

Measuring Teachers' Enjoyment, Anger, and Anxiety:
The Teacher Emotions Scales (TES)

Anne C. Frenzel^a

University of Munich, Germany

Reinhard Pekrun^a

University of Munich, Germany

Thomas Goetz^b

University of Konstanz, Germany & Thurgau University of Teacher Education, Switzerland

Lia M. Daniels^c

University of Alberta, Canada

Tracy L. Durksen^c

University of Alberta, Canada

Betty Becker-Kurz^a

University of Munich, Germany

Robert M. Klassen^d

University of York, United Kingdom

Author Note

^aDepartment of Psychology, University of Munich, Leopoldstrasse 13, 80802 Munich, Germany; Emails: frenzel@psy.lmu.de, pekrun@psy.lmu.de; betty.becker@psy.lmu.de

^bDepartment of Empirical Educational Research, Universitaetsstrasse 10, University of Konstanz, Konstanz, Germany, 78457 Konstanz and Thurgau University of Teacher Education, Kreuzlingen, Switzerland; Email: thomas.goetz@psy.lmu.de

^cDepartment of Educational Psychology, University of Alberta, 6-102 Education North, Edmonton, AB T6G 2G5, Canada; Emails: lia.daniels@ualberta.ca, tdurksen@ualberta.ca

^dDepartment of Education, University of York, York YO10 5DD, United Kingdom; Email: robert.klassen@york.ac.uk.

Correspondence concerning this article should be addressed to Anne Frenzel, Department of Psychology, University of Munich, Leopoldstr. 13, 80802 Munich, Germany, Email frenzel@psy.lmu.de, Telephone +49-89-2180-72557, Fax +49-89-2180-72509

Abstract

The emotions of teachers are considered relevant not only for their own well-being but also for the functioning of classrooms. Nevertheless, research on teacher emotions has been slow to emerge, and scales for their assessment via self-report are generally lacking. In the present research we developed four-item scales for three emotions considered most relevant in the context of teaching: enjoyment, anger, and anxiety (Teacher Emotions Scales, TES). Based on data of 944 teachers, we tested German and English language versions of the TES for reliability, internal and external validity, and cross-language equivalence, while exploring the utility of both a general and a student-group specific variant. All scales proved to be highly reliable, and confirmatory factor analysis supported internal validity by showing that three-factor models (enjoyment, anger, and anxiety) were superior to single-factor or two-factor (positive vs. negative affect) models. The external validation analyses provided consistent evidence for theoretically meaningful relations with teachers' general affect, burnout, job satisfaction, and teacher self-efficacy. These findings were robust across multiple studies. In addition, consistent relationships with student ratings of teaching behaviors were found. Analyses of measurement invariance revealed that the English and the German language versions were fully structurally equivalent and displayed metric invariance.

Keywords: teacher enjoyment, teacher anger, teacher anxiety, instrument validation, cross-language measurement invariance analysis

Measuring Teachers' Enjoyment, Anger, and Anxiety: The Teacher Emotions Scales (TES)

Emotions abound in classrooms. While there is an extensive body of research on teacher burnout, a clinical syndrome with strong links to emotional experiences – specifically, feelings of emotional exhaustion (Ghanizadeh & Jahedizadeh, 2015; Vandenberghe & Huberman, 1999), teachers' experiences of specific, discrete emotions such as enjoyment, anger, or anxiety have received relatively little research attention (Frenzel, 2014; Keller, Frenzel, Goetz, Pekrun, & Hensley, 2014; Schutz & Zembylas, 2009). It is striking that much of the current scientific knowledge on discrete teacher emotions rests on narrative, often single-case data, whereas quantitative data are scarce (Frenzel, 2014; Sutton & Wheatley, 2003). These existing studies, however, provide evidence that teachers' emotions are relevant outcomes, with clear links to overall teacher well-being, burn-out risk and retention in the teaching profession (e.g., Chang, 2009). Moreover, they are proposed to be relevant for students because teacher emotions are linked with teaching quality and the established bonds with students (Frenzel, 2014; Hagenauer & Volet, 2014b; Hargreaves, 2000; Klassen, Perry, & Frenzel, 2012; Sutton & Wheatley, 2003). Research also indicates that teachers' affective experiences are broadly diversified and that omnibus-constructs such as feeling generally emotionally exhausted or depleted are insufficient to describe teachers' emotional lives in detail (Hagenauer & Volet, 2014a; Taxer & Frenzel, 2015a). Furthermore, there is consensus that the teaching job involves intense efforts from teachers to regulate their internal and expressive components of emotions according to the display rules prevalent in teaching contexts – a phenomenon referred to as emotional labor (e.g., Grandey, 2003). However, so far little is known about the typically reported overall levels of experienced emotions of teachers. This knowledge will be essential to depict whether a teacher is up- or downregulating their emotions, and to advance our understanding of when and why teachers engage in emotional labor, and with what implications. Therefore, we present the Teacher

Emotions Scales (TES) as a tool which can be used to measure those baseline scores of typically reported levels of experienced emotions during teaching.

One reason for the conspicuous lack of quantitative studies on discrete teacher emotions is likely the lack of established scales with demonstrated measurement qualities. In this paper, we present a newly developed self-report instrument for the assessment of teachers' emotional experiences. We present two language versions (German and English) of the TES and provide data on their reliability, validity, and cross-language equivalence. Multiple studies were conducted in order to demonstrate replicability of findings.

1.1 Conceptual Framework for Emotions

In defining teachers' emotions, we follow an established multi-componential conceptualization as brought forward for example by Scherer (e.g., 1984, 2009). This implies that emotions are understood as encompassing cognitive, physiological, motivational, and expressive components. As such, an emotional episode will not only be characterized by its affective feelings (e.g., anxious), but typically will be accompanied by specific thoughts (e.g., "I might be hurt"), bodily-physiological changes (e.g., release of adrenal-medullary hormones), action tendencies including approach versus avoidance and fight versus flight (e.g., the urge to leave the situation), and expressive behavior (e.g., wide open eyes, cowered posture).

In explicating how emotions are caused within teachers, we follow a framework which is closely linked to the componential definition of emotions, namely appraisal theory (e.g., Moors, Ellsworth, Scherer, & Frijda, 2013). Appraisal theory posits that it is not an event itself that causes an emotion, but rather the individual's judgment (i.e., appraisal) of the event. Several appraisal dimensions have been suggested (including e.g. event novelty, goal congruence, or controllability) which, in combination, are assumed to determine the intensity and quality of the emotional reaction. This componential, appraisal-based framework to defining and understanding the causes of emotions provided the theoretical backbone for our

deliberations regarding the construction of our instrument as well as the assumptions we made regarding external validity of the TES.

1.2 Rationales for Test Construction

For constructing a self-report instrument aimed at assessing emotional experiences, some key conceptual issues have to be addressed (Ekman & Davidson, 1994). We considered three key issues when constructing the TES: (1) trait versus state emotions, (2) discrete emotions versus a dimensional approach, (3) the selection of the target constructs, and (4) the level of specificity for assessment.

1.2.1 State versus trait. In emotion research, it is common to differentiate between state versus trait approaches (for a conceptual overview, see e.g., Lazarus, 1994), with important implications for assessment. In a state approach to measuring emotions, participants' momentary experiences are assessed in situ, asking participants to report about their affective experiences as they occur in a given moment. For example, researchers may ask, *How do you feel right now?* and require participants to indicate their agreement with items such as, *I feel happy/anxious/angry*. In a trait approach to measuring emotions, participants' judgments regarding the "typical" levels of their emotional experiences are assessed retrospectively. In this instance, researchers may ask, *How do you generally feel?* and require participants to indicate their agreement with items such as, *I generally felt happy (or anxious or angry)*.

For the TES as presented herein, we chose a trait approach, for several reasons. First of all, past research has shown that people's reports of their trait emotions predict decision-making more so than reports of their state emotions (Conner & Barrett, 2012; Wirtz, Kruger, Scollon, & Diener, 2003). As such, it can reasonably be assumed that teachers' decisions with respect to their long-term professional development (including intentions to quit, or engaging in further education), but also their health (e.g., burnout) are systematically linked with teachers' trait emotions, rather than their momentary emotional episodes. In line

with this, a large majority of the existing research on teacher emotions is trait- rather than state-based and thus, there seems to be a need for trait-based teacher emotion scales. For the assessment of students' achievement emotions, such trait-based scales have been successful in past research (Pekrun, Goetz, Frenzel, & Perry, 2011; Pekrun et al., 2004). However, the instrument can also be used to assess state emotions by adapting the items and instructions accordingly.

1.2.2 Discrete emotions versus dimensional approach. Emotions, when considered through dimensions, are viewed as psychological experiences characterized by specific values along different dimensions. The most common considerations have been valence and arousal (e.g., Tellegen, Watson, & Clark, 1999; Watson, Wiese, Vaidya, & Tellegen, 1999), where researchers for example ask participants how they feel on a scale from “unpleasant” to “pleasant,” (valence) and on a scale from “calm” to “jittery” (arousal). Within this approach, participants' affective experiences are thus described by their location in a valence-arousal space.

By contrast, the discrete emotions approach relies on a more fine-grained categorization that is reflected in semantic fields for emotion in natural languages (Scherer, 2013), in distinct nonverbal behaviors (e.g., facial expression, cf. Ekman, 1982) and in specific appraisal patterns associated with different emotional experiences (e.g., Smith & Lazarus, 1993). Here participants' emotions would be assessed by their ratings of intensity or frequency for distinct emotional experiences (e.g., *I am anxious*; *I feel bored*).

Our multi-componential, appraisal-based conceptualization of emotions implied that we used a discrete approach for the TES. This approach is sensitive to emotional experiences which may be close in the valence-arousal space but quite different with respect to their subjective phenomenology, their underlying appraisal patterns, and their associated motivational tendencies and nonverbal behaviors. For example, anger and anxiety are both negative in valence and high in arousal, but the appraisal patterns differ. Anger appraisal is

typically ‘goal obstacle’ coupled with ‘other accountability’ (e.g., Kuppens, Van Mechelen, Smits, & De Boeck, 2003), whereas anxiety typically occurs when a situation is appraised as ‘potentially threatening’ and having a ‘low personal capability to control the situation’ (e.g., Pekrun, 2006). Furthermore, anger is typically associated with fight tendencies whereas anxiety is associated with flight tendencies (Carver & Harmon-Jones, 2009; Lerner & Keltner, 2001). Therefore, in order to account for the fine-grained differences between various emotions when considering their multi-componential nature and underlying appraisal patterns, we designed the TES such that it targets discrete emotions.

1.2.3 The TES emotions: Selection and conceptualization. As a consequence of choosing a discrete emotions approach, a selection had to be made as to which discrete emotions to include. In order to ensure instrument economy and reliability, we sought to construct multi-item scales for a small number of discrete emotions (enjoyment, anger, and anxiety), both positive and negative in valence. We based our decision to include enjoyment, anger, and anxiety on (1) frequency and relevance criteria, and (2) the clear distinctness of selected emotions with respect to their componential definition, their appraisal patterns, their subjective phenomenology, and their semantic meaning even for laypersons.

With respect to frequencies of emotions in human life in general, enjoyment (or happiness) and anger outrank all other discrete emotions, as demonstrated by a quasi-representative survey reported by Scherer, Wrانik, Sangsue, Tran, and Scherer (2004). Reviews of the qualitative literature on the teaching context in particular, revealed that enjoyment can be considered the most salient positive emotion (Frenzel, 2014; Sutton & Wheatley, 2003). Anger, in turn, proves to be the most prominent negative emotion for teachers as documented in qualitative and narrative research on teachers’ emotions (Chang, 2013; Frenzel, 2014; Sutton, 2007; Sutton & Wheatley, 2003). Confirming this, Hagenauer and Volet’s (2014a) interviews with university teachers revealed joy, happiness, and satisfaction as the most frequently mentioned positive emotions, and annoyance, frustration,

and anger as the most frequently mentioned negative emotions. Among the limited studies that used real-life assessment methods to explore frequencies of discrete emotions during teaching, enjoyment was also revealed as the most frequent positive emotion, and anger as the most frequent negative emotion (Carson, 2006; Frenzel & Goetz, 2007; Goetz et al., 2015; Keller, Frenzel, et al., 2014).

