UNIVERSITÄT AUGSBURG
Universitätsbibliothek

OPEN ACCESS
OPUS AUGSBURG

# Applying control-value theory for examining multiple emotions in L2 classrooms: validating the Achievement Emotion Questionnaire –Second Language Learning

**Kaiqi Shao, Kristina Stockinger, Herbert W. Marsh, Reinhard Pekrun**

**Angaben zur Veröffentlichung / Publication details:**

Shao, Kaiqi, Kristina Stockinger, Herbert W. Marsh, and Reinhard Pekrun. 2026. "Applying control-value theory for examining multiple emotions in L2 classrooms: validating the Achievement Emotion Questionnaire –Second Language Learning." *Language Teaching Research* 30 (1): 288–316. https://doi.org/10.1177/13621688221144497.

**Nutzungsbedingungen / Terms of use:**

**licgercopyright**

*Dieses Dokument wird unter folgenden Bedingungen zur Verfügung gestellt: / This document is made available under these conditions:*
***Deutsches Urheberrecht***
*Weitere Informationen finden Sie unter: / For more information see:*
*https://www.uni-augsburg.de/de/organisation/bibliothek/publizieren-zitieren-archivieren/publiz/*

Check for updates

*Article*

LANGUAGE TEACHING RESEARCH

Language Teaching Research
2026, Vol. 30(1) 288–316
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/13621688221144497
journals.sagepub.com/home/ltr
SAGE

# Applying control-value theory for examining multiple emotions in L2 classrooms: Validating the Achievement Emotions Questionnaire – Second Language Learning

**Kaiqi Shao**
Hangzhou Dianzi University, China

**Kristina Stockinger**
University of Augsburg, Germany

**Herbert W. Marsh**
Australian Catholic University, Australia; University of Oxford, UK

**Reinhard Pekrun**
University of Essex, UK; University of Munich, Germany; Australian Catholic University, Australia

## Abstract

This study used Pekrun's control-value theory (CVT) as a framework to validate the Achievement Emotion Questionnaire – Second Language Learning (AEQ-L2L) for assessing eight student emotions in second language (L2) learning. We tested and validated the instrument in two samples using three waves of data, with a total number of 1,021 Chinese university students. Item and scale statistics indicate the AEQ-L2L is reliable. Moreover, single- and multi-group confirmatory factor analysis supported the hierarchical four-factor component structure of L2 emotions in and across the two student groups. Furthermore, multi-model comparison indicated that language emotions are best represented by considering both the distinctiveness of the eight examined emotions and their component structure. Finally, latent correlation analysis demonstrated the external validity

**Corresponding author:**
Kaiqi Shao, School of Foreign Studies, Hangzhou Dianzi University, Baiyang Road 1158, Hangzhou, Zhejiang 310018, China
Email: shaokaiqi@aliyun.com

of the AEQ-L2L in terms of linkages with appraisals, motivation and language achievement. Findings are discussed by highlighting the unique value of the AEQ-L2L for investigating a broad range of emotions in language education.

**Keywords**
achievement, control-value theory, emotion, foreign language, learning, measurement

## I Introduction

Second and foreign language (L2) classrooms can evoke many different positive and negative emotions, such as enjoyment of learning, hope, pride, anxiety, shame, anger, hopelessness, or boredom. These emotions can have a profound influence on students' language learning and achievement by directing cognitive and attentional processes, sustaining or undermining interest and motivation, promoting different types of learning strategies, and facilitating or impeding students' self-regulation of learning (Shao et al., 2019). Despite the pivotal role played by these diverse emotions in L2 learning (MacIntyre & Gregersen, 2012; Swain, 2013), language researchers have, to date, examined a fairly limited number of emotions, with L2 anxiety (see Zhang, 2019), and enjoyment (Dewaele et al., 2019) more recently, having received by far the most attention. One key factor contributing to this paucity of research on a broader range of emotions is the lack of theoretically sound and empirically valid instruments for measuring L2 learners' different emotional experiences (Sudina, 2021).

In educational research, the control-value theory of achievement emotions (CVT; Pekrun, 2006, 2018) has been acknowledged as a solid framework for understanding multiple emotions experienced in academic contexts. The concomitantly developed Achievement Emotions Questionnaire (AEQ) has been validated and used extensively for measuring a wide range of student emotions related to different domains of learning (math, psychology, sports, etc.; e.g. Forsblom et al., 2022; Frenzel et al., 2007; Peixoto et al., 2015; Pekrun et al., 2011). However, validation work on the AEQ in the field of L2 learning is still largely lacking, despite increasing interest and need for the use of valid instruments to examine multiple emotions in second language acquisition (SLA; Davari et al., 2022).

Consistent with the component process model of emotions prominent in contemporary emotion science (Scherer, 2009; Scherer & Moors, 2019), CVT views emotions as sets of interrelated psychological processes, whereby affective, cognitive, physiological, and motivational components are of primary importance (Pekrun et al., 2011). The construction of the AEQ was based on this multi-component conception of emotion, which is consistent with state-of-the-art approaches to measuring test anxiety (Lowe, 2018; Zeidner, 2007). Emotion researchers generally concur with the advantages of this approach to investigating antecedents and consequences of emotions and to targeting emotion interventions according to individuals' emotion profiles (Pekrun, 2018; Scherer & Moors, 2019). Nevertheless, the majority of existing L2 emotion measures adopted a unidimensional approach (e.g. Dewaele & MacIntyre, 2014; Horwitz et al., 1986; Teimouri, 2018b), with very few notable exceptions focusing exclusively on anxiety (e.g. Cheng, 2017).

The present study takes steps towards bridging this theory-method gap by using CVT as a framework and adapting the AEQ (Pekrun et al., 2005, 2011) to the domain of language learning. This makes it possible to examine the component structure of a broader range of emotions in L2 learning and the relations of these emotions with their antecedents and outcomes. The adapted instrument (AEQ-L2L) taps into eight different emotions: enjoyment, hope, pride, anger, anxiety, shame, hopelessness, and boredom occurring before, during and after situations of L2 learning. Employing a prospective design, we tested the psychometrics of the AEQ-L2L by examining the instrument's item and scale statistics, internal validity in terms of component structures and correlations between emotions, and external validity in terms of links with appraisals, motivation and L2 achievement in two cohorts of students of English as a foreign language (EFL) at the tertiary level of English education in China.

## *1 An overview of research, theory, and measurement of L2 emotions*

Research on emotions and L2 learning can be traced back to the early 1960s, developing from a neglected area to a rapidly growing field of inquiry (Zhang, 2019). Earlier studies focused almost exclusively on language anxiety and adopted a situation-specific approach to defining L2 learners' anxiety as a distinct complex of self-perceptions, beliefs, feelings, and behaviors that arises from the uniqueness of the language learning process in the classroom (Horwitz et al., 1986; MacIntyre & Gardner, 1994). Following this conceptualization, Horwitz et al. (1986) developed the Foreign Language Classroom Anxiety Scale (FLCAS), which aims to measure three aspects of L2 anxiety: communication apprehension; test anxiety; and fear of negative evaluation. The development of the FLCAS has promoted a host of studies examining antecedents and outcomes of L2 anxiety (see Horwitz, 2010; Zhang, 2019). The majority of the findings, to date, suggest that anxiety impairs L2 learners' cognitive processes (MacIntyre & Gardner, 1994), motivation (Garrett & Young, 2009), self-regulation (Bown & White, 2010), language achievement (Aida, 1994), and willingness to communicate (Khajavy et al., 2018). Research has also shown that antecedents such as demographic variables (Dewaele et al., 2008), personality (Lou et al., 2022), and task-related competence and value appraisals (Phung, 2017; Sardegna et al., 2018) are significant predictors of L2 anxiety.

While the FLCAS has laid the foundation for L2 anxiety research, the theoretical basis and validity of the scale have recently been challenged (Park, 2014; Shao et al., 2019). The development of the FLCAS was mainly grounded in the authors' teaching and clinical experience, prior literature on learning anxiety and data from student interviews, rather than in contemporary appraisal theories of emotions (Pekrun, 2006; Scherer, 2009). Moreover, in studies using exploratory factor analysis (EFA) or confirmatory factor analysis (CFA), two-, three-, or four-factor solutions emerged for the FLCAS, with various labels used to denote the factors (e.g. Aida, 1994; Cheng et al., 1999; Park, 2014). However, among the three dimensions of the FLCAS originally proposed by Horwitz et al. (1986), these studies supported only the communication anxiety dimension. Recently, a short form of the FLCAS was developed which demonstrated satisfactory psychometric properties compared with previous evidence for the full FLCAS (Botes et al., 2022). However, the short form considers L2 anxiety as a unidimensional

construct, thus not allowing to differentiate between different components of anxiety (Lowe, 2018; Scherer & Moors, 2019). Furthermore, more than half of the 33 items in the FLCAS represent constructs which are apparently distinct from anxiety, such as self-efficacy, confusion, boredom, or social comparison (see Shao et al., 2019). Each of these variables have their unique research tradition in education. The FLCAS represents a mix of constructs; it measures more than its name denotes (for a similar critique of early test anxiety scales, see Nicholls, 1976). The lack of a clear theoretical foundation, indistinct factor structure, and problematic items pose validity issues for the future use of the FLCAS, and empirical research cannot be demonstrated to be effective if the target variables are not measured in a valid manner (Flake & Fried, 2020).

Investigating the role of emotions other than anxiety, especially positive emotions such as enjoyment, in L2 learning was promoted by MacIntyre and Gregersen's (2012) introduction of positive psychology into L2 learning. Following Fredrickson's (2001) broaden-and-build theory, Dewaele and MacIntyre (2014) designed a Foreign Language Enjoyment Scale (FLES) aiming to measure students' enjoyment of L2 learning when they are creative, acquire new knowledge, complete L2 tasks, and experience a positive classroom environment. The FLES assesses both social and private aspects of L2 enjoyment. Stimulated by this instrument development, a number of studies have been conducted to examine a wide range of predictors and outcomes of L2 enjoyment as well as the dynamic interactions among these variables (Jin & Zhang, 2021; Saito et al., 2018; Teimouri, 2018b). Findings have generally reported positive associations between enjoyment and L2 learners' motivation, and achievement (Elahi Shirvan et al., 2021; Papi & Khajavy, 2021), and negative associations between L2 enjoyment and anxiety (Elahi Shirvan & Taherian, 2018; Pan & Zhang, 2021). At the class level, recent studies have demonstrated that a classroom climate characterized by peer and teacher support (Khajavy et al., 2018), positive teacher characteristics (Dewaele et al., 2019), and positive peer emotions (Shao & Parkinson, 2021) is positively related to students' L2 enjoyment, and negatively related to their L2 anxiety.

