Learning and Individual Differences 22 (2012) 190–201

Contents lists available at ScienceDirect

Learning and Individual Differences

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

Measuring students' emotions in the early years: The Achievement Emotions Questionnaire-Elementary School (AEQ-ES)

Stephanie Lichtenfeld [a,*], Reinhard Pekrun [a], Robert H. Stupnisky [b], Kristina Reiss [c], Kou Murayama [a]

[a] University of Munich, Germany
[b] University of North Dakota, USA
[c] Technische Universität München, Germany

ARTICLE INFO

Article history:
Received 16 July 2010
Received in revised form 1 December 2010
Accepted 28 April 2011

Keywords:
Achievement emotion
Enjoyment
Boredom
Test anxiety
Control-value theory
Learning
Academic achievement
Elementary school

ABSTRACT

This article reports about the development and validation of a measurement instrument assessing elementary school students' achievement emotions (Achievement Emotions Questionnaire-Elementary School, AEQ-ES). Specifically, the instrument assesses students' enjoyment, anxiety, and boredom pertaining to three types of academic settings (i.e., attending class, doing homework, and taking tests and exams). Scale construction was based on Pekrun's (2006) control-value theory of achievement emotions. The instrument was tested using samples from German and American elementary school classrooms. The results of Study 1 (German sample) corroborate the reliability and structural validity of the new emotion measure. Moreover, they show that students' achievement emotions were linked with their control and value appraisals as well as their academic performance, thus supporting the external validity of the measure as well as propositions of Pekrun's (2006) control-value theory of achievement emotions. Study 2 (American sample) corroborated the cross-cultural equivalence of the measure and the generalizability of findings across the German and American samples. Implications for research on achievement emotions and educational practice are discussed.

Emotions are ubiquitous in achievement settings. Students may feel proud about good grades, worry that they don't understand course material, get angry about a teacher who treats them unfairly, or feel bored when dealing with a topic in which they are not interested. Despite the diversity of emotions students experience in their courses on a daily basis (Pekrun, Goetz, Titz & Perry, 2002), research on different achievement emotions has been slow to emerge (Schutz & Pekrun, 2007). This is especially puzzling because emotions have a large impact on students' motivation, learning, and performance, as well as on their health and well-being (Pekrun, 2006). Specifically, whereas research on emotions in older students has been growing in recent years (Efklides & Volet, 2005; Linnenbrink, 2006; Linnenbrink-Garcia & Pekrun, 2011; Schutz & Pekrun, 2007), empirical evidence on the achievement emotions experienced by elementary school children is largely lacking to date.

One possible reason for the lack of research on elementary students' achievement emotions is the limited number of measurement instruments. Exceptions include instruments assessing elementary school students' achievement anxiety, such as scales on test anxiety (Weinert & Helmke, 1997) and mathematics anxiety (e.g., Grützemann, 2003; Suinn, Taylor, & Edwards, 1988). In contrast, established measures capturing a range of emotions in addition to anxiety are lacking. Therefore, the primary goal of the present research was to construct an instrument that measures a variety of distinct emotions in elementary school students. Pekrun's control-value theory of achievement emotions (Pekrun, 2000, 2006; Pekrun, Frenzel, Goetz, & Perry, 2007) served as the theoretical framework for constructing and validating the instrument.

1. Previous research on elementary students' achievement emotions

Historically, achievement emotion researchers primarily pursued two lines of research: test anxiety studies, and studies on the attributional antecedents of achievement emotions (Hembree, 1988; Seipp, 1991; Zeidner, 1998). Research on test anxiety focused primarily on high school and college students; in contrast, studies investigating this construct in the early elementary school years are scarce. For instance, in Hembree's (1988) comprehensive meta-analysis of research on test anxiety, only one study was listed on 11- and 12-year old students and none on children younger than 11 years of age. Likewise, in a meta-analysis on the relationship between anxiety and achievement in mathematics, studies conducted in the early elementary grades were found to be rare (Ma, 1999).

The second commonly pursued line of emotion research addressed the attributional antecedents of achievement emotions. In his

* Corresponding author.
E-mail address: lichtenfeld@psy.lmu.de (S. Lichtenfeld).

1041-6080/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.lindif.2011.04.009

attributional theory of achievement emotions, Weiner (1985) proposed that emotions such as pride, shame, guilt, etc., depend on students' causal attributions of success and failure outcomes. Based on Weiner's (1985) theory, Stipek and Gralinski (1991) conducted a study in the elementary school context that investigated gender differences in third graders' pride and shame in mathematics. The results indicated that girls felt less proud about success and were more concerned about public humiliation (i.e., shame) after failure in mathematics than boys.

An additional line of emotions research concerns positive emotions in the elementary school years. Helmke (1993), for example, investigated the development of enjoyment of learning from kindergarten to grade five in mathematics and German. Results from this longitudinal study showed that enjoyment of learning decreased across the elementary school years, but did so while maintaining a relatively high level. In sum, with few exceptions, research on this topic is largely lacking, especially so with regard to studies examining a range of different achievement emotions.

2. The control-value theory of achievement emotions

Pekrun's (2006) control-value theory of emotions served as a theoretical framework for the present research. This theory integrates basic principles from attributional theories of achievement emotions (e.g., Weiner, 1985), expectancy-value approaches to emotions (Pekrun, 1992; Turner & Schallert, 2001), theories of perceived control (Patrick, Skinner, & Connell, 1993; Perry, 1991), and models involving the effects of emotions on learning and performance (Fredrickson, 2001; Pekrun et al., 2002; Zeidner, 1998). As part of the theory, Pekrun (2006) proposed a taxonomy that describes emotions along three dimensions: valence (positive vs. negative), level of activation (activating vs. deactivating), and object focus in terms of being related to either achievement activities (e.g., learning) or achievement outcomes (i.e., success and failure). For example, the experience of enjoyment during a specific class may be characterized as a positive, activating, activity-related emotion. In contrast, anxiety when facing an exam is considered a negative, activating, outcome-related emotion.

In his theory, Pekrun (2000, 2006) proposes that control and value appraisals serve as critical antecedents of achievement emotions. Control-related appraisals involve individuals' achievement-related competence beliefs, expectancies and attributions. Value appraisals refer to the perceived value of an activity or outcome (e.g., the perceived importance of success). Perceived controllability and the positive subjective value of achievement activities are expected to evoke positive activity emotions, such as enjoyment of learning, and reduce negative activity emotions, such as boredom and anger. While boredom has traditionally been assumed to be caused by a lack of challenge (Csikszentmihalyi, 1975), more recent studies have found that boredom is indeed related to a low self-concept of ability (Goetz, Pekrun, Hall, & Haag, 2006) and low perceived control (Pekrun, Goetz, Daniels, Stupnisky, & Perry, 2010). The theory further proposes that low perceived controllability and the negative subjective value of failure outcomes elicit negative outcome emotions such as anxiety, hopelessness, or shame. For instance, a student who anticipates failing an important exam and feels incapable of passing it will experience failure-related anxiety (Pekrun, 1992). In sum, the control-value theory posits that control and value appraisals serve as proximal antecedents of specific, discrete achievement emotions.

3. Examining elementary school students' specific emotions

3.1. Focus on enjoyment, anxiety, and boredom

The present research focused on examining three achievement emotions that are of primary importance in achievement settings, namely, enjoyment, anxiety, and boredom. We analyzed the structures and achievement outcomes of these emotions with regard to a specific domain (mathematics) and across two cultural contexts. The three emotions were selected because they are frequently experienced in achievement settings (Csikszentmihalyi & Larson, 1987; Helmke, 1993; Pekrun, Goetz, Frenzel, & Perry, 2011; Pekrun et al., 2002). Furthermore, the emotions provide a representation of the major dimensions of Pekrun's (2006) taxonomy of achievement emotions. Specifically, they represent both activity-related emotions (enjoyment and boredom) and outcome-related emotions (anxiety), both positive (enjoyment) and negative emotions (boredom and anxiety), and both activating (enjoyment, anxiety) and deactivating emotions (boredom). Furthermore, achievement emotions can be related to different academic settings, such as attending class, studying, and taking tests and exams. As these settings vary in respect to their functions, demands, and social structures, emotions may likewise differ across these settings. For example, students who enjoy attending class may not necessarily enjoy the challenge of an exam. As a consequence, measurements investigating achievement emotions should distinguish between emotions related to these different academic settings (Pekrun et al., 2011).

3.2. Domain specificity of emotions: focus on mathematics

Previous research has shown that control- and value-related constructs such as self-concepts of ability, achievement expectancies, and interest are best examined when considered in specific subject domains (e.g., Bong, 2001; Marsh, 1986). In turn, emotions depending on control and value appraisals should also be considered at the domain-specific level. Empirical evidence supports this proposition in terms of weak between-domain relations of various achievement emotions (see Goetz et al., 2006; Goetz, Frenzel, Pekrun, Hall, & Lüdtke, 2007). As a consequence, in the present research we constructed an instrument that assesses emotions in a domain-specific way; specifically, the questionnaire measured students' emotions in mathematics.

