Foreign language learning boredom: Refining its measurement and determining its role in language learning

Chengchen Li[1] , Enhao Feng[1], Xian Zhao[2] and Jean-Marc Dewaele[3,4]

[1]School of Foreign Languages, Huazhong University of Science and Technology, China; [2]School of Cultures, Languages and Linguistics, University of Auckland, New Zealand; [3]UCL Institute of Education, University College London, UK and [4]Birkbeck, University of London, UK
Corresponding author: Chengchen Li; Email: lichengchen@hust.edu.cn

Abstract

This cross-sectional and longitudinal study sought to refine the measurement of foreign language learning boredom (FLLB) and examine its links with overall/skill-specific second and foreign language (L2) achievement (curriculum-based course exam scores) and proficiency (Cambridge English test scores). In Substudy 1, we developed and validated an 11-item *Foreign Language Learning Boredom Scale–Short Form* among secondary and tertiary English as a foreign language (EFL) learners in China (n_1 = 2,223, n_2 = 504, n_3 = 934, n_4 = 1,109). The scale showed sound psychometric properties (i.e., construct/criterion/convergent/discriminant/predictive validity, reliability, and measurement invariance across time and groups). In Substudy 2, structural equation modeling results show that FLLB had a consistent modest negative effect on overall and skill-specific (vocabulary and grammar, listening, reading, and writing) L2 achievement and proficiency (n_4 = 1,109). Substudy 3 was a 12-month three-semester longitudinal investigation (n_4 = 1,109). Cross-lagged panel modeling results show that L2 achievement predicted subsequent FLLB negatively, while FLLB did not predict subsequent L2 achievement.

Introduction

Boredom has been recognized as an important affective individual difference factor in second and foreign language (L2) learning (Dewaele, Botes, & Greiff, 2023; Li, 2021; Pawlak et al., 2020). Yet, more empirical evidence is urgently needed to determine its exact role for the reasons below. Firstly, most extant studies on foreign language learning boredom (FLLB) have used scores in curriculum-based language course exams (e.g., midterm/end-term exams) or self-perceived L2 proficiency as an indicator of L2 achievement rather than more objective and widely used international language proficiency test results (e.g., Dewaele, Botes, & Meftah, 2023; Li & Han, 2022; Li & Li, 2023), which would allow more solid, comparable, and generalizable

findings. Secondly, many prior studies have examined the predictive effects of FLLB on overall L2 achievement (e.g., Li & Li, 2023). Still, very few have investigated its specific role in different L2 skills (i.e., listening, speaking, reading, and writing). The focus on the four L2 skills is important because the emotional experiences (e.g., intensity and frequency of emotions and how emotions arise and impact other learning outcomes) and their nomological network (i.e., how emotions are linked with learner-internal/-external factors and outcomes) may vary across L2 skills (Li et al., 2024). Indeed, different skills differ from each other in their linguistic goals, cognitive demands, visibility, recursiveness, evanescence, time constraints, and inter-activeness (Li, Li et al., 2023). Lastly, most prior studies have utilized cross-sectional designs (e.g., Zhao & Wang, 2023), neglecting the fact that learner emotions and L2 evolve.

Against this background, the current study is primarily concerned with the following two core questions: 1) How does FLLB relate to overall and skill-specific L2 achievement and proficiency? 2) How do FLLB and L2 achievement relate to each other over time? Before addressing the major questions, we firstly set out to refine the existing measurement of FLLB, given that different conceptualizations and measures of a construct may lead to inconsistent and incomparable findings (Borsboom, 2006). Specifically, the 32-item *Foreign Language Learning Boredom Scale* (Li, Dewaele et al., 2023) is to be reduced to the *Foreign Language Learning Boredom Scale–Short Form* (*FLLBS–SF*) with psychometric properties assessed to facilitate future research.

Literature review

Foreign language learning boredom: Conceptualization and measurement

Boredom has been neglected in L2 research until recently (e.g., Li, 2021; Pawlak et al., 2020). For any new construct, its research rationale, definition, conceptualization, and measurement need to be solid to allow and encourage further empirical explorations. In this regard, Pawlak et al. (2020) and Li, Dewaele et al. (2023) took the initiative, but the time has come to advance.

Pawlak et al. (2020) first argued that boredom itself justifies the emerging investigations due to its associative negative symptoms, including learners' distractions, demotivation, and dissatisfaction. The authors then explored the underlying structure of boredom in an English as a foreign language (EFL) context for 107 English majors in Poland using an exploratory factor analysis (EFA). They identified a two-factor structure representing the 23-item scale, *Boredom in Practical English Language Classes–Revised*: (1) Disengagement, monotony, and repetitiveness, and (2) Lack of satisfaction and challenge. This seminal study extends L2 emotion research by including and measuring boredom.

The pioneering scale developed by Pawlak et al. (2020) shows some limitations. Firstly, the process of generating the original item pool remains largely unknown, undermining methodological transparency and replicability. Secondly, the scale did not consider out-of-classroom boredom that learners experience outside the formal instructed classroom learning environment (Pawlak et al., 2023) and general trait boredom that learners are inclined to experience across circumstances (Li et al., 2023). However, both types could potentially differentiate an individual's emotional profile of boredom and its role in L2 learning. Indeed, L2 learning is not restricted to classroom settings, and boredom could also arise after class, such as when doing