In addition to enjoyment and anger, we included anxiety because this emotion has received considerable research interest most generally, but also in the teaching context (Chang, 2009; Darby, 2008; Hart, 1987; Payne & Manning, 1990; Sutton & Wheatley, 2003) – most likely due to its high relevance with respect to physical health and psychological well-being (e.g., Schonfeld et al., 1997). Confirming the relevance of anxiety, in Hagenauer and Volet's (2014a) recent interview study with university teachers, feelings of insecurity, worry/concern, and anxiety/nervousness were also mentioned as frequently by the participants as anger and annoyance. Other discrete emotions including sadness, disappointment, pride, shame, or boredom were mentioned by far less frequently.

Apart from their high relevance and frequency, we also chose to include enjoyment, anger, and anxiety into our TES because they are clearly separable with respect to their componential definition, their appraisal patterns, their subjective phenomenology, and their semantic meaning even for laypersons. Feelings of enjoyment should be characterized by high subjective pleasantness, approach motivational tendencies, open facial and gestural expressions, and caused by appraisals that involve goal congruence and goal conduciveness, coupled with appraisals of controllability and personal agency. The semantic field for this emotion should contain: joy/enjoy, fun, happy/happiness, and enthusiasm/enthusiastic. Anger should be characterized by unpleasant feelings, motivational tendencies that involve “fighting” (while it is debatable whether this is part of the aversive motivational system or approach-related; Carver & Harmon-Jones, 2009), aggressive facial and gestural expressions, and appraisals of goal blockage and other-accountability. The semantic field for this emotion

should contain: anger/angry, annoy/annoyed, mad, and frustrate. Anxiety, in turn, should be characterized by unpleasant feelings, coupled with avoidance motivational tendencies, worry cognitions, defensive facial and gestural expressions, and appraisals of threat and low personal coping capabilities. The semantic field for this emotion should contain: anxiety/anxious, tense, nervous, worry/worried, and uneasy.

1.2.4 Degree of specificity. When assessing trait affective experiences, researchers have adopted varying levels of specificity, ranging from entirely general (“In my life, I generally feel ...”), to context specific (e.g., “During work, I feel...”), or activity-specific (e.g., “When writing papers, I feel...”; see also Goetz, Hall, Frenzel, & Pekrun, 2006). Which level of specificity is optimal is a complex question that has been debated heatedly, for example, for the construct of (teacher) self-efficacy (Klassen, Tze, Betts, & Gordon, 2011). From the perspective of internal validity, assessing constructs at a more general level is debatable when the construct in fact strongly varies across contexts (see Frenzel, Becker-Kurz, Pekrun, & Goetz, 2015). In turn, “microscopically operationalized” measures have been criticized for low practical utility (Pajares, 1996).

With the goal of developing a measure for emotions related to teaching, we have already decided to address a rather specific context, namely teaching. Nevertheless, even within this context, it might make sense to differentiate even further. For students’ achievement emotions, it has been shown that they are considerably domain-specific; that is, they strongly vary within students depending on the subject covered in class (Goetz, Frenzel, Pekrun, & Hall, 2006; Goetz, Frenzel, Pekrun, Hall, & Lüdtke, 2007; Goetz, Pekrun, Hall, & Haag, 2006). As such, for teachers’ emotions as well, it might make sense to differentiate between different domains. Indeed, Frenzel and colleagues (2015) have shown mean levels of emotional experiences systematically vary within teachers depending on the subject of instruction. In addition, for teaching contexts in particular, there is empirical evidence that the group of students taught also seems to play an important role for emotional experiences

during instruction (Frenzel et al., 2015; Kunter, Frenzel, Nagy, Baumert, & Pekrun, 2011).

This research has shown that there is systematic mean-level variability of emotional experiences within teachers caused by different groups of students taught, and that class characteristics (e.g., disruption and student enjoyment as reported by the students) are systematically related to teacher-reported levels of emotional experiences. Therefore, it was concluded that teacher emotional experiences are specific to the particular group of students taught and that this specificity should be taken into account when assessing teachers' experiences during teaching.

Given that research on the optimal level of specificity for the assessment of teacher emotions is scarce, we decided to explore this issue as part of our instrument validation efforts. To this end, we constructed two different variants (general and specific) of each item. General was captured through items such as "I generally enjoy teaching." For the specific variant, we chose to identify the group of students taught because that appears to be a particularly meaningful distinction for teachers. The corresponding item in the specific variant would thus be, "I enjoy teaching *these students*". Overall, this resulted in six scales, one general and one student-group specific scale for each of the three emotions. As such, our analyses also aimed to validate the usefulness of the TES at different levels of specificity (general versus group-specific).

1.3 Scale Development and Translation

Items were initially developed in German, with formulations inspired by teacher narrative reports documented in qualitative studies (Bibby, 2002; Chang, 2009; Darby, 2008; Hargreaves & Tucker, 1991; Hargreaves, 1998, 2005; Hart, 1987; Payne & Manning, 1990; Prawat, Byers, & Anderson, 1983; Sutton, 2007; Sutton & Harper, 2009; Sutton & Wheatley, 2003; Thienel, 1988; van Veen, Seegers, & van de Ven, 2005; Zembylas, 2003). In addition, we were influenced by instruments for the general assessment of emotions (e.g., the Positive and Negative Affect Schedule Watson, Clark, & Tellegen, 1988) and instruments targeting

emotions in the context of learning and achievement, including the State-Trait Anxiety Inventory (Spielberger, 1983), the Test Emotions Questionnaire (Pekrun et al., 2004), and the Achievement Emotions Questionnaire (Pekrun et al., 2011).

From our initial item pool (37 items), items were selected for preliminary versions of the scales using expert judgment and criteria of semantic redundancy. Selection of items for the final scales was based on exploratory factor analyses and item and scale statistics from preliminary versions. Items for the final scales (four items for each for the three emotions) were selected according to convergent item validity (i.e., high factor loadings on the relevant scale) as well as divergent item validity (i.e., low factor loadings on other emotion scales).

Next, the German TES was translated into English independently by two different bilingual experts in the field of emotion research (one native German-speaker fluent in English, one native English-speaker fluent in German). Differences in their translations were discussed until an agreed-upon preliminary English version was created. Items were then back-translated by a third bilingual expert (native German-speaker fluent in English), and a final revision of the English scales was conducted within the team to ensure content-related item equivalence.

1.4 Strategy for Scale Analysis

The present research sought to investigate the reliability, internal and external test validity of the TES, while also considering potential biases of the answers to our scales due to social desirability (see Campbell, 1960, on the importance of ruling out substantial relationships with trait-irrelevant method factors including social desirability as part of instrument validation). Concerning the internal structure of the TES, our decision to integrate three discrete emotions for the TES implied that the instrument should reflect a three-factor structure (see e.g. Slaney & Maraun, 2008, on the importance to lay out the expected formal structure of a test as a prerequisite to testing its internal test validity). Therefore, we assessed internal test validity of the TES by evaluating whether the resultant data could be better

described by a three-factor model than a two-factor model (i.e., differentiating only positive vs. negative affect) or a single-factor model (i.e., denoting one-dimensional emotion factor; see Figure 1 for a visual depiction of the three models). To compare these competing models, we used confirmatory factor analysis.

Regarding external linkages, we examined the relations between teacher emotions and teacher self-reported general affect, burnout, job satisfaction, self-efficacy, and student ratings of teaching behaviors. Specifically, we included scales related to three commonly agreed upon basic dimensions of teaching quality, namely, cognitive quality, effective classroom management, and supportive climate (e.g. Baumert et al., 2010; Klieme, Pauli, & Reusser, 2009; Pianta & Hamre, 2009). Overall, we expected positive relationships between enjoyment and positive affect, job satisfaction, and teacher self-efficacy, as well as between anger/anxiety and negative affect and burnout. In turn, we expected negative relationships between enjoyment, negative affect and burnout, as well as between anger/anxiety and positive affect, job satisfaction and teacher self-efficacy. We expected these relationships to be sufficiently high to indicate validity of the TES, yet low enough to warrant conceptual separation between those established phenomena and the three discrete teacher emotions assessed by the TES.

Drawing upon general psychological insights into the effects of human emotions on behavior (such as, from positive psychology and mood research, see e.g. Fredrickson & Branigan, 2005; Isen, 2008; Mitchell & Phillips, 2007), and on corresponding frequently mentioned claims that teacher emotions are linked with teaching quality (Frenzel, 2014; Hagenauer & Volet, 2014b; Hargreaves, 2000; Sutton & Wheatley, 2003) we also proposed that there were systematic links between teachers' emotions and their instructional behavior. Thus, we expected enjoyment to be positively related to desirable teaching behaviors (such as clarity of instruction, withitness, or teacher caring), and negatively related to undesirable teaching behaviors (such as fast paced instruction, or teacher disrespect). For anger and anxiety, in

turn, we expected positive relationships with undesirable teaching behaviors and negative relationships with desirable teaching behaviors.

1.5 Research Questions

In sum this research aimed at answering the following questions:

- 1) Are the two variants of the TES reliable, as documented in analyses of internal consistency?
- 2) Are the two variants of the TES internally valid, as documented by confirmatory factor analyses (CFA) showing that three-factor models (i.e., three latent emotion factors) fit the data better than either two-factor (positive vs. negative affect) or single emotion factor models?
- 3) Are the two variants of the TES externally valid, as documented by theoretically meaningful relations with teacher self-reported general affect, burnout, job satisfaction, teaching self-efficacy, and student ratings of teaching behaviors?
- 4) Are the German and English language versions of the TES equivalent, as documented by measurement invariance across German and North American teacher samples?
- 5) Do the findings on reliability, internal and external validity replicate across studies, specifically across German and North American teacher samples?

Four studies were conducted to answer these questions. In Study 1, we examined the reliability and internal validity of the TES with a sample of German teachers from different types of schools (research questions 1 and 2). Study 2 aimed to investigate external validity, using two separate samples, a primary and a secondary school teacher sample (research question 3). Study 3 included a sample of Canadian teachers from various school types to validate the English version of the TES (research questions 4 and 5).

2. Study 1: Reliability and Internal Validity

Study 1 was designed to provide data on the reliability and internal structure of the general and student-group specific variants of the German TES.

2.1 Method

2.1.1 Sample and procedure. This study's sample consisted of 414 teachers (257 female; 155 male; 2 unspecified) from the German state of Bavaria. Participants ranged in age from 25 to 64 years ($M = 43$ years, $SD = 11$ years) with between 1 and 42 years of teaching experience ($M = 16$ years; $SD = 11$ years). Of these teachers, the majority worked in the German three-tiered secondary state school system (33.3% low, 25.6% middle, and 17.2% high) while 23.9% worked at primary school.

Contacts to schools were made through principals. Once permission was received, we placed the surveys in the individual mailboxes of each teacher. The survey cover letter asked for their voluntary participation with assurance of full confidentiality of their data. Instructions provided asked the teachers to fill out the survey and to return the completed survey in a sealed envelope to the secretary for pick-up by the research team.

2.1.2 Measures. Both the general and student-group specific variant of the German TES were administered in this study (see Appendix A for list of items). Items were answered on a four-point Likert Scale labeled with *strongly disagree* (*trifft nicht zu*), *disagree* (*trifft eher nicht zu*), *agree* (*trifft eher zu*), *strongly agree* (*trifft zu*). We used a 4-point scale to avoid a middle category participants can choose simply to indicate "I don't know/I don't want to think about this" (e.g., Kulas & Stachowski, 2009), and to reduce the influence of response sets such as social desirability (e.g., Garland, 1991). The 4-point scale also made it possible to meaningfully verbally label each of the answers, which has been shown to increase the psychometric quality of scales (Weng, 2004). Within the questionnaire, all 12 items for the three emotions were presented in random order. Demographic items were added between the general and the group-specific variants of the TES within the questionnaire.

While the general scales began with a general introduction (*Below you find a list of statements describing your experiences as a teacher. Please indicate your personal response to each of these statements by circling the number that best represents your answer*) the student-group specific scales were introduced with two different introductory sentences, depending on the teaching policy that was pursued in the corresponding school. If a classroom teacher policy was pursued (i.e., one teacher teaches all or most of the subjects to only one group of students per school year; in Germany this applies to primary and low-track secondary schools), the student-group specific scales were introduced with “When answering the items below, please think of the students in your class this year.” In contrast, if a subject teaching policy was pursued (i.e., teachers teach one or more subjects to several different groups of students per school year; in Germany this applies middle- and high-track secondary schools), the student-group specific items were introduced with, “When answering the items below, please think of your typical Tuesday morning. Envision yourself walking into your first class on Tuesday morning and think of the students in this particular class when answering the following items” (see Jiang, Spölte, & Lupescu, 2015, for a similar procedure to contextualize teacher ratings).