The FLES has played a pivotal role in promoting research on L2 enjoyment. However, a closer inspection reveals several critical measurement problems with this scale as well. As the underpinning theory of the FLES, the broaden-and-build theory (Fredrickson, 2001) is an emotion theory that is primarily built for examining the functions of positive emotions in relation to psychophysiological health (e.g. resilience, happiness, brain functioning and health behaviors) rather than investigating emotions aroused in the educational context (see Shao, Nicholson, et al., 2020). Moreover, studies using CFA to test the construct validity of the FLES have produced inconsistent findings regarding the number of factors underlying the L2 enjoyment scale, reporting one-, two-, or three-factor structures (see Botes et al., 2021). While a recent study employing sophisticated statistics has shown preliminary support for a three-dimensional structure of a short-form of the FLES (Botes et al., 2021), these prior validation studies generally focused on personal and social aspects of L2 learners' enjoyment. These were based on the original authors' experiential knowledge and lack a clear theoretical rationale. In this research, it also remains unclear why enjoyment should comprise three dimensions but anxiety only one (Botes et al., 2021, 2022). Given that both enjoyment and anxiety are multi-component emotions comprising affective, cognitive,

motivational, physiological, and behavioral-expressive facets, it seems more reasonable to assume that they should show similar dimensional structures. Furthermore, the 21 items of the instrument assess learners' multiple emotions (enjoyment, pride, and boredom), attitude toward L2 learning mistakes, development of L2 identity and skills, perceptions of the classroom environment (teachers, peers, group cohesion), etc. Only two items directly measure enjoyment (items 4 and 12). As such, similar to the FLCAS, most items refer to other constructs rather than the target emotion of enjoyment that the scale purports to assess (see also Arar & Tannenbaum, 2021; Pawlak et al., 2020). Further, the item stems of the FLES items refer to individual student-level variables ('I') in some of the items, but to classroom-level variables ('teacher/peers/we') in others. As such, the scale confounds enjoyment at the student and classroom levels, leading to measurement errors (Shao & Parkinson, 2021; for the importance of distinguishing levels of effects in L2 research, see Khajavy et al., 2018). These measurement problems have led other L2 researchers to develop their own measures of L2 enjoyment (e.g. Khajavy et al., 2018; Teimouri, 2018a).

Apart from the widely used FLCAS and FLES, a few instruments have recently been developed to assess other discrete emotions in L2 learning. For instance, Pawlak et al. (2020) developed a Foreign Language Boredom Scale aiming to probe L2 learners' experience of boredom in the classroom. Preliminary research has demonstrated that boredom, as one of the most frequently experienced emotions in L2 classrooms, may have no less detrimental effects than anxiety on L2 learning. Teimouri (2018b) designed two scenario-based scales for examining links of guilt and shame with L2 learners' motivation and proficiency. Results showed that shame negatively correlated with motivation and language achievement, whereas guilt was positively related to motivation and achievement. Additionally, Ross (2015) developed a questionnaire which was adapted from the Achievement Emotions Questionnaire (AEQ; Pekrun et al., 2005, 2011) to explore a wider range of other discrete L2 emotions, including enjoyment, hope, pride, happiness, excitement, frustration, boredom, embarrassment, and fear. To our knowledge, previous studies that examined multiple emotions in the L2 context all used the AEQ (Alamer & Lee, 2019; Davari et al., 2022; Shao, Pekrun, et al., 2020). Findings of this research have demonstrated the unique function and value of each emotion in L2 learners' language acquisition process. This highlights the need for instruments that are reliable, valid, and measure multiple discrete emotions to satisfy the growing interest in examining the diverse emotions that play a crucial role in SLA.

In developing such instruments, two fundamental issues pertaining to the appraisal antecedents and the multi-componential nature of emotions require consideration. First, contemporary emotion scientists generally agree that cognitive appraisals are the primary causes for emotions to be instigated (for a review, see Scherer & Moors, 2019). Appraisals reflect individuals' evaluation of their environment, and changes in appraisals lead to changes in emotions (Scherer, 2009). Second, emotion is not considered a static, unitary construct but a dynamic process comprising different component processes. As noted earlier, affective, cognitive, physiological, motivational, and expressive components are thought to play a central role in emotions (Scherer, 2009). This view aligns with an appraisal-driven componential approach to investigating and measuring emotions (Scherer & Moors, 2019). These conceptual and methodological advancements in the

appraisal theory of emotions have been adapted in educational research (Pekrun et al., 2005, 2011), particularly in test anxiety studies (Lowe, 2018; Zeidner, 2007), to develop multidimensional measures for investigating students' emotions.

However, rather than following the appraisal theory of emotion, most existing L2 emotion instruments were constructed on the basis of a uni- rather than multi-component approach to measuring emotions. As a notable exception, Cheng (2004, 2017) developed L2 anxiety scales based on the multi-componential theory of anxiety for measuring learners' physiological, cognitive, and behavioral facets of anxiety in L2 writing (Cheng, 2004) and later in all four L2 skills (Cheng, 2017). A few recent studies have confirmed the usefulness of multi-componential measures of learners' L2 anxiety (for a review, see Cheng, 2017; Kutuk et al., 2020). Nevertheless, no studies to date have explored the multi-component structure of emotion for a broader range of L2 emotions. It is foreseeable that the field of SLA research will need theoretical approaches and measurement instruments that have the capacity to account for the multifaceted nature of emotional experiences, connect emotions to L2 learners' cognitive appraisals (Davari et al., 2022), and align L2 emotion research with the larger research field of emotions in education (Shao, Pekrun, et al., 2020).

## 2 Control-value theory and the achievement emotions questionnaire

*a Control-value theory of achievement emotions (CVT).* Research on achievement emotions has begun to flourish within the field of educational psychology over the past twenty years (Loderer et al., 2020). Achievement emotions refer to affective arousal that is tied directly to individuals' perceptions of control over, and personal value of, achievement activities or achievement outcomes (Pekrun, 2006). In a series of studies, Pekrun et al. (2002) identified a number of emotions that were frequently reported by students in academic settings, including enjoyment, hope, pride, relief, anger, anxiety, hopelessness, shame, and boredom. These emotions can be classified using a three-dimensional taxonomy that combines the dimensions valence (positive vs. negative), activation (activating vs. deactivating), and object focus (activity vs. outcome). Outcome emotions are further distinguished into prospective and retrospective emotions.

The eight achievement emotions investigated in the present study can be defined with reference to this taxonomy. (1) Enjoyment is a positive activating emotion pertaining to current achievement-related activities (e.g. a fun classroom activity), while (2) boredom is a negative deactivating emotion experienced, for instance, when a classroom activity is perceived as dull. (3) Hope and (4) anxiety are prospective positive and negative activating emotions tied to possible (i.e. uncertain) success and failure outcomes, respectively. (5) Hopelessness is a prospective deactivating outcome emotion targeting expected (i.e. certain) failure or unattainable success on a future achievement outcome. (6) Pride is a retrospective activating emotion related to personal success on a past outcome, while its negative counterpart (7) shame concerns past failure. Finally, (8) anger is a negative activating emotion that can target either achievement-related activities that are perceived as aversive (e.g. a difficult or tedious classroom activity), or a retrospective outcome (e.g. a test result perceived as unfair).

The present conceptualization of achievement emotions is grounded in CVT which builds on the assumption that control and value appraisals are proximal antecedents of emotions experienced by students in achievement settings (Pekrun, 2006). Perceived control refers to the extent to which students believe that they can exert causal influence over the learning process and its outcomes. Perceived value refers to the importance students attribute to learning activities and outcomes (Pekrun & Perry, 2014). Value can be positive or negative, and can target different facets; e.g. intrinsic value: valuing English because it is interesting; achievement value: valuing attaining success and avoiding failure in an English exam; utility value: valuing the task because it is useful for one's future career. Higher levels of perceived control and positive value are posited to elicit positive emotions such as enjoyment, hope, or pride, whereas lower levels of control and higher levels of negative value are expected to elicit negative emotions such as anger, anxiety, shame, or hopelessness. Boredom can be triggered by either low control (over-challenge) or high control (under-challenge) and lack of value. CVT is not only congruent with the mainstream cognitive appraisal theory of emotions (Scherer & Moors, 2019), but also integrates propositions from multiple emotion theories in achievement contexts (see Pekrun et al., 2011).

Adherence to the component process models of emotions (Scherer, 2009), CVT defines achievement emotions as multifaceted phenomena that comprise various component processes as explained earlier, including subjective feelings, cognitions, motivational tendencies, and physiological and expressive processes (Pekrun, 2006). For example, a student experiencing pre-exam anxiety may feel nervous (affective component), worry about failure (cognitive component), want to flee the exam (motivational component), have sweaty palms (physiological component), and display an anxious facial expression (expressive component). From a measurement perspective, this multicomponent conception of emotions suggests that achievement emotions are best modeled as hierarchically organized structures, with the different emotion components represented as first-order factors, and the emotion itself as a second-order factor. However, because it is difficult to capture expression in self-report instruments, the expressive component is usually underrepresented (see Figure 1, adapted from Pekrun et al., 2011).

Furthermore, CVT suggests that the effects of achievement emotions on learning and performance depend on the interplay of several mediating mechanisms, such as students' motivation, strategy use, and regulation of learning (Pekrun et al., 2002). Emotions can influence students' interest in learning material and intrinsic motivation to learn, as well as their extrinsic motivation related to the attainment of success or the avoidance of failure (Pekrun, 2006). Specifically, positive activating emotions such as enjoyment, hope, and pride are thought to promote both intrinsic and extrinsic motivation and support self-regulation, thus positively affecting academic performance (Peixoto et al., 2015). Conversely, negative deactivating emotions, such as hopelessness and boredom, are posited to reduce motivation and self-regulation of learning, implying negative effects on performance. Negative activating emotions, such as anger, anxiety, and shame, are presumed to undermine intrinsic motivation and self-regulation, but may trigger extrinsic motivation to invest effort to avoid failure (Pekrun et al., 2011).

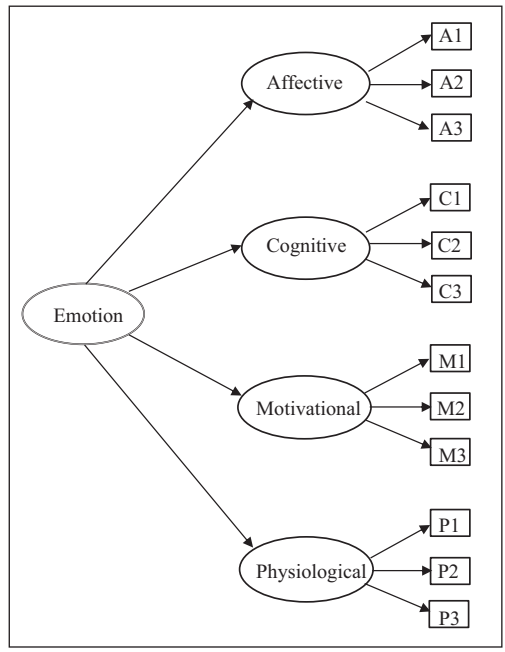


**Figure 1.** Hierarchical model of component structures of achievement emotions.

As a result, these emotions can have variable effects on students' learning, although negative effects on overall performance are expected to outweigh any benefits for most students (Pekrun, 2018).

