). Furthermore, individual levels of students' school satisfaction were positively related to their individual school grades (standardized β = .05, $p < .001$), but not to their individual standardized achievement test scores.

Discussion

Previous research has clearly documented that teachers' emotional exhaustion carries negative consequences for the teachers themselves in terms of job satisfaction (e.g., Skaalvik & Skaalvik, 2010), physical complaints (e.g., Hakanen et al., 2006), intentions to leave the profession (e.g., Leung & Lee, 2006), motivation (e.g., Schaufeli & Salanova, 2007), and self-efficacy (e.g., Evers et al., 2002). However, little research has investigated the possible effects of teachers' emotional exhaustion on students' educational outcomes. To address this gap, the present study aimed to examine the relation between teachers' emotional exhaustion and students' cognitive (achievement) and non-cognitive (competence self-perceptions, perceptions of teacher support, and school satisfaction) educational outcomes. For this purpose, doubly latent ML-SEM models were used to achieve a proper disaggregation of the processes occurring at the classroom, versus individual student level.

Our results showed that classes taught by teachers reporting higher levels of emotional exhaustion tended to present lower average levels of academic achievement. This relation was found to be more pronounced when achievement was assessed with standardized achievement tests than when it was assessed using school grades. Hence, teachers' emotional exhaustion appeared to be more strongly related to their ability to properly prepare students to meet curriculum standards which are assessed by standardized achievement tests. Teachers suffering from high levels of emotional exhaustion might lack the resources to provide high-quality instruction (Chang, 2009; Klusmann et al., 2008b), leading the students to miss the acquisition of potentially crucial learning contents pre-specified as part of the curriculum.

Compared to standardized achievement test scores, school grades were found to be less negatively associated with teachers' levels of emotional exhaustion. This result is not surprising given that the school grades involve more subjective components. Nonetheless, this study still revealed a negative relation between teachers' levels of emotional exhaustion and school grades. Many mechanisms may explain this result. Despite their more subjective nature, school grades do still partially reflect students' accomplishments and learning, which is likely to be affected as a result of teachers' emotional exhaustion. However, the "grading-on-a-curve" phenomenon (Marsh et al., 2014), which leads to similar distributions of grades across classes irrespective of students' absolute levels of attainments, may explain why teachers' levels of emotional exhaustion are less strongly related to school grades than to standardized achievement test scores.

Furthermore, compared to standardized achievement test scores, school grades depend to a larger extent on teachers' individual evaluation and grading practices and preferences (e.g., leniency, severity, arbitrariness) (Brookhart, 1993, 1994; Cizek, Fitzgerald, & Rachor, 1995;

McMillan et al., 2002). Thus, teachers can also influence students’ school grades by adapting exams, tests, and marking standards. In this context, emotionally exhausted teachers might apply stricter evaluation practices to compensate for their suboptimal teaching practices and to encourage students to rely on higher levels of self-initiated learning. Alternatively, teachers with high levels of emotional exhaustion might be more careless and arbitrary in their evaluations, making it harder for students to anticipate how to meet teachers’ expectations and to get good grades. Teachers with high levels of emotional exhaustion might also come to rely on more lenient evaluation practices to prevent additional problems and complaints arising from bad grades, or as a form of “excuse” for their inadequate teaching practices. This last mechanism, however, would result in a positive relation between teachers’ levels of emotional exhaustion and school grades. Finally, school grades are known to be influenced, at least in part, by students’ behavior and efforts in the classroom, in addition to their absolute levels of learning (Brookhart, 1993, 1994; McMillan et al., 2002; Zimmerman et al., 2013). Interestingly, students’ behaviors in the classroom may themselves induce higher levels of emotional exhaustion in teachers (e.g. Hoglund et al., 2015), suggesting possible reciprocal relations that would need to be more thoroughly investigated in future studies.

Overall, our results showed that teachers’ levels of emotional exhaustion tended to be related to lower levels of school grades and standardized achievement test scores, suggesting that this relation might be at least partly due to the possible negative effects of teachers’ emotional exhaustion on student learning. However, teachers’ emotional exhaustion may also influence teachers’ grading practices in an unpredictable manner (e.g., leniency, severity, etc.). Longitudinal studies allowing for the consideration of different time lags between measures of teachers’ emotional exhaustion and school grades might be able to provide more precision in this regard. Relations between measures of teachers’ emotional exhaustion and school grades taken closer in time might more likely reflect the influence of both mechanisms (i.e., students’ learning and teachers’ grading practices), whereas measures taken further apart in time are more likely reflect the negative impact of teachers’ emotional exhaustion on students’ learning than its impact on teachers’ grading practices. In addition, these studies would do well to incorporate direct measures of teachers’ grading practices as well as of teachers’ tendency to be more lenient, severe or arbitrary.

Beyond achievement, teachers’ level of emotional exhaustion was found to be negatively associated with students’ perceptions of teacher support and school satisfaction. The negative relation between teachers’ emotional exhaustion and students’ perceptions of teacher support might potentially be explained by the observation that teachers with high levels of emotional exhaustion do not have sufficient and adequate resources to build up positive and supportive student–teacher relations (Chang, 2009; Jennings & Greenberg, 2009; Yoon 2002). Given that student–teacher relations and teacher support have been found to serve as an important determinant of school satisfaction (DeSantis et al., 2006; Huebner & Gilman, 2006; Ito & Smith, 2006; Suldo & Shaffer, 2008), these findings might also explain the negative relation between teachers’ levels of emotional exhaustion and students’ levels of school satisfaction.

Teachers’ levels of emotional exhaustion did not seem to influence students’ competence self-perceptions. So far, research has shown students’ competence self-perceptions to be highly dependent on students’ levels of achievement (Marsh & Craven, 2006). The finding that teachers’ levels of emotional exhaustion were negatively related to students’ levels of achievement without affecting their competence self-perceptions suggests that some mechanism may offset students’ tendency to attribute their lower levels of achievement to internal causes (i.e., low levels of ability). Students may rather attribute their lower levels of achievement to the exposure to less stimulating lessons and inadequate instruction provided by emotionally exhausted teachers (Chang, 2009; Klusmann et al., 2008b). Alternatively, school grades remain a highly salient source of comparative feedback for students regarding their own standing within the class and therefore present a key determinant of students' competence self-perceptions (Marsh et al., 2014; Möller et al., 2009). In this context, the "grading-on-a-curve" phenomenon in the allocation of school grades might also partly explain the absence of significant association between teacher levels of emotional exhaustion and classroom levels of competence self-perceptions. Indeed, the "grading-on-a-curve" component of school grades suggests that the relative standing of students within their own class remains relatively independent of the quality of instructional practices used in that class (Marsh et al., 2014). Consequently, students' experiences of inadequate classroom instructions obtained by emotionally exhausted teachers might not affect their relative standing within their own class, therefore explaining the non-significant relation observed between teachers' levels of emotional exhaustion and students' competence self-perceptions.