2.2 Results and Discussion

2.2.1 Descriptive statistics and reliability of the TES. Table 1 shows means, standard deviations, and internal consistencies (Cronbach’s Alphas) for standardized items of the general and student-group specific variants of the TES. Mean ratings for enjoyment were relatively high (> 3 on the four-point scale), and anger and anxiety ratings were relatively low (< 2) for both the general and the student-group specific scales. However standard deviations were sufficiently large to preclude ceiling or floor effects. Internal consistencies of the general scales were good, with Cronbach’s Alphas ranging between .70 (anxiety) and .77 (enjoyment). For the student-group specific scales, internal consistencies were very good, with Cronbach’s Alphas approaching .90.

Manifest intercorrelations of the enjoyment, anger, and anxiety scales are depicted in Table 2. As expected, enjoyment was negatively related to both anger and anxiety (r s ranging around $-.40$ for general and $-.65$ for student-group specific scales), which in turn were positively related to each other ($r = .42$ for general and $.75$ for student-group specific scales). Additionally, for each emotion, the general and student-group specific scales proved to be quite highly positively correlated with one another (r s ranging around $.50$).

2.2.2. Internal validity of the TES. We inspected the internal validity of the TES by means of confirmatory factor analyses using the *Mplus* software Package Version 7.31 (Muthén & Muthén, 1998-2010). Since fit indices are differentially sensitive to different types of model misspecification (Hu & Bentler, 1998), we inspected a range of fit indices including the comparative fit index (CFI), the root-mean-square error of approximation (RMSEA), and the standardized root-mean-square residual (SRMR). We applied the typical rules of thumb proposing that model fit is acceptable with a CFI of close to $.95$ or higher, an RMSEA and SRMR of close to $.08$ or lower (see Browne & Cudeck, 1993; Hu & Bentler, 1999), while bearing in mind that some consider these cut-off criteria as too strict, particularly for complex data structures with large sample sizes (Marsh, Hau, & Wen, 2004).

As can be seen in Table 3, the three-factor model for the student-group specific scales had a very good fit, and also the fit of the three-factor model for the general scales was acceptable (with only the CFI failing the recommended fit criteria). All other models did not meet the recommended model fit criteria for at least two of the three fit indices. Since the three models were nested, we also performed chi-square difference tests (by inspecting whether the p -value for $\Delta\text{Chi-square}/\Delta\text{df}$ was significant) for the comparison of the two-factor versus single-factor model, as well as the two-factor versus three-factor models. Each of these comparisons was highly significant ($p < .01$). This underscored the superiority of the three-factor models over the single- or two-factor models. Our confirmatory factor analyses also yielded latent correlations between the three emotion factors (see Table 2). Due to correction

for unreliability, these correlations were generally high (almost .60 for the general scales, and reaching above .80 for the student-group specific scales), however they were still low enough to warrant conceptual separation between the three discrete emotions as assessed with the TES.

3. Study 2: External Validity

Study 2 was designed to explore the external linkages of the German TES with teacher self-reported variables as well as student-reported perceptions of teacher behavior. This study had a focus on the particular teacher-class relationship and therefore used only the student-group specific scales. Two separate samples were recruited for this study. In the first sample, linkages with general affect as well as burnout and social desirability were explored. With the second sample, we sought to replicate these external linkages, while additionally exploring relationships with job satisfaction and self-efficacy, and with teaching behaviors as observed by students.

3.1 Method

3.1.1. Samples and procedures. Sample 1 included 85 primary school teachers with 1 to 40 years of teaching experience ($M = 18$ years, $SD = 12$ years) from the German state of Bavaria. Participants were predominantly female (95.3%; this corresponds with the gender rate among primary teachers in Bavaria; Bavarian State Ministry of Education, 2013) and ranged in age from 23 to 64 years ($M = 43$ years; $SD = 12$ years). Sample 2 stems from a large-scale study on emotion processes in the classroom that involved both teachers and students. The study design implied a 1:1 match between teachers and one single group of students they taught in that school year, thus data are available on the student-group specific teaching behaviors of the teachers in this sample, as measured by aggregated student perceptions. It included teachers ($N = 68$) with their students ($N = 1,566$) from middle (55%) and upper track secondary schools (45%) in the German state of Bavaria. The majority of teachers (76.8%) were female with an average age of 40 years (Min/Max = 24/65 years; $SD =$

12 years) with an average of 18 years teaching experience (Min/Max = 1/42 years; $SD = 12$ years). The student sample consisted of 57.1% female participants from grades 5 through 10 (Mean age = 13.4 years, Min/Max= 10/20 years, $SD = 1.6$ years).

3.1.2 Measures. Participants from the first sample were asked to fill out the student-group specific variant of the TES as well as several other scales for the external validation of the emotion scales. The TES items were again presented in a random order. Positive and negative affect were measured with the full-length German version of Watson and colleagues' (1988) "Positive And Negative Affect Schedule" (PANAS, Krohne, Egloff, Kohlmann, & Tausch, 1996), and the three facets of teacher burnout (emotional exhaustion, depersonalization, personal accomplishment) were measured by Enzmann and Kleiber's (1989) German translation of Maslach's burnout inventory (Maslach & Jackson, 1986). To assess social desirability, we used Musch, Brockhaus and Broeder's (2002) German translation of Paulhus' (1994) "Balanced Inventory of Desirable Responding" (BIDR), which comprises subscales for self-deception and impression management. With the exception of the impression management subscale ($\text{Alpha} = .51$), the internal consistencies for the measurement of these validating constructs were satisfactory to good (see Appendix B).

In sample 2, positive and negative affect were again assessed as measured with Krohne et al.'s (1996) German full-length version of Watson et al.'s (1988) Positive and Negative Affect Schedule (PANAS) and emotional exhaustion measured with Maslach and Jackson's (1986) burnout inventory (German adaptation by Böhm-Kasper, Bos, Jaeckel, & Weishaupt, 2000), and social desirability, which for reasons of space limit in the teacher questionnaire was measured only with the impression management subscale (Musch et al., 2002). Additionally, in this sample, teacher job satisfaction adapted from Böhm et al. (2000) and teacher self-efficacy as measured by a student-group specific adaptation of the Teachers' sense of efficacy scale (Tschannen-Moran & Hoy, 2001) was included. The internal consistencies (presented through Appendix B) for these scales were satisfactory to good.

In addition, in sample 2, we assessed different facets of teacher behavior using established scales from two German large-scale studies on classroom processes (Professional Competence of Teachers, Cognitively Activating Instruction, and Development of Students' Mathematical Literacy, COACTIV, cf. Brunner et al., 2006; and Project for the analysis of learning mathematics, PALMA, cf. Pekrun et al., 2007). The scales referred to three commonly agreed-upon basic dimensions of teaching quality: cognitive quality of instruction, effective classroom management, and supportive climate (Baumert et al., 2010; Klieme et al., 2009; Pianta & Hamre, 2009). Specifically, for cognitive quality of instruction, we assessed student perceptions of clarity of instruction (e.g., "Our teacher can explain things really well") and variety in instruction (e.g., "Our teacher provides varied instruction"). For classroom management, we assessed withitness as a desirable behavior (e.g., "Our teacher always knows exactly what's going on in class") and fast-paced instruction as an undesirable one (e.g., "Our teacher gives us little time to reflect on questions. For supportive climate, we assessed acceptance of errors (e.g., "Our teacher is patient when a student gives a wrong answer"), teacher caring (e.g., "Our teacher is available if I need to talk to him"), and teacher support after failure (e.g., "If I had a bad grade, my teacher tries to find the reasons together with me") as desirable teaching behaviors, and teacher disrespect (e.g., "Our teacher treats students condescendingly") as an undesirable teaching behavior. As presented through Appendix C, all scales consisted of 3 or 4 items and showed acceptable internal consistency. More importantly, students' answers were highly homogeneous within classes, as indicated by high second-order intra-class-correlations (ICC(2); see Appendix C, cf. Lüdtke, Trautwein, Kunter, & Baumert, 2006, for the calculation and use of the ICC(2) for assessing reliabilities of class-aggregated ratings).

3.2 Results and Discussion

3.2.1 Descriptive statistics and reliability of the TES. Means, standard deviations, and internal consistencies (i.e., Cronbach's Alphas) for standardized items for Study 2 are

included in Table 1. As found in Study 1, average ratings for enjoyment were generally high (> 3 on the four-point scale), and average anger and anxiety ratings were generally low (however, the average anger scores exceeded 2 for the primary school teachers in the first sample of Study 2). Standard deviations were again sufficiently large to preclude ceiling or floor effects. Internal consistencies were again high for enjoyment and anger in the first sample (primary school teachers) and anger in the second sample (secondary school teachers), with Cronbach's Alphas above .80. They were a little weaker for anxiety in the first sample and enjoyment and anxiety in the second sample, though still above .70. The pattern of correlations between the enjoyment, anger, and anxiety scales was similar to the Study 1 findings (see Table 2). Again, enjoyment was negatively related to both anger and anxiety in both samples. Anger and anxiety were positively related to each other in the first, but not in the second sample.

3.2.2 External validity. To explore the external linkages of the TES, we inspected the correlations between enjoyment, anger, and anxiety and the external constructs that were included in the two samples (see Table 4).

3.2.2.1 Correlations with other teacher variables. Across the two samples and in line with expectations, we found consistent small to medium-sized relationships between teaching enjoyment and general positive affect, and between anger as well as anxiety and general negative affect. With respect to the three dimensions of burnout, data from sample 1 revealed substantial and consistent relationships with teacher emotions (negative correlations for enjoyment; positive correlations for anger and anxiety), with only one exception: anxiety was unrelated to depersonalization. In sample 2, these links between teacher emotions and emotional exhaustion were fully replicated, even though the relationships were weaker than in the first sample.

Furthermore, in the second sample we found significant (yet small) negative relationships of teaching anxiety with job satisfaction and self-efficacy. Inconsistent with

expectations, job satisfaction and teaching efficacy proved to be unrelated to both enjoyment and anger.

Finally, in sample 1, all three emotions were unrelated to impression management, and both enjoyment and anger proved to be unrelated to self-deception. Anxiety showed a weak negative correlation with self-deception. In sample 2, a small positive correlation between enjoyment and impression management was observed, whereas anger and anxiety were unrelated to impression management.

3.2.2.1 Correlations with student ratings. Table 5 presents the bivariate correlations between enjoyment, anger, and anxiety during teaching and group-aggregated student ratings of their teachers' behaviors. Overall, the results clearly corroborated assumptions that the experience of enjoyment during teaching was significantly positively correlated with the desirable teaching behaviors assessed in our study (clarity of instruction, variety in instruction, acceptance of errors, teacher caring, and support after failure) and significantly negatively related to the more undesirable teacher behaviors (fast paced instruction and teacher disrespect towards students). Also in line with expectations, teacher anger was significantly positively correlated with those undesirable teaching behaviors, and negatively related to student perceptions of desirable teacher behaviors with respect to teaching quality (specifically, variety in instruction) and to the quality of their relationship with the teacher (teacher caring and support after failure). For anxiety, all correlations were in the expected directions, yet only the negative relationship between teacher anxiety and acceptance of errors reached significance (likely due to small sample size).

In order to further strengthen the evidence for the incremental validity of the TES over more general affect scales, we also inspected the partial correlations between teacher-reported emotions and student-reported teacher behaviors, controlling for teacher-reported positive affect (for enjoyment) and negative affect (for anger and anxiety). As seen through Table 5, the partial correlational pattern was virtually identical with the bivariate correlation pattern,

with two exceptions. For enjoyment, the partial correlation between enjoyment and clarity of instruction, teacher caring, and teacher support after failure were still substantial, but clearly smaller than the bivariate correlations. In addition, the relationship between anger and student-perceived withitness was increased to a value as high as .30 when controlling for general negative affect. This finding is inconsistent with our expectation, since we had proposed that withitness was a desirable teaching behavior that would correlate positively with enjoyment, but negatively with anger and anxiety. However, our finding rather implies that teachers who reportedly experience more anger in the classroom, above and beyond experiencing more negative affect generally, are perceived by their students as more closely monitoring the classroom, while teacher enjoyment and anxiety proved to be unrelated with student-perceived withitness.