3.3. Achievement outcomes

In his control-value theory Pekrun (2006) suggests that positive activating emotions generally improve academic achievement by promoting task-related attention, strengthening motivation, and enhancing use of flexible learning strategies. For instance, enjoying a particular school subject will direct attention toward related tasks, which consequently leads to better student performance. Several studies provide empirical evidence for the positive enjoyment-performance link (e.g., Helmke, 1993; Pekrun et al., 2002). In contrast, negative deactivating emotions such as boredom have been found to impair motivation and self-regulation of learning, leading to shallow information processing and poor student performance (Pekrun et al., 2010). Negative activating emotions such as anxiety may also impair interest and intrinsic motivation; however, they may enhance extrinsic motivation to invest effort and avoid failures. Consequently, the effects of negative activating emotions on achievement outcomes are more variable (see e.g., Pekrun et al., 2002), although the negative impact of these emotions on overall academic achievement are generally believed to outweigh any advantageous effects (Hembree, 1988; Pekrun et al., 2007). Taken together, we hypothesized positive relations of enjoyment with students' mathematics achievement and negative relations for students' boredom and anxiety.

3.4. Cultural context

Elementary school students' emotions in mathematics may also differ across different cultural contexts. For example, the German and the American elementary school systems traditionally differ in a number of aspects that may influence students' emotions, the most salient difference being the relative emphasis on *tracking* or *ability*

grouping. In the United States, tracking typically takes place within schools. One form of tracking is between-classes, wherein high achieving students take advanced courses and low achieving students take remedial classes. Another form of tracking is within-class tracking, implying that students are taught in different groups within the classroom depending on their achievement level. These forms of ability grouping are common across the early elementary grades in the United States (Entwisle & Alexander, 1993; Hallinan & Sørensen, 1983; Pallas, Entwisle, Alexander, & Stluka, 1994; Rosenholtz & Simpson, 1984).

Alternatively, in Germany the entire school system is traditionally based on tracking with few exceptions. At the end of Grade 4 or 6 when students are about 10 or 12 years of age they are selected to enter different secondary school tracks. The selection process is based on students' performance in elementary school. Across the German states, number and quality of these tracks vary considerably (Baumert, Trautwein, & Artelt, 2003; Mintrop, 1997). The present study was run in an area with the most common tracking system, which is called the "tripartite" tracking system. In this tracking system, students are selected to enter either the least academically demanding track called "Hauptschule", the intermediate "Realschule" track, or the college-bound track called "Gymnasium". The selection into the different school tracks is important for students' subsequent educational and professional career. Those in the lower tracks typically enter a dual system that combines part-time education at vocational schools with on-the-job training. Students from Hauptschule generally receive relatively lower paid jobs, while Realschule students typically aspire to more skilled occupations. Those students who pass the final Gymnasium examination (Abitur) get the opportunity to enter higher education (Maaz, Trautwein, Lüdtke, & Baumert, 2008).

Clearly, elementary school achievement outcomes have more profound consequences for German than American students, which in turn may increase the subjective value of these outcomes for German students. Additionally, there is evidence that American elementary schools have instituted more self-esteem enhancing and less critical feedback practices compared to German elementary schools (Little, Oettingen, Stetsenko, & Baltes, 1995). This may also influence elementary students' achievement emotions, particularly those related to exams and test taking.

4. The present research

The present research had three primary aims. The first aim was to develop an instrument measuring students' enjoyment, anxiety, and boredom in the early elementary school years (Achievement Emotions Questionnaire-Elementary School, AEQ-ES). This instrument can be used to assess these emotions in various school subjects; however, in the present research the focus was on students' enjoyment, anxiety, and boredom in mathematics. The instrument was analyzed using samples from German (Study 1) and American (Study 2) elementary school classrooms. Specifically, item and scale statistics including Cronbach's alpha reliability, internal test validity, and external test validity (see Slaney & Maraun, 2008, for the distinction of internal vs. external test validity) were examined.

The second aim of our research was to test central propositions of the control-value theory of achievement emotions in early elementary school students (Pekrun, 2006). This was accomplished by examining the links between our emotion measures and critical antecedents and outcomes of achievement emotions. Specifically, correlations of elementary school students' emotions with control and value appraisals as well as students' achievement outcomes were calculated.

The third aim of our research was to establish the construct equivalence and cross-cultural utility of the German and American versions of the AEQ-ES. Specifically, we examined the measurement invariance of the instrument across the two samples using multigroup confirmatory factor analysis.

5. Study 1

In Study 1 we examined the psychometric properties of the AEQ-ES in a sample of German elementary school students. Moreover, we examined linkages between achievement emotions, students' appraisals, and their academic achievement.

5.1. Method

5.1.1. Participants and procedure

The study involved students of two age groups from 30 different elementary schools in Germany. The sample represented a wide range of students in terms of ability and socio-economic background, including 678 second-grade students (345 females) from 30 classrooms and 687 third-grade students (330 female) from 30 classrooms. Students without achievement data ($n = 48$) and those who didn't complete the entire survey ($n = 128$) were excluded from the analysis. The final breakdown of the sample was 594 second graders (mean age = 9.05 years, range 7 to 11, $SD = .71$) and 595 third graders (mean age = 10.10 years, range 8 to 12, $SD = .72$).[1]

Students completed the study instruments during regular classroom hours. On the first day of testing, a one-hour mathematics achievement test was administered. On the following day, students responded to the AEQ-ES, items pertaining to students' perceived control and value in mathematics, as well as several demographic items. Before answering the questionnaire, students were assured of the confidentiality of their responses. The participants were asked to express their personal opinion and judgments, and were told that there were no "right" or "wrong" answers. All items were read to the students to ensure that every student understood the questions.

5.1.2. Measures

5.1.2.1. Achievement emotions. The scales of the AEQ-ES were adapted from the Achievement Emotions Questionnaire (Pekrun et al., 2011). The primary goal of the adaptation was to create an instrument tailored to the cognitive and language ability levels of elementary school students. The AEQ-ES consists of 28 items (nine enjoyment, 12 anxiety, and seven boredom items) that are organized in eight scales. Enjoyment and anxiety are measured with three scales pertaining to experiencing the emotion when attending class, doing homework, and taking tests and exams. Boredom is measured with only two scales, relating to class time and doing homework, as boredom is not typically experienced in relation to taking tests and exams (see Appendix A for sample items). Items are answered on a 5-point Likert scale anchored by five graphical displays of faces showing increasing emotional intensity. To ensure that both boys and girls could identify themselves with the faces, there were different versions for male and female students using male and female faces, respectively (see Appendix B).

In the present study, the second-grade students only answered scales regarding class- and learning-related emotions as German students lack experience with taking tests and receiving grades during the first two school years. In the third grade sample, all eight scales were administered.

[1] The subset of students who did not provide responses to all emotion items, or did not take part in the achievement test in second (12.4%, n = 84) or third grade (13.4%, n = 92), did not differ from the rest of the sample. Paired samples t-tests and chi-square tests indicated that students with missing data did not differ with respect to age, gender, and sociocultural variables including the number of books at home and the hours of TV watched per day (ps > .05). Only students with complete responses to the emotion scales were included in the final analyses to reduce error from missing responses (Graham, 2009).

5.1.2.2. Parents' judgment of students' emotions. Parents were asked to indicate how much enjoyment, anxiety, and boredom they thought their child would experience in mathematics on a 5-point Likert scale using single items for each emotion (e.g., "I feel that my child enjoys math"; second grade sample: enjoyment: $M = 3.98$, $SD = 1.02$; anxiety: $M = 1.60$, $SD = .99$; boredom: $M = 1.60$, $SD = .93$; third grade sample: enjoyment: $M = 3.92$, $SD = 1.03$; anxiety: $M = 1.65$, $SD = 1.01$; boredom: $M = 1.59$, $SD = .91$).

5.1.2.3. Perceived control and value. Students were asked to rate their perceived control and value in mathematics on a 5-point Likert scale. Perceived control was operationalized with three items (e.g., "I can solve math problems well"; second grade sample: $M = 4.05$, $SD = .87$, $\alpha = .88$; third grade sample: $M = 3.92$, $SD = .80$, $\alpha = .87$). Perceived value was assessed with the single item "I think math is important" (second grade sample: $M = 4.46$, $SD = .97$; third grade sample: $M = 4.41$, $SD = .91$).

5.1.2.4. Mathematics achievement. Students' mathematics achievement was measured with a standardized mathematics test (second grade sample: $M = 9.26$, $SD = 3.59$, $\alpha = .77$; third grade sample: $M = 11.08$, $SD = 4.60$, $\alpha = .81$). The test contained 18 items for second graders and 24 items for third graders measuring students' arithmetical and geometrical competencies. It was run in every Bavarian second and third grade classroom on the same day. Typically the teacher of the class administers the test; however, the sample for this study was assessed by externally trained testing personnel in order to ensure objectivity of the test taking procedure.

In addition, teachers were asked to provide students' mathematics grades. Third grade teachers provided the students' grades as documented in the last progress report. Second grade teachers were asked to judge students' math performance based on their performance during the previous school year, as students did not receive any grades until after the end of second grade. Grades in Germany range from 1 (excellent) to 6 (poor). In order to facilitate interpretation, the scores were reverse coded (1 = poor, 6 = excellent; second grade sample: $M = 4.62$, $SD = 1.05$; third grade sample: $M = 4.38$, $SD = 1.02$).