In sum, the current results clearly showed that teachers' emotional exhaustion is important to consider, and related to key educational outcomes among students. Yet, further studies are needed to examine and disentangle the various mechanisms underlying the observed relations. In this context, it is crucial to consider whether students are able to detect teachers' emotional exhaustion and the possible effects of these perceptions on students' outcomes. So far, research has supported the assertion that students are able to reliably rate teachers' behaviors in the classroom, as well as the quality of their instructional practices, and that these ratings present significant relations with students' levels of achievement and motivation (Hattie, 2009; Kunter, Baumert, & Köller, 2007; Marsh et al., 2012; Seidel & Shavelson, 2007). Thus, students might be able to perceive, if not teachers' levels of emotional exhaustion, at least the impact of emotional exhaustion on teachers' instructional practices (Chang, 2009; Klusmann et al., 2008b). Hence, teachers' instructional practices and classroom management might be important factors to consider when attempting to explain the relations between teachers' emotional exhaustion and students' outcomes.

Over and above the relations between teachers' emotional exhaustion and student outcomes existing at the classroom level (L2), the present study replicated previous results showing positive relations located at the individual level (L1) between students' competence self-perceptions (Marsh & Craven, 2006), school satisfaction (Huebner & Gilman, 2006), and perceptions of teacher support (Hamre & Pianta, 2001, 2005; Hughes et al., 2008; Martin & Downson, 2009), and their individual levels of achievement (although the effect of school satisfaction was limited to school grades). In other words, students with higher levels of competence self-perceptions and school satisfaction as well as students who felt that they received more support from their teachers tended to perform better both on standardized achievement tests and school grades.

Although the burnout syndrome is known to encompass three components (emotional exhaustion, depersonalization, and reduced personal accomplishment; Maslach et al., 2001), the present study only focused on emotional exhaustion. This selection is theoretically reasonable given the conceptualization of emotional exhaustion as the core symptom of burnout, and the first to emerge in its development (Maslach et al., 2001). It is noteworthy that many recent studies of teacher burnout have made a similar decision to focus only on emotional exhaustion (Dicke et al., 2015; Klusmann et al., 2008a, 2008b). Nonetheless, future research should aim to more formally assess how the relations differ across other components of the burnout syndrome, or how they translate to alternative conceptualizations of burnout (e.g., the job demands-resources model of burnout proposed by Demerouti et al., 2011). For instance, it might be reasonable to assume negative associations between depersonalization and student–teacher relations as depersonalization is characterized as "an attempt to put distance between oneself and service recipients by actively ignoring the qualities that make them unique and engaging people" (Maslach et al., 2001, p. 403).

Although this study relied on a large and representative sample of the German student population, it remained limited by its cross-sectional design which did not allow us to disentangle the directionality of the observed relations, or to assume causality. Further explorations of these issues would benefit from longitudinal research designs, or even experimental designs aimed at reducing teachers' levels of emotional exhaustion. Given that the present study was restricted to 4^{th} grade students from Germany and their teachers, it might also be interesting to consider students and teachers from other countries, cultures, and educational systems. Similarly, it might be worthwhile to focus on other teacher populations. For example, special education teachers might be of particular interest since they have been found to be of a particular risk for developing burnout due to the high demands associated with this work (e.g., Zabel & Zabel, 2001). Moreover, it would be worthwhile to consider a greater variety of grade levels and school subjects (not only German language as in the present study) as the relation between teachers' emotional exhaustion and students' educational outcomes may vary as a function of the specific instructional practices implemented in various school grades and subjects. Additionally, that some of the relations might have been underestimated by a possible range restriction in teachers' ratings of emotional exhaustion given that the teachers presenting the highest levels of emotional exhaustion are more likely to have been on medical leave at the time of the data collection. The possible effect of this phenomenon should be explored in future longitudinal studies where it would be possible to consider evolution in teachers' levels of emotional exhaustion.

In addition, future research should also take into account a greater range of students' outcomes, such as positive affect (i.e., liking and enjoyment of learning; e.g., Marsh et al., 2013), proactive classroom behaviors (i.e., active participation and engagement; e.g., Fredricks, Blumenfeld, & Paris, 2004), or sense of school belonging (e.g., Ma, 2007; Pittman & Richmond, 2007). The significance of teachers' emotional exhaustion for students' educational outcomes might further vary as a function of students' characteristics such as age or gender. Here, it might also be informative to consider the possibility of moderation. For instance, teachers' emotional exhaustion might be more deteriorative for some students (e.g., for students with a high need for teacher affiliation) compared to others.

The results of this study have important practical implications. Since teachers' emotional exhaustion has been found to be associated with more negative outcomes for the teachers themselves (e.g., Hakanen et al., 2006; Leung & Lee, 2006; Skaalvik & Skaalvik, 2010), as well as for the exposed students (in the present study), efforts should be made toward the prevention of teacher burnout, or to support teachers already suffering from burnout to recover more quickly and efficiently. In general, although burnout interventions should address the individual by strengthening coping strategies, individual resources, and individual behaviors (Maslach & Goldberg, 1998), it is equally important to consider the situational and organization context. Individual-oriented approaches might benefit from research on individual characteristics known to be related to lower levels of teacher burnout such as the development of teachers' professional knowledge (Dicke et al., 2015), self-efficacy (Zimmerman, 1995), social and emotional competence (Chan, 2006; Jennings & Greenberg, 2009), or emotion-regulation ability (Brackett, Palomera, Mojsa-Kaja, Reyes, & Salovey, 2010). Complementary interventions targeting schools and educational policies should rather consider ways to protect teachers from unreasonable workloads (Hakanen et al., 2006; Klusmann et al., 2008a; Kokkinos, 2007; Schaufeli & Bakker, 2004), to create more positive school climates (Grayson & Alvarez, 2008), and to enhance support among colleagues and from supervisors (Burke, Greenglass, & Schwarzer, 1996; Fernet, Guay, Senécal, & Austin, 2012; Hoglund et al., 2015). Finally, it is also relevant for interventions targeting teacher burnout not to focus solely on teachers' professional lives in terms of skills (e.g., professional knowledge) or conditions (e.g. workload and support), but to also consider interventions to address the likely impact of burnout on teachers' private lives (e.g., marital problems, alcohol problems; Burke et al., 1996), for example by fostering social support from family and friends (Greenglass, Fiksenbaum, & Burke, 1994; Tuxford & Badley, 2014). It might also be of high importance to raise the awareness of new or future teachers regarding the signs, determinants, and consequences of teacher burnout and the availability of support systems to help them cope with professional work demands.

References

Baker J.A., Dilly, L.J., Aupperlee J.L., & Patil, S.A. (2003). The development of school satisfaction: schools as psychologically healthy environments. *School Psychology Quarterly, 18*, 206-221.

Bliese, P.D. (2000). Within- group agreement, non-independence, and reliability: Implications for data aggregation. In K. J. Klein & S.W. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations* (pp. 349-381). San Francisco: Jossey-Bass.