homework, massive open online courses, or virtual language classes within the curriculum or extracurricular mobile learning (Li, 2021; Li et al., 2024; Pawlak et al., 2022). Recognizing out-of-classroom boredom is thus urgently needed, especially in the current era of technology-enhanced L2 learning, as it is part of the boredom in relation to L2 learning (Li et al., 2024; Pawlak et al., 2022). In addition, as noted, some learners are more likely to feel bored across life domains, not only in L2 learning but also in other subjects (e.g., maths) or gaming (Li et al., 2023). That is, the general trait of boredom contributes to the experience of L2-specific boredom and becomes part of it. Thirdly, the factor structure identified with EFA should have been further confirmed in a different group of participants using confirmatory factor analysis (CFA), ideally followed by assessment of validity (e.g., criterion/convergent/discriminant/predictive validity), reliability (internal consistency and test–retest reliability), and measurement invariance (across groups and time). Fourthly, the sample size was relatively small for instrument validation. Although there is no consensus on the required sample size for instrument validation, it is without doubt that the sample "should be sufficiently large to eliminate subject variance" (DeVellis 2016, p. 130) and to represent the target population better (DeVellis 2016). Boateng et al. (2018) recommended ten respondents for each scale item as a minimum sample size (there are 23 items in the *Boredom in Practical English Language Classes— Revised*), while Hair et al. (2019) recommended 200 cases as the required size. Lastly, the scale was developed and validated among English majors. Their academic selves, achievement goals, and value appraisal systems, which are assumed as distal and proximal antecedents of academic emotions (Pekrun, 2006), are inherently distinct from those of non-English majors. This may limit the applicability of this scale to non-English majors.

Taking into account the concerns of Pawlak et al.'s (2020) scale mentioned above, Li, Dewaele et al. (2023) moved a step forward. Before developing and validating the *Foreign Language Learning Boredom Scale* (*FLLBS*), the authors first explored the existence of boredom and defined it in a Chinese EFL context. They found that boredom was one of the most frequently experienced emotions of Chinese non-English majors. It manifested itself with sleepiness, inattention, mind wandering, and disengagement. The authors defined FLLB as "a negative, deactivating achievement emotion arising from ongoing learning activities or tasks" (p. 3) based on the three-dimensional taxonomy of the control–value theory in educational psychology (Pekrun, 2006). Based on such a preliminary understanding and conceptualization of FLLB, *FLLBS* was then developed and validated among a relatively large sample (n = 2,223) following a series of systematic psychometric tests. Consequently, Li, Dewaele et al. (2023) identified a seven-factor model representing a psychometrically sound 32-item *FLLBS*: "Foreign Language Learning Classroom Boredom, Underchallenging Task Boredom, PowerPoint Presentation Boredom, Homework Boredom, Teacher Dislike Boredom, General Learning Trait Boredom, and Overchallenging or Meaningless Task Boredom" (p. 4).

The full version and different short versions of the *FLLBS* have been applied in heterogeneous L2 contexts. Firstly, its full version has been mainly used and validated in diverse groups of L2 learners in China, such as 868 university EFL students (Li, 2022), 954 secondary EFL learners in a southeastern rural area (Li & Li, 2023), 504 ethnic minority and Han EFL students in a northwestern rural area (Zhao & Wang, 2023), 517 FL majors (Zhang, 2022), and 348 Chinese-as-a-second-language international students in an online context (Chen et al., 2022). Secondly, its eight-item *Foreign Language Classroom Boredom Subscale* has been applied in

more diverse FL groups, such as 1,555 urban and 600 rural EFL students from central China (Ma et al., 2023), 1,205 urban and 616 rural EFL students from southeastern China (Li & Li, 2024), Chinese university EFL students in online contexts (*n* = 348, Li & Dewaele, 2020; *n* = 880, Wang & Li, 2022), 168 Arab and Kurdish EFL learners in both in-person and emergency remote teaching contexts (Dewaele, Albakistani, & Kamal Ahmed, 2024), 118 English majors from Thailand (Apridayani & Waluyo, 2022), and 332 FL (English, French, and Spanish) learners across educational levels from the UK, China, and Italy (Dewaele, Botes & Greiff, 2023). Lastly, some researchers extracted several items from the *FLLBS*. For example, Wang et al. (2023) extracted 21 items out of the 32 items, three items for each factor, and Zhao et al. (2023) extracted five items.

There is an empirical trend using shorter versions of the *FLLBS*. Nevertheless, the existing shorter versions of *FLLBS* have several limitations. Firstly, the eight-item *Foreign Language Classroom Boredom Subscale* only partially presents FLLB since the subscale is restricted to the classroom. As mentioned earlier, one of the advantages of the full *FLLBS* is that it includes both in-class and out-of-class boredom, and both L2-specific and domain-general boredom. Secondly, for those extracted versions, it remains unclear how the items were selected and if the extracted versions are representative, valid, reliable, and invariant across groups (Wang et al., 2023; Zhao et al., 2023).

Foreign language learning boredom: Links with L2 achievement and proficiency

FLLB has also been theorized to have debilitating effects on learners' motivation (e.g., demotivation), engagement (e.g., pseudo-/superficial/procedural engagement), cognition (e.g., distraction, short attention span, and superficial information processing), strategy use (e.g., less self-regulation), behavior (e.g., withdrawal), and flow experience in class, which further impairs L2 achievement and proficiency (Dewaele et al., 2023; Kruk & Zawodniak, 2017; Li, Dewaele et al., 2023; Li et al., 2024; Pawlak et al., 2020). Underpinned by the control–value theory of achievement emotions (Pekrun, 2006), which posits the dynamic reciprocal linkages between emotions and their antecedents and outcomes, we further assume that FLLB and L2 achievement are bidirectionally linked to each other over time.

Prior empirical studies have explored the links between FLLB and L2 achievement (commonly operationalized as scores in curriculum-based language course exams such as midterm/end-term exams) or self-perceived L2 proficiency in diverse L2 contexts (e.g., Li & Li, 2023; Liu & Wang, 2023; Zhao & Wang, 2023). The findings are mixed. In addition, very few have examined the links between FLLB and L2 proficiency measured with international language proficiency tests (e.g., Tsang & Dewaele, 2023).