4. Study 3: The English TES

Study 3 was designed to replicate findings on the reliability, internal structure, and external validity of the TES with a Canadian sample using the English version of the TES. In addition, beyond replication, this study sought to assess measurement invariance across the German and English versions of the TES.

4.1 Method

4.1.1 Sample and procedure. We recruited 377 (75.9% female) Canadian teachers over two consecutive days while in attendance at a city-wide mandated teacher convention. Teachers were approached in the convention center and invited to voluntarily complete a paper-and-pencil survey requiring about 15 minutes. More than half of participants (53.6%) were teaching in primary schools (kindergarten to Grade 6) and 44.8% were teaching middle or high school (1.6% unspecified) with an average of 13 years of teaching experience (Min/Max = 0.5/45 years; $SD = 10$ years). Participants ranged in age from 21 to 68 years ($M = 40$ years, $SD = 11$ years), and most (93.7%) were Non-Aboriginal (1.3% Aboriginal; 5% not specified).

4.1.2 Measures. Participants were asked to fill out both the general and the student-group specific variants of the TES. Additionally, teachers were asked to answer the short version of the PANAS scales measuring positive and negative affect (Watson et al., 1988), and the three subscales of Maslach and Jackson's (1986) teacher burnout inventory. The Marlowe-Crown social desirability scale (Fischer & Fick, 1993; Strahan & Gerbasi, 1972) was used to assess social desirability. In the questionnaire, demographic information was assessed first, followed by two separate sections that assessed the PANAS and social desirability. The general TES items and the items of the burnout inventory were then presented (in random order). In the last section of the questionnaire, teachers were asked to judge the group-specific TES items (also presented in random order). Reliabilities for these scales were satisfactory to good (see Appendix B).

4.2 Results and Discussion

4.2.1 Descriptive statistics and reliability. Means, standard deviations, and internal scale consistencies (i.e., Cronbach's Alphas) for standardized items were presented through Table 1. Again, average ratings for enjoyment were generally high (> 3 on the four-point scale), and average anger and anxiety ratings were generally low (< 2). However, standard deviations were again sufficiently large to preclude ceiling or floor effects. Cronbach's Alpha was .73 for the general teaching enjoyment scale, and ranged above .80 for the general anger and anxiety scale. For the student-group specific scales, Cronbach's Alphas were again even higher (reaching .80 or higher) for all three emotions.

Furthermore, the pattern of intercorrelations between the enjoyment, anger, and anxiety scales was equivalent with that observed for the German TES (see Table 2). Again, enjoyment was negatively related to both anger and anxiety (*r*s ranging around -.50 for general and -.60 for student-group specific scales), which in turn were positively related to each other (.50 for general and .69 for student-group specific scales). Furthermore, the positive correlations

between the general and student-group specific scales were a little higher than in the German sample (r s ranging above .70).

4.2.2 Internal validity. The three-factor model for both the general and the student-group specific variants had a good fit (see Table 3). All other models did not meet the recommended model fit criteria for at least two of the three fit indices. We again performed chi-square difference tests (by inspecting whether the p -value for $\Delta\text{Chi-square}/\Delta\text{df}$ was significant) for the comparison of the nested two-factor versus single-factor model, as well as the two-factor versus three-factor models. Each of these comparisons was highly significant ($p < .01$). This underscored the superiority of the three-factor models over the single- or two-factor models for the English TES. Our confirmatory factor analyses also yielded latent correlations between the three emotion factors (see Table 2). As with the German TES, due to correction for unreliability these latent correlations were rather high (.68 and higher), but low enough to warrant conceptual separation between the three discrete emotions measured.

4.2.3 Cross-language equivalence. We inspected measurement invariance of the TES across the Canadian sample (this study) and the German sample (Study 1) by applying a commonly used stepwise procedure (e.g., Chen, 2007; Van de Schoot, Lugtig, & Hox, 2012). This procedure involves inspecting a (1) configural invariance model which assumes that the same item is associated with the same factor in each group, (2) metric invariance model that additionally assumes equivalence of item factor loadings, and (3) scalar invariance model that additionally assumes equality of item intercepts. Configural invariance is required to conclude that two language versions measure an equivalent set of latent variables. Invariance of factor loadings is necessary to be able to compare relationships with other variables across the two language versions. Additional invariance of intercepts is required to compare latent means across groups (e.g., Chen, 2007; Cheung & Rensvold, 2002; Van de Schoot et al., 2012).

In order to judge the severity of loss of model fit when imposing equality constraints, we used Chen's (2007) recommendations. Accordingly, with adequate sample sizes (as in our

studies) for testing loading invariance, a change of .01 or lower in CFI, supplemented by a change of .015 or lower in RMSEA or a change of .030 or lower in SRMR would indicate noninvariance; for testing intercept invariance, a change of close to .010 or lower in CFI, supplemented by a change of close to .015 or lower in RMSEA or a change of close to .010 or lower in SRMR would indicate noninvariance.

Overall, the configural invariance models for both the general and the student-group specific scales had an acceptable fit, indicating that the factor structure can be assumed to be equivalent across the two languages (see Table 6). When imposing equality constraints for the loadings, the loss of model fit was tenable for both the general and the student-group specific scales. For the general scales, Δ RMSEA and Δ SRMR were clearly below the recommended thresholds, and Δ CFI was just above .01. For the student-group specific scales, metric invariance could be established based on all three indicators. However, constraining intercepts to invariance resulted in an untenable loss of fit for both the general and the student-group specific scales. To explore the severity of the lack of intercept invariance for each emotion separately, we ran three partial invariance models, setting each emotion's intercepts to equality at a time (see Models 3a to 3c in Table 6). This revealed that the lack of intercept invariance was particularly pronounced for anger and anxiety, and less pronounced for enjoyment.

A closer inspection of the intercept differences at the item level revealed that for enjoyment, those differences were all of similar size and inconsistent in direction. For anxiety though, there was a general tendency for intercepts in the English version to be higher as compared to the German version, with two items sticking out in particular. Those were, "I am often worried that my teaching isn't going so well" and "Preparing to teach often causes me to worry." These items share the term "worry" which in English thus seems to be more easily agreed with than the German "sich Sorgen machen." Finally, for anger, two items had clearly higher intercepts in the German than in the English language version ("I often have reason to

be angry while I teach these students” and “Sometimes I get really mad at these students”). Thus the terms “to be angry” and “get mad” seem to be harder to agree with than the German terms “sich ärgern” and “sauer werden.” By contrast, one anger item had a clearly higher intercept in the English language version (“Teaching these students frustrates me”), thus “frustrate” seems to be more easily agreed with than “frustrieren” in German.

4.2.4 External validity. Correlations between the general and student-group specific TES and the external constructs are included in Table 4. The pattern of correlations between the teacher emotions and general affect and burnout was strikingly similar to that obtained in the German samples, and the patterns of relationships of the external constructs were equivalent to those of the general and student-group specific English variants of the TES. There were again small to medium-sized relationships between teaching enjoyment and general positive affect, and substantial and consistent relationships between the emotions and all three burnout dimensions (negative correlations for enjoyment; positive correlations for anger and anxiety). In addition, social desirability proved to be weakly positively related to enjoyment, and weakly negatively related to anxiety and anger.

5. General Discussion

The current research provided evidence on the reliability and validity of a newly developed self-report instrument to assess three emotions considered most relevant in the context of teaching, generally or specific to a student group, in two languages (German and English): enjoyment, anger, and anxiety. So far, instruments for the quantitative assessment of discrete teacher emotions have been grossly lacking. As such, this research provides a first systematic measurement instrument on teacher emotions. The instrument is easy to administer, shows metric invariance across its German and English language versions, has strong psychometric properties in terms of reliability and internal validity, and demonstrates strong external validity. These findings were robust across multiple samples, as discussed below.

5.1 Reliability and Internal Validity

All variants of the TES, including both the general and the student-group specific scales and both the German and English language versions, demonstrated acceptable to very good reliabilities in terms of Cronbach's Alphas, with values ranging between .70 and .92. In addition, the confirmatory factor analyses revealed more detailed information on the internal validity of the different variants of the TES. Specifically, our discrete emotion approach proved to be valid in that the three-factor model (including separate latent variables for enjoyment, anger, and anxiety) was clearly superior to either the single-factor model (with a unidimensional "emotion" factor) or the two-factor model (with a positive versus negative affect factor; see Figure 1 for the three models). As such, the TES is clearly suitable for assessing teaching enjoyment, anger and anxiety as distinct constructs. The three emotions are correlated, yet they are clearly conceptually and statistically separable. Furthermore, the three emotions show differential relationships with other constructs.

Acknowledging that the level of specificity for a self-report instrument like the TES is an important issue (e.g., Klassen et al., 2011; Pajares, 1996, on this debate for the construct of self-efficacy), we further sought to explore the utility of the TES at two levels of specificity. Drawing upon findings that the particular group of students seems to be an important contextual factor for how teachers feel during teaching (Frenzel et al., 2015; Kunter et al., 2011), we constructed both general teaching scales and student-group specific variants of the TES. Overall, both variants of the TES display good psychometric properties and a consistent internal three-factor structure. Correlations between the general and student-group specific scales were generally high (around .50 in the German sample, and around .70 in the Canadian sample). The internal consistencies (Cronbach's Alphas) of the student-group specific scales and the correlations between these scales tended to be higher than those of the general scales, and the fit of three-factor CFA models tended to be better for the general scales than for the student-group specific scales, particularly for the German language TES.

As such, attending to the social context in terms of the specific student group indeed seems to be a quite meaningful way to contextualize teacher emotional experiences during teaching. Future studies may also call for a contextualization that takes into account the particular subject taught. To this end, it will be easy to adapt the TES for creating subject specific variants (e.g., “I enjoy teaching Math”; “I often have reasons to be angry while I teach English”; “I feel tense and nervous when I teach Chemistry”), in case researchers seek to consider specific subjects in their research. Indeed, for students’ achievement emotions, it has been shown that they are considerably domain-specific (Goetz, Frenzel, et al., 2006; Goetz et al., 2007; Goetz, Pekrun, et al., 2006), and also for teachers, there seems to be considerable within-person variability due to the subject taught (Frenzel et al., 2015).

5.2 Cross-Language Equivalence

For both the general and the student-group specific scales, the findings suggest both configural invariance and metric invariance (i.e., equality of factor loadings) across the German and English versions of the TES. As such, it can be concluded that the two language versions measure an equivalent set of latent variables, and relationships with other variables can be compared across samples that used the two language versions.

We could, however, not establish invariance of intercepts across the German and English language versions of the TES. This implies that mean-level comparisons across samples that used the two language versions should be done cautiously (Chen, 2007; Cheung & Rensvold, 2002). Overall, the cross-language equivalence of the TES is satisfactory, considering that translations for content as delicate as emotions are particularly challenging (Fontaine, Scherer, & Soriano, 2013; Spielberger, 2006).

5.3 External Validity

The external validation provided evidence that the TES shows consistent external validity in terms of correlations with related constructs including general positive and negative affect, burnout, job satisfaction, and teacher self-efficacy. As expected, teaching

enjoyment was positively linked with general positive affect, and negatively linked with general negative affect, emotional exhaustion, lack of accomplishment, and depersonalization. In contrast, teaching anger and anxiety were negatively linked with general positive affect, and positively linked with general negative affect, emotional exhaustion, lack of accomplishment, and depersonalization. Generally, these relationships were consistent in their direction, yet small enough in size to warrant clear conceptual separation of discrete emotions as experienced by teachers and those related constructs. Contrary to expectations, we found that neither enjoyment nor anger were systematically linked with job satisfaction or teacher self-efficacy. However, teaching anxiety showed clear negative relationships with these constructs. This suggests that teacher anxiety, despite overall rather low means, is an important construct to consider because of its relationship with important teacher outcomes.