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

Brackett, M.A., Palomera, R., Mojsa-Kaja, J., Reyes, M.R., & Salovey, P. (2010). Emotion-regulation ability, burnout, and job satisfaction among British secondary-school teachers. *Psychology in the Schools, 47*, 406-417.

Brookhart, S.M. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement, 30*, 123-142.

Brookhart, S.M. (1994). Teachers' grading: Practice and theory. *Applied Measurement in Education, 7*, 279-301.

Brouwers, A., & Tomic, W. (2000). A longitudinal study of teacher burnout and perceived self-efficacy in classroom management. *Teaching & Teacher Education, 16*, 239-253.

Burke, R.J., Greenglass, E.R., & Schwarzer, R. (1996). Predicting teacher burnout over time: Effects of work stress, social social support, and self-doubts on burnout and its consequence. *Anxiety, Stress, and Coping, 9*, 261-275.

Byrne, B.M. (1994). Burnout: Testing for the validity, replication, and invariance of causal structure across elementary, intermediate, and secondary teachers. *American Educational Research Journal*, *21*, 645-673.

Chan, D.W. (2006). Emotional intelligence and components of burnout among secondary school teachers in Hong Kong. *Teaching and Teacher Education, 22*, 1042-1054.

Chang, M.-L. (2009). An appraisal perspective of teacher burnout: Examining the emotional work of teachers. *Educational Psychology Review, 21*, 193-218.

Chen, F.F. (2007). Sensitivity of goodness of fit indices to lack of measurement invariance. *Structural Equation Modeling*, *14*, 464-504.

Cheung, G.W., & Rensvold, R.B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233-255.

Cizek, G.J., Fitzgerald, S.M., & Rachor, R.E. (1995). Teachers' assessment practices: Preparation, isolation and the kitchen sink. *Educational Assessment, 3*, 159-179.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.

Cordes, C.L., & Dougherty, T.W. (1993). A review and an integration of research on job burnout. *Academy of Management Review, 18*, 621-656.

Craven, R.G., Marsh, H.W., & Debus, R.L. (1991). Effects of internally focused feedback and attributional feedback on the enhancement of academic self-concept. *Journal of Educational Psychology, 83*, 17-27.

Cropanzano, R., Rupp, D.E., & Byrne, Z.S. (2003). The relationship of emotional exhaustion to work attitudes, job performance, and organizational citizenship behaviors. *Journal of*

Applied Psychology, 88, 160-169.

Demerouti, E., Bakker, A.B., Nachreiner, F., & Schaufeli, W.B. (2001). The job demands–resources model of burnout. *Journal of Applied Psychology, 86*, 499-512.

DeSantis-King, A., Huebner, E.S., Suldo, S.M., & Valois, R.F. (2006). An ecological view of school satisfaction in adolescence: Linkages between social support and behavior problems. *Applied Research in Quality of Life, 1*, 279-295.

Dicke, T., Parker, P.D., Holzberger, D., Kunina-Habenicht, O., Kunter, M., & Leutner, D. (2015). Beginning teachers' efficacy and emotional exhaustion: Latent changes, reciprocity, and the influence of professional knowledge. *Contemporary Educational Psychology, 41*, 62-72.

Elmore, G.M. & Huebner, E.S. (2010). Adolescents' satisfaction with school experiences: Relationships with demographics, attachment relationships, and school engagement behavior. *Psychology in the Schools, 47*, 525- 537.

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

Enders, C.K., & Bandalos, D.L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling, 8*, 430-457.

Erikson, R., Goldthorpe, J.H., & Portocarero, L. (1979). Intergenerational class mobility in three Western European societies. *Birtish Journal of Sociology, 30,* 415-441.

Evers, W.J.G., Brouwers, A., & Tomic, W. (2002). Burnout and self-efficacy: A study on teachers' beliefs when implementing an innovative educational system in the Netherlands. *British Journal of Educational Psychology, 72*, 227-243.

Evers, W.J.G., Tomic, W., & Brouwers, A. (2004). Burnout among teachers. *School Psychology International, 25*, 131-148.

Fernet, C., Guay, F., Senécal, C., & Austin, S., (2012). Predicting intraindividual changes in teacher burnout: The role of perceived school environment and motivational factors. *Teaching and Teacher Education, 28*, 514-525.

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.

Friedman, I.A. (2000). Burnout in teachers: Shattered dreams of impeccable professional performance. *Journal of Clinical Psychology, 56*, 595-606.

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.

Goddard, R., Hoy, W.K., & Woolfolk Hoy, A. (2000). Collective teacher efficacy: Its meaning, measure, and impact on student achievement. *American Educational Research Journal, 37*, 479-507.

Graham, J. W. (2012). *Missing data: Analysis and design.* New York: Springer

Grayson, J.L, & Alvarez, H.K. (2008). School climate factors relating to teacher burnout: A mediator model. *Teaching and Teacher Education, 24*, 1349-1363.

Green, D.E., Walkey, F.H., & Taylor, A.J.W. (1991). The three-factor structure of the Maslach Burnout Inventory. *Journal of Social Behavior and Personality, 6*, 453-472.

Greenglass, E.R., Fiksenbaum, L., & Burke, R. (1994). Relationship between social support and burnout in teachers. *Journal of Social Behavior and Personality, 9*, 219-230.

Gregory, A., & Weinstein, R.S. (2004). Connection and regulation at home and in school: Predicting growth in achievement for adolescents. *Journal of Adolescent Research, 19,* 405-427.

Halbesleben, J.R.B, & Bowler, W.M. (2007). Emotional exhaustion and job performance: The mediating role of motivation. Journal of Applied Psychology, 92, 93-106.

Hakanen, J.J., Bakker, A.B., & Schaufeli, W.B. (2006). Burnout and work engagement among teachers. *Journal of School Psychology, 43*, 495-513.

Hamre, B.K., & Pianta, R.C. (2001). Early teacher–child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development, 72*, 625-638.

Hamre, B.K., & Pianta, R.C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Development*, *76*, 949-967.

Hamre, B.K., Pianta, R.C., Downer, J.T., & Mashburn, A.J. (2008). Teachers' perceptions of conflict with young students: Looking beyond problem behaviors. *Social Development, 17*, 115-136.

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

Hedges, L.V., & Hedberg, E.C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational evaluation and policy analysis*, *29*, 60-87.

Hoglund, W.L.G., Klingle, K.E., & Hosan, N.E. (2015). Classroom risks and resources: Teacher burnout, classroom quality and children's adjustment in high needs elementary schools. *Journal of School Psychology, 53*, 337-357.

Hornberg, S., Bos, W., Buddeberg, I., Potthoff, B., & Stubbe, T. C. (2007). Anlage und Durchführung von IGLU 2006. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert & R. Valtin (Eds.), *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (pp. 21-45). Münster: Waxmann.