5.2. Results and discussion

5.2.1. Item and scale statistics

Findings on response distributions, item-total correlations, and reliabilities of the AEQ-ES scales are reported in Table 1. Both the anxiety and boredom score distributions were positively skewed, indicating that these emotions are somewhat uncommon in elementary school students. Cronbach's alpha internal reliability thresholds are typically set at .70 (Nunnally, 1978) or .80 (Henson, 2001). The current coefficients ranged from .71 to .93 for the second grade sample and from .73 to .92 for the third grade sample, indicating that reliability was sufficient to excellent for all emotions, in each specific achievement situation, and in both grades. Also, part-whole corrected item-total correlations were excellent, with none of the correlations falling below the .30 threshold. Finally, the AEQ-ES scales appeared to yield sufficient variation in responses, and results support that even children at this young age are able to reliably answer items regarding different emotions.

5.2.1.1. Internal validity: factor structures of emotions. Confirmatory factor analyses (CFAs) were conducted on the achievement emotion items using AMOS 17.0 (Arbuckle, 2006) to examine if the instrument displays internal validity with regard to its presumed internal structures. The analyses were conducted using covariance matrices, and the solutions were generated on the basis of maximum-likelihood estimation. As recommended by Hoyle and Panter (1995), in addition to the traditional chi-square test we used several different indexes to evaluate model fit including the standardized root mean squared residual (SRMR), the comparative fit index (CFI), the Tucker–Lewis index (TLI), and the root-mean square error of approximation (RMSEA). The following criteria were used to evaluate the adequacy of model fit: CFI ≥ .90, TLI ≥ .90 (Browne & Cudeck, 1993), and SRMR < .10 (Kline, 2005). RMSEAs below .05 are thought to indicate good fit, RMSEAs between .05 and .08 reasonable fit, and RMSEAs between .08 and .10 mediocre fit (Browne & Cudeck, 1993).

Three alternative CFA models were tested competitively. Model A assumed one positive and one negative emotion factor, in which positive and negative emotion items loaded on two separate latent factors, respectively. Model B postulated three latent factors, in which items from each emotion loaded on separate latent factors (enjoyment, anxiety, and boredom). Model C was constructed as a hierarchical model postulating eight latent primary factors with the items for each of the eight AEQ-ES scales as indicators, and three latent secondary emotion factors (see Fig. 1).

In the second grade sample, models A and B showed a poor fit (Model A: $\chi^2[169, N = 594] = 1326.93$, $p < .01$, SRMR = .072, CFI = .85 TLI = .84, RMSEA = .107; Model B: $\chi^2[167, N = 594] = 908.83$, $p < .01$, SRMR = .056, CFI = .91, TLI = .89, RMSEA = .087). Likewise, Models A

Table 1
Item and scale statistics in Study 1.

	No. of items	Possible range	Observed range	*M*	*SD*	Skewness	Mean $r_{i(t-i)}$[a]	Alpha
Second grade sample								
Class-related emotions								
Enjoyment	4	4–20	4–20	16.06	4.57	−1.02	.83	.93
Anxiety	4	4–20	4–20	6.29	3.33	1.84	.60	.78
Boredom	4	4–20	4–20	6.96	4.28	1.58	.77	.90
Learning-related emotions								
Enjoyment	2	2–10	2–10	7.12	2.64	−0.52	.57	.73
Anxiety	3	3–15	3–15	4.78	2.69	1.79	.53	.71
Boredom	3	3–15	3–15	5.29	3.37	1.54	.72	.85
Third grade sample								
Class-related emotions								
Enjoyment	4	4–20	4–20	15.36	4.37	−0.77	.82	.92
Anxiety	4	4–20	4–20	5.82	2.65	2.06	.59	.77
Boredom	4	4–20	4–20	6.85	3.84	1.61	.79	.91
Learning-related emotions								
Enjoyment	2	2–10	2–10	6.31	2.54	−0.13	.63	.77
Anxiety	3	3–15	3–15	4.97	2.61	1.74	.56	.73
Boredom	3	3–15	3–15	5.49	3.02	1.33	.64	.79
Test Emotions								
Enjoyment	3	3–15	3–15	10.10	3.62	−0.22	.73	.86
Anxiety	5	5–25	5–25	9.79	5.01	1.18	.70	.87

[a] Median of part-whole corrected item-total correlations.

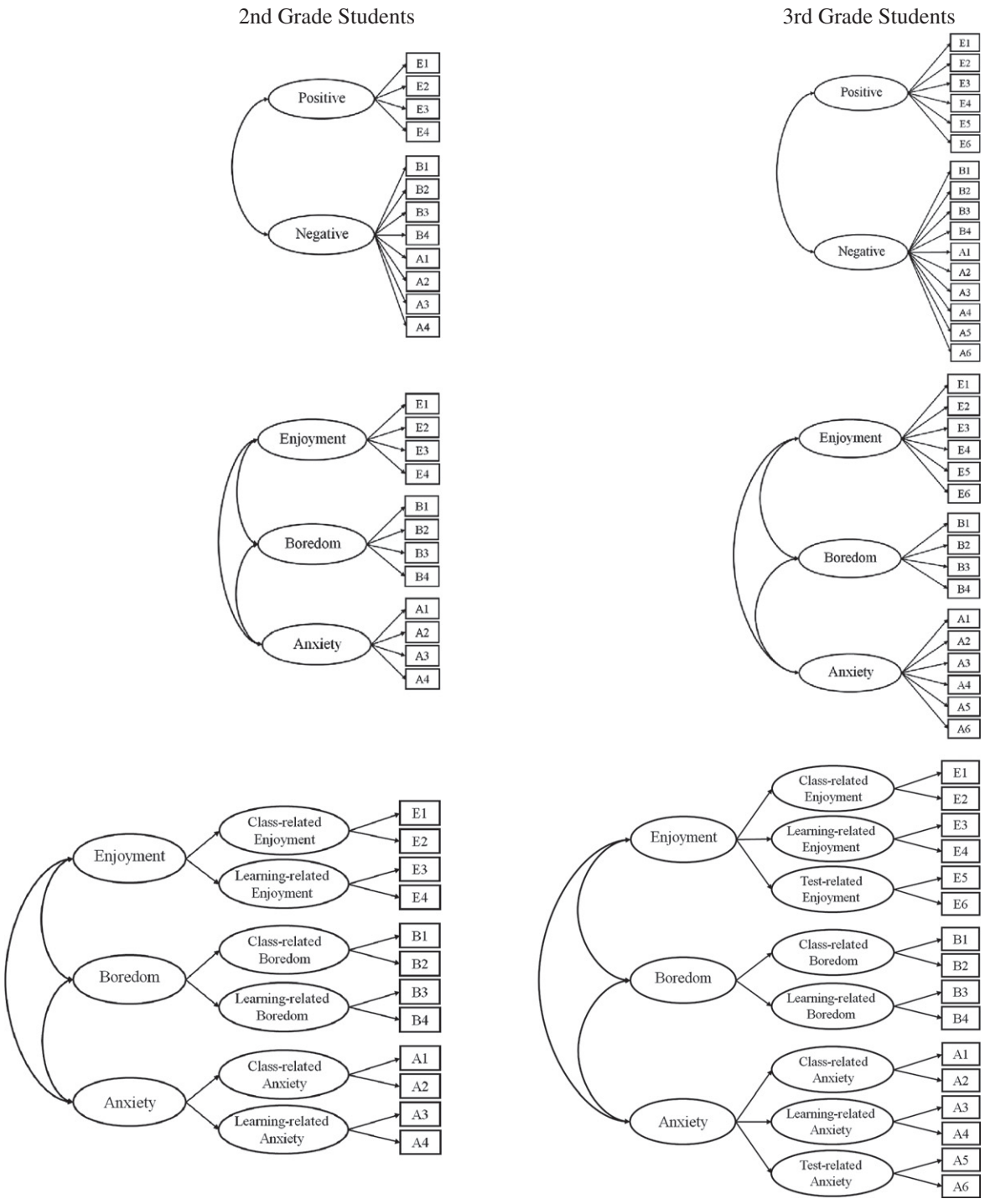


Fig. 1. SEM models for structures of achievement emotions. Models on the left and right represent models for second and third grade students, respectively. The factor structures of models are depicted, not the exact number of items per factor (see method section for exact item numbers). E1/E2, B1/B2, A1/A2 denote class-related enjoyment, boredom, and anxiety; E3/E4, B3/B4, A3/A4 denote learning-related enjoyment, boredom, and anxiety; E5/E6, A5/A6, denote test-related enjoyment and anxiety, respectively. Upper models: Model A = two-factor model consisting of separate latent factors representing the positive and negative emotion dimensions. Middle models: Model B = three-factor model consisting of separate latent factors representing each of the three distinct emotions. Lower models: Model C = hierarchical second-order model distinguishing between different settings for each emotion.

and B showed a poor fit in the third grade sample (Model A: $\chi^2[349, N = 594] = 3131.23$, $p < .01$, SRMR = .087, CFI = .75, TLI = .73, RMSEA = .116; Model B: $\chi^2[347, N = 595] = 1889.92$, $p < .01$ SRMR = .063, CFI = .86, TLI = .85, RMSEA = .087). In contrast, the hierarchical Model C, which accounted for both the different emotions and the different achievement settings, best described the structure of the instrument (second grade: $\chi^2(162, N = 594) = 683.34$, $p < .01$, SRMR = .047, CFI = .93, TLI = .92, RMSEA = .074; third grade: $\chi^2(339, N = 594) = 1223.49$, $p < .01$, SRMR = .057, CFI = .92, TLI = .91, RMSEA = .066).[2] Factor loadings were larger than .61 in the sample of second graders and larger than .66 in the third grade sample, and all

[2] To examine the potential influence of statistical biases resulting from non-normality of distributions, we repeated all the structural equation analyses using maximum-likelihood method with robust standard errors (Asparouhov & Muthen, 2005). The results of all these analyses were very similar to the current results and provided no changes to our conclusions.