Finally, we found consistent relationships with student ratings of teaching behaviors, even when controlling for general positive and negative affect. It is important to note that this is particularly strong evidence for external validity since these findings are not threatened by single-source bias. Overall, these findings corroborate claims that “happier teachers do provide better teaching” (Taxer & Frenzel, 2015a, p. 86), because there were positive relationships between teacher enjoyment and desirable teaching behaviors including clarity of instruction, variety in instruction, acceptance of errors, teacher caring and support after failure. Moreover, we found negative relationships between teacher enjoyment and more undesirable teacher behaviors including fast paced instruction and teacher disrespect towards students. Additionally, our data point to the fact that the more anger teachers report, the less variation their students experience in their instruction, and the less caring and support their students experience to receive from them, but the more their students perceive the instruction to be fast paced and the teacher to display disrespect. Counter to expectations, anger also was positively related to witness as perceived by the students, which we would have categorized as a desirable teaching behavior. This effect was revealed particularly once

general negative affect was controlled for. As such, it seems to be the specific nature of anger (but not any other, related negative emotions such as anxiety, disappointment, or hopelessness) which positively relates to student-perceived withitness. We speculate that when teachers express their experiences of anger to students, students perceive these teachers as “strict”, and maybe therefore also feel closely monitored by them. This interpretation is in line with existing teacher interview data which suggests that anger expressions in the classroom can help to enforce rules or to make the seriousness of an infraction clear to students (Sutton, Mudrey-Camino, & Knight, 2009; Winograd, 2003).

Relationships between teacher anxiety and student ratings of desirable versus undesirable teacher behaviors were similarly consistent, however comparably small in size. In sum, our data proposes that teaching anxiety is a threat more to teachers’ well-being rather than to the quality of their teaching. In contrast, teaching enjoyment is positively linked with effective teaching, but not necessarily with being satisfied with the teaching job or feeling particularly efficacious when teaching. Future research will be needed to replicate and refine these findings, particularly with respect to their causal ordering and long-term effects. It is important to note that these findings on the external validity of the TES were robust across multiple studies and two cultural contexts (German versus Canadian). We conclude that both the English and the German language versions of the TES demonstrate external validity.

Finally, in some of our studies, we observed a tendency for enjoyment to correlate slightly positively and anger and anxiety to correlate slightly negatively with social desirability. Similar findings have been reported for student reports about anxiety (e.g. Zeidner, 1998) and measures of general well-being (Huebner, Laughlin, Ash, & Gilman, 1998). For well-being it has been suggested that positive correlations with social desirability measures do not necessarily suggest a bias, but rather reflect an underlying personality characteristic that contributes to positive life satisfaction (Diener, Sandvik, Pavot, & Gallagher, 1991). Overall, the present findings suggest that social desirability is not a serious

threat to the validity of the TES. Yet, these findings are consistent with the idea that there are emotion display rules for teachers, for example that emotions considered ‘too strong’ should generally not be displayed, and that positive emotions should be expressed, whereas negative emotions should be suppressed. The existence of such display rules for teacher emotions has been addressed in qualitative research (Schutz, Cross, Hong, & Osbon, 2007; Sutton, 2004; Winograd, 2003; Zembylas, 2005), but, to the authors’ knowledge, not yet tested quantitatively.

5.4 Limitations and Implications for Future Research

For the TES as presented herein, we chose a trait approach, which should be suitable to assess teachers’ emotions in a non-intrusive and economical way. We are aware of and point any potential user of the instrument to the fact that such trait-based retrospective judgments may be influenced not only by the “true scores” of participants’ emotional experiences, but also by their beliefs about their emotions (Goetz, Bieg, Lüdtke, Pekrun, & Hall, 2013; Robinson & Clore, 2002). Nevertheless, we are convinced that such trait-like emotional experiences are highly relevant for general well-being and also have important behavioral implications (e.g., teachers’ intentions to quit, Klassen & Chiu, 2011) – in fact, some researchers have argued that trait emotions matter even more than state emotions, for example with respect to decision making (Wirtz et al., 2003).

Notably, trait-based scales can easily be adapted for state assessments (Gogol et al., 2014), and we suggest that the TES can be adapted in a similar fashion. For example, teachers would be asked “How do you feel at the moment?” and corresponding TES items would then be, “I enjoy teaching/I feel enthusiastic/I feel annoyed/I have reasons to get angry/I feel tense and nervous/I am worried that my teaching isn’t going so well.” Such in situ approaches to assessing teacher emotions have been applied successfully in scattered past studies (Carson, Weiss, & Templin, 2010; Goetz et al., 2015; Keller, Chang, Becker, Goetz, & Frenzel, 2014).

A less intrusive, yet still more highly situated approach to assessing teacher emotions would be a teaching diary, which involves teachers judging their emotions right after a lesson (e.g., Becker, Keller, Goetz, Frenzel, & Taxer, 2015; Frenzel et al., 2015). The TES can also easily be adapted for use in such diary studies, by introducing the items with “In the past lesson...” and then adding for example, “I enjoyed teaching/I had reasons to become angry/I was tense and nervous”, et cetera.

Additionally, when constructing questionnaires on emotional experiences, it is challenging to tease apart inner feelings from shown behaviors. In that respect, our item “I teach with enthusiasm” might be slightly ambiguous, potentially addressing either feeling enthusiastic, or behaving enthusiastically in terms of a teaching style (see also Kunter et al., 2011). Given that we have repeatedly found in our studies that this item – both in its German and English versions – showed consistently high item-total correlations with the enjoyment scale, we suspect that most respondents interpret the term “enthusiasm” in this context more as an affective experience than as a behavior. Nevertheless, we are convinced that teasing apart affective experiences from displayed behaviors is meaningful particularly for the construct of teacher enthusiasm (Keller, Woolfolk Hoy, Goetz, & Frenzel, in press; Taxer & Frenzel, 2015b). Researchers who focus on this distinction in their research will either want to abandon the “enthusiasm” item from the enjoyment scale, or adapt it into “I experience enthusiasm when I teach” to focus explicitly on the feeling component of enthusiasm.

Another limitation of the present study is that it included only teachers from western, individualistic cultures (Germany and Canada). While we were able to show that the German and English language versions of the TES are largely equivalent, future research will have to show if translations into other languages can also be successful, particularly into languages that are spoken in eastern, collectivistic cultures (see Frenzel, Thrash, Pekrun, & Goetz, 2007, about attempts to translate the Achievement Emotions Questionnaire for students into Chinese). Provided that there was measurement equivalence, using the TES among teachers

from eastern versus western countries could reveal if the mean levels of reported enjoyment, anger, and anxiety differ between teachers from those cultures, and it would be interesting to explore if differing emotion display rules for classrooms contribute to those differences.

Finally, it is worth mentioning that future research may want to address a broader range of discrete emotions for teachers. While the TES address three highly relevant and frequent emotions teachers experience, there are other emotions with most likely lower frequency that nevertheless might deserve future research attention. We believe that pity would be an intriguing candidate, since despite its clearly pro-social, empathetic connotation, attribution theory would imply that it has detrimental effects for students' self-concepts if teachers express pity about students' poor performance (Butler, 1994; Graham, 1984; Rustemeyer, 1984). In addition, the emotion of boredom also seems to be underestimated for teaching contexts. While this emotion is barely ever mentioned by teachers in interview studies, it was shown to be experienced with surprisingly high frequency in studies that used the experience sampling method to explore teachers' emotional lives (Carson, 2006; Goetz et al., 2015; Keller, Chang, et al., 2014). Further discrete emotions worth exploring might be pride, which teachers reported to rather frequently express in the classroom, and disappointment, which they reported to deliberately hide, according to a recent quantitative study of teachers' emotion regulation in the classroom (Taxer & Frenzel, 2015a).

To conclude, the general and student-group specific variants of the TES are economic, reliable, and valid instruments in both German and English. As such, they stand to make an important contribution to advancing research on teachers' emotions by providing robust quantitative measures that can expand the current body of primarily qualitative knowledge on teacher emotions (Frenzel, 2014; Sutton & Wheatley, 2003). We hope that the existence of the TES inspires future research directed at exploring relations between teachers' emotions and their coping and emotion-regulation. For example, it will be intriguing to explore how teachers' reports of their enjoyment, anger, and anxiety relate with their efforts to engage in

antecedent-focused emotion regulation (i.e., regulation before the emotion is generated) and response-focused emotion regulation (i.e., after the emotion is generated, cf. Gross, 1998). As such, the TES should also be an effective tool for evaluating the success of intervention studies directed at emotion regulation and coping for teachers. Additionally, the TES provides a tool to measure baseline scores of typically reported levels of experienced emotions during teaching, and thus will allow explorations of when and why teachers engage in emotional labor, and with what implications. This type of research is particularly important and timely because teachers' emotions are proving to be highly relevant not only for their own well-being but also for the functioning of classrooms.

References

- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., & al., e. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, *47*, 133-180
- Bavarian State Ministry of Education. (2013). *Schule und Bildung in Bayern 2013: Statistische Übersichten [School and education in Bavaria in 2013: Statistical overview]*. Munich, Germany: Bavarian State Ministry of Education.
- Becker, E. S., Keller, M. M., Goetz, T., Frenzel, A. C., & Taxer, J. L. (2015). Antecedents of teachers' emotions in the classroom: An intraindividual approach. *Frontiers in Psychology*, *6*, 635. doi:10.3389/fpsyg.2015.00635
- Bibby, T. (2002). Shame: an emotional response to doing mathematics as an adult and a teacher. *British Educational Research Journal*, *28*(5).
- Böhm-Kasper, O., Bos, W., Jaeckel, S., & Weishaupt, H. (2000). *Skalenhandbuch zur Belastung von Schülern und Lehrern. Das Erfurter Belastungs-Inventar. Erfurter Materialien und Berichte zur Entwicklung des Bildungswesens (Bd. 2)*. Erfurt: Universität Erfurt.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.
- Brunner, M., Kunter, M., Krauss, S., Klusmann, U., Baumert, J., Blum, W., . . . Tsai, Y.-M. (2006). Die professionelle Kompetenz von Mathematiklehrkräften: Konzeptualisierung, Erfassung und Bedeutung für den Unterricht. Eine Zwischenbilanz des COACTIV-Projekts. In M. Prenzel & L. Allolio-Näcke (Eds.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms* (pp. 54–82). Münster, Germany: Waxmann.
- Butler, R. (1994). Teacher communications and student interpretations: Effects of teacher responses to failing students on attributional inferences in two age groups. *British Journal of Educational Psychology*, *64*, 277-294.
- Campbell, D. T. (1960). Recommendations for APA test standards regarding construct, trait, or discriminant validity. *American Psychologist*, *15*, 546-553.
- Carson, R. L. (2006). *Exploring the episodic nature of teachers' emotions as it relates to teacher burnout*. (PhD Dissertation), Purdue University, IN.
- Carson, R. L., Weiss, H. M., & Templin, T. J. (2010). Ecological momentary assessment: a research method for studying the daily lives of teachers. *International Journal of Research & Method in Education*, *33*(2), 165-182.
- Carver, C. S., & Harmon-Jones, E. (2009). Anger is an approach-related affect: Evidence and implications. *Psychological Bulletin*, *135*, 183-204.
- Chang, M. L. (2009). An Appraisal Perspective of Teacher Burnout: Examining the Emotional Work of Teachers. *Educational Psychology Review*, *21*, 193-218.
- Chang, M. L. (2013). Toward a theoretical model to understand teacher emotions and teacher burnout in the context of student misbehavior: Appraisal, regulation and coping *Motivation and Emotion*, *37*, 799-817. doi:10.1007/s11031-012-9335-0
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*, 464-504. doi:10.1080/10705510701301834
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233-255.
- Conner, T. S., & Barrett, L. F. (2012). Trends in ambulatory self-report: the role of momentary experience in psychosomatic medicine. *Psychosomatic Medicine*, *74*, 327-337. doi:10.1097/PSY.0b013e3182546f18