FLLB was found to be a negative predictor of overall L2 achievement in most prior relevant studies. Özsaray and Eren (2018), for example, found that boredom had a large negative predictive effect on English achievement of undergraduate students from Turkey ($\beta = -.50$, $p < .05$). Smaller negative predictive effects on English achievement were found among Moroccan English learners from secondary schools, universities, and language centers ($\beta = -.14$, $p < .01$) (Dewaele, Botes, & Meftah, 2023) and English-as-an-L3 (third language) and Chinese-as-an-L2 ethnic minority students from China ($\beta = -.17$, $p < .01$) (Zhao & Wang, 2023). Li and Li (2023) further confirmed the small negative achievement effect of FLLB among secondary EFL students from rural China. Notably, they also revealed the limited durability of such achievement impact: The

effect faded over time (the first week: $\beta = -.14$, $p < .001$; the fifth week: $\beta = -.11$, $p < .001$; the ninth week: nonsignificant).

FLLB was also found to be negatively correlated with self-perceived English proficiency. Liu and Wang (2023), for example, found a small-to-medium negative predictive effect among Chinese secondary students ($r = -.28$, $p < .001$). Similar results were found in an online L2 context among Chinese university learners ($\beta = -.29$, $p < .001$) (Li & Han, 2022).

By contrast, some other studies found no significant predictive effects of FLLB on L2 achievement/proficiency. For example, Dewaele, Botes, and Greiff (2023) found no significant predictive effect of FLLB on English achievement among international students. In two other studies among primary school EFL learners from Hong Kong, Tsang and Dewaele (2023) and Yeung et al. (2023) found that boredom did not significantly predict overall English proficiency or skill-specific English achievement (reading and writing).

Motivation of the current study

To determine the role of foreign language learning boredom in L2 learning

More empirical evidence is urgently needed to determine the exact role of FLLB. Firstly, and surprisingly, international language proficiency tests (e.g., the Cambridge English Assessment, International English Language Testing System [IELTS], Test of English as a Foreign Language [TOEFL], and Oxford Placement Test) have been scarcely used in L2 boredom (also enjoyment) research, which is in sharp contrast with the extensive use of curriculum-based course exams (e.g., end-term/midterm exams) (e.g., Li & Li, 2023; Zhao & Wang, 2023). However, L2 proficiency is of paramount importance to L2 learners as "an index of the comprehension and production abilities that L2 learners develop across linguistic domains (e.g., lexical competence, grammatical competence, discourse competence) and modalities (spoken and written) to communicate" (Tremblay, 2011, p. 340). Although course exams are the primary ways to assess curriculum implementation, teaching effectiveness, and students' mastery of skills and knowledge in the period of coursework (Li et al., 2024), they are regional and administered at different levels of the education system within a country (e.g., national, municipal, and school levels). Thus, they are limited in external validity (e.g., applicability, transferability, and generalizability of relevant findings and implications) (Li et al., 2024). In addition, unlike established L2 proficiency tests, most course exams were administered with few considerations about their reliability and construct validity. Secondly, very few studies have taken a skill-specific approach to examine the distinctive roles of FLLB (and other emotions) in impacting overall L2 and different L2 skills (i.e., listening, speaking, reading, and writing) (e.g., Yeung et al., 2023), leaving this area largely uncharted. Such a skill-specific approach to L2 learner emotion is warranted because learners' emotions may vary across language skills (subsystems) as a result of the variations in terms of linguistic goals, cognitive demands, visibility, recursiveness, evanescence, time constraints, self-paced learning, interactiveness, and social pressure involved in different skills (Li, Li et al., 2023; Li et al., 2024). Lastly, the widely used cross-sectional designs cannot capture the dynamic nature of FLLB and L2 achievement. Moreover, longitudinal investigations reveal causal relationships (Kenny, 1979).

All in all, the current study aims to provide a fuller picture of the role of FLLB in L2 learning more effectively by 1) using diverse measurements (both curriculum-based course exam scores and international L2 proficiency test scores), 2) considering both

general L2 and specific L2 skills, and 3) utilizing both cross-sectional and longitudinal research designs.

To develop a short but sound measure of foreign language learning boredom

As reviewed previously, there is a clear empirical preference for shorter versions of the *FLLBS*. The rationale for reducing the items in the existing *FLLBS* to a minimum number is evident. Firstly, lengthy questionnaires may impair the initial willingness to participate in and complete the questionnaire (Galesic & Bosnjak, 2009; Rolstad et al., 2011). The shorter the questionnaire is, the more respondents start, and the fewer respondents drop out (Galesic & Bosnjak, 2009). Secondly, lengthy questionnaires potentially cause a response burden (Dörnyei & Dewaele, 2022; Galesic & Bosnjak, 2009). The shorter the questionnaire is, the more likely respondents are to maintain their attention and interest, which contributes to a higher response rate and better response quality (Galesic & Bosnjak, 2009; Rolstad et al., 2011). Thirdly, the longer the questionnaire, the more time it takes to complete it, and the more likely the questionnaire will be considered and declined as an obstacle in practice (Fowler, 2014; Rolstad et al., 2011). Fourthly, although it is expected to use multiple items that are semantically similar but phrased differently to measure the same construct or dimension of a construct in scales, unnecessarily identical items not only increase the length of the scale but also may impact participants' affective states (e.g., impatience, fatigue, and hastiness) in response, which may further undermine both the construct validity and reliability of the survey results (Clark & Watson, 1995).

To sum up, short questionnaires have the advantages of being convenient, user-friendly, and feasible and tend to have higher response rates and quality (Dörnyei & Dewaele, 2022). Nevertheless, as identified earlier, either the existing eight-item *Foreign Language Classroom Boredom Subscale* or other extracted versions of the *FLLBS* have their conceptual or methodological limitations. To this end, the current study aims to develop and validate a *Foreign Language Learning Boredom Scale–Short Form (FLLBS–SF)*. It involved a series of systematic psychometric processes (from item selection to scale validation, including reliability assessment, validity assessment, and invariance assessment), following consistent guidelines and stringent statistical criteria, and utilizing data from a large integrated sample with heterogeneous backgrounds (e.g., age, education level, linguistic background, region, ethnicity, and economic status) over different time points.