*b Construction and validation of the AEQ.* Based on CVT, measures of test anxiety, and several exploratory studies, Pekrun et al. (2002, 2005, 2011) developed the Achievement Emotions Questionnaire (AEQ). The instrument assesses nine emotions that are most commonly reported by students in three common academic settings: attending class, doing homework, and taking tests. Emotions experienced in these situations can differ due to different social structures and functions of these settings (e.g. enjoyment of classroom instruction vs. enjoying the challenge of taking an exam). Within each situation, the items are ordered in three blocks assessing emotional experiences before, during, and after the respective situation. Following the multi-component conceptualization of emotions, the items of the AEQ assess affective, cognitive, motivational, and physiological facets of each of the emotions assessed.

The initial item pool yielded more than 1,500 items. The selection of items for the final scales was based on item and scale statistics, criteria of semantic redundancy, and expert judgment. Convergent (i.e. high factor loadings on the relevant emotion scale) and divergent (i.e. low factor loadings on other emotion scales) validity of the items was established (see Pekrun et al., 2011). The final instrument consists of 232 items and 24 scales which can be used to assess trait, course-specific, and state emotions in achievement settings by modifying the instructions accordingly. Moreover, the AEQ can be used

to measure achievement emotions experienced in different domains such as math, sport, or language learning (Pekrun et al., 2005) through adapting the instructions to different subjects. This is in line with CVT's proposition that achievement emotions are organized in domain-specific ways, as evidenced by research documenting that levels of emotions differ across domains (Goetz et al., 2007).

The psychometric properties of AEQ have been validated across different cultures, age groups, genders, and learning domains and situations (e.g. Frenzel et al., 2007; Goetz et al., 2007; Lichtenfeld et al., 2012; Peixoto et al., 2015). Although these validation studies differ in methodologies, student populations, and numbers of emotions and items extracted from the AEQ, the studies have yielded generally consistent results demonstrating that the psychometric properties of the AEQ can be empirically substantiated, and that both the entire AEQ and its subscales are valid and reliable measures of achievement emotions. Furthermore, findings have also shown that there are positive correlations between emotions of the same valence and negative correlations between emotions of opposite valence. In studies examining the external validity of the AEQ, it has been found that there are theoretically meaningful relations between students' achievement emotions and their control and value appraisals, motivation, self-regulation, learning strategies and performance in accordance with the propositions of CVT (e.g. Forsblom et al., 2021; Lichtenfeld et al., 2012; Peixoto et al., 2015; Pekrun et al., 2011).

However, these studies have predominantly focused on students' math-related or domain-general emotions. The AEQ has only recently been employed to measure achievement emotions in the field of L2 learning to cater to the need for reliable and valid instruments measuring multiple emotions in this domain (Davari et al., 2022; Shao, Pekrun, et al., 2020; Starkey-Perret et al., 2018; Wang et al., 2021). Davari et al. (2022) examined the validity of a short version of the class-related AEQ for measuring eight emotions in an Iranian EFL context. Results substantiate the psychometric properties of the adapted AEQ and its structural invariance across genders and learning contexts. Moreover, Starkey-Perret et al. (2018) investigated the impact of different teaching approaches on seven L2 emotions in a French middle school context. Results showed that the adapted short form of AEQ is a reliable and valid instrument for measuring different emotions in the L2 context; a two-facets structure with emotions and learning situations modeled as latent variables had the best fit to the data. Although these findings provide promising support for the utility of AEQ for investigating multiple emotions in L2 learning, several open issues remain unresolved. First, these two studies did not examine the validity of the component structure of achievement emotions assumed by the AEQ. Second, they did not probe the relations between appraisal antecedents, L2 emotions, and motivation and achievement outcomes as posited by CVT. Third, these two studies used short forms of the AEQ comprising reduced numbers of items, which can compromise reliability and validity.

## III The present study

Drawing on CVT (Pekrun, 2006, 2018), the primary goal of this study was to test the psychometric properties of the learning-related scales of the AEQ (AEQ-L2L) for measuring eight distinctive emotions in L2 learning: enjoyment, hope, pride, anxiety, anger,

shame, boredom, and hopelessness. To this end, item and scale statistics, component structures (i.e. affective, cognitive, motivational, physiological emotion components), and latent correlations with appraisal antecedents, motivation, and L2 achievement were examined among two groups (teacher track vs. non-teacher track) of EFL students preparing for an L2 exam. The analyses were guided by the following research questions:

- Research question 1: What are the item and scale statistics of the AEQ-L2L for the two cohorts of EFL students?
- Research question 2: Is the AEQ-L2L internally valid in terms of representing the four-factor component structure for all eight emotion scales in and across the two student groups?
- Research question 3: Is the AEQ-L2L internally valid in terms of relations between the eight L2 emotions taken together with their four components for the two cohorts of EFL students?
- Research question 4: Is the AEQ-L2L externally valid in terms of relations of the emotions with appraisals, motivation, and achievement in the two student groups?

## IV Method

### *1 Participants and procedure*

The sample included two cohorts of (total $N = 1{,}021$) college students[1] studying English as their major at a foreign language studies university in Southeastern China. The two groups of students differed in terms of their future career choices. The teacher track cohort consisted of 471 students (gender: 76 males, 393 females, 2 unspecified; age: $M = 18.72$, $SD = .70$) who intended to become middle or primary school English teachers and took specialized L2 pedagogical courses. The non-teacher track cohort consisted of 550 students (gender: 50 males, 498 females, 2 unspecified; age: $M = 19.66$, $SD = .76$) who planned to take on various English-related occupations other than teaching (e.g. translation, international business, tourism) and attended major-related courses as well. Such a distinction is typical for the majority of English-major students in China. We examined measurement invariance of the AEQ-L2L across these two meaningful groups. The gender distribution in the sample is typical for English major students in China. All participants were enrolled in a comprehensive English course that aims to enhance students' reading, speaking, writing, and listening skills in English to fulfill their degree requirements.

Students were informed about the general purpose and the voluntary nature of participating in this research by their teachers at the beginning of the autumn semester. Participants were asked to complete the measures in three different assessments. Demographics, appraisals, and motivational variables were assessed in the sixth week of the semester. Achievement emotions were measured in the 17th week of the semester, when students were preparing for their exam. In the 18th week, participants completed the final course exam. This prospective design provided a clear temporal ordering of all measures while controlling for prior achievement. Institutional Review Board approval was obtained, and the research was conducted in accordance with the American Educational Research Association ethical principles regarding research with human participants.

## 2 Instrumentation

In this study, students' L2 emotions, appraisals, and motivation were measured using self-report scales. The scales comprised a total of 102 items which were administered in two separate sections to avoid survey fatigue. They were translated into Chinese and back-translated by a team of six bilingual researchers. All items were assessed on a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The items are presented in Appendix S1 in the supplemental materials. Students took approximately 10 minutes and 15 minutes to answer the first set (Time 1, appraisals and motivation) and second set (Time 2, emotions) of scales, respectively. Students' L2 achievement was assessed using their final exam grades. Achievement scores and prior English achievement were obtained from the head teacher of the course at the end of the semester (instruments are available on IRIS).

*a Achievement emotions.* The AEQ-L2L tested in the present study was derived from the learning-related emotion scales of the AEQ (Pekrun et al., 2011). It consists of 74 items pertaining to four components (affective, cognitive, motivational, physiological) of eight emotions (enjoyment, hope, pride, anger, anxiety, shame, hopelessness, and boredom) in language learning. As items in the AEQ do not contain any references to specific domains, they can be used to measure students' emotions in different subject domains. For the purpose of our investigation, we adapted the AEQ instructions by asking students to report on their emotional experiences prior to, during, and after studying for the end-of-course English language exam. We used the same stems: 'Before/while/after studying for this English exam . . .' to contextualize the items. The questionnaire was piloted with 118 students to check the wording and internal consistency of the adapted instrument. One or two items for each emotion scale were dropped due to low response rates and relatively low item-total correlations ($r_{it} < .50$) and factor loadings ($\lambda < .40$; Byrne, 2011) in the pilot sample. Sixty-three items were retained for the final investigation.

*b Control and value appraisals.* Students' perceived control in L2 learning was assessed using the eight-item version of Perry et al.'s (2001) Perceived Academic Control Scale. The scale assesses students' self-perceived ability to influence their academic performance ($M = 3.01/3.33$, $SD = .63/.62$, 95% CI [2.96, 3.07]/[3.27, 3.38], $\alpha = .86/.82$ for teacher track/non-teacher track). Items were adapted to reflect perceived control in the current English course. Students' perceived English course value was assessed via eight items adapted from the Task Value Questionnaire (Pekrun & Meier, 2011). The scale measures students' intrinsic, attainment, and utility value ($M = 3.67/3.36$, $SD = .53/.62$, 95% CI [3.63, 3.72]/[3.30, 3.40], $\alpha = .76/.80$). Items were adapted to assess the perceived value of preparing for the English exam.

*c Motivation variables.* The Intrinsic Goal Orientation, Extrinsic Goal Orientation, and Self-regulated Learning scales from the Motivated Strategies of Learning Questionnaire (MSLQ; Pintrich et al., 1991) were chosen as indicators of L2 motivation. We adapted items in these scales to the L2 context. The Intrinsic Goal Orientation scale measures

students' intrinsic motivation toward L2 learning based on interest and curiosity ($M$=3.37/3.36, $SD$=.76/.68, 95% CI [3.29, 3.43]/[3.30, 3.42, $\alpha$=.74/.71] for teacher track/non-teacher track). The Extrinsic Goal Orientation scale measures students' extrinsic motivation for L2 learning related to getting good grades ($M$=3.41/3.58, $SD$=.77/.74, 95% CI [3.34, 3.49]/[3.51, 3.64], $\alpha$=.73/.73). The Self-regulated Learning scale is a measure of students' overall regulation of effort to learn the target language ($M$=3.03/3.01, $SD$=.64/.62, 95% CI [2.97, 3.10]/[2.95, 3.06], $\alpha$=.76/.72). Each scale consists of four items and the scores were summed to form the intrinsic motivation, extrinsic motivation, and self-regulated learning indexes.