Table 2
Manifest intercorrelations of AEQ-E scales in study 1.

	Second grade			Third grade		
	Enjoyment	Anxiety	Boredom	Enjoyment	Anxiety	Boredom
			Correlations within settings			
Class-related emotions						
Anxiety	−.48		−.49			
Boredom	−.76	.57	−.70	.43		
Learning-related emotions						
Anxiety	−.45		−.43			
Boredom	−.61	.65	−.60	.55		
Test-related emotions	–					
Anxiety				−.60		
			Correlations across settings			
Class vs. learning	.72	.64	.75	.69	.66	.71
Class vs. test	–	–	–	.70	.65	–
Learning vs. test	–	–	–	.62	.66	–

Note. For boredom, class-related and learning-related boredom were assessed only.
$p<.01$ for all coefficients.

of them were significant ($ps<.01$). In sum, confirmatory factor analyses corroborated the proposed internal structure of the AEQ-ES and indicated that children distinguish not only between different emotions, but also between these emotions across different academic achievement settings.

5.2.1.2. Internal validity: correlations between emotion scales. We computed Pearson product-moment intercorrelations between the emotion scale scores using SPSS. As can be seen from Table 2, enjoyment correlated negatively with both anxiety and boredom in all three settings. The correlation between the two negative emotions, in contrast, was positive. The positive correlation between boredom and anxiety might indicate that students of this age think of boredom more as an over-challenging rather than an under-challenging emotion (Acee et al., 2010; Pekrun et al., 2010). In addition, examining the correlations of each emotion across the different achievement settings indicates that all three emotions were positively correlated across settings. For example, children who reported more enjoyment in math class also experienced more enjoyment when doing homework and taking tests in mathematics. In sum, the scale correlations corroborate the internal validity of the instrument.

5.2.1.3. External validity: linkages with parents' judgments, students' appraisals, and achievement outcomes. To examine the external validity of the instrument, the scale scores were correlated with parents' judgments of how much enjoyment, anxiety, and boredom they thought their child experienced in mathematics. Overall, the answers given by the children correlated positively with their parents' views (see Table 3). More specifically, the correlations between parents' judgments and their child's judgments of enjoyment and anxiety showed convergence. The correlations between parents' and children's reports of boredom were smaller, perhaps due to boredom being an emotion that is less openly expressed as compared with enjoyment or anxiety. For all three emotions, correlations between parents' and children's views were higher in the third than in the second grade. This may suggest that students provide more valid responses to the AEQ-ES as they become older. Another possible explanation would be that children better communicate their achievement-related emotions to their parents in third grade.

Additionally, the linkages between emotions and students' appraisals and mathematics achievement were analyzed. In line with expectations, the correlations revealed that children who stated that they enjoyed mathematics also reported more perceived control and value in this domain. In contrast, boredom related negatively to perceived control and value which is consistent with previous findings (Pekrun et al., 2010). Likewise, our results provide support for the proposition that anxiety correlates negatively with perceived control (for similar findings, see e.g., Pekrun, Goetz, Titz, & Perry, 2004; Zeidner, 1998). Interestingly, anxiety was negatively related to the perceived value of mathematics as well. This may be due to the fact that children who experience anxiety in a subject may start to devalue the subject in order to feel less distressed. Also, in interpreting this finding it should be noted that anxiety primarily relates to the importance of failure (Pekrun, 2006), which was not assessed in the present study.

The relationships between emotions and mathematics achievement were also in line with expectations. With few exceptions, the relationships between the enjoyment scales and both test achievement and grades were significantly positive. Alternatively, the relationships were negative for anxiety and boredom with achievement.

5.2.1.4. Gender differences. Analyses of variance (ANOVAs) were conducted to investigate mean level differences between boys' and girls' emotions in mathematics. To analyze the relevance of these differences – particularly in light of the large sample size – effect sizes

Table 3
Manifest correlations of achievement emotions with parents' judgments, appraisals, and achievement in Study 1.

		Appraisals		Achievement	
Emotion	Parents' judgments	Academic control	Task value	Test score	Grade
Second grade sample					
Enjoyment	.44**	.62**	.48**	.17**	.23**
	.34**	.46**	.36**	.10*	.11*
Anxiety	.37**	−.53**	−.29**	−.37**	−.31**
	.26**	−.52**	−.34**	−.28**	−.26**
Boredom	.16**	−.48**	−.44**	−.13**	−.24**
	.14**	−.46**	−.37**	−.19**	−.17**
Third grade sample					
Enjoyment	.47**	.65**	.54**	.15**	.23**
	.36**	.48**	.41**	.00	.09*
	.42**	.61**	.41**	.22**	.28**
Anxiety	.49**	−.52**	−.27**	−.29**	−.32**
	.38**	−.53**	−.29**	−.24**	−.27**
	.40**	−.54**	−.26**	−.30**	−.35**
Boredom	.17**	−.44**	−.47**	−.11**	−.16**
	.18**	−.37**	−.41**	−.06	−.12**

Note. Within each block, upper/middle/lower coefficients are for class-, learning-, and test-related emotions, respectively. For boredom, class-related and learning-related boredom were assessed only. In second grade, class-related and learning-related emotions were assessed only. Grades were reverse coded (1 = poor, 6 = excellent).
* $p<.05$.
** $p<.01$.

(Cohen's *d*) of the differences were calculated, in addition to the *F*-values of the ANOVAs (see Table 4). For second grade students, boys reported more class-related enjoyment, less class-related anxiety, and less learning-related boredom than girls. Also, second grade girls reported less perceived control in mathematics than boys. This negative emotional pattern was more apparent in the sample of third-grade girls. In this age group, girls stated that they experienced less enjoyment and more anxiety across the three achievement settings. Moreover, girls reported more class-related boredom and less perceived control and value than boys. These results are in line with previous findings on gender differences in appraisals (e.g., Eccles, Wigfield, Harold, & Blumenfeld, 1993; Wigfield et al., 1997) and emotions (see Frenzel, Pekrun, & Goetz, 2007; Helmke, 1993) related to mathematics. The effects can be considered small in size (Cohen, 1988); however, it should be noted that these differences were larger in the third than in the second grade sample.

5.3. Conclusion

In sum, the present results suggest that our attempt at constructing psychometric scales to measure achievement emotions in elementary school students was successful. Descriptive statistics, confirmatory factor analyses, and relationships with external variables document the reliability, internal validity, and external validity of the instrument. The findings show that students at this age can differentiate between discrete emotions, and furthermore between these emotions across different academic settings. Additionally, they indicate that students' self-reported emotions show moderate convergence with their parents' perceptions of these emotions, as well as meaningful relationships with control- and value-related appraisals and achievement outcomes.

6. Study 2

Study 2 examined the psychometric properties of the AEQ-ES in a sample of American elementary school students. Furthermore, we sought to establish the cross-cultural equivalence of the measure across German and American students by comparing the English version of the AEQ-ES with the German version.

Table 4
Mean gender differences in Study 1.

	M		*SD*			
Emotion	Girls	Boys	Girls	Boys	*F*-value	Cohen's *d*
Second grade sample						
Enjoyment	15.45	16.71	4.60	4.46	11.47	−.28**
	6.93	7.32	2.67	2.60	3.26	−.15
Anxiety	6.67	5.90	3.64	2.93	7.86	.23**
	4.91	4.63	2.79	2.58	1.67	.11
Boredom	7.17	6.74	4.28	4.28	1.52	.10
	5.58	4.98	3.56	3.15	4.73	.18*
Perceived Control	3.83	4.29	.92	.75	45.18	−.55**
Perceived Value	4.45	4.48	1.02	.91	.16	−.03
Third grade sample						
Joy	14.39	16.28	4.61	3.93	29.18	−.44**
	6.02	6.58	2.56	2.49	7.50	−.23**
	9.11	11.04	3.55	3.44	45.41	−.55**
Anxiety	6.34	5.32	2.93	2.25	22.78	.39**
	5.30	4.66	2.80	2.38	9.19	.25**
	11.06	8.58	5.40	4.29	38.49	.51**
Boredom	7.43	6.30	4.01	3.59	13.02	.30**
	5.69	5.30	3.14	2.89	2.46	.13
Perceived Control	3.73	4.09	.79	.76	31.65	−.47**
Perceived Value	4.31	4.49	.97	.84	6.03	−.21*

Note. Within each block, upper/middle/lower coefficients are for class-, learning-, and test-related emotions, respectively. For boredom, class-related and learning-related boredom were assessed only. In second grade, class-related and learning-related emotions were assessed only.
* $p<.05$.
** $p<.01$.