The current study

Research questions and hypotheses

The following research questions (RQs) guided the present study:

1. Is the newly developed *Foreign Language Learning Boredom Scale–Short Form (FLLBS–SF)* reliable, valid, and invariant?
2. How does FLLB predict overall and skill (subsystem)-specific L2 achievement and proficiency?
3. How are FLLB and L2 achievements related to each other over three consecutive semesters in the span of a year?

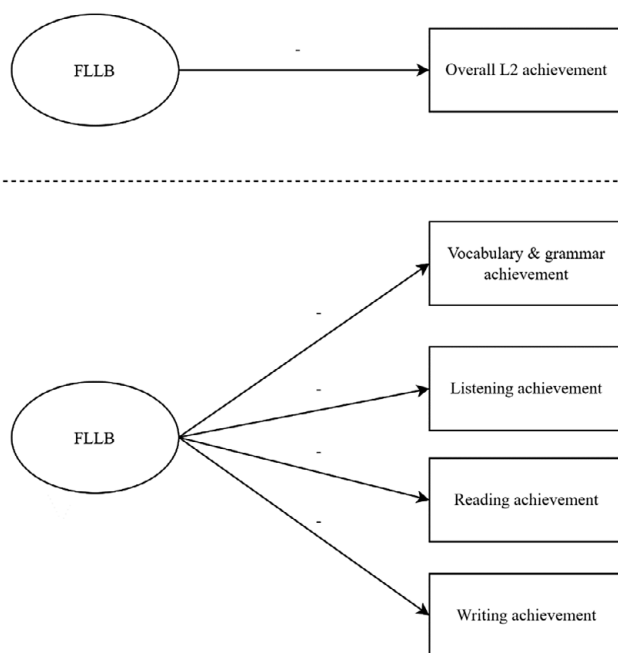For RQ2 and RQ3, we proposed the following hypotheses (Hs):

Figure 1. The hypothesized associations between FLLB and overall/skill (subsystem)-specific L2 achievement.

H1: FLLB would negatively predict overall L2 academic achievement and achievement in specific skills or subsystems (i.e., vocabulary and grammar, listening, reading, and writing) (see Figure 1);

H2: FLLB would negatively predict overall L2 proficiency and skill-specific proficiency (listening, reading, and writing) (see Figure 2);

H3: FLLB and L2 achievement would negatively and reciprocally predict each other across time (see Figure 3).

The first hypothesis was proposed based on the emotion–achievement link assumption of control–value theory and relevant empirical literature (e.g., Dewaele, Botes, & Meftah, 2023; Li & Han, 2022; Li & Li, 2023). Regarding the second hypothesis, there are very few theoretical assumptions/predictions or empirical investigations specifically addressing the effects of emotions on (overall/skill-specific) L2 proficiency. However, as reviewed, extant theoretical assumptions (e.g., the emotion–achievement link assumption of control–value theory) and empirical evidence (Li, Dewaele et al., 2023; Pawlak et al., 2020) suggest that FLLB has a significant impact on learning outcomes such as cognition, motivation, engagement, and behavior, which are linked to L2 proficiency. The second hypothesis is thus that FLLB could impair L2 proficiency. In addition, although L2 proficiency is inherently different from L2 achievement, they are interconnected, especially in FL learning contexts where L2 proficiency is mainly developed in instructed L2 learning within a particular curriculum, whose effectiveness is typically assessed with course scores (a commonly used indicator of L2 achievement). The third hypothesis was underpinned by the assumption of the control–value theory on the reciprocal and dynamic relationships between achievement emotions and
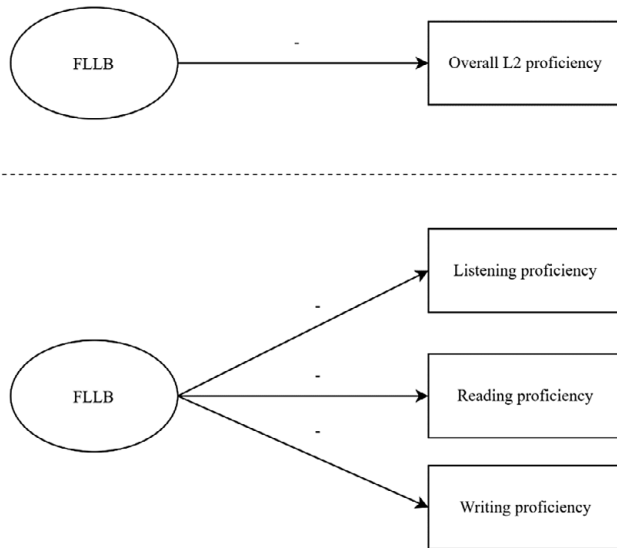
Figure 2. The hypothesized associations between FLLB and overall/skill-specific L2 proficiency.
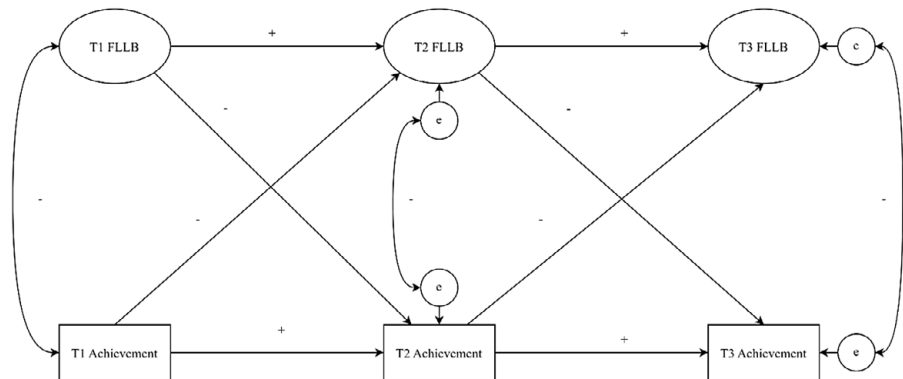

Figure 3. The hypothesized longitudinal associations between FLLB and L2 achievement.

academic achievement, which has been insufficiently corroborated in L2 learner emotion research (Pekrun, 2006).