- Darby, A. (2008). Teachers' emotions in the reconstruction of professional self-understanding. *Teaching and Teacher Education, 24*, 1160-1172.
- Diener, E., Sandvik, E., Pavot, W., & Gallagher, D. (1991). Response artifacts in the measurement of subjective well being. *Social Indicators Research, 24*, 35-36.
- Ekman, P. (1982). *Emotion in the human face*. New York: Cambridge University Press.
- Ekman, P., & Davidson, R. J. (Eds.). (1994). *The nature of emotion. Fundamental questions*. New York: Oxford University Press.
- Enzmann, D., & Kleiber, D. (1989). *MBI/EK - Maslach Burnout Inventory - deutsche Fassung*. Heidelberg, Germany: Asanger.
- Fischer, D. G., & Fick, C. (1993). Measuring social desirability: short forms of the Marlowe-Crowne social desirability scale. *Educational Psychological Measurement, 52*, 417-424.
- Fontaine, J., Scherer, K., & Soriano, C. (Eds.). (2013). *Components of emotional meaning. A sourcebook*. Oxford: OUP.
- Fredrickson, B. L., & Branigan, C. (2005). Positive emotions broaden the scope of attention and thought-action repertoires. *Cognition & Emotion, 19*, 313-332.
doi:doi:10.1080/02699930441000238
- Frenzel, A. C. (2014). Teacher emotions. In E. A. Linnenbrink-Garcia & R. Pekrun (Eds.), *International Handbook of Emotions in Education* (pp. 494-519). New York: Routledge.
- Frenzel, A. C., Becker-Kurz, B., Pekrun, R., & Goetz, T. (2015). Teaching this class drives me nuts! – Examining the person and context specificity of teacher emotions. *PLoS ONE*. doi:10(6): e0129630.doi:10.1371/journal.pone.0129630
- Frenzel, A. C., & Goetz, T. (2007). Emotionales Erleben von Lehrkräften beim Unterrichten [Teachers' emotional experiences during teaching]. *Zeitschrift für Pädagogische Psychologie, 21*, 283-295.
- Frenzel, A. C., Thrash, T. D., Pekrun, R., & Goetz, T. (2007). Achievement emotions in Germany and China: A cross-cultural validation of the Academic Emotions Questionnaire–Mathematics. *Journal of Cross-Cultural Psychology, 38*, 302-309.
- Garland, R. (1991). The mid-point on a rating scale: Is it desirable? *Marketing Bulletin, 2*, 66-70.
- Ghanizadeh, A., & Jahedizadeh, S. (2015). Teacher burnout: A review of sources and ramifications. *British Journal of Education, Society, and Behavioural Sciences, 6*, 24-39.
- Goetz, T., Becker, E. S., Bieg, M., Keller, M. M., Frenzel, A. C., & Hall, N. C. (2015). The glass half empty: How emotional exhaustion affects the state-trait discrepancy in self-reports of teaching emotions. *PLoS ONE, 10*(9), e0137441.
doi:10.1371/journal.pone.0137441
- Goetz, T., Bieg, M., Lüdtke, O., Pekrun, R., & Hall, N. C. (2013). Do girls really experience more anxiety in mathematics? *Psychological Science, 24*, 2079-2087.
- Goetz, T., Frenzel, A. C., Pekrun, R., & Hall, N. C. (2006). The domain specificity of academic emotional experiences. *Journal of Experimental Education, 75*(1), 5-29.
- Goetz, T., Frenzel, A. C., Pekrun, R., Hall, N. C., & Lüdtke, O. (2007). Between- and within-domain relations of students' academic emotions. *Journal of Educational Psychology, 99*, 715-733. doi:10.1037/0022-0663.99.4.715
- Goetz, T., Hall, N. C., Frenzel, A. C., & Pekrun, R. (2006). A hierarchical conceptualization of enjoyment in students. *Learning and Instruction, 1* - 16.
- Goetz, T., Pekrun, R., Hall, N. C., & Haug, L. (2006). Academic emotions from a social-cognitive perspective: antecedents and domain specificity of students' affect in the context of Latin instruction. *British Journal of Educational Psychology, 76*(2), 289-308.

- Gogol, K., Brunner, M., Goetz, T., Martin, R., Ugen, S., Fischbach, A., . . . Preckel, F. (2014). 'My questionnaire is too long!' The assessments of motivational-affective constructs with three-item and single-item measures. *Contemporary Educational Psychology, 39*, 188-205.
- Graham, S. (1984). Communicating sympathy and anger to black and white children: the cognitive (attributional) consequences of affective cues. *Journal of Personality and Social Psychology, 47*(1), 40-54.
- Grandey, A. A. (2003). When "the show must go on": surface acting and deep acting as determinants of emotional exhaustion and peer-related service delivery. *Academy of Management Journal, 46*, 86-96.
- Gross, J. J. (1998). Antecedent- and response-focused emotion regulation: Divergent consequences for experience, expression, and physiology. *Journal of Personality and Social Psychology, 74*, 224-237.
- Hagenauer, G., & Volet, S. E. (2014a). 'I don't think I could, you know, just teach without any emotion': exploring the nature and origin of university teachers' emotions. *Research Papers in Education, 29*(2), 240-262.
- Hagenauer, G., & Volet, S. E. (2014b). Teacher-student relationship at university: an important yet under-researched field. *Oxford Review of Education, 40*(3), 370-388.
- Hargreaves, A., & Tucker, E. (1991). Teaching and guilt: Exploring the feelings of teaching. *Teaching and Teacher Education, 7*(5/6), 491-505.
- Hargreaves, A. (1998). The emotional practice of teaching. *Teaching and Teacher Education, 14*, 835-854.
- Hargreaves, A. (2000). Mixed emotions: Teachers' perceptions of their interactions with students. *Teaching and Teacher Education, 16*, 811-826.
- Hargreaves, A. (2005). The Emotions of Teaching and Educational Change. In A. Hargreaves (Ed.), *Extending Educational Change* (pp. 278-295): Springer.
- Hart, N. I. (1987). Student teachers' anxieties: four measured factors and their relationships to pupil disruption in class. *Educational Research, 29*, 12 — 18.
- Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure analysis: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*, 424-453.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55. doi:10.1080/10705519909540118
- Huebner, E. S., Laughlin, J. E., Ash, C., & Gilman, R. (1998). Further validation of the multidimensional students' life satisfaction scale. *Journal of Psychoeducational Assessment, 15*, 118-134.
- Isen, A. M. (2008). Some ways in which positive affect influences decision making and problem solving. In M. Lewis, J. M. Haviland-Jones, & L. Feldman Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 548-573). New York: Guilford Press.
- Jiang, J. Y., Sparte, S. E., & Luppescu, S. (2015). Teacher Perspectives on Evaluation Reform: Chicago's REACH Students. *Educational Researcher, 44*(2), 105-116. doi:10.3102/0013189x15575517
- Keller, M. M., Chang, M.-L., Becker, E. S., Goetz, T., & Frenzel, A. C. (2014). Teachers' emotional experiences and exhaustion as predictors of emotional labor in the classroom: An experience sampling study. *Frontiers in Psychology, 5*. doi:1442.doi:10.3389/fpsyg.2014.01442
- Keller, M. M., Frenzel, A. C., Goetz, T., Pekrun, R., & Hensley, L. (2014). Exploring teacher emotions: A literature review and an experience sampling study. In P. W. Richardson, S. Karabenick, & H. M. G. Watt (Eds.), *Teacher Motivation: Theory and Practice* (pp. 69-82). New York: Routledge.

- Keller, M. M., Woolfolk Hoy, A. E., Goetz, T., & Frenzel, A. C. (in press). Teacher enthusiasm: Reviewing and redefining a complex construct. *Educational Psychology Review*. doi:10.1007/s10648-015-9354-y
- Klassen, R. M., & Chiu, M. M. (2011). The occupational commitment and intention to quit of practicing and pre-service teachers: Influence of self-efficacy, job stress, and teaching context. *Contemporary Educational Psychology, 36*, 114-129. doi:10.1016/j.cedpsych.2011.01.002
- Klassen, R. M., Perry, N. E., & Frenzel, A. C. (2012). Teachers' relatedness with students: An underemphasized component of teachers' basic psychological needs. *Journal of Educational Psychology, 104*, 150-165. doi:10.1037/a0026253
- Klassen, R. M., Tze, V. M. C., Betts, S., & Gordon, K. A. (2011). Teacher efficacy research 1998-2009: Signs of progress or unfulfilled promise? *Educational Psychology Review, 23*, 21-43.
- Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras study: investigating effects of teaching and learning in Swiss and German mathematics classrooms
In T. Janik & T. Seidel (Eds.), *The power of video studies in investigating teaching and learning in the classroom* (pp. 137-160). Münster, Germany: Waxmann.
- Krohne, H. W., Egloff, B., Kohlmann, C.-W., & Tausch, A. (1996). Untersuchungen mit einer deutschen Version der "Positive and Negative Affect Schedule" (PANAS). *Diagnostica, 42*, 139-156.
- Kulas, J. T., & Stachowski, A. A. (2009). Middle category endorsement in odd-numbered Likert response scales: Associated item characteristics, cognitive demands, and preferred meanings. *Journal of Research in Personality, 43*, 489-493.
- Kunter, M., Frenzel, A. C., Nagy, G., Baumert, J., & Pekrun, R. (2011). Teacher enthusiasm: Dimensionality and context specificity. *Contemporary Educational Psychology, 36*(4), 289-301. doi:10.1016/j.cedpsych.2011.07.001
- Lazarus, R. S. (1994). The stable and the unstable in emotion. In P. Ekman & R. J. Davidson (Eds.), *The Nature of Emotion. Fundamental Questions* (pp. 79-85). Oxford: Oxford University Press.
- Lerner, J. S., & Keltner, D. (2001). Fear, anger, and risk. *Journal of Personality and Social Psychology, 81*(1), 146.
- Lüdtke, O., Trautwein, U., Kunter, M., & Baumert, J. (2006). Reliability and agreement of student ratings of the classroom environment - A reanalysis of TIMSS data. *Learning Environments Research, 9*, 215-230.
- Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*, 320-341.
- Maslach, C., & Jackson, S. E. (1986). *Maslach Burnout Inventory manual*. Palo Alto: Consulting Psychologists Press.
- Mitchell, R. L. C., & Phillips, L. H. (2007). The psychological, neurochemical and functional neuroanatomical mediators of the effects of positive and negative mood on executive functions. *Neuropsychologia, 45*, 617-629.
- Moors, A., Ellsworth, P., Scherer, K. R., & Frijda, N. (2013). Appraisal theories of emotion: State of the art and future development. *Emotion Review, 5*, 119-124.
- Musch, J., Brockhaus, R., & Bröder, A. (2002). Ein Inventar zur Erfassung von zwei Faktoren sozialer Erwünschtheit. *Diagnostica, 48*(121-129).
- Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus user's guide* (6 ed.). Los Angeles, CA: Muthén & Muthén.

- Pajares, M. F. (1996). *Assessing self-efficacy beliefs and academic outcomes: The case for specificity and correspondence*. Paper presented at the Annual Meeting of the American Educational Research Association, New York.
- Paulhus, D. L. (1994). *Balanced Inventory of Desirable Responding: Reference manual for BIDR version 6*. University of British Columbia. Vancouver, Canada.
- Payne, B. D., & Manning, B. H. (1990). The effect of self-instructions on preservice teacher's anxiety about teaching. *Contemporary Educational Psychology Review*, *15*, 261-267.
- Pekrun, R., Goetz, T., Frenzel, A. C., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The achievement emotions questionnaire (AEQ). *Contemporary Educational Psychology*, *36*, 36-48.
doi:10.1016/j.cedpsych.2010.10.002
- Pekrun, R., Goetz, T., Perry, R. P., Kramer, K., Hochstadt, M., & Molfenter, S. (2004). Beyond test anxiety: Development and validation of the test emotions questionnaire (TEQ). *Anxiety, Stress and Coping: An International Journal*, *17*, 287-316.
- Pekrun, R., vom Hofe, R., Blum, W., Frenzel, A. C., Goetz, T., & Wartha, S. (2007). Development of mathematical competencies in adolescence: The PALMA longitudinal study. In M. Prenzel (Ed.), *Studies on the educational quality of schools. The final report on the DFG Priority Programme* (pp. 17-37). Münster, Germany: Waxmann.
- Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: standardized observation can leverage capacity *Educational Researcher*, *38* 109-119
- Prawat, R., Byers, J., & Anderson, A. H. (1983). An attributional analysis of teachers' affective reactions to student success and failure. *American Educational Research Journal*, *20*, 137-152.
- Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin*, *128*(6), 934-960.
doi:10.1037/0033-2909.128.6.934
- Rustemeyer, R. (1984). Selbsteinschätzung eigener Fähigkeit - vermittelt durch die Emotionen anderer Personen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *16*(2), 149-161.
- Scherer, K. R. (1984). On the nature and function of emotion: a component process approach. In P. Ekman (Ed.), *Approaches to emotion*. Hillsdale: Erlbaum.
- Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and Emotion*, *23*, 1307-1351.
doi:10.1080/02699930902928969
- Scherer, K. R. (2013). Measuring the meaning of emotion words: A domain-specific componential approach. In J. Fontaine, K. Scherer, & C. Soriano (Eds.), *Components of emotional meaning. A sourcebook*. Oxford: OUP.
- Scherer, K. R., Wranik, T., Sangsue, J., Tran, V., & Scherer, U. (2004). Emotions in everyday life: Probability of risk factors, appraisal and reaction patterns. *Social Science Information*, *43*, 499-570.
- Schonfeld, W. H., Verboncoeur, C. J., Fifer, S. K., Lipschutz, R. C., Lubeck, D. P., & Buesching, D. P. (1997). The functioning and well-being of patients with unrecognized anxiety disorders and major depressive disorder. *Journal of Affective Disorders*, *43*(2), 105-119. doi:[http://dx.doi.org/10.1016/S0165-0327\(96\)01416-4](http://dx.doi.org/10.1016/S0165-0327(96)01416-4)
- Schutz, P. A., Cross, D. I., Hong, J. Y., & Osbon, J. N. (2007). Teacher identities, beliefs, and goals related to emotions. In P. A. Schutz & R. Pekrun (Eds.), *Emotion in education*. London: Elsevier.
- Schutz, P. A., & Zembylas, M. (Eds.). (2009). *Advances in teacher emotion research: The impact on teachers' lives*. New York: Springer.