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.

Huebner, E.S., & McCullough, G. (2000). Correlates of school satisfaction among adolescents. *Journal of Educational Research, 93*, 331-335.

Huebner, E.S., & Gilman, R. (2006). Students who like and dislike school. *Applied Research in Quality of Life, 1*, 139-150.

Hughes, J.N., Luo,W., Kwok, O-M., & Loyd, L.K. (2008). Teacher-student support, effortful engagement, and achievement: A 3-year longitudinal study. *Journal of Educational Psychology, 100*, 1-14.

Ito, A., & Smith, D.C. (2006). Predictors of school satisfaction among Japanese and U.S. youth. *Community Psychologist, 38*, 19-21.

Jennings, P.A., & Greenberg, M.T. (2009). The prosocial classroom: Teacher social and emotional competence in relation to student and classroom outcomes. *Review of Educational Research, 79*, 491-525.

Klem, A.M., & Connel, J.P. (2004). Relationship matter: Linking teacher support to student engagement and achievement. *Journal of School Health*, *74*, 262-273.

Klusmann, U., Kunter, M., Trautwein, U., Lüdtke, O. & Baumert, J. (2008a). Engagement and emotional exhaustion in teachers: Does the school context make a difference? *Applied Psychology: An International Review*, *57*, 127-151.

Klusmann, U., Kunter, M., Trautwein, U., Lüdtke, O., & Baumert, J. (2008b). Teachers' occupational well-being and quality of instruction: The important role of self-regulatory patterns. *Journal of Educational Psychology, 100*, 702-715.

Kokkinos, C.M. (2007). Job stressors, personality and burnout in primary school teachers. *British Journal of Educational Psychology, 77*, 229-243.

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

Ladd, G.W., Buhs, E.S., & Seid, M. (2000). Children's initial sentiments about kindergarten: Is school liking an antecedent of early classroom participation and achievement? *Merrill-Palmer Quarterly, 46,* 255-279.

Larsen, R. (2011). Missing data imputation versus full information maximum likelihood with second level dependencies. *Structural Equation Modeling*, *18*, 649-662.

Leung, D., & Lee, W. (2006). Predicting intention to quit among Chinese teachers: Different predictability of burnout components. *Anxiety, Stress & Coping*, *19,* 129-141.

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

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

Lüdtke, O., Marsh, H.W., Robitzsch, A., & Trautwein, U. (2011). A 2x2 taxonomy of multilevel latent contextual model: Accuracy-bias trade-offs in full and partial error correction models. *Psychological Methods*, *16*, 444-467.

Ma, X. (2003). Sense of belonging to school: Can schools make a difference? *Journal of Educational Research*, *96*, 340-349.

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., Abduljabbar, A.S., Abu-Hilal, M., Morin, A.J.S., Abdelfattah, F., Leung, K.C., Xu, M.K., Nagengast, B., & Parker, P. (2013). Factor structure, discriminant and convergent validity of TIMSS math and science motivation measures: A comparison of USA and Saudi Arabia. *Journal of Educational Psychology*, *105*, 108-128.

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.

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.

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.

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

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.

Marsh, H.W., Trautwein, U., Lüdtke, O., Baumert, J., & Köller, O. (2007). The Big-Fish-Little-Pond Effect: Persistent negative effects of selective high schools on self-concept after graduation. *American Educational Research Journal*, *44*, 631-669.

Martin, A.J., & Dowson, M. (2009). Interpersonal relationships, motivation, engagement, and achievement: Yields for theory, current issues, and educational practice. *Review of Educational Research 79*, 327-365.

Martin, M.O., Mullis, I.V.S., & Kennedy A.M. (2007). *PIRLS 2006 Technical Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Maslach, C. & Goldberg J. (1998). Prevention of burnout: New perspectives. *Applied and Preventive Psychology*, *7*, 63-74.

Maslach, C., & Jackson, S. E. (1981). The measurement of experienced burnout. *Journal of Occupational Behaviour, 2,* 99-113.

Maslach, C., Jackson, S.E., & Leiter, M.P. (1996). *Maslach burnout inventory manual* (3rd ed.). Mountain View, CA: CPP.

Maslach, C., Schaufeli, W.B., & Leiter, M.P. (2001). Job burnout. *Annual Review of Psychology*, *52,* 397-422.

McMillan, J.H., Myran, S., & Workman, D. (2002). Elementary teachers‘ classroom assessment and grading practices. *The Journal of Educational Research, 95*, 203-213.

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.

Moore, J.E. (2000). Why is this happening? A causal attribution approach to work exhaustion consequences. *Academy of Management Review, 25,* 335-349.

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.

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.

Muthén, L.K., & Muthén, B.O. (2014). *Mplus User's Guide, Version 7.2.* Los Angeles, CA: Muthén & Muthén.

Patrick, H., & Ryan, A.M. (2005). Identifying adaptive classrooms: Dimensions of the classroom social environment. In K.A. Moore & L.H. Lippman (Eds.), *What do children need to flourish? Conceptualizing and measuring indicators of positive development* (pp. 271-287). New York: Springer

Pittman, L.D., & Richmond, A. (2007). Academic and psychological functioning in late adolescence: The importance of school belonging. *Journal of Experimental Education, 75*, 270-290.

Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Reddy, R., Rhodes, J.E., & Mulhall, P. (2003). The influence of teacher support on student adjustment in the middle school years: A latent growth curve study. *Development and Psychopathology, 15*, 19-138.

Ryan, A.M., & Patrick, H. (2001). The classroom social environment and changes in adolescents' motivation and engagement during middle school. *American Educational Research Journal,* 38, 437–460.

Schaarschmidt, U., Kieschke, U., & Fischer, A. (1999). Beanspruchungsmuster im Lehrerberuf [Patterns of teachers' occupational stress]. *Psychologie in Erziehung und Unterricht, 46*, 244-268.

Schaufeli, W.B., & Bakker, A.D. (2004). Job demands, job resources, and their relationship with burnout and engagement: A multi-sample study. *Journal of Organizational Behavior, 25,* 293-315.

Schaufeli, W.B., & Salanova, M. (2007). Efficacy or inefficacy, that's the question: burnout and work engagement, and their relationships with efficacy beliefs. *Anxiety, Stress, & Coping, 20,* 177-196.

Schwarzer, R., & Hallum, S. (2008). Perceived teacher self-efficacy as a predictor of job stress and burnout: Mediation. *Applied Psychology, 57*, 152-171.

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.

Shin, T., Davidson, M.L., & Long, J.D. (2009). Effects of missing data methods in structural equations modeling with nonnormal data. *Structural Equation Modeling, 16,* 70–98.

Shoemaker, D.M. (1973). *Principles and procedures of multiple matrix sampling*. Cambridge, MA: Ballinger.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*, 107-120.

Skaalvik, E.M., & Skaalvik, S. (2007). Dimensions of teacher self-efficacy and relations with strain factors, perceived collective teacher efficacy, and teacher burnout. *Journal of Educational Psychology*, *99*, 611-625.