*d Foreign language achievement.* Students' L2 achievement was measured by their scores on the final English course exam. The test used in the exam was developed and marked by the course teachers based on the course textbook. The text measures students' reading comprehension, vocabulary and grammar knowledge, translation skills, and writing ability. In the Chinese education system, exam scores range from 0 to 100. The actual range of scores in the present study was 42 to 94. In this test, the reading section contained two cloze-tests (30 items, 20% of the overall score) and two passage reading tasks (10 items, 20%); the vocabulary and grammar section included 20 sentence completion items (20%); the translation section comprised 10 sentences to be translated from Chinese into English (20%); and the writing section contained a short essay of 120 to 180 words (20%). Experts in the school examination panel evaluated the content validity and the structure of the achievement test. The reliability of the test was satisfactory ($\alpha$=.83/.82 for the teacher track/non-teacher track samples). The translation and writing sections of the test were evaluated based on the scoring criteria of the national college English test (CET) of China. The achievement scores on the subtests were combined by the respective course instructors and expressed in a summative score ($M$=63.23/66.42, $SD$=8.43/8.79, 95% CI [62.38, 64.01]/[65.63, 67.17]). Students who passed the exam (i.e. scores $\geqslant$ 60) received 3 course credits. The test took 150 minutes and was administered by the course instructors. It is not possible to make the complete test available because it is proprietary to the university. However, we have included a simulated test extracted from the exam paper item pool including sample items for each section in Appendix S2 (in the supplemental materials).

## 3 Data analysis

Structural equation modeling (SEM) was conducted using M*plus* 8.0 (Muthén & Muthén, 1998–2017) to test the hypothesized structure of the AEQ-L2L. In SEM, study variables can be modeled as latent constructs that are corrected for measurement error (Byrne, 2011). As such, SEM allows for more accurate estimation of students' L2 emotions, their interrelations, and their relations with appraisals, motivation, and achievement, as compared with research using manifest variables (e.g. Saito et al., 2018). Moreover, multigroup SEM allows for simultaneous model fitting for two or more samples, which is an ideal technique for examining model invariance across the two cohorts of EFL students in this study. As participants in the two groups came from multiple classrooms (16 classes for teacher track; 18 classes for non-teacher track), the data had a nested structure (with

**Table 1.** Items and scale statistics of the Achievement Emotion Questionnaire – Second Language Learning (AEQ-L2L) scales for both student samples.

| Emotion | Items | Range | *M* | 95% CI | SD | Skewness | *M* $r_{i(t-i)}$* | α |
|---|---|---|---|---|---|---|---|---|
| Enjoyment | 10 | 19–50/19–50 | 3.29/3.11 | [3.23, 3.35] | .59/.62 | .49/.23 | .57/.61 | .86/.88 |
| Hope | 6 | 7–30/9–30 | 3.15/3.07 | [3.09, 3.20] | .66/.67 | .47/.31 | .63/.66 | .85/.86 |
| Pride | 7 | 10–35/12–35 | 3.32/3.26 | [3.26, 3.39] | .69/.66 | .54/.12 | .62/.63 | .86/.83 |
| Anger | 8 | 8–36/8–39 | 2.24/2.36 | [2.18, 2.30] | .63/.66 | .26/.66 | .61/.59 | .86/.85 |
| Anxiety | 8 | 8–35/8–39 | 2.63/2.66 | [2.58, 2.69] | .62/.58 | -.03/.30 | .58/.58 | .87/.85 |
| Shame | 8 | 8–36/8–40 | 2.58/2.47 | [2.53, 2.70] | .65/.61 | .17/.40 | .53/.52 | .81/.81 |
| Hopelessness | 8 | 8–34/8–35 | 2.14/2.04 | [2.09, 2.21] | .64/.59 | .28/.59 | .65/.63 | .88/.87 |
| Boredom | 8 | 8–38/8–37 | 2.21/2.48 | [2.15, 2.28] | .64/.65 | .25/.35 | .53/.56 | .81/.84 |

*Notes*. * Mean of part–whole corrected item-total correlations. The left entry is the parameter from the teacher track ($N = 471$) and the right entry is the parameter from the non-teacher track ($N = 550$).

students at Level 1 nested in classes at Level 2). This nesting was taken into account by using the 'type = complex' command in M*plus* to correct parameter estimates. We used the full information maximum likelihood (FIML) approach to handle missing data, and the robust maximum likelihood estimator (MLR) to deal with possible non-normally distributed data.

Following Hu and Bentler's (1999) recommendations, we used both absolute and incremental fit indexes to evaluate the model fit, including the comparative fit index (CFI), the Tucker–Lewis index (TLI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). CFIs $\geqslant .95$, TLIs $\geqslant .95$, RMSEAs $\leqslant .06$, and SRMRs $\leqslant .08$ are thought to indicate good fit, $.95 \geqslant$ CFIs $\geqslant .90$, $.95 \geqslant$ TLIs $\geqslant .90$, RMSEAs between .06 and .08 reasonable fit, and RMSEAs between .08 and .10, SRMRs between .08 and .10 mediocre fit. As the chi-square value $\chi^2$ is sensitive to sample size, leading to biased rejection of models, it was not used as an indicator in the analyses considering the large sample in the present investigation (Byrne, 2011).

## V Results

### *1 Item and scale statistics*

Table 1 presents descriptive statistics of the AEQ-L2L scales for each of the two samples. The ranges, means, and standard deviations indicate sufficient variations of scores for all emotions. Mean item scores of each emotion scale did not substantially deviate from a normal distribution. Notably, students generally experienced stronger positive emotions than negative emotions when learning for the English exam. Multiple group-mean comparison using one-way ANOVAs showed that the differences between the average means of the three positive emotions and the five negative emotions were significant for both teacher track ($F(1, 6) = 40.25$, $p = .001$) and non-teacher track ($F(1, 6) = 27.11$, $p = .002$) EFL students. Moreover, the items had excellent part–whole corrected item-total correlations with $r_{it} > .50$ for all items and scales. Furthermore, the reliability coefficients (α)

were above .80 for all scales and above .85 for 5 of the 8 scales, indicating good to excellent reliabilities for the AEQ-L2L. Finally, the majority of the parameters were similar for each emotion between the two samples, which further supports the generalizability of the scale characteristics of the AEQ-L2L.

## 2 Internal validity: Component structures of the AEQ-L2L

To examine the presumed four-component structure of each emotion scale in the AEQ-L2L, we used confirmatory factor analysis (CFA; Byrne, 2011) to competitively test three structural models (for a similar procedure, see Pekrun et al., 2011). Model 1 was a one-factor model assuming one latent emotion factor, with all items loading on this factor. Model 2 was a four-factor component model consisting of four separate latent factors representing the four emotion components, with the items representing the components used as indicators for these factors. Model 3 was a hierarchical model with affective, cognitive, motivational, and physiological items used as indicators for four separate first-order latent component factors, and the four first-order factors loading on a second-order factor representing the emotion (Figure 1). We tested the fit of the three models for each of the eight L2 emotion scales separately for the two samples (Mplus 8.0; Muthén & Muthén, 1998–2017). Following Alamer and Marsh's (2022; Alamer, 2022) recommendation, we also employed the exploratory structural equation model (ESEM) to examine the model fit for each of the eight L2 emotion scales. However, due to scale properties of the AEQ-L2L and technical limitation of the M*plus* package with ESEM models,[2] these models did not converge. Therefore, we retained the three CFA models for comparison.

As can be seen from Table 2, model fit of the four-factor and hierarchical component factor models was good for all eight emotion scales in both groups (except for RMSEA which was reasonable for the hope and pride scales in the non-teacher track sample). In contrast, the one-factor models demonstrated a less than optimal fit for 9 of the scales in terms of TLI being below .95 or RMSEA above .06. This indicates that the proposed component structure of the AEQ-L2L provides a better representation of L2 emotions than a one-factor model. Furthermore, the fit of the hierarchical models is similar to the fit of the four-factor component models. This finding demonstrates that nesting the four-component factors under the overarching second-order emotion factor is empirically supported, especially considering that the hierarchical models involved estimation of one additional latent factor. Given our strong theoretical justification which proposes that the four first-order latent factors represent four components rather than merely subscales of L2 emotions (Lowe, 2018; Pekrun, 2006; Scherer, 2009), we adopted the hierarchical models as a basis for further analyses.

Based on these group-specific baseline models, a series of multi-group confirmatory factor analyses (CFA) were conducted to test the invariance of the hierarchical four-component factor models across both cohorts. We fit increasingly restrictive CFA models to the data by setting factor loadings, item intercepts, factor variances/covariances, and latent means to be equal across both cohorts (Byrne, 2011), and examined model fit under each condition. Given that the chi-square difference ($\Delta\chi^2$) is sensitive to sample size, we instead report ΔCFI along with ΔRMSEA and ΔSRMR for evaluating differences in fit between the models. We adopted Cheng's (2007) cut-off criteria, with a loss of fit of $\Delta CFI \geqslant .01$, $\Delta RMSEA \geqslant .015$, and $\Delta SRMR \geqslant .03$ (for loading

**Table 2.** Emotion component structures of the Achievement Emotion Questionnaire – Second Language Learning (AEQ-L2L) scales: fit statistics of confirmatory factor analysis (CFA) for each student group.

| Emotion | Model | Teacher track ($N = 471$) | | | | | Non-teacher track ($N = 550$) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | *df* | CFI | TLI | RMSEA | SRMR | *df* | CFI | TLI | RMSEA | SRMR |
| Enjoyment | 1 | 34 | .95 | .93 | .066 | .044 | 34 | .96 | .95 | .062 | .036 |
| | 2 | 29 | .97 | .96 | .051 | .038 | 29 | .98 | .98 | .035 | .032 |
| | 3 | 32 | .97 | .96 | .054 | .034 | 32 | .98 | .98 | .035 | .036 |
| Hope | 1 | 9 | .97 | .96 | .069 | .037 | 9 | .97 | .95 | .066 | .034 |
| | 2 | 6 | .99 | .97 | .056 | .018 | 6 | .98 | .95 | .068 | .031 |
| | 3 | 6 | .99 | .97 | .056 | .018 | 6 | .98 | .95 | .062 | .034 |
| Pride | 1 | 14 | .95 | .92 | .090 | .040 | 14 | .97 | .93 | .076 | .036 |
| | 2 | 11 | .99 | .97 | .054 | .028 | 11 | .97 | .94 | .066 | .035 |
| | 3 | 11 | .99 | .98 | .050 | .028 | 11 | .97 | .95 | .063 | .035 |
| Anger | 1 | 20 | .95 | .93 | .080 | .038 | 20 | .97 | .96 | .056 | .030 |
| | 2 | 14 | .99 | .98 | .041 | .020 | 14 | .98 | .97 | .054 | .023 |
| | 3 | 16 | .99 | .99 | .029 | .021 | 16 | .98 | .97 | .050 | .024 |
| Anxiety | 1 | 20 | .98 | .97 | .039 | .037 | 20 | .97 | .96 | .052 | .030 |
| | 2 | 14 | .99 | .99 | .021 | .022 | 14 | .98 | .97 | .041 | .025 |
| | 3 | 18 | .99 | .98 | .031 | .035 | 18 | .98 | .96 | .044 | .026 |
| Shame | 1 | 20 | .94 | .92 | .064 | .041 | 20 | .97 | .96 | .050 | .032 |
| | 2 | 17 | .97 | .95 | .051 | .032 | 17 | .97 | .96 | .052 | .031 |
| | 3 | 18 | .97 | .95 | .049 | .031 | 18 | .97 | .96 | .046 | .031 |
| Hopelessness | 1 | 20 | .98 | .97 | .045 | .036 | 20 | .96 | .94 | .071 | .037 |
| | 2 | 14 | .99 | .98 | .041 | .021 | 14 | .98 | .96 | .059 | .028 |
| | 3 | 16 | .99 | .98 | .042 | .024 | 16 | .98 | .96 | .058 | .032 |
| Boredom | 1 | 20 | .98 | .98 | .035 | .034 | 18 | .98 | .97 | .046 | .032 |
| | 2 | 14 | .99 | .99 | .029 | .021 | 14 | .98 | .97 | .042 | .028 |
| | 3 | 16 | .99 | .99 | .024 | .023 | 16 | .98 | .97 | .044 | .023 |

*Notes*. Model 1 = general one-factor model. Model 2 = four-factor component model. Model 3 = hierarchical component model. CFI = comparative fit index. TLI = Tucker–Lewis index. RMSEA = root mean square error of approximation. SRMR = standardized root mean square residual.

invariance) or ⩾.01 (for intercept and residual invariance) being regarded as indicating non-invariance.