6.1. Method

6.1.1. Participants and procedure

The study was run in four different elementary schools in Minnesota (USA). The sample consisted of 163 third-grade students (95 females; mean age = 8.69 years, range 7 to 10, SD = .42) from a range of socio-economic backgrounds. Participants were predominantly white Caucasians ($n = 161$) and only 2 students were African Americans.[3]

As in Study 1, students completed the questionnaire during regular classroom hours. Before completing the questionnaire, students were asked to express their personal opinion and judgments, were informed that there were no right or wrong answers, and were assured of the confidentiality of their responses. All items were read aloud to the students.

6.1.2. Measures

6.1.2.1. Translation of measures. The translation was carried out in accordance with Brislin's (1980) recommendations for pragmatic translation of psychometric instruments. The original German version of the AEQ-ES, as well as the perceived control and value items, were first translated into English by the first author of the present study and then translated back to the source language by a research assistant as a check for consistency. Next, the original and back-translated versions were compared to ensure equivalence of content.

6.1.2.2. Achievement emotions. The English version of the AEQ-ES was used to assess achievement emotions. As in the German AEQ-ES, the measurement consisted of eight scales pertaining to experiencing different emotions (enjoyment, anxiety, and boredom) when attending class, doing homework, and taking tests and exams[4]; once again, boredom during tests was not assessed. Items were answered on a 5-point Likert scale anchored by the same male and female graphical displays of faces showing increasing emotional intensity as in Study 1.

6.1.2.3. Parents' judgment of students' emotions. Parents were again asked to indicate how much enjoyment, anxiety, and boredom they thought their child would experience in mathematics using single items for each emotion (enjoyment: $M = 3.69$, $SD = .99$; anxiety: $M = 1.88$, $SD = .94$; boredom: $M = 2.01$, $SD = .99$).

6.1.2.4. Perceived control and value. Perceived control ($M = 11.58$, $SD = 2.93$, $\alpha = .85$) and value ($M = 4.27$, $SD = 1.14$) were assessed with English versions of the same items as in Study 1.

6.1.2.5. Mathematics achievement. To assess students' mathematics achievement, teachers were asked to provide students' grades of the last progress report. The grades were converted into numbers (F = 1, A+ = 13; $M = 9.54$, range 1 to 13, $SD = 2.42$).

6.2. Results and discussion

6.2.1. Item and scale statistics

Descriptive statistics, item-total correlations, and reliabilities of the AEQ-ES scales are reported in Table 5. As in the German sample, both the anxiety and boredom scales showed positively skewed distributions, indicating that these emotions are likewise less pronounced in American elementary school students. Cronbach

[3] As in Study 1, students who failed to answer the complete survey (n = 3) were excluded from the sample. The final sample consisted of 160 students.

[4] Although students in the US Grade III sample were younger than the Grade II students in the German sample, the US students still received grades, so we felt that they should likewise be able to answer questions related to taking tests and exams. Alternatively, German second graders in the present study had not received any grades during their first two years of schooling, and have less experience with taking tests and exams as compared to students in the US sample.

Table 5
Item and scale statistics in Study 2.

	No. of items	Possible range	Observed range	M	SD	Skewness	Mean $r_{i(t-i)}$ [a]	Alpha
Class-related emotions								
Enjoyment	4	4–20	4–20	15.13	5.17	−0.73	.88	.95
Anxiety	4	4–20	4–20	6.22	3.47	2.22	.68	.84
Boredom	4	4–20	4–20	7.07	4.35	1.62	.82	.92
Learning-related emotions								
Enjoyment	2	2–10	2–10	6.35	2.77	−0.12	.64	.78
Anxiety	3	3–15	3–15	4.76	2.75	1.88	.58	.75
Boredom	3	3–15	3–15	5.66	3.85	1.52	.86	.93
Test emotions								
Enjoyment	3	3–15	3–15	10.68	4.16	−0.48	.82	.91
Anxiety	5	5–25	5–25	8.70	4.78	1.80	.69	.86

[a] Median of part-whole corrected item-total correlations.

alphas ranged from .75 to .95, indicating sufficiently high reliability for all three emotions in each achievement setting. Also, part-whole corrected item-total correlations for all items were excellent as none of the correlations fell below the .30 threshold. Overall, the AEQ-ES scales yielded sufficient variation and reliability in the American sample, thus replicating the Study 1 findings for the German sample.

6.2.2. Internal validity: factor structures of emotions

As in Study 1, CFAs were conducted with the achievement emotion items to examine the internal validity of the instrument. Again, three alternative models were tested competitively: Model A assuming one positive and one negative emotion factor, Model B including three separate latent emotion factors (enjoyment, anxiety, and boredom), and Model C as a hierarchical model with eight latent primary factors for the three emotions in the different achievement situations and three latent secondary emotion factors. As in Study 1, neither Model A (χ^2[349, N = 163] = 1480.07, p < .01, SRMR = .102, CFI = .72 TLI = .70, RMSEA = .142) nor Model B (χ^2[347, N = 163] = 1015.31, p < .01, SRMR = .066, CFI = .84, TLI = .82, RMSEA = .110) yielded a good fit to the data. The findings again confirmed the hypothesized hierarchical Model C, which accounted for the different emotions within the various achievement situations (χ^2[339, N = 163] = 721.45, p < .01, SRMR = .062, CFI = .91, TLI = .90, RMSEA = .084). All of the factor loadings were larger than .54 and significant (ps < .01). In sum, confirmatory factor analyses supported the eight-factor structure found in Study 1 and confirmed that elementary school children distinguish not only between different emotions, but also between these emotions across different achievement settings.

6.2.3. Internal validity: correlations between emotion scales

The relationships between the scale scores were again analyzed using Pearson product-moment correlations (see Table 6). As in Study 1, enjoyment displayed negative correlations with both anxiety and boredom, whereas the two negative emotions correlated positively. Likewise, the intercorrelations of each emotion across the different achievement settings were positive, indicating, for instance, that those children who reported more anxiety in math class also experienced more anxiety when doing homework or taking tests in mathematics. Overall, scale validities support the findings from Study 1 and corroborate the internal validity of the English AEQ-ES scales.

6.2.4. External validity: linkages with parents' judgments, students' appraisals, and achievement

To investigate the external validity of the English AEQ-ES, children's scale scores were again correlated with parents' judgments regarding how much enjoyment, anxiety, and boredom they thought their child would experience in mathematics. The expected positive correlations between parents' judgments and their child's self-reported emotions were again seen in the American sample, although convergence was less pronounced than in Study 1. Specifically, the strongest positive correlations were between parents' and their child's judgments of anxiety and enjoyment, whereas the correlations were smaller for boredom.

The correlations between American students' emotions and appraisals displayed a pattern that was consistent with those in the German sample (see Table 7). Students' enjoyment correlated positively with both perceived control and value. In contrast, anxious as well as bored children reported less perceived control and value in mathematics. As anticipated, enjoyment correlated positively with math achievement, whereas correlations were negative for anxiety and boredom.

6.2.5. Gender differences

Mean level differences between boys' and girls' emotions were examined using ANOVAs, and effect sizes (Cohen's d) were calculated to evaluate the practical significance of any differences. Interestingly, for the American sample the only significant gender difference was in learning-related boredom, as boys (M = 6.37, SD = 4.54) reported more boredom than girls (M = 5.16, SD = 3.20, F[1, 159] = 3.97, p < .05, d = −.32) similar to the findings for North American university students' boredom reported by Pekrun et al. (2010; Study 5). Also, boys and girls did not differ with regard to perceived control, F(1, 159) = .38, p = .54, or value, F9(1, 159) = .11, p = .75. Perceived control and value are central antecedents for students' emotions (Pekrun, 2006). The fact that boys and girls report equal levels of perceived control and value may therefore account for the similar emotional patterns across gender in the American sample.

6.2.6. Construct equivalence across the German and US samples

We tested if the structures of the American and German versions of the AEQ-ES scales were equivalent across the two samples. Measurement invariance is seen as a necessary precondition to compare scores between different groups. If measurement invariance requirements are not met, scores represent qualitatively different constructs and

Table 6
Manifest intercorrelations of AEQ-ES scales in Study 2.

	Enjoyment	Anxiety	Boredom
Correlations within settings			
Class-related emotions			
Anxiety	−.42		
Boredom	−.72	.59	
Learning-related emotions			
Anxiety	−.37		
Boredom	−.72	.51	
Test-related emotions			
Anxiety	−.55		
Correlations across settings			
Class vs. learning	.74	.69	.79
Class vs. test	.79	.70	–
Learning vs. test	.73	.71	–

Note. For boredom, class-related and learning-related boredom were assessed only. p < .01 for all coefficients.

Table 7
Manifest correlations of achievement emotions with parents' judgments, appraisals, and math grade in Study 2.