Research design

The three RQs were answered in three substudies. Study 1 aimed to develop and validate the *Foreign Language Learning Boredom Scale–Short Form* (*FLLBS–SF*) based on different datasets within diverse EFL learning contexts (see Figure 4). Study 2 adopted a cross-sectional design, and the *FLLBS–SF* was applied to assess the links between FLLB and (overall and skill-specific) L2 achievement and proficiency (see Figures 1 and 2). Study 3 was longitudinal in design, and the *FLLBS–SF* was applied to investigate the longitudinal associations between FLLB and English achievement (see Figure 3).
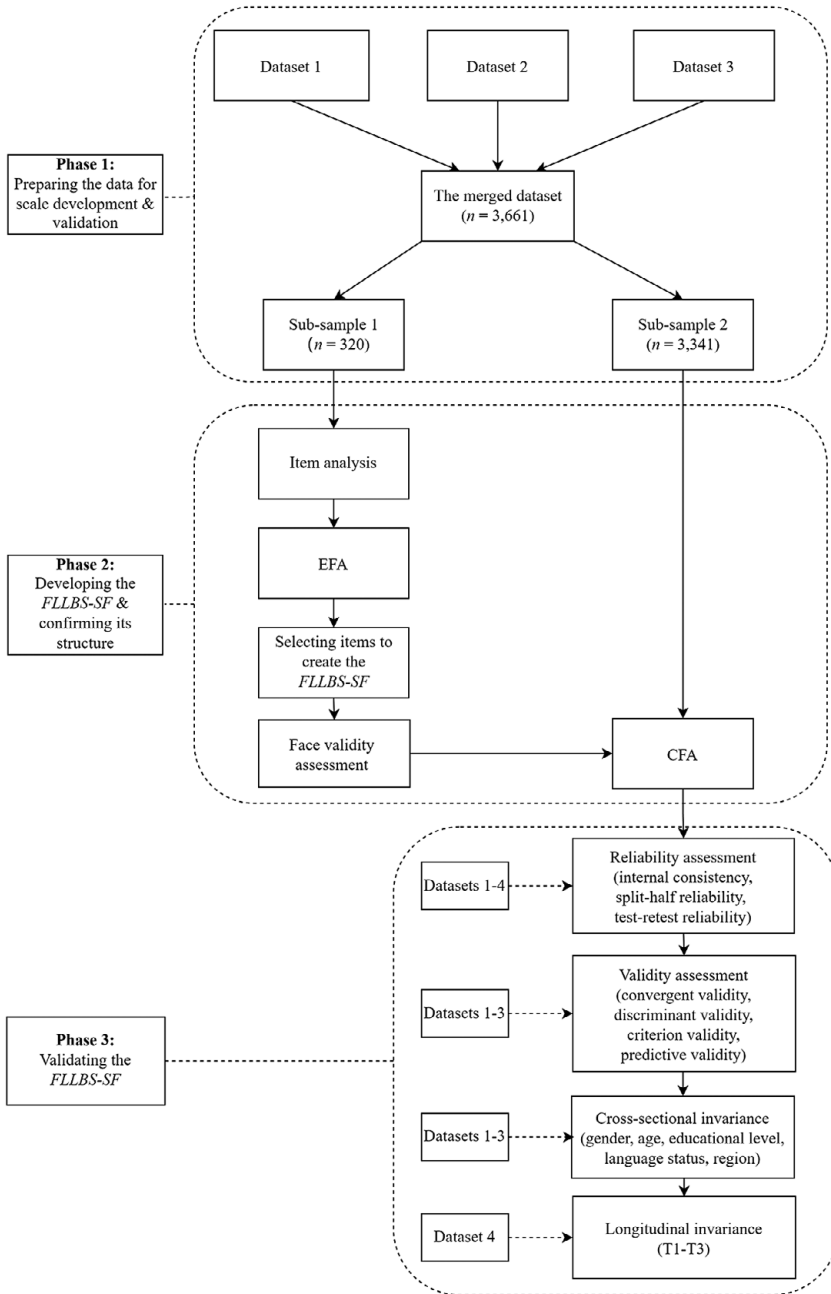
Figure 4. Roadmap of Study 1.
Note. EFA = exploratory factor analysis, CFA = confirmatory factor analysis; T1–T3 = Time points 1–3.

Table 1. Dataset Information

Dataset	1	2	3	4
Design	Cross–sectional	Cross–sectional	Cross–sectional	Cross–sectional and longitudinal
First application	Li, Dewaele et al. (2023)	Zhao & Wang (2023)	Current study	Current study
Time of data collection	2019	2022	May–June 2022	T1: May–June 2021, T2: December 2021–January 2022, T3: May–June 2022
Location	Urban	Rural	Urban and rural	Rural
L1 (%)	Chinese (100%)	Chinese (41.07%)/ minority languages (58.93%)	Chinese (100%)	Chinese (100%)
L2	English	English/ Chinese	English	English
L3	–	–/English	–	–
N	2,223	504	934	1,109
Mean age (SD)	18.33 (2.15)	14.29 (2.07)	13.49 (.88)	13.50 (.77)
Male (%)	901 (40.53%)	256 (50.80%)	473 (50.60%)	696 (62.80%)
Female (%)	1,089 (48.99%)	248 (49.20%)	461 (49.40%)	413 (37.20%)
Education	Tertiary	Secondary	Secondary	Secondary
Observed emotions	FLLB, FLE, FLCA, BS, and ABE	FLLB, FLE	FLLB, FLE, and FLCA,	FLLB, FLE, and FLCA (T1–T3)
English achievement	–	End–term exam	End–term exam	End–term exam (T1–T3)
L2 proficiency	–	–	–	*Cambridge English Test* (T2)

Note. FLE = foreign language enjoyment, FLCA = foreign language classroom anxiety, BS = boredom susceptibility, ABE = academic boredom in English; T1–T3 = Time points 1–3.