- Slaney, K. L., & Maraun, M. D. (2008). A proposed framework for conducting data-based test analysis. *Psychological Methods, 13*(4), 376-390. doi:10.1037/a0014269
- Smith, C. A., & Lazarus, R. S. (1993). Appraisal components, core relational themes, and the emotions. *Cognition and Emotion, 7*(3/4), 233-269.
- Spielberger, C. D. (1983). *Manual for the State-Trait Anxiety Inventory (STAI)*. Palo Alto, CA: Consulting Psychologists Press.
- Spielberger, C. D. (2006). Cross-Cultural assessment of emotional states and personality traits. *European Psychologist, 11*, 297-303.
- Strahan, R., & Gerbasi, K. C. (1972). Short, homogeneous versions of the Marlowe-Crowne social desirability scale. *Journal of Clinical Psychology, 28*, 191-193.
- Sutton, R. E. (2004). Emotional regulation goals and strategies of teachers. *Social Psychology of Education, 7*, 379-398.
- Sutton, R. E. (2007). Teachers' anger, frustration, and self-regulation. In P. A. Schutz & R. Pekrun (Eds.), *Emotion in education* (pp. 251-266). San Diego: Academic Press.
- Sutton, R. E., & Harper, E. (2009). Teachers' emotion regulation. *Springer International Handbooks of Education, 21*, 389-401. doi:10.1007/978-0-387-73317-3_25
- Sutton, R. E., Mudrey-Camino, R., & Knight, C. C. (2009). Teachers' emotion regulation and classroom management. *Theory Into Practice, 48*(2), 130-137. doi:10.1080/00405840902776418
- Sutton, R. E., & Wheatley, K. F. (2003). Teachers' emotions and teaching: A review of the literature and directions for future research. *Educational Psychology Review, 15*, 327-358.
- Taxer, J. L., & Frenzel, A. C. (2015a). Facets of teachers' emotional lives: A quantitative investigation of teachers' genuine, faked, and hidden emotions. *Teaching and Teacher Education, 42*, 78-88. doi:dx.doi.org/10.1016/j.tate.2015.03.003
- Taxer, J. L., & Frenzel, A. C. (2015b). *Fake enthusiasm, does it exist? Discovering differing levels of expressed enthusiastic behaviours and experienced enjoyment among teachers*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.
- Tellegen, A., Watson, D., & Clark, L. A. (1999). On the dimensional and hierarchical structure of affect. *Psychological Science, 10*, 297-309.
- Thienel, A. (1988). Lehrerwahrnehmungen und-gefühle in problematischen Unterrichtssituationen. *Zum Einfluß von Wahrnehmungen und Emotionen auf das differentielle Erleben und Verhalten von Lehrern*. Frankfurt.
- Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*, 783-805.
- Van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology, 9*, 486-492. doi:10.1080/17405629.2012.686740
- van Veen, K., Seegers, P., & van de Ven, P.-H. (2005). Teacher's identity, emotions and commitment to change: A case study into the cognitive-affective processes of a secondary school teacher in the context of reforms. *Teaching and Teacher Education*.
- Vandenberghe, R., & Huberman, A. M. (Eds.). (1999). *Understanding and preventing teacher burnout. A sourcebook of international research and practice*. New York: Cambridge University Press.
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology, 54*, 1063-1070.
- Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology, 76*, 820-838.

- Weng, L. J. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement, 64*, 956-972.
- Winograd, K. (2003). The functions of teacher emotions: The good, the bad, and the ugly. *Teachers College Record, 105*, 1641-1673.
- Wirtz, D., Kruger, J., Scollon, C. N., & Diener, E. (2003). What to Do on Spring Break?: The Role of Predicted, On-Line, and Remembered Experience in Future Choice. *Psychological Science, 14*(5), 520-524. doi:10.1111/1467-9280.03455
- Zeidner, M. (1998). *Test anxiety: The state of the art*. New York: Plenum.
- Zembylas, M. (2003). Emotions and teacher identity: A poststructural perspective. *Teachers and teaching: Theory and practice, 9*, 213-238.
- Zembylas, M. (2005). Discursive practices, genealogies, and emotional rules: A poststructuralist view on emotion and identity in teaching. *Teaching and Teacher Education, 21*, 935-948. doi:10.1016/j.tate.2005.06.005

Table 1
Means, Standard Deviations, and Internal Consistencies (Cronbach's Alphas) for the TES

	General Scales			Student-Group Specific Scales		
	Enjoy- ment	Anger	Anxiety	Enjoy- ment	Anger	Anxiety
Means						
Study 1	3.39	1.88	1.44	3.24	1.85	1.43
Study 2 (sample 1)	^a	^a	^a	3.25	2.05	1.54
Study 2 (sample 2)	^a	^a	^a	3.39	1.49	1.34
Study 3	3.30	1.63	1.82	3.30	1.67	1.72
Standard Deviations						
Study 1	.50	.52	.44	.67	.71	.57
Study 2 (sample 1)	^a	^a	^a	.67	.69	.51
Study 2 (sample 2)	^a	^a	^a	.49	.51	.39
Study 3	.44	.56	.60	.48	.60	.60
Cronbach's Alphas						
Study 1	.77	.73	.70	.90	.89	.89
Study 2 (sample 1)	^a	^a	^a	.92	.86	.74
Study 2 (sample 2)	^a	^a	^a	.72	.80	.70
Study 3	.73	.80	.81	.80	.87	.87

Note. Study 1: $N = 414$ German teachers, various school types; Study 2, sample 1: $N = 85$ German primary school teachers; Study 2, sample 2: $N = 68$ German secondary school teachers; Study 3: $N = 377$ Canadian teachers, various school types.

^a Only the student-group specific scales were administered in this study.

Table 2
Intercorrelations of Enjoyment, Anger and Anxiety

	General			Student-Group Specific		
	Enjoyment 1	Anger 2	Anxiety 3	Enjoyment 4	Anger 5	Anxiety 6
Studies 1 and 3						
1	---	-.44** (-.58**)	-.43** (-.58**)	<u>.51**</u>	-.30**	-.35**
2	-.51** (-.68**)	---	.42** (.58**)	-.32**	<u>.52**</u>	.40**
3	-.51** (-.70**)	.50** (.71**)	---	-.30**	.35**	<u>.53**</u>
4	<u>.71**</u>	-.51**	-.43**	---	-.69** (-.85**)	-.65** (-.73**)
5	-.50**	<u>.74**</u>	.51**	-.62** (-.76**)	---	.75** (.81**)
6	-.50**	.50**	<u>.75**</u>	-.57** (-.70**)	.69** (.78**)	---
Study 2						
4				---	-.72**	-.59**
5				-.50**	---	.67**
6				-.41**	.08	---

Note. In the upper part of the table, intercorrelations for Study 1 ($N = 414$ German teachers, various school types) are reported above the diagonal, and intercorrelations for Study 4 ($N = 377$ Canadian teachers, various school types) are reported below the diagonal. Correlations pertaining to the general/group-specific emotion link are underlined. Latent correlations as obtained within the CFAs are shown in parentheses. In the lower part of the table, intercorrelations for the first sample of Study 2 ($N = 85$ German Primary School Teachers) reported above the diagonal, and intercorrelations for the second sample of Study 2 ($N = 68$ German Secondary School Teachers) are reported below the diagonal. With the latter sample, only the student-group specific scales were used.

** $p < .01$.

Table 3
Internal Validity of the TES: One-, Two-, and Three-Factor Models

Model	Germany (<i>N</i> = 414, Study 1)					Canada (<i>N</i> = 377, Study 3)				
	χ^2	df	CFI	RMSEA	SRMR	χ^2	df	CFI	RMSEA	SRMR
<i>General Scales</i>										
1. Single Emotion Factor Model	408	54	.737	.126	.083	343	54	.815	.120	.072
2. Two-factor Model: Positive vs. Negative Affect	301	53	.815	.106	.070	267	53	.863	.104	.062
3. Three-factor Model: Enjoyment, Anger, Anxiety	204	51	.886	.085	.065	140	51	.943	.068	.044
<i>Student-Group Specific Scales</i>										
1. Single Emotion Factor Model	510	54	.848	.156	.069	335	54	.868	.119	.062
2. Two-factor Model: Positive vs. Negative Affect	270	53	.916	.117	.053	228	53	.918	.095	.047
3. Three-factor Model: Enjoyment, Anger, Anxiety	188	51	.958	.081	.040	102	51	.976	.052	.030

Table 4
External Validity of the TES: Correlations with Related Teacher Constructs

	Enjoyment			Anger			Anxiety		
	Study 2 (sample 1)	Study 2 (sample 2)	Study 3	Study 2 (sample 1)	Study 2 (sample 2)	Study 3	Study 2 (sample 1)	Study 2 (sample 2)	Study 3
General Affect: pos	.37**	.41**	.39**/.31**	-.20	-.16	-.22**/-.19**	-.10	-.15	-.25**/-.23**
General Affect: neg	-.25*	-.23	-.27**/-.24**	.31**	.29*	.33**/.28**	.30**	.36**	.45**/.38**
Emotional Exhaustion	-.51**	-.25*	-.48**/-.45**	.45**	.27*	.50**/.49**	.40**	.33**	.48**/.44**
Depersonalization	-.42**	^a	-.48**/-.52**	.36**	^a	.63**/.49**	.06	^a	.50**/.56**
Lack of Accomplishment	-.56**	^a	-.67**/-.61**	.42**	^a	.42**/.52**	.32**	^a	.50**/.48**
Job Satisfaction	^a	.07	^a	^a	.03	^a	^a	-.25*	^a
Teacher Self-Efficacy	^a	.17	^a	^a	-.13	^a	^a	-.29*	^a
Social Desirability	.13/-.03	.24*	.21**/.22**	-.04/.09	-.15	-.26**/-.22*	-.25*/.05	-.03	-.21**/-.17**

Note. Study 2, sample 1: $N = 85$ German primary school teachers; Study 2, sample 2: $N = 68$ German secondary school teachers; Study 3: $N = 377$ Canadian teachers [various school types]. For Study 2, correlations for self-deception/impression management social desirability subscales are shown. For Study 3, correlations with general/specific emotion scales are shown.

^aThis construct was not assessed in the corresponding study.

* $p < .05$. ** $p < .01$.

Table 5
External Validity of the TES: Relations with Student Ratings of Teacher Behavior

Student Ratings	Bivariate Correlation			Partial correlation		
	Enjoy- ment	Anger	Anxiety	Enjoy- ment ^a	Anger ^b	Anxiety ^b
Clarity of Instruction	.34**	-.22	-.16	.23	-.18	-.11
Variety in Instruction	.45**	-.31*	-.22	.37**	-.28*	-.18
Withitness	.12	.20	-.22	-.07	.31*	-.17
Fast Paced Instruction	-.38*	.32**	.22	-.41**	.32*	.22
Acceptance of Errors	.38*	-.23	-.27*	.35**	-.21	-.26*
Teacher Caring	.41**	-.28*	-.23	.31*	-.24	-.19
Teacher Support after Failure	.41**	-.29*	-.19	.32**	-.25*	-.15
Teacher Disrespect	-.37*	.32**	.23	-.36**	.31*	.21

Note. Sample: $N = 68$ German secondary school teachers and their $N = 1,566$ students.