Emotion	Appraisals		Task value	Math grade
	Parents' judgments	Academic control		
Enjoyment	.35**	.53**	.46**	.14
	.30**	.41**	.36**	.07
	.24**	.54**	.44**	.12
Anxiety	.23**	−.41**	−.36**	−.24**
	.25**	−.38**	−.28**	−.14
	.16**	−.43**	−.28**	−.20**
Boredom	.20**	−.38**	−.42**	−.23**
	.17**	−.33**	−.37**	−.15

Note. Within each block, upper/middle/lower coefficients are for class-, learning-, and test-related emotions, respectively. For boredom, class-related and learning-related boredom were assessed only.
** $p<.01$
* $p<.05$.

comparisons are difficult to interpret (Hoyle, 2005). As Horn (1991) stated, "Without evidence of measurement invariance, the conclusions of a study must be weak" (p. 119). To examine measurement invariance, a series of multigroup CFAs were conducted and freely estimated models were compared with models containing increasingly stringent constraints.

As a first step, we specified a configural invariance model (Model 1) in which the same pattern of fixed and free factor loadings is specified for each group (Horn, McArdle, & Mason, 1983; Steenkamp & Baumgartner, 1998). This model is supposed to be the "minimal condition for factorial invariance" (Marsh, 1993, p. 851) and serves as the baseline model for comparison with all subsequent models (Hong, Malik, & Lee, 2003). As specified in Study 1, the hierarchical model (Model C) served as the baseline model for comparison across the two samples. Metric invariance was tested by constraining item loadings to be identical across groups (Model 2). Next, the scalar invariance model was tested involving invariance of the item intercepts in addition to metric invariance (Model 3). In the second-order factor invariance model, the loadings of second-order factors on first-order factors were set to be equal across the groups (Model 4). The factor covariance invariance model additionally set the factor covariances to be equal across the groups (Model 5). Finally, we tested the factor variance invariance model (Model 6), in which factor variances were constrained to be equal across the groups, in addition to the constraints described for the previous models. When we compared the six nested models we evaluated the differences in fit using the CFI and the TLI. A loss of fit of $\Delta CFI > .01$ and $\Delta TLI > .05$ was regarded as substantial (Cheung & Rensvold, 2002; Little, 1997).

Loss of fit associated with the addition of each of the constraints was negligible for each of the proposed models (see Table 8). This result suggests that all of the parameters tested (factor loadings, intercepts, means, and covariances) were sufficiently equivalent across the German and American samples. In sum, the non-significant differences between the models indicate that the structure of the AEQ-ES is fully equivalent across the German and American samples. By implication, the findings provide evidence that item and scale scores of the AEQ-ES can be interpreted in the same way across the German and English versions of the instrument.

6.2.7. Mean differences in emotions between the German and US samples

Comparing the emotion scale statistics between countries, we found that means were different for the test-related emotions only. Specifically, German students reported significantly more test-related anxiety ($M = 9.79$, $SD = 5.01$) compared to American students ($M = 8.70$, $SD = 4.78$; $t[754] = 2.47$, $p < .05$). In addition, we found that German students indicated having marginally less test-related enjoyment ($Ms = 10.10$ and 10.68, $SDs = 3.62$ and 4.16 for German and US students, respectively; $t[754] = -1.74$, $p < .10$). There were no significant mean differences for any of the students' class-related or learning-related emotions, and no differences with respect to their perceived control and value.

In interpreting these differences, it should be taken into account that the sample size between the two samples differed considerably, implying that the results should be interpreted cautiously. A plausible interpretation is that students are selected into different tracks after elementary school in Germany, implying that test taking may be more threatening and less enjoyable for German elementary school students.

7. General discussion

The present research had three aims, all of which were accomplished with the current studies. First, we aimed at constructing an instrument measuring students' emotions in the elementary school years. Second, we sought to provide evidence for Pekrun's (2006) control-value theory in early elementary school students by investigating the relationships of students' achievement emotions with their control and value appraisals, as well as their academic performance. Finally, we sought to support cross-cultural equivalence of the emotion constructs as measured by the AEQ-ES scales.

7.1. Construction and psychometric quality of the AEQ-ES

The results of the present research provide support for the reliability and validity of the AEQ-ES. Specifically, reliabilities were good or excellent for all emotion scales in each of the two studies. Moreover, the correlations between the emotion scales provide support for the internal validity of the measure and are consistent with related findings for university students (Pekrun et al., 2011). Results of structural equation modeling showed that enjoyment, anxiety, and boredom differed across the three types of academic settings involved (i.e., attending class, doing homework, and taking tests and exams), which indicates that elementary school children can distinguish between emotions in different achievement settings. External validity of the AEQ-ES scales was also documented in positive correlations between AEQ-ES scores and parents' judgments of students' emotions.

Table 8
Tests of invariance of the measurement model and latent means: model fit indexes.

	Model fit								
Model	χ^2	df	SRMR	CFI	TLI	RMSEA	$\Delta\chi^2$	ΔCFI	ΔTLI
1. Configural invariance	1946.34	678	.057	.92	.91	.050	–	–	–
2. Metric invariance	1999.44	698	.056	.91	.91	.050	53.10	.00	.00
3. Scalar invariance	2115.17	726	.056	.91	.90	.050	115.73	.01	.00
4. Invariance of second order factors	2132.48	731	.057	.91	.90	.050	17.31	.00	.00
5. Invariance of factor covariances	2147.97	734	.059	.91	.90	.051	15.49	.00	.00
6. Invariance of second order factor variances	2155.70	737	.059	.91	.90	.051	7.73[a]	.00	.00

Note. *df* = degrees of freedom; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error of approximation. All other *p* values for the chi-square tests are less than .01.
[a] The *p* value for this chi-square test is not significant.

The majority of previous research on achievement emotions has focused only on anxiety or on the attributional antecedents of these emotions (Bong, 2009; Stipek & Gralinski, 1991; Zeidner, 1998). Research assessing emotions more broadly focused almost exclusively on student populations beyond elementary school. A potential reason for this gap in the literature is the lack of instruments designed to assess elementary school students' emotions. The current results therefore represent a critical step forward, as researchers can now study young students' emotions and their impact in a more sophisticated manner.

7.2. Appraisal antecedents and achievement outcomes of emotions

In line with recent findings on achievement emotions experienced by college students (Pekrun et al., 2010, 2011), our research provides evidence corroborating aspects of Pekrun's (2006) control-value theory of achievement emotions regarding the linkages between achievement emotions, students' appraisals, and their academic achievement. Specifically, enjoyment correlated positively with both perceived control and task value, and boredom and anxiety were negatively related to these variables.

Furthermore, the study findings support the hypothesis that achievement emotions are significantly related to students' academic performance. In line with previous research (Helmke, 1993), enjoyment correlated positively with mathematics achievement. In contrast, both anxiety and boredom were negatively related to mathematics achievement. These results are in line with the control-value theory of achievement emotions (Pekrun et al., 2002; Pekrun et al., 2011) as well as the extensive literature on test anxiety (see Hembree, 1988; Zeidner, 1998).

Interestingly, correlations between emotions and achievement, specifically the correlations for enjoyment and anxiety, were higher in the German than in the American sample. However, care should be taken when interpreting these differences, because grades in the two samples were based on different scales and sample size differed considerably between the two samples. One possible reason for this cross-cultural difference is that German students receive grades in the context of high-stakes testing early in elementary school, thus making performance more salient for German students.

7.3. Applicability of the AEQ-ES across cultures and school subjects

Using multigroup confirmatory factor analysis the present results support the measurement invariance of the German and American versions of the AEQ-ES. Cross-cultural equivalence was also suggested by the similarity of relationships between achievement emotions, control and value beliefs, and achievement in mathematics. These findings provide evidence for the applicability of the instrument with both German and American students, as well as the utility of the AEQ-ES for cross-cultural research. The fact that German students reported significantly more test-related anxiety than American students further suggests that cross-cultural work may be particularly useful in determining environmental and cultural influences on students' emotions.

While the present study investigated students' emotions across two samples from Western cultures, it would also be valuable to compare the present results with findings from samples of other cultures, such as collectivist (e.g., Asian) cultures. A recent study found that achievement emotions differ considerably between Chinese and German middle school students, with Chinese children experiencing higher levels of enjoyment, pride, anxiety, and shame, and German students reporting higher levels of anger (Frenzel, Thrash, Pekrun, & Goetz, 2007). Consequently, there may also be differences in elementary school students' achievement emotions across individualistic and collectivistic cultures.

The present studies provide support for the usefulness of the AEQ-ES scales to assess emotions in mathematics. However, we anticipate that the instrument can also be used to assess emotions in other school subjects or academic domains by adapting the items accordingly. This can be done by simply exchanging the term "mathematics" with terms related to other subjects in each item. Consequently, an important task for future research would be to assess the applicability of the instrument in subject domains other than mathematics.

8. Limitations and conclusions

Some limitations should be considered when interpreting the present results and in designing future research. First, while the present study is consistent with Pekrun's (2006) control-value theory and supports the external validity of the AEQ-ES, the study design was correlational, thus precluding any firm causal conclusions regarding the linkages between students' control and value beliefs, their achievement emotions, and their academic achievement. As yet, longitudinal studies on these linkages are largely lacking. One of the few exceptions is the study by Helmke (1993), which investigated cross-lagged effects of enjoyment and achievement across the elementary school years. The findings showed that enjoyment had positive effects on achievement, but that achievement reciprocally influenced students' enjoyment. Future studies should address the reciprocal linkages between elementary school students' achievement emotions, their appraisals, and their academic achievement.