As shown in Table 3, the configural model (Model 0), which only fixes the pattern of parameters to be invariant across groups, had good fit for all eight emotions, indicating configural invariance of the baseline model. Next, we constrained the item loadings to be equal across groups (Model 1; metric invariance). Comparison of Model 1 with Model 0 resulted in no significant decrease of fit for any scales. In contrast, additionally constraining item intercepts to be invariant (scalar invariance) led to a loss of fit for all eight emotions (Model 2 vs. Model 1). To locate the source of non-invariance for each

**Table 3.** Test of model invariance of the Achievement Emotion Questionnaire – Second Language Learning (AEQ-L2L) across both student samples.

| Emotion | Model | Model 0 | Model 1 | Model 2 | Model2P | Model 3 | Model 4 |
|---|---|---|---|---|---|---|---|
| | Index | | | | | | |
| Enjoyment | CFI/RMSEA | .98/.039 | .98/.042 | .93/.067 | .97/.043 | .97/.043 | .97/.042 |
| | SRMR | .035 | .047 | .078 | .055 | .058 | .052 |
| Hope | CFI/RMSEA | .98/.066 | .98/.063 | .96/.070 | .97/.065 | .97/.064 | .97/.061 |
| | SRMR | .029 | .042 | .053 | .048 | .050 | .045 |
| Pride | CFI/RMSEA | .98/.059 | .98/.057 | .96/.059 | .98/.055 | .97/.054 | .97/.056 |
| | SRMR | .033 | .042 | .048 | .045 | .049 | .048 |
| Anger | CFI/RMSEA | .99/.020 | .99/.020 | .95/.057 | .98/.033 | .98/.033 | .99/.024 |
| | SRMR | .043 | .040 | .056 | .042 | .044 | .036 |
| Anxiety | CFI/RMSEA | .98/.039 | .98/.034 | .94/.059 | .97/.040 | .97/.040 | .98/.038 |
| | SRMR | .028 | .033 | .050 | .039 | .046 | .043 |
| Shame | CFI/RMSEA | .97/.051 | .97/.049 | .95/.058 | .96/.051 | .96/.050 | .96/.052 |
| | SRMR | .032 | .042 | .055 | .046 | .046 | .046 |
| Hopelessness | CFI/RMSEA | .98/.051 | .98/.049 | .96/.064 | .97/.050 | .97/.053 | .97/.053 |
| | SRMR | .028 | .041 | .058 | .044 | .053 | .052 |
| Boredom | CFI/RMSEA | .98/.037 | .98/.037 | .95/.060 | .97/.044 | .97/.043 | .98/.037 |
| | SRMR | .028 | .039 | .083 | .047 | .047 | .041 |

*Notes.* Model 0 = configural model (no invariance imposed). Model 1 = invariant factor loadings. Model 2 = invariant factor loadings and invariant intercepts. Model 2P = invariant factor loadings and partially invariant intercepts. Model 3 = invariant factor loadings, partially invariant intercepts, and invariant factor variances and covariances. Model 4 = invariant factor loadings, partially invariant intercepts, invariant factor variances and covariances, and latent means. CFI = comparative fit index. RMSEA = root mean square error of approximation. SRMR= standardized root mean square residual.

emotion, we examined modification indices (MI) with a critical value of ⩾10. Following the recommendation to release one parameter at a time (Byrne, 2011), Model 2 was rejected in favor of Model 2P, a partial-invariance model in which the intercepts for particular items were freed to vary across groups. There was no substantial decrease of fit comparing Model 2P with Model 1. Based on these partial invariance models, we continued to test invariance by imposing equality constraints on factor variances and covariances (Model 3). The difference of fit between Model 2P and Model 3 was non-substantial for all eight emotions. Finally, we examined invariance at the level of latent means for each emotion (Model 4). This was done by fixing the latent factor means for one group to zero; this group then operated as a referent group against which latent means for the other group were compared (Byrne, 2011). As shown in Table 2, Model 4 exhibited a slightly better fit, albeit non-substantially, than Model 2P for all eight emotions. This indicates that the latent means representing levels of emotional experience were similar for both cohorts. In sum, these results provide reasonable support for the measurement invariance and structural invariance of the AEQ-L2L across two different student groups.

**Table 4.** Latent intercorrelations among second language (L2) emotions of the Achievement Emotion Questionnaire – Second Language Learning (AEQ-L2L) scales.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1 Enjoyment | – | .78 | .81 | −.49 | −.48 | −.41 | −.44 | −.50 |
| 2 Hope | .79 | – | .76 | −.42 | −.46 | −.46 | −.54 | −.52 |
| 3 Pride | .82 | .77 | – | −.42 | −.40 | −.34 | −.37 | −.48 |
| 4 Anger | −.48 | −.53 | −.45 | – | .68 | .60 | .67 | .71 |
| 5 Anxiety | −.28 | −.48 | −.47 | .60 | – | .73 | .69 | .71 |
| 6 Shame | −.31 | −.46 | −.36 | .69 | .72 | – | .68 | .62 |
| 7 Hopelessness | −.47 | −.52 | −.35 | .59 | .59 | .68 | – | .78 |
| 8 Boredom | −.46 | −.52 | −.44 | .81 | .62 | .63 | .74 | – |

*Notes*. Coefficients below/above the diagonal are for teacher track ($N = 471$) and non-teacher track ($N = 550$), respectively. $p < .001$ for all coefficients.

## *3 Internal validity: Relations between L2 emotions and their four components*

To test the theoretical conception of emotions underlying the AEQ-L2L, it is also useful to distinguish between the eight discrete L2 emotions. We used the factor scores from the above four-component models of the emotions to estimate the correlations between all eight L2 emotions. As shown in Table 4, positive L2 emotions – enjoyment, hope, and pride – were positively correlated in each of the two samples. Similarly, negative L2 emotions – anger, anxiety, shame, hopelessness, and boredom – were positively related to one another. In contrast, the correlations between the positive L2 emotions, on the one hand, and the negative L2 emotions, on the other, were moderately negative. Overall, these findings indicate that the eight L2 emotions measured by the AEQ-L2L are clearly separable. This is particularly supported by the correlations for emotions which might be assumed to constitute opposite ends of a bipolar continuum, such as enjoyment and boredom, or hope and hopelessness, which demonstrated no more than moderately negative correlations.

To more fully probe the distinctiveness of each emotion construct as well as the component structure of the AEQ-L2L at scale-level across all eight L2 emotions, we constructed four hypothesized models involving different arrangements of relations between the different emotions and their components, and tested them competitively for each cohort of students (see Figure 2). Scores of each component scale assessing the eight L2 emotions in the AEQ-L2L served as manifest indicators in each model.[3] Model A was a two-factor model assuming that the relations between the eight L2 emotions can be reduced to two general latent variables, positive and negative emotion. Model B included eight latent factors made up of the eight discrete L2 emotions assessed by the AEQ-L2L. Model C was a four-component model comprised of four latent factors representing the four components of emotions addressed by the AEQ-L2L. Model D sought to fully represent the two-facet structure of the AEQ-L2L by simultaneously taking the eight discrete emotions and the four components into account.

Following Marsh's (1993; see also Lichtenfeld et al., 2022; Pekrun, Marsh, Elliot, et al., 2022 in press; Pekrun, Marsh, Suessenbach, et al., 2022) recommendations, a correlated uniqueness approach was employed to construct this model. The eight L2