Skaalvik, E.M., & Skaalvik, S. (2010). Teacher self-efficacy and teacher burnout: A study of relations *Teaching and Teacher Education, 26,* 1059-1069.

Spilt, J.L., Hughes, J.N., Wu, J., & Kwok, O. (2012). Dynamics of teacher-student relationships: Stability and change across elementary school and the influence on children's academic success. *Child Development, 83,* 1180-1195.

Suldo, S.M., & Shaffer, E.J. (2008). Looking beyond psychopathology: The dual-factor model of mental health in youth. *School Psychology Review, 37,* 52-68.

Trautwein, U., Lüdtke, O., Marsh, H.W., Köller, O., & Baumert, J. (2006). Tracking, grading, and student motivation: Using group composition and status to predict self-concept and interest in 9th mathematics. *Journal of Educational Psychology*, *98*, 788-806.

Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W.K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68,* 202-248.

Tuxford, L.M., & Bradley, G.L. (2014). Emotional job demands and emotional exhaustion in teachers. *Educational Psychology*. Online first: doi:10.1080/01443410.2014.912260

Vandenberghe, R., & Huberman, A. M. (Eds.). (1999). *Understanding and preventing teacher burnout: A sourcebook of international research and practice*. Cambridge, England: Cambridge University Press.

Wright, T.A., & Bonett, D.G. (1997). The contribution of burnout to work performance. *Journal of Organizational Behavior*, *18*, 491-499.

Wright, T.A., & Cropanzano, R. (1998). Emotional exhaustion as a predictor of job performance and voluntary turnover. *Journal of Applied Psychology, 83,* 486-493.

Yoon, J.S. (2002). Teacher characteristics as predictors of teacher-student relationships: Stress, negative affect, and self-efficacy. *Social Behavior and Personality: An international journal,* 30, 485-493.

Zabel, R.H., & Zabel, M.K. (2001): Revisiting burnout among special education teachers: Do age, experience, and preparation still matter? *Teacher Education and Special Education,* 24, 128-139.

Zimmerman, B.J. (1995). Self-efficacy and educational development. In A. Bandura (Ed.), *Self-efficacy in changing societies* (pp. 202-231). New York: Cambridge University.

Zimmermann, F., Schütte, K., Taskinen, P., & Köller, O. (2013). Reciprocal effects between adolescent externalizing problems and measures of achievement. *Journal of Educational Psychology, 105,* 747-761

Zullig, K.J., & Huebner, E.S., & Patton, J.M. (2011). Relationships among school climate domains and school satisfaction. *Psychology in the Schools, 48,* 133-145.

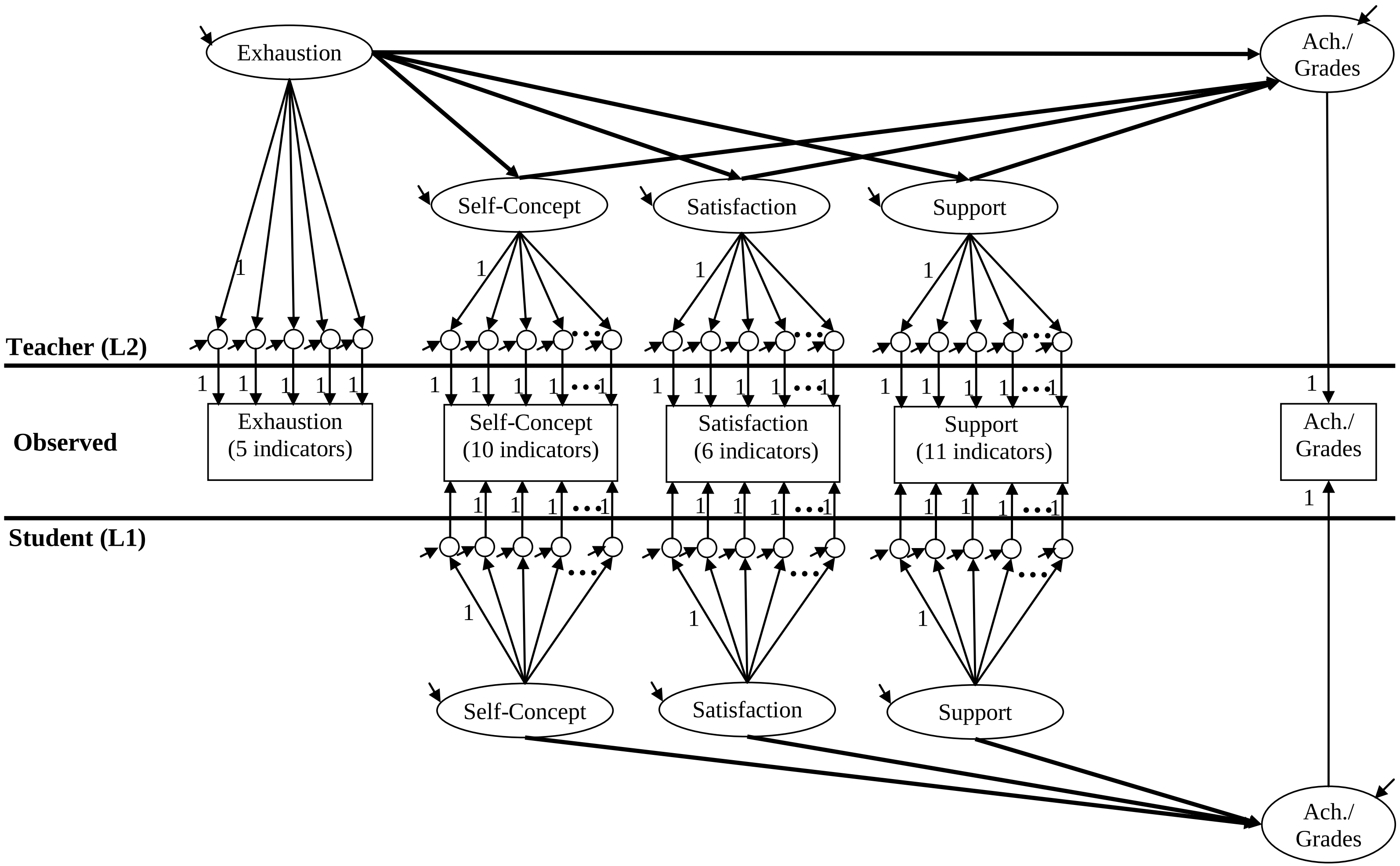

Figure 1. Doubly latent model tested in the current study.
Note. Factor correlations (L1 and L2 correlations between academic self-concept, school satisfaction, and perceptions of teacher support) are not shown in the figures. For parsimony, only 5 indicators are depicted for constructs including more than five indicators, although all indicators are used in the model.