Second, self-report instruments are susceptible to response biases (Campbell & Fiske, 1959; Graham, Collins, Donaldson, & Hansen, 1993; Schwarz, 1999; Stone et al., 2000), and correlational linkages may be inflated by common method variance caused by these biases (Borman, 1991; Donaldson, Thomas, Graham, Au, & Hansen, 2000; Spector, 1994). Research on personality judgment suggests that the self holds unique advantages and disadvantages as a judge of personality (Funder, 1999; John & Robins, 1993, 1994). As stated by Abe (2004, p. 338), "The self has access to internal thoughts and feelings and other 'privileged information' which is not available to the external observer. On the other hand, the self is also more ego-involved than others in its assessment of itself and has greater difficulty viewing itself objectively". In the present research we aimed at addressing the limitations of self-report by validating children's responses to the AEQ-ES items by means of parents' judgments of their child's emotions in the present research. Future research should also link elementary school students' self-reports of emotions to physiological indicators and implicit measures of emotions.

Finally, it is important to consider the implications of the present research for educational practice. Our findings imply that the AEQ-ES can be used to assess elementary school students' achievement emotions and may be a useful instrument not only for scientific research, but also for teachers and educators with applied purposes. The linkages of students' emotions with achievement outcomes at this very early age further highlight the necessity of research on achievement emotions in younger age groups. Students who are not performing well may adopt detrimental emotional patters, which in turn may impair their academic achievement. To gain knowledge how to prevent such a vicious cycle, it seems highly important to investigate emotions and implement interventions designed to improve students' emotions early in the elementary school years. For this purpose, the instrument developed in this research may serve important diagnostic functions.

In closing, the present research illustrates the utility of attending to achievement emotions in the early elementary school years. We believe that studying emotions in this early age group is an important step for researchers, scientists, and educators alike. We hope that the instrument developed and evaluated in the present study serves as a catalyst for future research in this important area of inquiry.

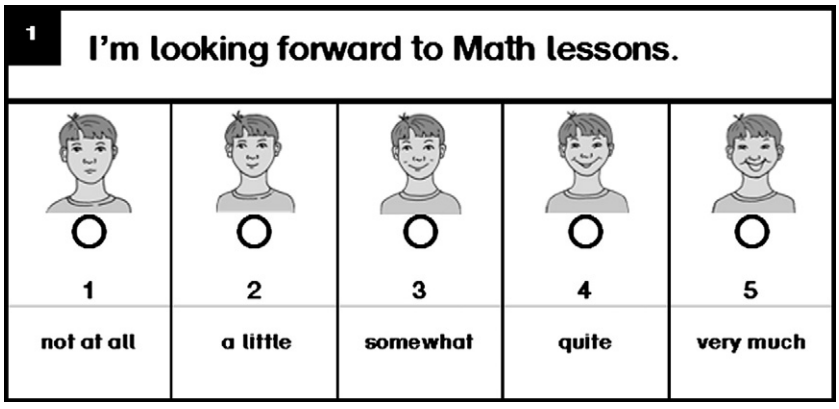
Appendix A

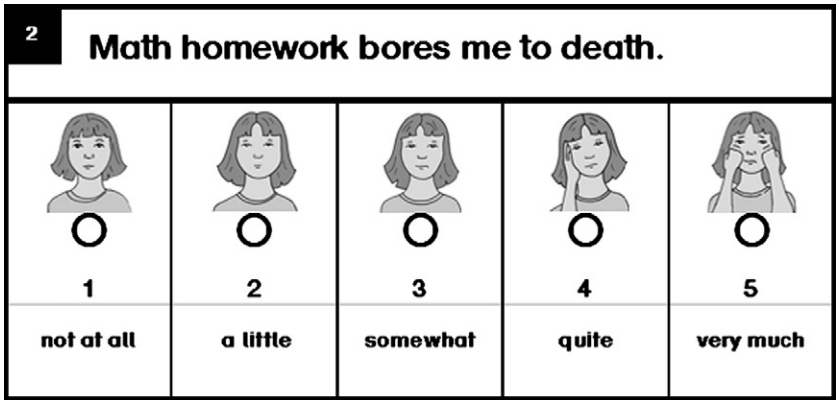
Emotions Questionnaire–Elementary School (AEQ-E): Scales and Sample Items

Class-related emotions	
Enjoyment	I enjoy math class.
Anxiety	When I think about math class, I get nervous.
Boredom	I find math class so boring that I would rather do something else.
Learning-related emotions	
Enjoyment	When I do math homework, I am in a good mood.
Anxiety	When I do math homework, I worry if I will ever understand it.
Boredom	Math homework bores me to death.
Test emotions	
Enjoyment	I look forward to math tests.
Anxiety	I get very nervous during math tests.

Appendix B

Sample Items from the AEQ-E (version for girls and boys)





References

Abe, J. A. (2004). Shame, guilt, and personality judgment. *Journal of Research in Personality*, *38*, 85–104. doi:10.1016/S0092-6566(03)00055-2.

Acee, T. W., Kim, H., Kim, H. J., Kim, J. -I., Chu, H. -N. R., Kim, M., et al. (2010). Academic boredom in under- and over-challenging situations. *Contemporary Educational Psychology*, *35*, 17–27. doi:10.1016/j.cedpsych.2009.08.002.

Arbuckle, J. L. (2006). *(Version 17.0) [Computer program].* Chicago: SPSS.

Asparouhov, T., & Muthen, B. (2005). Multivariate statistical modeling with survey data. *Proceedings of the Federal Committee on Statistical Methodology (FCSM) research conference*. http://www.fcsm.gov/05papers/Asparouhov_Muthen_IIA.pdf.

Baumert, J., Trautwein, U., & Artelt, C. (2003). School environments — Institutional conditions of learning and instruction. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, & U. Schiefele (Eds.), *PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (pp. 261–331). Opladen, Germany: Leske+Budrich.

Bong, M. (2001). Between- and within-domain relations of academic motivation among middle and high school students: Self-efficacy, task value, and achievement goals. *Journal of Educational Psychology*, *93*, 23–34. doi:10.1037//0022-0663.93.1.23.

Bong, M. (2009). Age-related differences in achievement goal differentiation. *Journal of Educational Psychology*, *101*, 879–896. doi:10.1037/a0015945.

Borman, W. C. (1991). Job behavior, performance, and effectiveness. In L. M. Hough, & M. D. Dunnette (Eds.), *Handbook of industrial and organizational psychology, Vol. 2* (pp. 271–326). (2nd ed). Palo Alto, CA: Consulting Psychologists Press.

Brislin, R. W. (1980). Translation and content analysis of oral and written materials. In H. C. Triandis, & J. W. Berry (Eds.), *Handbook of cross-cultural psychology*, *Vol. 2*. (pp. 389–444). Boston, MA: Allyn & Bacon.

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). Beverley Hills, CA: Sage.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105. doi:10.1037/h0046016.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233–255. doi:10.1207/s15328007sem0902_5.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety.* San Francisco, CA: Jossey-Bass.

Csikszentmihalyi, M., & Larson, R. W. (1987). Validity and reliability of the experience — Sampling method. *Journal of Nervous and Mental Diseases*, *9*, 526–536 doi: 0022-3018/87/1759-0526$02.00/0.

Donaldson, S. I., Thomas, C. W., Graham, J. W., Au, J. G., & Hansen, W. B. (2000). Verifying drug abuse prevention program effects using reciprocal best friend reports. *Journal of Behavioral Medicine*, *23*, 585–601. doi:10.1023/A:1005559620739.

Eccles, J. S., Wigfield, A., Harold, R. D., & Blumenfeld, P. (1993). Age and gender differences in children's self- and task perceptions during elementary school. *Child Development*, *64*, 830–847. doi:10.2307/1131221.

Efklides, A., & Volet, S. (Eds.). (2005). Feelings and emotions in the learning process [special issue]. *Learning and Instruction*, *15*. (pp. 377–515).

Entwisle, D. R., & Alexander, K. L. (1993). Entry into school: The beginning school transition and educational stratification in the United States. *Annual Review of Sociology*, *19*, 401–423. doi:10.1146/annurev.so.19.080193.002153.

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American Psychologist*, *56*, 218–226. doi:10.1037//0003-066X.56.3.218.

Frenzel, A. C., Pekrun, R., & Goetz, T. (2007a). Girls and mathematics — A "hopeless" issue? A control-value approach to gender differences in emotions towards mathematics. *European Journal of Psychology of Education*, *22*, 497–514. doi:10.1007/BF03173468.

Frenzel, A. C., Thrash, T., Pekrun, R., & Goetz, T. (2007b). Achievement emotions in Germany and China. A cross-cultural validation of the academic emotions questionnaire-mathematics. *Journal of Cross Cultural Psychology*, *38*, 302–309. doi:10.1177/0022022107300276.

Funder, D. C. (Ed.). (1999). *Personality judgment: A realistic approach to person perception*. San Diego, CA: Academic Press.