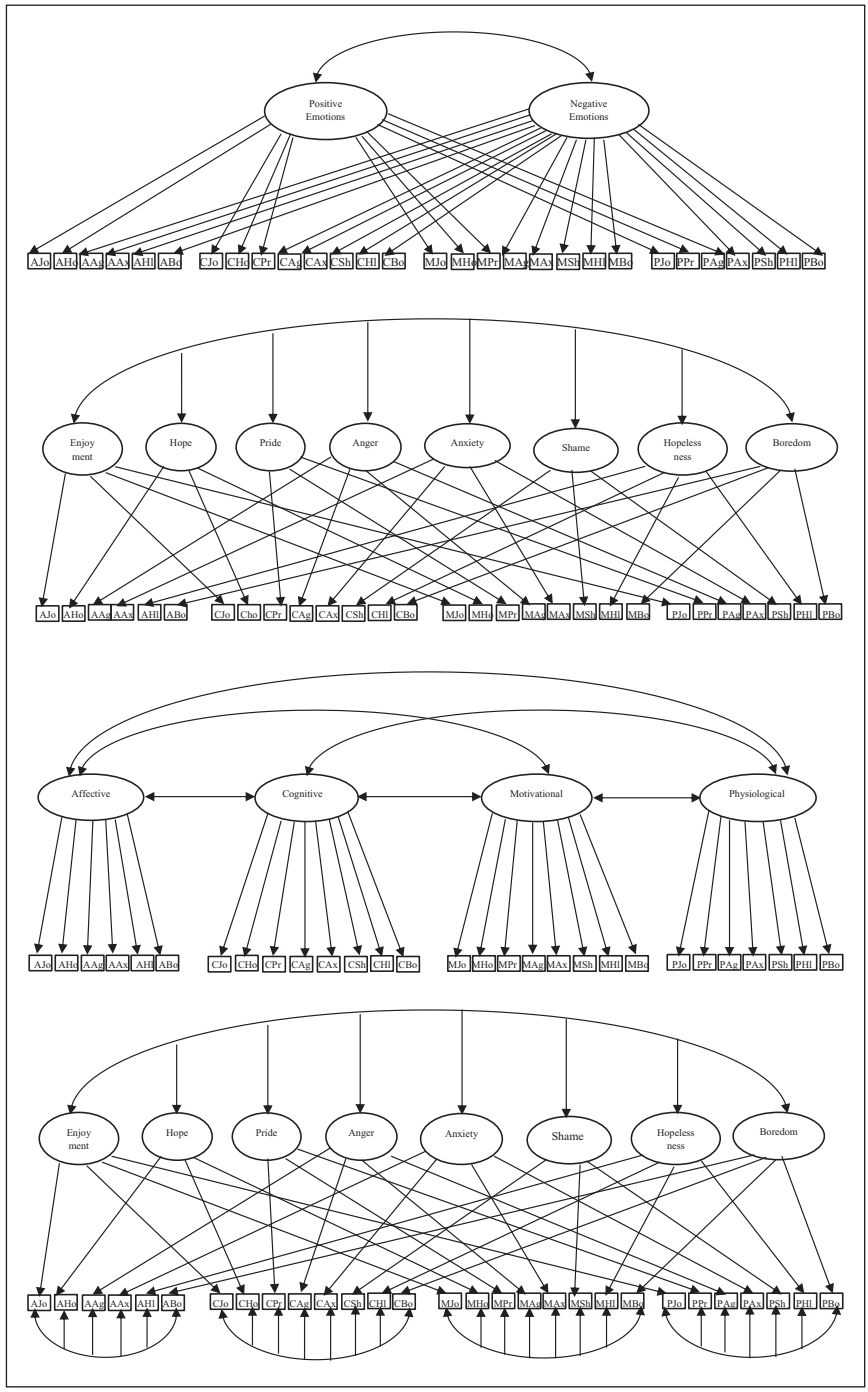


Figure 2. *(Continued)*

**Figure 2.** Structural equation modeling (SEM) models for relations between second language (L2) emotions and their four components.
*Notes*. Upper panel: Model A (positive vs. negative two-factor model). Upper-middle panel: Model B (eight emotion-factor model). Lower-middle panel: Model C (four component-factor model). Lower panel: Model D (emotion × component-factor model). C, A, M, and P denote cognitive, affective, motivational, and physiological components of emotions, respectively. Jo = enjoyment. Ho = hope. Pr = pride. An = anger. Ax = anxiety. Sh = Shame. Hl = hopelessness. Bo = boredom.

**Table 5.** Latent correlations of second language (L2) emotions with appraisals, motivation, and achievement.

| Emotion | Appraisals | | Motivation | | | Achievement |
|---|---|---|---|---|---|---|
| | Control | Value | Intrinsic | Extrinsic | Regulation | L2 grades |
| Enjoyment | .57/.58 | .39/.47 | .33/.31 | .21/.20 | .44/.43 | .43/.40 |
| Hope | .63/.58 | .30/.39 | .27/.29 | .18/.18 | .45/.44 | .48/.35 |
| Pride | .56/.53 | .32/.40 | .32/.25 | .24/.19 | .41/.44 | .47/.36 |
| Anger | −.35/−.38 | −.14/−.09 | −.15/−.23 | −.07/−.06 | −.27/−.24 | −.33/−.13 |
| Anxiety | −.36/−.41 | −.01/−.01 | −.14/−.20 | .01/.02 | −.26/−.26 | −.33/−.18 |
| Shame | −.28/−.42 | −.09/−.10 | −.16/−.17 | .03/.05 | −.18/−.27 | −.21/−.21 |
| Hopelessness | −.42/−.47 | −.13/−.17 | −.14/−.22 | −.13/−.13 | −.28/−.30 | −.38/−.33 |
| Boredom | −.35/−.46 | −.19/−.21 | −.16/−.26 | −.11/−.12 | −.29/−.35 | −.32/−.26 |

*Notes*. Correlations are partial correlations controlling for prior achievement. The left entry is the parameter from teacher track ($N = 471$) and the right entry is the parameter from non-teacher track ($N = 550$). $p < .05/.01/.001$ for $|r| > 0.10/.12/.15$ for teacher track. $p < .05/.01/.001$ for $|r| > 0.09/.11/.14$ for non-teacher track.

emotions were represented by eight latent factors, and the contributions of the four components were taken into account by letting the uniqueness of items representing the same component correlate across emotions. The positive versus negative emotion two-factor model (Model A) showed a poor fit to the data, with CFI = .82/.80, TLI = .81/.78, RMSEA = .054/.056, and SRMR = .061/.062 for teacher/non-teacher track students. The fit for the eight-emotion model (Model B) was better, albeit not satisfactory either, with CFI = .86/.88, TLI = .85/.87, RMSEA = .045/.044, and SRMR = .061/.054. The fit for the four-component factor model (Model C) was even poorer, with CFI = .72/.75, TLI = .70/.73, RMSEA = .063/.063, and SRMR = .100/.094. In contrast, the two-facet, emotion × component model (Model D) showed a good fit for both groups, with CFI = .94/.95, TLI = .91/.93, RMSEA = .035/.034, and SRMR = .049/.044.

## *4 External validity: Linkages of L2 emotions with appraisals, motivation, and achievement*

To test the external validity of the AEQ-L2L, we examined the correlations of students' L2 emotions with control-value appraisals, motivation, and language exam grades. Students' prior achievement was controlled in the analysis. Supporting the propositions of CVT (Pekrun, 2006, 2018), L2 emotions, appraisals, motivation, and achievement

were clearly linked (Table 5), with different groups of emotions demonstrating different patterns of relations. Specifically, positive activating L2 emotions (enjoyment, hope, and pride) correlated positively and significantly with control and value appraisals, intrinsic motivation, extrinsic motivation, self-regulated learning, and exam scores in both teacher and non-teacher students. In contrast, the opposite patterns of correlations were observed for negative deactivating L2 emotions (hopelessness and boredom). Furthermore, negative activating L2 emotions (anger, anxiety, and shame) correlated negatively and significantly with perceived control and value (except for anxiety), intrinsic motivation, self-regulated learning, and language grades across the two groups of students, but none of them showed significant correlations with students' extrinsic motivation targeting course achievement in the present samples. Interestingly, the overall magnitude of correlations between the eight emotions and their antecedents and outcomes was stronger for positive as compared with negative emotions for both cohorts of students.

## VI Discussion

To meet the increasing need for refined theory and valid instruments for investigating multiple emotions in L2 learning, this study sought to use CVT for testing the psychometric properties of the AEQ-L2L in university EFL classrooms and to examine the component structure of eight emotions in the domain of language learning. Research question 1 targeted item and scale statistics of the AEQ-L2L. The ranges, means, and standard deviations of the scales showed that the eight emotions considered in the present study were commonly experienced by L2 learners when studying for an L2 exam, with substantial variation across learners. These results are consistent with recent studies of emotions in the L2 learning context and substantiate the need to investigate multiple L2 emotions (Davari et al., 2022; Wang et al., 2021). The good to excellent item-total correlations and reliability coefficients in both samples demonstrate that the scales of the AEQ-L2L are psychometrically robust measures of students' emotions in L2 learning. Importantly, the results suggest that students actually experienced more positive emotions than negative emotions in their L2 learning. This is in line with theory and research of positive psychology in L2 learning (MacIntyre et al., 2019; Saito et al., 2018), demonstrating that positive emotions are important constituents of L2 learning and require more empirical attention than they have received in the past.

Research question 2 addressed the internal validity of the AEQ-L2L by examining the component structure of the eight language emotion scales in and across two cohorts of EFL students. Model fit parameters of group-specific CFAs demonstrated a high degree of fit for models representing the hierarchical component structure of L2 emotions compared with the general one-factor models. Furthermore, multi-group CFAs attested to the invariance of the four-component structure of the AEQ-L2L scales across two independent samples even at the most restrictive statistical level (invariance of latent means). These results provide strong evidence on the structural validity of the AEQ-L2L in assessing the eight L2 emotions. Consistent with CVT (Pekrun, 2006), L2 emotions are thus best conceptualized as hierarchically organized constructs consisting of affective, cognitive, motivational, and physiological components. The findings align with recent research supporting the adoption of a multi-componential approach to measuring

different facets of L2 anxiety (Cheng, 2017; Kutuk et al., 2020), and extend them by showing that this multidimensional conceptualization is suited to represent a wide range of other discrete L2 emotions as well. The present results are also in line with appraisal theory of emotions (Scherer & Moors, 2019) and test anxiety research (Lowe, 2018) supporting multi-componential approaches to conceptualize emotions.

Research question 3 explored the internal validity of the AEQ-L2L by assessing the relations between the eight L2 emotions together with their four components. Results of latent correlation analyses confirmed that the eight L2 emotions as measured by the AEQ-L2L indeed represent distinct emotion constructs. This highlights the importance and usefulness for L2 researchers to move beyond language anxiety and enjoyment and attend to a broader range of emotions experienced by students in language learning (Davari et al., 2022; Teimouri, 2018b). It is time to treat both negative and positive L2 emotions in a more differentiated manner and analyse diverse emotions as unique drivers of language learning. In interpreting the substantial correlations between some of the same-valenced L2 emotions (enjoyment and pride; anger and boredom), it is important to note that latent coefficients are corrected for measurement error and thus represent the highest possible estimates for these relations compared with manifest correlations.

Moreover, model fit indexes for alternative models of the total set of emotions and their components suggest that it is best to consider both the distinctiveness of these emotions and their internal component structure to achieve an adequate representation of L2 learners' emotional experiences. On the one hand, these findings imply that reducing emotional experience to the dimension of valence by classifying emotional states as either positive or negative without any further differentiation does not sufficiently capture the qualities of students' emotional experiences in language learning (Garrett & Young, 2009). On the other hand, it is also insufficient to only distinguish between multiple L2 emotions and neglect the basic components that constitute these emotions (Cheng, 2017). The hierarchical component structure of emotions should be considered while examining the uniqueness of each emotion construct. Overall, these findings provide further support for the construct validity of the AEQ-L2L for investigating multiple emotions in L2 learning.

Research question 4 focused on the external validity of the AEQ-L2L by exploring correlations of students' L2 emotions with their appraisals, motivation, and achievement. Consistent with CVT (Pekrun, 2006, 2018), we found positive relations between positive activating emotions and control-value appraisals, motivational variables, and achievement, and negative relations between negative deactivating emotions and these antecedents and outcomes. These results are in line with previous research investigating the correlates of appraisal antecedents and motivation and achievement outcomes of L2 anxiety and enjoyment (Khajavy et al., 2018; Phung, 2017; Saito et al., 2018; Sardegna et al., 2018), but go beyond previous findings by demonstrating that these linkages extend to other discrete L2 emotions as well.

Negative activating emotions showed negative correlations with appraisals, motivation, and achievement, except for their correlations with extrinsic motivation which were non-significant. These findings are congruent with CVT's proposition that negative activating emotions can exert variable effects on students' learning processes and outcomes due to their positive influence of stimulating extrinsic motivation for some

students, although the overall effects of these emotions on academic performance are prone to be negative (Pekrun et al., 2011). The findings are also consistent with qualitative studies suggesting that negative emotions such as anger, anxiety, and shame can serve as language learning motivators for some students under certain circumstances (Swain, 2013), which may explain the null correlations in the present investigation. Regarding the non-significant correlation between perceived value and anxiety, CVT (Pekrun, 2006) proposes that the negative value of failure amplifies negative outcome emotions such as anxiety and shame. The present study used scores from a value scale representing a combined assessment of different facets of value, which may have attenuated the correlations between perceived value and these L2 emotions.