^a controlling for general positive affect

^b controlling for general negative affect

Table 6
Tests of Measurement Invariance for the German versus English Language Version of the TES

Model	χ^2	df	CFI	RMSEA	SRMR	Δ CFI	Δ RMSEA	Δ SRMR
General Scales								
1. Configural Invariance	344	102	.917	.078	.056	--	--	--
2. Metric Invariance	388	111	.905	.080	.069	.012	.002	.013
3. Scalar Invariance	917	120	.726	.130	.131	.179	.050	.062
3a. Invariance of all loadings and enjoyment intercepts	552	115	.850	.098	.094	.055	.018	.025
3b. Invariance of all loadings and anger intercepts	832	115	.753	.126	.129	.152	.046	.060
3c. Invariance of all loadings and anxiety intercepts	506	115	.866	.093	.113	.039	.013	.044
Student-Group Specific Scales								
1. Configural Invariance	291	102	.965	.069	.035	--	--	--
2. Metric Invariance	326	111	.960	.070	.058	.005	.001	.023
3. Scalar Invariance	770	120	.880	.118	.085	.129	.041	.027
3a. Invariance of all loadings and enjoyment intercepts	390	115	.949	.078	.066	.011	.008	.008
3b. Invariance of all loadings and anger intercepts	696	115	.893	.114	.090	.067	.044	.032
3c. Invariance of all loadings and anxiety intercepts	402	115	.947	.080	.087	.013	.010	.029

Note. The delta values are calculated relative to the preceding model in the stepwise measurement invariance test procedure. Metric invariance involves equal loadings, scalar invariance involves equal loadings plus equal item intercepts.

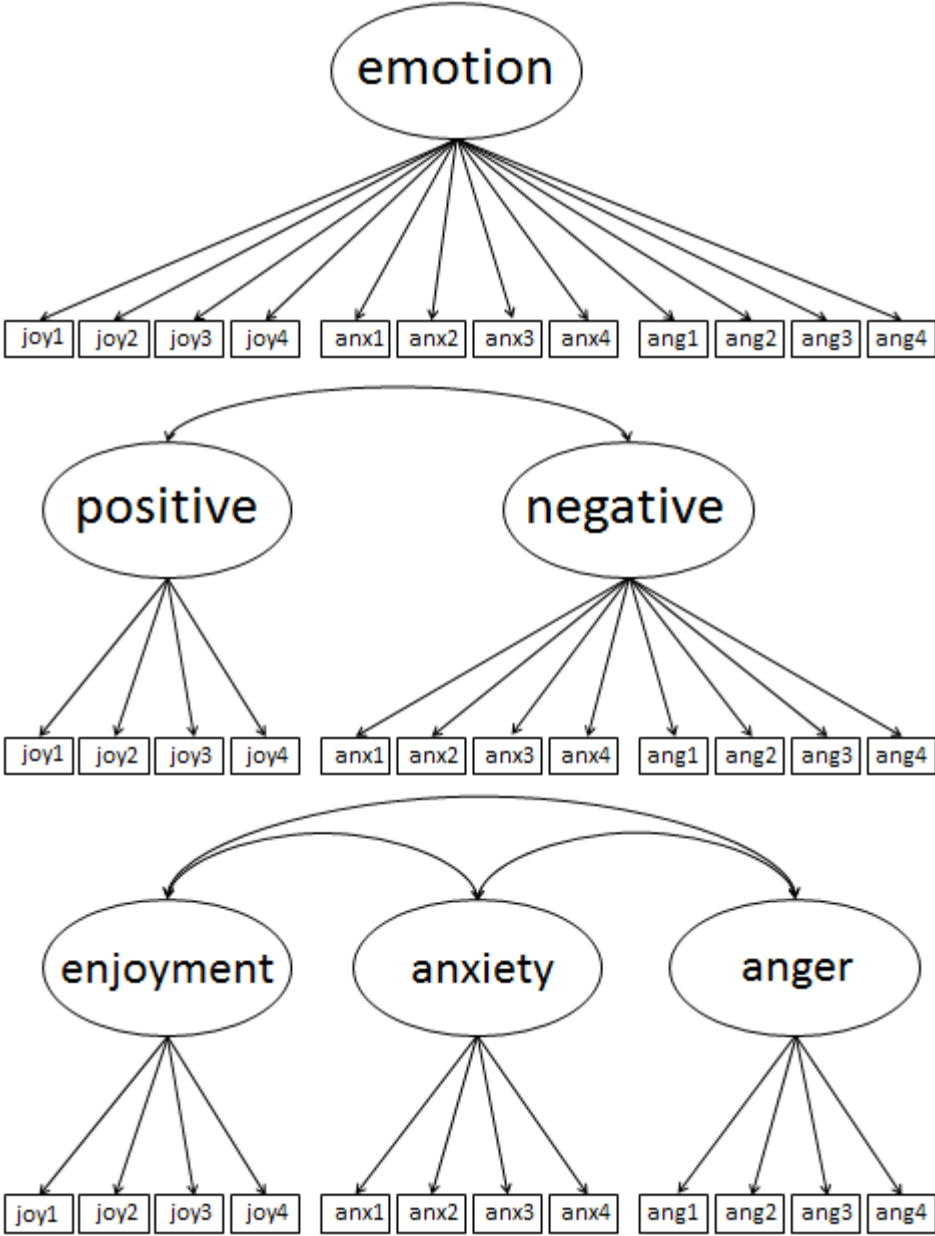


Figure 1. Unidimensional, two-factor, and three-factor models (depicted in the upper, middle, and lower parts of the figure, respectively).

Appendix A: English and German Items of the TES

		English	German
		General Scales	
	Short name		
Instruction		Below you find a list of statements describing your experiences as a teacher. Please indicate your personal response to each of these statements by circling the number that best represents your answer.	Im Folgenden finden Sie eine Reihe von Aussagen, die sich auf Ihre Erfahrungen als Lehrkraft beziehen. Bitte kreuzen Sie diejenige Antwort an, die Ihr persönliches Erleben am besten beschreibt.
Enjoyment	joy1	I generally enjoy teaching.	Im Allgemeinen macht mir Unterrichten Freude.
	joy2	I generally have so much fun teaching that I gladly prepare and teach my lessons.	Im Allgemeinen macht mir Unterrichten so viel Spaß, dass ich den Unterricht gerne vorbereite und durchführe.
	joy3	I often have reasons to be happy while I teach.	Während des Unterrichts habe ich oft Grund, mich zu freuen.
	joy4	I generally teach with enthusiasm.	Im Allgemeinen unterrichte ich mit Begeisterung.
Anger	ang1	I often have reasons to be angry while I teach.	Während des Unterrichts habe ich oft Grund, mich zu ärgern.
	ang2	I often feel annoyed while teaching.	Während des Unterrichts bin ich oft genervt.
	ang3	Sometimes I get really mad while I teach.	Beim Unterrichten werde ich gelegentlich richtig sauer.
	ang4	Teaching generally frustrates me.	Im Allgemeinen frustriert mich das Unterrichten.
Anxiety	anx1	I generally feel tense and nervous while teaching.	Beim Unterrichten bin ich in der Regel angespannt und nervös.
	anx2	I am often worried that my teaching isn't going so well.	Ich mache mir oft Sorgen, dass das Unterrichten nicht so richtig klappt.
	anx3	Preparing to teach often causes me to worry.	Die Vorbereitung des Unterrichts bereitet mir Sorgen.
	anx4	I feel uneasy when I think about teaching.	Wenn ich an das Unterrichten denke, bin ich beunruhigt.
		Student-Group Specific Scales	
Instruction		When answering the items below, please think of the students in your class this year. If you teach more than one class, please think of your typical Tuesday morning. Envision yourself walking into your first class on Tuesday morning and think of the students in this particular class when answering the following items	Bei der Bearbeitung der folgenden Fragen beziehen Sie sich bitte auf die Klasse, die Sie zurzeit unterrichten. Sollten Sie mehrere Klassen unterrichten, denken Sie bitte an einen typischen Dienstag Morgen. Bitte beantworten Sie die Fragen in Bezug auf diejenige Klasse, die Sie dienstags morgens in der ersten Stunde unterrichten.
Enjoyment	joy_s1	I enjoy teaching these students.	In dieser Klasse macht mir Unterrichten Freude.
	joy_s2	I have so much fun teaching these students that I gladly prepare and teach my lessons.	In dieser Klasse macht mir Unterrichten so viel Spaß, dass ich den Unterricht gerne vorbereite und durchführe.
	joy_s3	I teach these students with enthusiasm.	In dieser Klasse unterrichte ich mit Begeisterung.

	joy_s4	I often have reason to be happy while I teach these students.	Während des Unterrichts in dieser Klasse habe ich oft Grund, mich zu freuen.
Anger	ang_s1	I often have reason to be angry while I teach these students.	Während des Unterrichts in dieser Klasse habe ich oft Grund, mich zu ärgern.
	ang_s2	I often feel annoyed while teaching these students.	Während des Unterrichts in dieser Klasse bin ich oft genervt.
	ang_s3	Sometimes I get really mad at these students.	In dieser Klasse werde ich gelegentlich richtig sauer.
	ang_s4	Teaching these students frustrates me.	In dieser Klasse frustriert mich das Unterrichten.
Anxiety	anx_s1	I feel tense and nervous while teaching these students.	Beim Unterrichten in dieser Klasse bin ich angespannt und nervös.
	anx_s2	I am often worried that my teaching isn't going so well with these students.	Ich mache mir oft Sorgen, dass das Unterrichten in dieser Klasse nicht so richtig klappt.
	anx_s3	Preparing to teach these students often causes me to worry.	Die Vorbereitung des Unterrichts in dieser Klasse bereitet mir oft Sorgen.
	anx_s4	I feel uneasy when I think about teaching these students.	Wenn ich an das Unterrichten in dieser Klasse denke, bin ich beunruhigt.

Note. We recommend presenting these 12 items in random order within the questionnaire. It is also possible to mix in other scale items that fit into the context (e.g., general TES items can be mixed with burnout items).

Appendix B: Reliabilities of Teacher-Level External Scales

	Cronbach's Alpha		
	Study 2 (sample 1)	Study 2 (sample 2)	Study 3
Social Desirability (Study 2, sample 1: 5 items each for self-deception/impression management; Study 2, sample 2: 5 items for impression management; Study 3: 10 items for global social desirability)	.61/.51	.64	.65
General affect: pos (10 items/Study 2, sample 2: 5 items)	.82	.81	.71
General affect: neg (10 items/Study 2, sample 2: 5 items)	.78	.74	.74
Emotional Exhaustion (9 items/Study 3: 5 items)	.85	.60	.86
Depersonalization (5 items)	.77	^a	.73
Lack of Accomplishment (8 items)	.69	^a	.73
Job Satisfaction (5 items)	^a	.84	^a
Teacher Self-Efficacy (24 items)	^a	.85	^a

Note. Study 2, sample 1: $N = 85$ German primary school teachers; Study 2, sample 2: $N = 68$ German secondary school teachers; Study 3: $N = 377$ Canadian teachers [various school types].

^a This construct was not assessed in the corresponding study.

Appendix C: Reliabilities of Student-Level External Scales (Study 4)

	Scale Reliability	
	Individual Level: Cronbach's Alpha	Group Level: Intraclass Correlation (2)
Clarity of Instruction ^a	.84	.89
Variety in Instruction ^a	.77	.85
With-it-ness ^a	.67	.83
Fast Paced Instruction ^a	.72	.83
Acceptance of Errors ^a	.65	.90
Teacher Caring ^b	.76	.86
Teacher Support after Failure ^a	.78	.84
Teacher Disrespect ^a	.82	.81

Note. Sample: $N = 1,566$ students; the ICC(2) was calculated based on an average of $n=23$ students per group (total 68 groups).

^a 3 items ^b 4 items