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., Pekrun, R., Hall, N. C., & Haag, 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*, 289–308. doi:10.1348/000709905X42860.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549–576. doi:10.1146/annurev.psych.58.110405.085530.

Graham, J. W., Collins, N. L., Donaldson, S. I., & Hansen, W. B. (1993). Understanding and controlling for response bias: Confirmatory factor analysis of multitrait-multimethod data. In R. Steyer, K. F. Wender, & K. F. Widamen (Eds.), *Psychometric Methodology* (pp. 585–590). Stuttgart: Gustav Fisher Verlag.

Grützemann, W. (2003). *Skalendokumentation "Mathematikspezifische Personmerkmale bei GrundschülerInnen der vierten Klassenstufe" Berichte zur Pädagogischen Psychologie, 6.* Halle-Wittenberg, Germany: Martin-Luther-Universität.

Hallinan, M. T., & Sørensen, A. B. (1983). The formation and stability of instructional groups. *American Sociological Review*, *48*, 838–851. doi:10.2307/2095329.

Helmke, A. (1993). Die Entwicklung der Lernfreude vom Kindergarten bis zur 5. *Klassenstufe. Zeitschrift für Pädagogische Psychologie*, *7*, 77–86.

Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research*, *58*, 47–77. doi:10.2307/1170348.

Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development*, *34*, 177–189.

Hong, S., Malik, M. L., & Lee, M. (2003). Testing configural, metric, scalar, and latent mean invariance across genders in sociotropy and autonomy using a non-Western sample. *Educational and Psychological Measurement*, *63*, 636–654. doi:10.1177/0013164403251332.

Horn, J. L. (1991). Comments on "Issues in factorial invariance". In L. M. Collins, & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 114–125). Washington, DC: American Psychological Association.

Horn, J. L., McArdle, J., & Mason, R. (1983). When is invariance not invariant: A practical scientist's look at the ethereal concept of factor invariance. *Southern Psychologist*, *1*, 179–188.

Hoyle, R. H. (2005). Applications of structural equation modeling in clinical and health psychology research. In J. Miles, & P. Gilbert (Eds.), *A handbook of research methods for clinical and health psychology* (pp. 261–278). New York, NY: Oxford University Press.

Hoyle, R. H., & Panter, A. T. (1995). Writing about structural equation models. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 158–176). Thousand Oaks, CA: Sage.

John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The Big Five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality*, *61*, 521–551. doi:10.1111/j.1467-6494.1993.tb00781.x.

John, O. P., & Robins, R. W. (1994). Accuracy and bias in self-perception: Individual differences in self-enhancement and the role of narcissism. *Journal of Personality and Social Psychology*, *66*, 206–219. doi:10.1037/0022-3514.66.1.206.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: Guilford Press.

Linnenbrink, E. A. (2006). Emotion research in education: Theoretical and methodological perspectives on the integration of affect, motivation, and cognition. *Educational Psychology Review*, *18*, 307–314. doi:10.1007/s10648-006-9028-x.

Linnenbrink-Garcia, L., & Pekrun, R. (Eds.). (2011). Students' emotions and academic engagement [Special issue]. *Contemporary Educational Psychology*, *36*, 1–70.

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, *32*, 53–76. doi:10.1080/10705519909540118.

Little, T. D., Oettingen, G., Stetsenko, A., & Baltes, P. B. (1995). Children's action-control beliefs about school performance: How do American children compare with German and Russian children? *Journal of Personality and Social Psychology*, *69*, 686–700. doi:10.1037/0022-3514.69.4.686.

Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for Research in Mathematics Education*, *30*, 520–540. doi:10.2307/749772.

Maaz, K., Trautwein, U., Lüdtke, O., & Baumert, J. (2008). Educational transitions and differential learning environments: How explicit between-school tracking contributes to social inequality in educational outcomes. *Child Development Perspectives*, *2*, 99–106. doi:10.1111/j.1750-8606.2008.00048.x.

Marsh, H. W. (1986). Verbal and math self-concept: An internal / external frame of reference model. *American Educational Research Journal*, *23*, 129–149. doi:10.2307/1163048.

Marsh, H. W. (1993). The multidimensional structure of academic self-concept: Invariance over gender and age. *American Educational Research Journal*, *30*, 841–860. doi:10.2307/1163206.

Mintrop, H. (1997). Retracking on a grand scale: Policy and pedagogy in the reform of eastern German secondary schools after the fall of socialism. *Journal of Education Policy*, *12*, 333–354. doi:10.1080/0268093970120503.

Nunnally, J. C. (1978). *Introduction to psychological measurement*. New York, NY: McGraw-Hill.

Pallas, A. M., Entwisle, D. R., Alexander, K. L., & Stluka, M. F. (1994). Ability-group effects: Instructional, social, or institutional? *Sociology of Education*, *67*, 27–46. doi:10.2307/2112748.

Patrick, B. C., Skinner, E. A., & Connell, J. P. (1993). What motivates children's behavior and emotion? Joint effects of perceived control and autonomy in the academic domain. *Journal of Personality and Social Psychology*, *65*, 781–791. doi:10.1037/0022-3514.65.4.781.

Pekrun, R. (1992). Expectancy-value theory of anxiety: Overview and implications. In K. Wrzesniewski, D. G. Forgays, & T. Sosnowski (Eds.), *Anxiety: Recent developments in cognitive, psychophysiological, and health research* (pp. 23–41). Washington, DC: Hemisphere Publishing Corp.

Pekrun, R. (2000). A social-cognitive, control-value theory of achievement emotions. In J. Heckhausen (Ed.), *Motivational psychology of human development: Developing motivation and motivating development* (pp. 143–163). New York, NY: Elsevier Science.

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, *18*, 315–341. doi:10.1007/s10648-006-9029-9.

Pekrun, R., Frenzel, A. C., Goetz, T., & Perry, R. P. (2007). The control-value theory of achievement emotions: An integrative approach to emotions in education. In P. A. Schutz, & R. Pekrun (Eds.), *Emotion in education* (pp. 13–36). San Diego, CA: Academic Press.

Pekrun, R., Goetz, T., Daniels, L. M., Stupnisky, R. H., & Perry, R. P. (2010). Boredom in achievement settings: Exploring control-value antecedents and performance outcomes of a neglected emotion. *Journal of Educational Psychology*, *102*, 531–549.

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*, *17*, 287–316. doi:10.1080/10615800412331303847.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, *37*, 91–106. doi:10.1207/S15326985EP3702_4.

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology*, *36*, 36–48.

Perry, R. P. (1991). Perceived control in college students: Implications for instruction in higher education. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research*, *Vol. 7*. (pp. 1–56) New York, NY: Agathon.

Rosenholtz, S. J., & Simpson, C. (1984). The formation of ability conceptions: Developmental trend or social construction? *Review of Educational Research*, *54*, 31–63. doi:10.2307/1170397.

Schutz, P. A., & Pekrun, R. (2007). Introduction to emotion in education. In P. A. Schutz, & R. Pekrun (Eds.), *Emotion in education* (pp. 3–10). San Diego, CA: Academic Press.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *The American Psychologist*, *54*, 93–105. doi:10.1037/0003-066X.54.2.93.

Seipp, B. (1991). Anxiety and academic performance: A meta-analysis of findings. *Anxiety Research*, *4*, 27–41. doi:10.1080/08917779108248762.

Slaney, K. L., & Maraun, M. D. (2008). A proposed framework for conducting data-based test analysis. *Psychological Methods*, *13*, 376–390. doi:10.1037/a0014269.

Spector, P. E. (1994). Using self-report questionnaires in OB research: A comment on the use of a controversial method. *Journal of Organizational Behavior*, *15*, 385–392.

Steenkamp, J. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*, 78–90. doi:10.1086/209528.

Stipek, D. J., & Gralinski, J. H. (1991). Gender differences in children's achievement-related beliefs and emotional responses to success and failure in mathematics. *Journal of Educational Psychology*, *83*, 361–371. doi:10.1037/0022-0663.83.3.361.

Stone, A. A., Turkkan, J. S., Bachrach, C. A., Jobe, J. B., Kurtzman, H. S., & Cain, V. S. (Eds.). (2000). *The science of self-report: Implications for research and practice*. Mahwah, NJ: Lawrence Erlbaum Associates.

Suinn, R. M., Taylor, S., & Edwards, R. W. (1988). Suinn mathematics anxiety rating scale for elementary school students (MARS-E): Psychometric and normative data. *Educational and Psychological Measurement*, *48*, 979–986. doi:10.1177/0013164488484013.

Turner, J. E., & Schallert, D. L. (2001). Expectancy-value relationships of shame reactions and shame resiliency. *Journal of Educational Psychology*, *93*, 320–329. doi:10.1037/0022-0663.93.2.320.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, *92*, 548–573. doi:10.1037/0033-295X.92.4.548.

Weinert, F. E., & Helmke, A. (Eds.). (1997). *Entwicklung im Grundschulalter*. Psychologie Verlags Union: Weinheim.

Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbreton, A. J. A., Freedman Doan, C., et al. (1997). Change in children's competence beliefs and subjective task values across the elementary school years: A 3-year study. *Journal of Educational Psychology*, *89*, 451–469.

Zeidner, M. (1998). *Test anxiety: The state of the art*. New York, NY: Plenum Press.