Notably, linkages with appraisals, motivation and achievement were stronger for positive than for negative L2 emotions. These findings align with recent research demonstrating that the strength of correlations between L2 emotions and their antecedents and outcomes may indeed be stronger for positive than for negative emotions (Saito et al., 2018; Wang et al., 2021). Thus, while acknowledging the essential role played by negative emotions in L2 learning, it is important to emphasize that positive emotions may show even stronger associations with language behaviors and achievement. Overall, the present findings support the external validity of AEQ-L2L for tapping into the correlates of different positive and negative L2 emotions and moving towards a more nuanced understanding of the emotional parameters of language learning.

## *1 Pedagogical implications*

The present study bears several practical implications for language educators. As our findings suggest that the AEQ-L2L can be used to reliably measure eight different learning-related L2 emotions, language teachers may apply this instrument as a diagnostic tool for assessing students' diverse emotion profiles and adjusting their teaching accordingly. For example, if many students feel bored by the learning materials, teachers may consider incorporating more stimulating activities (Loderer et al., 2020). Moreover, if some students report high levels of L2 anxiety, shame, or hopelessness when preparing for an upcoming exam, teachers may aim to emphasize students' progress toward their desired goals and foster intrinsic values of L2 learning, rather than comparing their performance to the performance of other students (Teimouri, 2018b). Importantly, the most effective way of promoting L2 learning may be through the cultivation of positive emotions such as enjoyment, hope, and pride, considering the stronger correlations between positively charged emotions and L2 outcomes in the present study as well as previous research (Khajavy et al., 2018). To achieve this, teachers should cultivate positive and supportive classroom climates, and express their own joy for language learning which may have contagious effects on students' corresponding positive emotions (Shao & Parkinson, 2021).

Moreover, the hierarchical component structure of the AEQ-L2L suggests that L2 teachers might need to tailor intervention programs to different components of students' emotions. For example, if a student experiencing pre-exam L2 anxiety is primarily due to cognitive worry, teachers may consider enhancing students' perception of control through using attributional retraining (Perry et al., 2001) that promotes effort attributions

and a growth mindset (Lou et al., 2022). In contrast, if a student often experiences high physiological arousal before L2 tests, teachers can use physical relaxation techniques such as aerobic exercises or background music to ease students' tension (Shao, Pekrun, et al., 2020). If a student feels anxious to begin L2 studying as a result of avoidance motivation, teachers may provide students with role models in L2 learning to promote their ideal-L2 selves and mastery orientation (Teimouri, 2018a). Of course, these interventions will have concomitant effects on all other components of L2 emotions given the high correlations between emotion components.

Furthermore, students can be instructed to use the AEQ-L2L as a self-evaluation tool to appraise the degree and changes of their emotions during L2 learning. Students may also be provided with different formats of the AEQ to self-assess their L2 emotions in different academic settings (test, class, learning) and across different temporal perspectives (trait, situation-specific, state). With sufficient guidance, students can be taught to apply a repertoire of intervention strategies and different study techniques (e.g. technology-based learning, role play, fictional talk, movies, or TV shows; see Shao, Nicholson, et al., 2020) to enhance pleasant and reduce unpleasant emotions in L2 learning. These insights can help guide efforts towards adaptive emotion regulation that enables students to stay motivated and feel positive about their learning, or trigger help-seeking behaviors.

## *2 Limitations and future directions*

The present study yielded robust evidence demonstrating the importance of considering multiple emotions in L2 learning, and the findings attest to the reliability and validity of the AEQ-L2L. Nevertheless, the study also has limitations, which can be used to suggest directions for future research. First, the AEQ-L2L focuses on emotions related to studying. However, language learners' achievement emotions may target different aspects of their academic activities. For example, a student who enjoys learning language in the classroom, or doing homework in the library, may not necessarily enjoy taking L2 exams. Thus, future research may consider L2 emotions in testing situations using the test-related emotion scales of the AEQ (Pekrun et al., 2011) and investigate whether students' emotional experiences may differ in this context, compared with their emotions during studying.

Second, the present study used the AEQ-L2L to measure students' emotions in relation to antecedents and outcomes when studying for a general L2 course exam. The exam provided a combined score for students' reading and writing proficiency in the English language. Previous research has shown that skill-specific anxieties played differential roles in the acquisition of different language skills (Cheng, 2004; Zhang, 2013). Therefore, future research is advised to complement the approach used in the present study by additionally developing skill-specific emotion scales (Cheng, 2017) and examining relations between students' skill-specific L2 emotions and their appraisals, motivation, and achievement.

Third, due to administrative restrictions, the nature of the data was largely correlational. As such, the findings provide a snapshot of the relations between emotions, their antecedents, and their outcomes, but do not capture the dynamics of emotion processes during the L2 learning process. Nevertheless, the findings on the links between emotions

and students' L2 achievement can be regarded as longitudinal because students' prior achievement was controlled. Future research may use more extended longitudinal designs involving multiple assessments, or experimental designs to disentangle causal relationships as well as possible reciprocal linkages between emotions, appraisals, motivation, and L2 achievement (Elahi Shirvan & Taherian, 2018).

Fourth, the perceived value scale used in the present research contains items measuring three facets of value (intrinsic value, attainment value, and utility value). Although these three facets correlate positively (Pekrun & Meier, 2011), it is important to note that they are conceptually different. For example, a student who likes reading English texts may not necessarily consider it important to get a good grade in the English exam. Therefore, it may be useful for future research to use separate scales to measure each value dimension in order to acquire a more nuanced understanding of the relations between different facets of value appraisals and discrete emotions in L2 classrooms.

## VII Conclusions

Language classrooms are filled with a broad spectrum of emotions, each of which deserves attention. Instead of only focusing on anxiety or enjoyment, we should also appreciate the uniqueness and dynamics of the many other emotions occurring in the L2 classroom, such as students' hope, pride, anger, shame, hopelessness, and boredom. The present research aimed to validate a self-report instrument, the AEQ-L2L, designed to measure eight emotions experienced by students in L2 learning environments. The internal validity of the AEQ-L2L was supported by confirming the component structure and distinctiveness of the eight L2 emotions, and the external validity of the scales was confirmed through substantial correlations between the emotions, appraisal antecedents, motivation, and learning outcomes. From an interdisciplinary perspective, our results suggest that it can be useful to adapt emotion theory and emotion instruments from educational psychology to the field of L2 learning. We believe the AEQ-L2L adapted in the present study may serve as a useful tool for L2 researchers to examine a broader range of emotions in language classrooms than has been considered before. We also hope that this research may inspire language researchers to not only use the AEQ-L2L, but to also to design further rigorous instruments that simultaneously consider the general nature of human emotion and the domain specificity of language learning. The development and application of instruments like the AEQ-L2L can help launch a new era of investigation which will ultimately benefit students' language development as well as their affective wellbeing.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by a grant from the Zhejiang Social Science Foundation, China, awarded to K. Shao (NO: 22NDJC084YB).

### ORCID iDs

Kaiqi Shao https://orcid.org/0000-0002-0082-7767
Reinhard Pekrun https://orcid.org/0000-0003-4489-3827

## Supplemental material

Supplemental material for this article is available online.

## Notes

1. One-hundred and thirty-eight participants were excluded prior to the analyses due to unfilled identification information or missing more than 20% of all items (Barry et al., 2013).
2. With the multi-component structure of each L2 emotion scale and the limited number of items representing each emotion component, ESEM models showed identification problems due to the high number of estimated parameters and resulting lack of degrees of freedom. Specifically, given that we estimated four first-order latent factors for all emotion components, ESEM included a substantially higher number of cross-loadings compared with standard CFA. Therefore, using ESEM for models representing emotion components would require a higher number of items per emotion component than available in the AEQ-L2L. More generally, ESEM is appropriately used when the fit of the more parsimonious CFA models do not adequately fit the data. However, for the present data, the fit of the more parsimonious CFA models was so good (e.g. CFIs .97 – .99; Table 3) and the factors so well defined that there was no need to pursue the corresponding ESEMs (Marsh et al., 2014).
3. As the L2 hope scale does not include physiological items and the L2 pride and shame scales contain only one affective item, these three component scales were not modeled.

## References

Aida, Y. (1994). Examination of Horwitz, Horwitz, and Cope's construct of foreign language anxiety: The case of students of Japanese. *The Modern Language Journal*, *78*, 155–168.

Alamer, A. (2022). Exploratory structural equation modeling (ESEM) and bifactor ESEM for construct validation purposes: Guidelines and applied example. *Research Methods in Applied Linguistics*, *1*, 100005.

Alamer, A., & Lee, J. (2019). A motivational process model explaining L2 Saudi students' achievement of English. *System*, *87*, 102133.

Alamer, A., & Marsh, H. (2022). Exploratory structural equation modeling in second language research: An applied example using the dualistic model of passion. *Studies in Second Language Acquisition*, *44*, 1477–1500.

Arar, E.A., & Tannenbaum, M. (2021). Pride, shame, and guilt: Socio-emotional aspects of the encounter between the Bedouin and Fellahi dialects. *International Journal of Bilingual Education and Bilingualism*, *24*, 1232–1246.

Barry, T.J., Hermans, D., Lenaert, B., Debeer, E., & Griffith, J.W. (2013). The eACS: Attentional control in the presence of emotion. *Personality and Individual Differences*, *55*, 777–782.

Botes, E., Dewaele, J.-M., & Greiff, S. (2021). The development and validation of the short form of the foreign language enjoyment scale. *The Modern Language Journal*, *105*, 858–876.

Botes, E., van der Westhuizen, L., Dewaele, J.-M., MacIntyre, P., & Greiff, S. (2022). Validating the short-form foreign language classroom anxiety scale (S-FLCAS). *Applied Linguistics*, *43*, 1006–1033.

Bown, J., & White, C.J. (2010). Affect in a self-regulatory framework for language learning. *System*, *38*, 432–443.

Byrne, B.M. (2011). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. Routledge / Taylor and Francis.

Cheng, F.F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*, 464–504.

Cheng, Y.-S. (2004). A measure of second language writing anxiety: Scale development and preliminary validation. *Journal of Second Language Writing*, *13*, 313–335.

Cheng, Y.-S. (2017). Development and preliminary validation of four brief measures of L2 language-skill-specific anxiety. *System*, *68*, 15–25.

Cheng, Y.-S., Horwitz, E. K., & Schallert, D. L. (1999). Language anxiety: Differentiating writing and speaking components. *Language Learning*, *49*, 417–446.

Davari, H., Karami, H., Nourzadeh, S., & Iranmehr, A. (2022). Examining the validity of the achievement emotions questionnaire for measuring more emotions in the foreign language classroom. *Journal of Multilingual and Multicultural Development*, *43*, 701–714.