Table 1
Goodness-of-fit Indices

Model	χ^2	df	CFI	TLI	RMSEA	Model description
Standardized Achievement Test Scores						
Model 1	6587.707*	831	.908	.899	.029	Multilevel Confirmatory Factor Analysis (CFA).
Model 2	6585.465*	855	.908	.902	.029	Multilevel CFA, factor loadings invariant across levels
Model 3	6686.420*	858	.906	.901	.029	Multilevel Predictive Model
School Grades						
Model 4	6478.894*	831	.901	.909	.029	Multilevel Confirmatory Factor Analysis (CFA).
Model 5	6472.253*	855	.910	.904	.028	Multilevel CFA, factor loadings invariant across levels
Model 6	6567.685*	858	.908	.903	.029	Multilevel Predictive Model

Note. * = $p \leq .05$; χ^2 = chi square test of model fit calculated under the Robust Maximum Likelihood (MLR) estimator; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation.

Table 2.

Main Effects from the Final Multilevel Predictive Models Presented in Figure 1

	Est. (S.E.)	Std. (S.E.)	ES (S.E.)
Model 3: Standardized achievement			
Classroom-level (L2) effects			
Teacher exhaustion → Achievement	-0.166 (0.055)**	-0.109 (0.036)**	-0.263 (0.087)**
Teacher exhaustion → Competence Self-perceptions	-0.026 (0.019)	-0.027 (0.020)	-0.057 (0.042)
Teacher exhaustion → School satisfaction	-0.070 (0.031)*	-0.079 (0.035)*	-0.174 (0.077)*
Teacher exhaustion → Perceptions of support	-0.079 (0.025)**	-0.092 (0.029)**	-0.195 (0.061)**
Competence self-perceptions → Achievement	0.782 (0.432)	0.128 (0.071)	0.310 (0.172)
School satisfaction → Achievement	0.412 (0.344)	0.089 (0.075)	0.217 (0.183)
Perceptions of support → Achievement	-0.481 (0.489)	-0.083 (0.086)	-0.201 (0.209)
Individual-level (L1) effects			
Competence self-perceptions → Achievement	0.404 (0.037)**	0.238 (0.022)**	0.578 (0.053)**
School satisfaction → Achievement	-0.047 (0.031)	-0.025 (0.016)	-0.060 (0.040)
Perceptions of support → Achievement	0.124 (0.027)**	0.065 (0.014)**	0.159 (0.035)**
Model 6: School grades			
Classroom-level (L2) effects			
Teacher exhaustion → Achievement	-0.080 (0.037)*	-0.054 (0.025)*	-0.115 (0.053)*
Teacher exhaustion → Competence self-perceptions	-0.023 (0.019)	-0.024 (0.020)	-0.051 (0.041)
Teacher exhaustion → School satisfaction	-0.069 (0.031)*	-0.078 (0.035)*	-0.172 (0.077)*
Teacher exhaustion → Student perceptions of support	-0.081 (0.025)**	-0.093 (0.029)**	-0.199 (0.061)**
Competence self-perceptions → Achievement	0.053 (0.183)	0.009 (0.031)	0.019 (0.066)
School satisfaction → Achievement	-0.051 (0.123)	-0.011 (0.027)	-0.024 (0.058)
Perceptions of support → Achievement	0.050 (0.170)	0.009 (0.030)	0.019 (0.064)
Individual-level (L1) effects			
Competence self-perceptions → Achievement	0.549 (0.044)**	0.336 (0.027)**	0.713 (0.058)**
School satisfaction → Achievement	0.083 (0.031)**	0.045 (0.017)**	0.095 (0.035)**
Perceptions of support → Achievement	0.136 (0.027)**	0.074 (0.015)**	0.157 (0.031)**

Note. * = $p \leq .05$;** = $p \leq .01$; Est. = unstandardized parameter estimate; S.E.: standard error of the estimate; Std. = standardized parameter estimate; ES: effect size.

Online Supplements for:

Relations between teachers’ emotional exhaustion and students’ educational outcomes

Table S1

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

	School satisfaction				Perceived teacher support				Competence self-perceptions				Teachers’ emotional exhaustion			
	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.)
Indicator 1	1.000	0.566	0.630	0.680	1.000	0.563	0.644	0.683	1.000	0.622	0.595	0.614				
	(0.000)	(0.016)	(0.019)	(0.018)	(0.000)	(0.013)	(0.022)	(0.015)	(0.000)	(0.017)	(0.030)	(0.021)				
Indicator 2	1.098	0.629	0.546	0.604	1.223	0.691	0.488	0.522	0.912	0.565	0.663	0.680				
	(0.024)	(0.016)	(0.019)	(0.021)	(0.032)	(0.012)	(0.019)	(0.017)	(0.025)	(0.017)	(0.031)	(0.019)				
Indicator 3	0.945	0.534	0.664	0.715	1.166	0.657	0.533	0.568	0.964	0.599	0.623	0.642				
	(0.028)	(0.015)	(0.019)	(0.017)	(0.033)	(0.010)	(0.014)	(0.013)	(0.027)	(0.017)	(0.027)	(0.020)				
Indicator 4	0.823	0.471	0.704	0.778	1.229	0.693	0.486	0.519	0.982	0.630	0.548	0.603				
	(0.046)	(0.017)	(0.019)	(0.016)	(0.034)	(0.011)	(0.017)	(0.015)	(0.036)	(0.016)	(0.029)	(0.020)				
Indicator 5	0.996	0.570	0.611	0.675	1.156	0.653	0.537	0.574	0.966	0.600	0.622	0.640				
	(0.051)	(0.016)	(0.020)	(0.019)	(0.029)	(0.011)	(0.016)	(0.014)	(0.031)	(0.017)	(0.029)	(0.020)				
Indicator 6	1.065	0.604	0.585	0.635	1.221	0.695	0.477	0.517	1.062	0.661	0.544	0.563				
	(0.053)	(0.016)	(0.020)	(0.019)	(0.032)	(0.011)	(0.018)	(0.016)	(0.034)	(0.016)	(0.030)	(0.022)				
Indicator 7					1.068	0.600	0.604	0.640	1.077	0.670	0.533	0.551				
					(0.029)	(0.012)	(0.017)	(0.014)	(0.033)	(0.013)	(0.024)	(0.018)				
Indicator 8					1.216	0.687	0.494	0.528	1.094	0.698	0.471	0.512				
					(0.036)	(0.010)	(0.016)	(0.014)	(0.037)	(0.014)	(0.025)	(0.019)				
Indicator 9					0.874	0.489	0.724	0.761	1.018	0.633	0.580	0.599				
					(0.030)	(0.011)	(0.015)	(0.011)	(0.034)	(0.015)	(0.025)	(0.019)				
Indicator 10					1.042	0.586	0.620	0.657	1.126	0.704	0.483	0.504				
					(0.033)	(0.011)	(0.016)	(0.013)	(0.036)	(0.012)	(0.022)	(0.017)				
Indicator 11					1.007	0.566	0.643	0.680								
					(0.032)	(0.012)	(0.019)	(0.013)								