Datasets and participants

Table 1 shows the four datasets used in the three substudies. Datasets 1 and 2 were used in prior studies by Li, Dewaele et al. (2023), and Zhao and Wang (2023), and the other two are novel ones. The entire integrated dataset consists of 4,770 students, 2,223 (46.60%) university students from Dataset 1, and 2,547 (53.40%) secondary students from Datasets 1–3. The data on the same variables (e.g., FLLB, FLE, and FLCA) under discussion were obtained with the same instruments across datasets, enabling the integration of the datasets (Isbell & Son, 2022). In Study 1, all the datasets were merged to develop and validate the *FLLBS–SF*. In Study 2, the cross-sectional part of Dataset 4 was used to examine the predictive effect of FLLB on overall and skill-specific L2 achievement and proficiency. In Study 3, the longitudinal part of Dataset 4 was used to explore the associations between FLLB and L2 achievement over three consecutive semesters.

Dataset 1 included 2,223 non-English majors from nine universities in China (M_{age} = 18.33, SD_{age} = 2.15). There were 901 (40.53%) men and 1,089 (48.99%) women. Their English proficiency was roughly at the B1 level in the Common European Framework of Reference for Languages (CEFR) as they were non-English majors who were going to take or had already passed the College English Test–Band 4 (passing the test is a graduation requirement for non-English majors in China), which is commonly aligned with the B1 level in CEFR (Jin et al., 2022).

Dataset 2 included 504 participants (50.80% men, 49.20% women). They were junior secondary school students in rural areas with a mean age of 14.29 (SD = 2.07). Over half of the participants (n = 297, 58.93%) were Chinese-as-an-L2 and English-as-an-L3 students from various minority groups (e.g., Tibetan, Tujia, and Mongolian). The rest were all Chinese L1 and English L2 speakers (n = 207, 41.07%). As regulated by the Chinese Ministry of Education (2022), the participants were all English beginners, roughly at the A1 level in the CEFR.

Dataset 3 contained 934 participants (473 men and 461 women) from rural and urban areas in east China. They were all junior secondary school students with a mean age of 13.49 (SD = .88). All the participants were Chinese L1 speakers learning English as a foreign language. Their proficiency was at the beginning level (roughly equivalent to A1 in the CEFR), as regulated by the Chinese Ministry of Education (2022).

Dataset 4 consisted of 1,109 junior secondary EFL students in a rural boarding school in eastern China. They were from 26 intact classes, among which only four classes lived with their parent(s), and the other 22 classes were all left-behind children whose parent(s) worked in remote urban regions. Their mean age was 13.50 (SD = .77). There were 696 men (62.80%) and 413 (37.20%) women. Participants' scores in the *Cambridge A2 Key for Schools English Test* (M = 45.63, SD = 15.23) indicate relatively low English proficiency, roughly at A1 in the CEFR.

Instruments

Table 2 presents the main instruments used in the three substudies: (1) Emotion scales measuring FLLB and its criterion variables (e.g., enjoyment and anxiety), (2) end-term English exams, and (3) English proficiency tests. All the scales were in Chinese and responded to on a 5-point Likert scale ranging from 1 "*strongly disagree*" to 5 "*strongly agree.*" All the instruments showed acceptable reliability across datasets.

Foreign Language Learning Boredom Scale (*FLLBS*). The original 32-item *FLLBS* (Li, Dewaele et al., 2023) was used to measure general boredom in relation to EFL learning with seven factors (see Literature Review).

The Chinese Foreign Language Enjoyment Scale (*CFLES*). The *CFLES* (Li et al., 2018) was used to measure student foreign language enjoyment, a criterion variable selected for FLLB based on literature (e.g., Li & Li, 2023). The scale contains eleven items measuring three factors: FLE–Private, FLE–Teacher, and FLE–Atmosphere.

Table 2. Reliability of Original Instruments in Different Datasets

Scale	k	Reliability			
		Dataset 1	Dataset 2	Dataset 3	Dataset 4
FLLBS	32	.95	.98	.99	T1–T3: .98, .98, .99
CFLES	11	.88	.94	.96	T1–T3: .95, .95, .97
FLCAS	6	.76	–	.85	T1–T3: .91, .83, .94
BSS	10	.74	–	–	–
ABES	3	.91	–	–	–
End–term exam	81	–	–	–	T1–T3: .93, .91, .92
Cambridge English Test	3	–	–	–	T2: .84

Note. K = the number of scale items, test items, or sections; T1–T3 = Time points 1–3; *FLLBS* = *Foreign Language Learning Boredom Scale*, *CFLES* = *Chinese Foreign Language Enjoyment Scale*, *FLCAS* = *Foreign Language Classroom Anxiety Scale*, *BSS* = *Boredom Susceptibility Scale*, *ABES* = *Academic Boredom in English Scale*.

Foreign Language Classroom Anxiety Scale (*FLCAS*). The short *FLCAS* was used to measure students' anxiety levels in foreign language classrooms, another criterion variable for FLLB (e.g., Li & Li, 2023). Li and Li (2023) truncated the six items from the eight-item *FLCAS* (Dewaele & MacIntyre, 2014), applying and validating it in a similar Chinese EFL context.

Boredom Susceptibility Scale (*BSS*). The *BSS* (Zuckerman, 1979) was used to elicit learners' general proneness to boredom, a criterion variable of FLLB (Li, Dewaele et al., 2023). The scale includes ten items.

Academic Boredom in English Scale (*ABES*). Students' academic boredom in English subject, a criterion variable of FLLB (Li, Dewaele et al., 2023), was measured via the three-item *ABES* that Li (2021) adapted from the *Achievement Emotions Questionnaire* (Pekrun et al., 2011).