Dewaele, J.-M., & MacIntyre, P.D. (2014). The two faces of Janus? Anxiety and enjoyment in the foreign language classroom. *Studies in Second Language Learning and Teaching*, *4*, 237–274.

Dewaele, J.-M., Magdalena, A.F., & Saito, K. (2019). The effect of perception of teacher characteristics on Spanish EFL learners' anxiety and enjoyment. *The Modern Language Journal*, *103*, 412–427.

Dewaele, J.-M., Petrides, K.V., & Furnham, A. (2008). Effects of trait emotional intelligence and sociobiographical variables on communicative anxiety and foreign language anxiety among adult multilinguals: A review and empirical investigation. *Language Learning*, *58*, 911–960.

Elahi Shirvan, M., & Taherian, T. (2018). Longitudinal examination of university students' foreign language enjoyment and foreign language classroom anxiety in the course of general English: Latent growth curve modeling. *International Journal of Bilingual Education and Bilingualism*, *24*, 31–49.

Elahi Shirvan, M., Taherian, T., Shahnama, M., & Yazdanmehr, E. (2021). A Longitudinal study of foreign language enjoyment and l2 grit: A latent growth curve modeling. *Frontiers in Psychology*, *12*, 720326.

Flake, J.K., & Fried, E.I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, *3*, 456–65.

Forsblom, L., Pekrun, R., Loderer, K., & Peixoto, F. (2022). Cognitive appraisals, achievement emotions, and students' math achievement: A longitudinal analysis. *Journal of Educational Psychology*, *114*, 346–367.

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.

Frenzel, A.C., Thrash, T., 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.

Garrett, P., & Young, R.F. (2009). Theorizing affect in foreign language learning: An analysis of one learner's responses to a communicative based Portuguese course. *The Modern Language Journal*, *93*, 209–226.

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.

Horwitz, E.K. (2010). Foreign and second language anxiety. *Language Teaching*, *43*, 154–167.

Horwitz, E.K., Horwitz, M., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, *70*, 125–132.

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

Jin, Y., & Zhang, L.J. (2021). The dimensions of foreign language classroom enjoyment and their effect on foreign language achievement. *International Journal of Bilingual Education and Bilingualism*, *24*, 948–962.

Khajavy, G.H., MacIntyre, P.D., & Barabadi, E. (2018). Role of the emotions and classroom environment in willingness to communicate: Applying doubly latent multilevel analysis in second language acquisition research. *Studies in Second Language Acquisition*, *3*, 605–624.

Kutuk, G., Putwain, D.W., Kaye, L., & Garrett, B. (2020). Development and validation of a new multidimensional language class anxiety scale. *Journal of Psychoeducational Assessment*, *38*, 649–658.

Lichtenfeld, S., Pekrun, R., Marsh, H.W., Nett, U.E., & Reiss, K. (2022). Achievement emotions and elementary school children's academic performance: Longitudinal models of developmental ordering. *Journal of Educational Psychology*. Epub ahead of print 2022. DOI: 10.1037/edu0000748

Lichtenfeld, S., Pekrun, R., Stupnisky, R.H., Reiss, K., & Murayama, K. (2012). Measuring students' emotions in the early years: The achievement emotions questionnaire – elementary school (AEQ-ES). *Learning and Individual Difference*, *22*, 190–201.

Loderer, K., Pekrun, R., & Lester, J.C. (2020). Beyond cold technology: A systematic review and meta-analysis on emotions in technology-based learning environments. *Learning and Instruction*, *70*, 101162.

Lou, N.M., Chaffee, K.E., & Noels, K.A. (2022). Growth, fixed, and mixed mindsets: Mindset system profiles in foreign language learners and their role in engagement and achievement. *Studies in Second Language Acquisition*, *44*, 607–632.

Lowe, P.A. (2018). An investigation into the psychometric properties of the test anxiety measure for college students. *Journal of Psychoeducational Assessment*, *36*, 322–336.

MacIntyre, P.D., & Gardner, R.C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning*, *44*, 283–305.

MacIntyre, P.D., & Gregersen, T. (2012). Emotions that facilitate language learning: The positive-broadening power of the imagination. *Studies in Second Language Learning and Teaching*, *2*, 193–213.

MacIntyre, P.D., Gregersen, T., & Mercer, S. (2019). Setting an agenda for positive psychology in SLA: Theory, practice, and research. *The Modern Language Journal*, *103*, 262–274.

Marsh, H.W. (1993). Academic self-concept: Theory, measurement, and research. In Suls, J.M. (Ed.), *The self in social perspective* (pp. 59–98). Lawrence Erlbaum.

Marsh, H.W., Morin, A.J., Parker, P.D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, *10*, 85–110.

Muthén, L.K., & Muthén, B.O. (1998–2017). M*plus user's guide*. Muthén & Muthén.

Nicholls, J.G. (1976). When a scale measures more than its name denotes: The case of the test anxiety scale for children. *Journal of Consulting and Clinical Psychology*, *44*, 976–985.

Pan, C., & Zhang, X. (2021). A longitudinal study of foreign language anxiety and enjoyment. *Language Teaching Research*. Epub ahead of print 18 February 2021. DOI: 10.1177/1362168821993341

Papi, M., & Khajavy, G.H. (2021). Motivational mechanisms underlying second language achievement: A regulatory focus perspective. *Language Learning*, *71*, 537–572.

Park, G.P. (2014). Factor analysis of the foreign language classroom anxiety scale in Korean learners of English as a foreign language. *Psychological Reports*, *115*, 261–275.

Pawlak, M., Kruk, M., Zawodniak, J., & Pasikowski, S. (2020). Investigating factors responsible for boredom in English classes: The case of advanced learners. *System*, *91*, 102–122.

Peixoto, F., Mata1, L., Monteiro, V., Sanches, C., & Pekrun, R. (2015). The achievement emotions questionnaire: Validation for pre-adolescent students. *European Journal of Developmental Psychology*, *12*, 472–481.

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.

Pekrun, R. (2018). Control-value theory: A social-cognitive approach to achievement emotions. In Liem, G.A.D., & D.M. McInerney (Eds.), *Big theories revisited 2: A volume of research on sociocultural influences on motivation and learning* (pp. 162–190). Information Age Publishing.

Pekrun, R., & Meier, E. (2011). Task value questionnaire. Unpublished manuscript, Department of Psychology, University of Munich, Munich, Germany.

Pekrun, R., & Perry, R.P. (2014). Control-value theory of achievement emotions. In Pekrun, R., & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 120–141). Taylor and Francis.

Pekrun, R., Goetz, T., & Perry, R.P. (2005). *Academic emotions questionnaire (AEQ): User's manual*. 2nd edition. Department of Psychology, University of Munich.

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*, 91–105.

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.

Pekrun, R., Marsh, H.W., Suessenbach, F., Frenzel, A.C., & Goetz, T. (2022). School grades and students' emotions: Longitudinal models of within-person reciprocal effects. *Learning and Instruction*, *83*, 101626.

Pekrun, R., Marsh, H.W., Elliot, A.J., et al. (2022, in press). A three-dimensional taxonomy of achievement emotions. *Journal of Personality and Social Psychology*.

Perry, R.P., Hladkyj, S., Pekrun, R., & Pelletier, S.T. (2001). Academic control and action control in the achievement of college students: A longitudinal field study. *Journal of Educational Psychology*, *93*, 776–789.

Phung, L. (2017). Task preference, affective response, and engagement in L2 use in a US University context. *Language Teaching Research*, *21*, 751–766.

Pintrich, P.R., Smith, D.A.F., Garcia, T., & McKeachie, W.J. (1991). *A manual for the use of the motivated strategies for learning questionnaire (MSLQ)*. Tech. Report No. 91-B-004. Board of Regents, University of Michigan.

Ross, A. (2015). An exploration of the emotions and motivation of tertiary English language learners in Australia. Unpublished doctoral thesis, University of Canberra, Canberra, Australia.

Saito, K., Dewaele, J., Abe, M., & In'nami, Y. (2018). Motivation, emotion, learning experience, and second language comprehensibility development in classroom settings: A cross-sectional and longitudinal study. *Language Learning*, *68*, 709–743.

Sardegna, V.G., Lee, J., & Kusey, C. (2018). Self-efficacy, attitudes, and choice of strategies for English pronunciation learning. *Language Learning*, *68*, 83–114.

Scherer, K.R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and Emotion*, *23*, 1307–1351.

Scherer, K.R., & Moors, A. (2019). The emotion process: Event appraisal and component differentiation. *Annual Review of Psychology*, *70*, 719–745.

Shao, K.Q., Nicholson, L.J., Kutuk, G., & Lei, F. (2020). Emotions and instructed language learning: Proposing a second language emotions and positive psychology model. *Frontiers in Psychology*. Epub ahead of print 25 August 2020. DOI: 10.3389/fpsyg.2020.02142/full

Shao, K.Q., & Parkinson, B. (2021). Social psychological accounts of peer emotion transfer in EFL classrooms: A doubly latent multilevel analysis. *Language Teaching Research*. Epub ahead of print 4 May 2021. DOI: 10.1177/13621688211011513

Shao, K.Q., Pekrun, R., Marsh, H.W., & Loderer, K. (2020). Control-value appraisals, achievement emotions, and foreign language performance: A latent interaction analysis. *Learning and Instruction*, *69*, 101356.

Shao, K.Q., Pekrun, R., & Nicholson, L.J. (2019). Emotions in classroom language learning: What can we learn from achievement emotion research? *System*, *86*, 102–121.

Starkey-Perret, R., Deledalle, A., Jeoffrion, C., & Rowe, C. (2018). Measuring the impact of teaching approaches on achievement-related emotions: The use of the achievement emotions questionnaire. *British Journal of Educational Psychology*, *88*, 446–464.

Sudina, E. (2021). Study and scale quality in second language survey research, 2009–2019: The case of anxiety and motivation. *Language Learning*, *71*, 1149–1193.

Swain, M. (2013). The inseparability of cognition and emotion in second language learning. *Language Teaching*, *46*, 195–207.

Teimouri, Y. (2018a). L2 selves, emotions, and motivated behaviors. *Studies in Second Language Acquisition*, *39*, 681–709.

Teimouri, Y. (2018b). Differential roles of shame and guilt in L2 learning: How bad is bad? *The Modern Language Journal*, *102*, 632–652.

Wang, H., Peng, A.Q., & Patterson, M.M. (2021). The roles of class social climate, language mindset, and emotions in predicting willingness to communicate in a foreign language. *System*, *99*, 102529.

Zeidner, M. (2007). Test anxiety in educational contexts: What I have learned so far. In Schutz, P.A., & R. Pekrun (Eds.), *Emotion in education* (pp. 165–184). Academic Press.

Zhang, X. (2013). Foreign language listening anxiety and listening performance: Conceptualizations and causal relationships. *System*, *41*, 164–177.

Zhang, X. (2019). Foreign language anxiety and foreign language performance: A meta-analysis. *The Modern Language Journal*, *103*, 763–781.