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

Table S2

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

	School satisfaction				Perceived teacher support				Competence self-perceptions				Teachers' emotional exhaustion			
	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.)
Indicator 1	1.000	0.915	0.012	0.163	1.000	0.828	0.018	0.315	1.000	0.954	0.003	0.090	1.000	0.671	0.563	0.550
	(0.000)	(0.037)	(0.005)	(0.067)	(0.000)	(0.037)	(0.004)	(0.061)	(0.000)	(0.039)	(0.002)	(0.074)	(0.000)	(0.043)	(0.056)	(0.058)
Indicator 2	1.098	0.867	0.024	0.248	1.223	0.948	0.007	0.101	0.912	0.971	0.002	0.058	0.960	0.642	0.605	0.588
	(0.024)	(0.031)	(0.006)	(0.054)	(0.032)	(0.023)	(0.003)	(0.043)	(0.025)	(0.027)	(0.001)	(0.052)	(0.091)	(0.044)	(0.077)	(0.057)
Indicator 3	0.945	0.849	0.021	0.279	1.166	0.944	0.007	0.109	0.964	0.945	0.003	0.107	1.124	0.758	0.431	0.426
	(0.028)	(0.031)	(0.004)	(0.052)	(0.033)	(0.027)	(0.003)	(0.052)	(0.027)	(0.059)	(0.004)	(0.112)	(0.130)	(0.041)	(0.064)	(0.063)
Indicator 4	0.823	0.652	0.055	0.574	1.229	0.974	0.003	0.050	0.982	0.601	0.051	0.639	1.016	0.694	0.513	0.519
	(0.046)	(0.034)	(0.007)	(0.044)	(0.034)	(0.021)	(0.003)	(0.041)	(0.036)	(0.102)	(0.023)	(0.123)	(0.129)	(0.048)	(0.068)	(0.067)
Indicator 5	0.996	0.804	0.032	0.353	1.156	0.903	0.012	0.185	0.966	0.960	0.002	0.078	0.848	0.575	0.670	0.669
	(0.051)	(0.034)	(0.006)	(0.055)	(0.029)	(0.026)	(0.003)	(0.048)	(0.031)	(0.042)	(0.002)	(0.080)	(0.084)	(0.043)	(0.067)	(0.049)
Indicator 6	1.065	0.904	0.015	0.183	1.221	0.898	0.014	0.193	1.062	0.975	0.002	0.050				
	(0.053)	(0.029)	(0.005)	(0.053)	(0.032)	(0.024)	(0.004)	(0.043)	(0.034)	(0.029)	(0.002)	(0.057)				
Indicator 7					1.068	0.881	0.013	0.224	1.077	0.998	0.000	0.005				
					(0.029)	(0.031)	(0.004)	(0.054)	(0.033)	(0.001)	(0.000)	(0.003)				
Indicator 8					1.216	0.950	0.006	0.097	1.094	0.693	0.039	0.520				
					(0.036)	(0.020)	(0.003)	(0.037)	(0.037)	(0.101)	(0.020)	(0.140)				
Indicator 9					0.874	0.770	0.021	0.407	1.018	0.967	0.002	0.064				
					(0.030)	(0.043)	(0.005)	(0.066)	(0.034)	(0.039)	(0.003)	(0.075)				
Indicator 10					1.042	0.873	0.013	0.238	1.126	0.935	0.005	0.125				
					(0.033)	(0.033)	(0.004)	(0.058)	(0.036)	(0.035)	(0.003)	(0.066)				
Indicator 11					1.007	0.854	0.015	0.271								
					(0.032)	(0.037)	(0.004)	(0.063)								

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

Table S3

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

	School satisfaction	Perceived teacher support	Academic self-concept	Achievement	Teachers' emotional exhaustion
Level 1					
School satisfaction	0.296 (0.018)*	0.132 (0.007)*	0.054 (0.007)*	0.025 (0.008)*	
Perceived teacher support	0.445 (0.017)*	0.298 (0.017)*	0.029 (0.006)*	0.043 (0.007)*	
Competence self-perceptions	0.163 (0.020)*	0.086 (0.017)*	0.375 (0.025)*	0.152 (0.010)*	
Achievement	0.053 (0.018)*	0.091 (0.014)*	0.291 (0.020)*	0.731 (0.014)*	
Level 2					
School satisfaction	0.060 (0.007)*	0.032 (0.004)*	0.009 (0.003)*	0.024 (0.012)	-0.023 (0.013)
Perceived teacher support	0.648 (0.054)*	0.040 (0.005)*	0.005 (0.003)	0.004 (0.010)	-0.031 (0.011)*
Competence self-perceptions	0.223 (0.075)*	0.151 (0.091)	0.030 (0.009)*	0.027 (0.014)	-0.010 (0.009)
Achievement	0.164 (0.078)*	0.036 (0.089)	0.267 (0.117)*	0.346 (0.071)*	-0.081 (0.027)*
Teachers' emotional exhaustion	-0.140 (0.075)	-0.232 (0.072)*	-0.083 (0.079)	-0.203 (0.059)*	0.461 (0.077)*

* $p < .05$.

Table S4

Factor Loadings and Item Uniquenesses of the Invariant Multilevel Confirmatory Factor Analytic Model (see Model 4, see Table 1): Level 1

	School satisfaction				Perceived teacher support				Competence self-perceptions				Teachers' emotional exhaustion			
	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.)
Indicator 1	1.000	0.562	0.634	0.684	1.000	0.563	0.644	0.683	1.000	0.624	0.592	0.611				
	(0.000)	(0.016)	(0.019)	(0.018)	(0.000)	(0.013)	(0.022)	(0.015)	(0.000)	(0.017)	(0.030)	(0.021)				
Indicator 2	1.097	0.625	0.551	0.610	1.224	0.691	0.488	0.522	0.910	0.566	0.662	0.679				
	(0.024)	(0.016)	(0.019)	(0.020)	(0.032)	(0.012)	(0.019)	(0.017)	(0.025)	(0.017)	(0.031)	(0.019)				
Indicator 3	0.947	0.532	0.666	0.717	1.166	0.657	0.533	0.568	0.961	0.599	0.622	0.641				
	(0.028)	(0.016)	(0.019)	(0.017)	(0.033)	(0.010)	(0.014)	(0.013)	(0.027)	(0.017)	(0.027)	(0.020)				
Indicator 4	0.833	0.474	0.702	0.776	1.229	0.694	0.486	0.519	0.979	0.630	0.548	0.603				
	(0.047)	(0.016)	(0.019)	(0.016)	(0.034)	(0.011)	(0.017)	(0.015)	(0.036)	(0.016)	(0.028)	(0.020)				
Indicator 5	1.008	0.573	0.608	0.671	1.156	0.653	0.537	0.573	0.963	0.600	0.622	0.640				
	(0.051)	(0.016)	(0.020)	(0.018)	(0.029)	(0.011)	(0.016)	(0.014)	(0.031)	(0.017)	(0.029)	(0.020)				
Indicator 6	1.078	0.608	0.581	0.631	1.221	0.695	0.477	0.517	1.061	0.663	0.542	0.561				
	(0.053)	(0.016)	(0.020)	(0.019)	(0.032	(0.011)	(0.018)	(0.016)	(0.033)	(0.016)	(0.030)	(0.022)				
Indicator 7					1.067	0.600	0.604	0.640	1.073	0.670	0.533	0.551				
					(0.029)	(0.012)	(0.017)	(0.014)	(0.032)	(0.013)	(0.024)	(0.018)				
Indicator 8					1.216	0.687	0.494	0.528	1.089	0.697	0.472	0.514				
					(0.036)	(0.010)	(0.016)	(0.014)	(0.036)	(0.014)	(0.025)	(0.019)				
Indicator 9					0.874	0.489	0.724	0.761	1.010	0.630	0.584	0.603				
					(0.030)	(0.011)	(0.015)	(0.011)	(0.034)	(0.015)	(0.025)	(0.019)				
Indicator 10					1.041	0.586	0.620	0.657	1.120	0.703	0.485	0.506				
					(0.033)	(0.011)	(0.016)	(0.013)	(0.036)	(0.012)	(0.022)	(0.017)				
Indicator 11					1.008	0.566	0.642	0.680								
					(0.032)	(0.012)	(0.019)	(0.013)								