End-term English exam. Participants' scores in their school-level end-term English exams were used as the index of their L2 achievement. Although the specific content of these exams differed across datasets, their structures were largely the same, containing sections for listening, vocabulary and grammar, reading, and writing, with the same maximum score of 120. For Dataset 2, we only obtained the global score. For Dataset 4, we obtained the scores for all 81 items in the sections on listening, vocabulary and grammar, reading, and writing, with the maximum scores of 20, 40, 40, and 20, respectively.

Cambridge A2 Key for Schools English Test. A practice version of the test was used to measure young school-age learners' English proficiency in Dataset 4. The maximum score of the test utilized in the current study was 85, with 25, 30, and 30 for the three sections on listening, reading, and two writing tasks (an email writing task and a picture-based story-telling task), respectively. The research site declined the original speaking section because it was not part of their curriculum or high-stake English exams in China. Participants' writing samples were assessed in terms of three dimensions of content (0–5 points), language (0–5 points), and organization (0–5 points) following the guidelines of the *Cambridge Writing Assessment Subscales* (see https://assets.cambridgeenglish.org/schools/CER%206647%20V1c%20JUL20_Teacher%20Guide%20for%20Writing%20A2%20Key%20for%20Schools.pdf).

Six English teachers were recruited from the research site to complete the (inter-) rating. The first author trained them systematically in three 60-minute sessions: 1) Familiarizing them with the rating rubrics, 2) showcasing them how to assess essays, and 3) guiding them to practice rating and having group discussions for challenging and confusing parts. Ten percent of the compositions were double-scored, and the inter-rater reliability for all six rating dimensions was acceptable (all r_s >. 70; Koo & Li, 2000).

Data collection and ethics

For Dataset 1 (Li, Dewaele et al., 2023) and Dataset 2 (Zhao & Wang, 2023), both of which had been used in prior publications, we obtained consent to use the data from both author teams. For Datasets 3 and 4, official approval was obtained from the first author's institution and the research sites. Written consent was obtained from participants and their guardians after they had been offered sufficient information on the nature, purpose, and duration of the project, their unconditional nonparticipation/withdrawal rights, rewards after the completion of the project, and data anonymization in research outcomes. Questionnaire surveys were then carried out via the online

questionnaire platform of WenJuanXing (https://www.wjx.cn/) in computer classes. English exams/proficiency tests were completed in classrooms.

Study 1: Development and validation of the *Foreign Language Learning Boredom Scale-Short Form* (FLLBS-SF)

Study 1 aimed to reduce the 32-item *FLLBS* (Li, Dewaele et al., 2023) to the *FLLBS–SF* and assess its psychometric properties utilizing Datasets 1–4. Following a series of steps recommended by Marsh et al. (2005), Study 1 involved three phases: 1) Preparing the data, 2) developing the *FLLBS–SF*, exploring, and confirming its structure, and 3) validating the *FLLBS–SF* (Figure 4).

Phase 1: Preparing the data for scale development and validation

Datasets 1–3 were first merged and then randomly divided into two subsamples. Subsample 1 was used to develop the *FLLBS–SF* and explore its structure via EFA (n = 320, ten times the number of the scale items; Hair et al. 2019). Subsample 2 was used to confirm the structure through CFA further (n = 3,341). *T*-tests showed no significant differences between the two subsamples in terms of their mean age, levels of boredom, enjoyment, and anxiety.

Phase 2: Developing the *FLLBS-SF* and confirming its structure

Phase 2 involved a series of procedures to develop the *FLLBS–SF* and explore and verify its underlying factor structure using data from Subsample 2 (n = 320). All these followed more stringent criteria or the same criteria used in the validation of the original *FLLBS* (Li, Dewaele et al., 2023).

Item analysis
Item analyses were conducted to assess the quality and measurement characteristics of the individual scale items (Reynolds et al., 2021). Specifically, we performed the item–total correlation analysis, inter-item analysis, and item discrimination analysis.

Item–total correlation analyses were first conducted to assess whether an item is measuring the same construct as the overall test measures, following a more stringent criterion (>. 40; Loiacono et al., 2002) than that used in Li, Dewaele et al. (2023) (>. 30). No items were eliminated after the assessment.

Then, the inter-item correlations were calculated to determine the consistency among items. Coefficients between. 30 and. 80 were considered acceptable (Field, 2013). Consequently, Items 4, 7, 13, 14, 15, 16, 17, 20, 24 and 28 were deleted.

Item discrimination analysis assessed how well the items can differentiate participants (Reynolds et al., 2021). Participants whose scores on the *FLLBS* fell within the upper and lower 27% (n = 85 and n = 86, respectively) constituted two comparison groups for independent sample *t*-tests (Kelley, 1939). Significant differences were detected on each item between the two groups, indicating no need for item deletion.

Exploratory factor analysis
To explore the factor structure underlying FLLB, EFAs were conducted after the assessment of Bartlett's test of sphericity and the Kaiser–Meyer–Olkin (KMO)
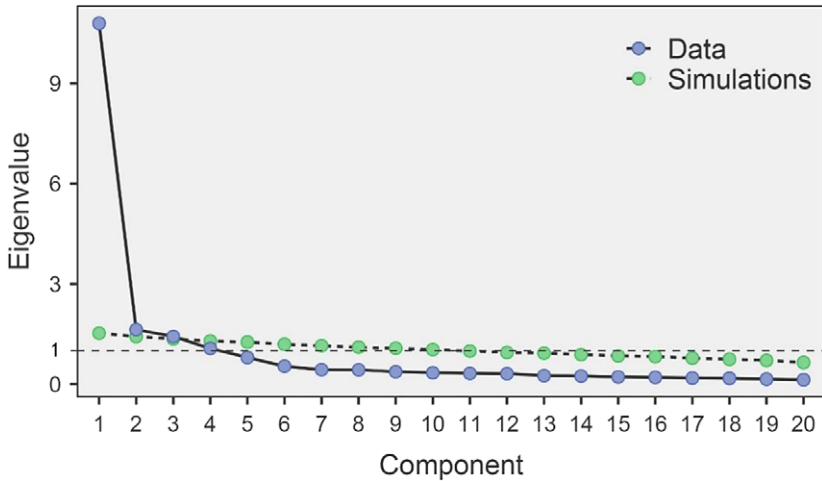
Figure 5. Scree plot and parallel analysis results.