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

Table S5

Factor Loadings and Item Uniquenesses of the Invariant Multilevel Confirmatory Factor Analytic Model (see Model 4, see Table 1): Level 2

	School satisfaction				Perceived teacher support				Competence self-perceptions				Teachers' emotional exhaustion			
	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.)
Indicator 1	1.000	0.910	0.012	0.172	1.000	0.827	0.018	0.316	1.000	0.951	0.003	0.096	1.000	0.667	0.568	0.555
	(0.000)	(0.037)	(0.005)	(0.068)	(0.000)	(0.037)	(0.004)	(0.061)	(0.000)	(0.039)	(0.002)	(0.075)	0.000	(0.043)	(0.056)	(0.058)
Indicator 2	1.097	0.860	0.025	0.261	1.224	0.948	0.007	0.101	0.910	0.969	0.002	0.061	0.968	0.644	0.602	0.585
	(0.024)	(0.031)	(0.006)	(0.054)	(0.032)	(0.023)	(0.003)	(0.043)	(0.025)	(0.028)	(0.002)	(0.054)	0.093	(0.044)	(0.077)	(0.057)
Indicator 3	0.947	0.845	0.021	0.286	1.166	0.943	0.007	0.111	0.961	0.942	0.003	0.113	1.127	0.755	0.435	0.430
	(0.028)	(0.031)	(0.004)	(0.052)	(0.033)	(0.027)	(0.003)	(0.052)	(0.027)	(0.060)	(0.004)	(0.113)	0.130	(0.041)	(0.063)	(0.062)
Indicator 4	0.833	0.657	0.054	0.569	1.229	0.975	0.003	0.049	0.979	0.598	0.051	0.643	1.026	0.696	0.509	0.515
	(0.047)	(0.033)	(0.007	(0.044)	(0.034)	(0.021)	(0.003)	(0.041)	(0.036)	(0.101)	(0.023)	(0.121)	0.129	(0.047)	(0.068)	(0.066)
Indicator 5	1.008	0.809	0.031	0.345	1.156	0.902	0.012	0.187	0.963	0.959	0.002	0.079	0.856	0.578	0.667	0.666
	(0.051)	(0.034)	(0.006)	(0.055)	(0.029)	(0.027)	(0.003)	(0.048)	(0.031)	(0.042)	(0.002)	(0.080)	0.086	(0.043)	(0.067)	(0.049)
Indicator 6	1.078	0.907	0.015	0.177	1.221	0.899	0.014	0.193	1.061	0.974	0.002	0.052				
	(0.053)	(0.029)	(0.005)	(0.052)	(0.032	(0.024)	(0.004)	(0.043)	(0.033)	(0.030)	(0.002)	(0.058)				
Indicator 7					1.067	0.880	0.013	0.225	1.073	0.998	0.000	0.005				
					(0.029)	(0.031)	(0.004)	(0.054)	(0.032)	(0.001)	(0.000)	(0.003)				
Indicator 8					1.216	0.950	0.006	0.097	1.089	0.689	0.039	0.526				
					(0.036)	(0.020)	(0.003)	(0.038)	(0.036)	(0.100)	(0.020)	(0.138)				
Indicator 9					0.874	0.769	0.021	0.409	1.010	0.966	0.002	0.066				
					(0.030)	(0.043)	(0.005)	(0.066)	(0.034)	(0.040)	(0.003)	(0.077)				
Indicator 10					1.041	0.871	0.014	0.241	1.120	0.937	0.005	0.122				
					(0.033)	(0.033)	(0.004)	(0.058)	(0.036)	(0.035)	(0.003)	(0.066)				
Indicator 11					1.008	0.854	0.015	0.270								
					(0.032)	(0.037)	(0.004)	(0.063)								

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

Table S6

Factor Covariances (above the diagonal), Factor Variances (in the diagonal), and Factor Correlations (below the diagonal) of the Invariant Multilevel Confirmatory Factor Analytic Model (see Model 4, see Table 1)

	School satisfaction	Perceived teacher support	Competence self-perceptions	Grades (achievement)	Teachers' emotional exhaustion
Level 1					
School satisfaction	0.293 (0.018)*	0.132 (0.007)*	0.053 (0.007)*	0.072 (0.008) *	
Perceived teacher support	0.446 (0.017)*	0.298 (0.017)*	0.028 (0.006)*	0.067 (0.007)*	
Academic self-concept	0.160 (0.019)*	0.083 (0.017)	0.377 (0.024)*	0.215 (0.012)*	
Grades (achievement)	0.141 (0.016)*	0.130 (0.013)*	0.371 (0.021)*	0.894 (0.015)*	
Level 2					
School satisfaction	0.059 (0.007)*	0.031 (0.004)*	0.009 (0.003)*	0.005 (0.006)	-0.023 (0.013)
Perceived teacher support	0.651 (0.054)*	0.040 (0.005)*	0.005 (0.003)	0.005 (0.005)	-0.032 (0.011)*
Academic self-concept	0.215 (0.074)*	0.146 (0.089)	0.030 (0.009)*	0.003 (0.005)	-0.008 (0.009)
Grades (achievement)	0.060 (0.072)	0.077 (0.081)	0.047 (0.089)	0.111 (0.012)*	-0.037 (0.016)
Teachers' emotional exhaustion	-0.139 (0.075)	-0.235 (0.072)	-0.073 (0.079)	-0.163 (0.067)*	0.455 (0.076)*

* $p < .05$.