measure of sampling adequacy via SPSS 27. The extraction method was principal component analysis with the oblique (promax) rotation. The final factor number was decided by jointly considering theoretical underpinnings and the following indices: the eigenvalue, the scree plot, and the parallel analysis (Hair et al., 2019). Firstly, factors with eigenvalues greater than one were retained. Secondly, an inflection point occurs in the scree plot when the eigenvalue transitions from a steep slope to a horizontal line. Factors before this point are suitable for retention. Lastly, factors with higher eigenvalues than values generated by parallel analysis were retained (Hair et al., 2019).

The KMO (KMO ≥. 80) and Bartlett's test ($p < .001$) results indicate that Subsample 1 was suitable for EFA. The first EFA was then conducted, revealing four factors with eigenvalues higher than 1. However, Items 18 and 19 were deleted due to their low factor loadings on all four factors (< .40).

Then, a second EFA was conducted. The eigenvalue, scree plot, and parallel analysis indicated four factors, one factor and three factors representing FLLB, respectively (Figure 5). The one-factor solution was abandoned because FLLB is theorized as a multidimensional construct (Li, Dewaele et al., 2023; Pawlak et al., 2020). The three-factor solution of parallel analysis was finally selected because parallel analysis offers more accurate and parsimonious results (e.g., Hair et al., 2019; Hayton et al., 2004) compared to the eigenvalue criterion, which has been criticized for being "amongst the least accurate methods" for factor retention decisions (Costello & Osborne, 2005, p.2).

Item selection

The items were selected for each identified factor based on the EFA results and theoretical considerations. We deleted items according to the following criteria (e.g., Awang, 2012; Clark & Watson, 1995; Hair et al., 2019).

1. Communality lower than .50;
2. Factor loadings on the focal factor lower than .60;

3. Cross-loadings higher than .30;
4. Conceptually inconsistent with other items under the same factors.

Items 21, 22, 23, 25, and 32 were deleted due to low factor loadings. Items 30 and 31 were also deleted due to their conceptual inconsistency with other items under the same factor. Resultantly, 13 items were retained in the scale.

Face validity assessment and the final FLLBS–SF

The initial version of *FLLBS–SF* was evaluated by 13 researchers in applied linguistics. Following Hair et al. (2019), they were invited to judge each scale item in terms of their (a) relevance in meaning (1 "*imprecisely measuring boredom*," 10 "*accurately measuring boredom*") and (b) clarity in the linguistic expression (1 "*very ambiguous*," 10 "*very clear*") on a 10-point scale. Then, they were asked a further open-ended question about whether/why certain items were problematic in measuring FLLB.

Experts' quantitative ratings on item relevance and clarity were both high (M_s > 7.30). However, they raised concerns in the open-ended questions about Items 2 and 5, particularly in terms of their conceptual inconsistency. The two items were deleted after a group discussion among authors. In the end, 11 items were maintained in the final *FLLBS–SF*, representing three factors (Table 3).

The three identified factors explained 56.93%, 11.84%, and 10.31% of the variance, respectively, and 79.08% in total. The three-factor structure represents FLLB as a multilevel and mixed emotional experience construed in language activities/tasks at the micro level, in language classrooms at the meso level, and in general learning at the macro level. Factor 1 is named Foreign Language Activity Boredom because the items concern boredom experienced in specific FL activities both in and out of class (e.g., English exercises). Factor 2 is named Foreign Language Classroom Boredom, and the items reflect boredom experienced in FL classrooms in general. Factor 3 is named General Learning Boredom because such boredom, although arising in L2 learning contexts, is shared in general learning as an emotional proneness ready to be brought to any subject.

Table 3. Factor Loadings of the Final *FLLBS–SF*

	Foreign language activity boredom	Foreign language classroom boredom	General learning boredom
1. The English class bores me.		.844	
3. My mind begins to wander in the English class.		.882	
6. Time is dragging on in English class.		.838	
8. I always think about what else I might be doing to kill the time rather than sitting in this English class.		.861	
9. An analysis of a long (English) text is really dreary.	.786		
10. It is really boring to repeat the (English) text after the modeling audio.	.857		
11. So many similar types of (English) exercises make me lose interest.	.948		
12. So much practice on the same (English–related) subject matter makes me restless.	.957		
26. I'm somebody who is not interested in studying.			.917
27. Studying (English) is dull in general.			.869
29. I'm forced to learn all the subjects, including English.			.893

Confirming the structure of FLLBS–SF

CFA was conducted to confirm the three-factor structure identified in EFAs, using the data of Subsample 2 via Mplus 8.0 (n = 3,341). The model fit indices and criteria were the root mean square error of approximation (RMSEA < .08), the standardized root mean square residual (SRMR < .08), the comparative fit index (CFI > .95), and the Tucker–Lewis index (TLI > .95) (Hu & Bentler, 1999). Additionally, standardized factor loadings higher than .70 were considered to be ideal (Hair et al., 2019), a higher standard than Li, Dewaele et al. (2023) (>. 50).

The results show good model fits (χ^2/df = 517.768/41, CFI = .984, TLI = .979, SRMR = .020, RMSEA [90% CI] = .059 [.055,.064]) and acceptable factor loadings (see Figure 6). We thus confirmed the three-factor structure representing the 11-item *FLLBS–SF*.
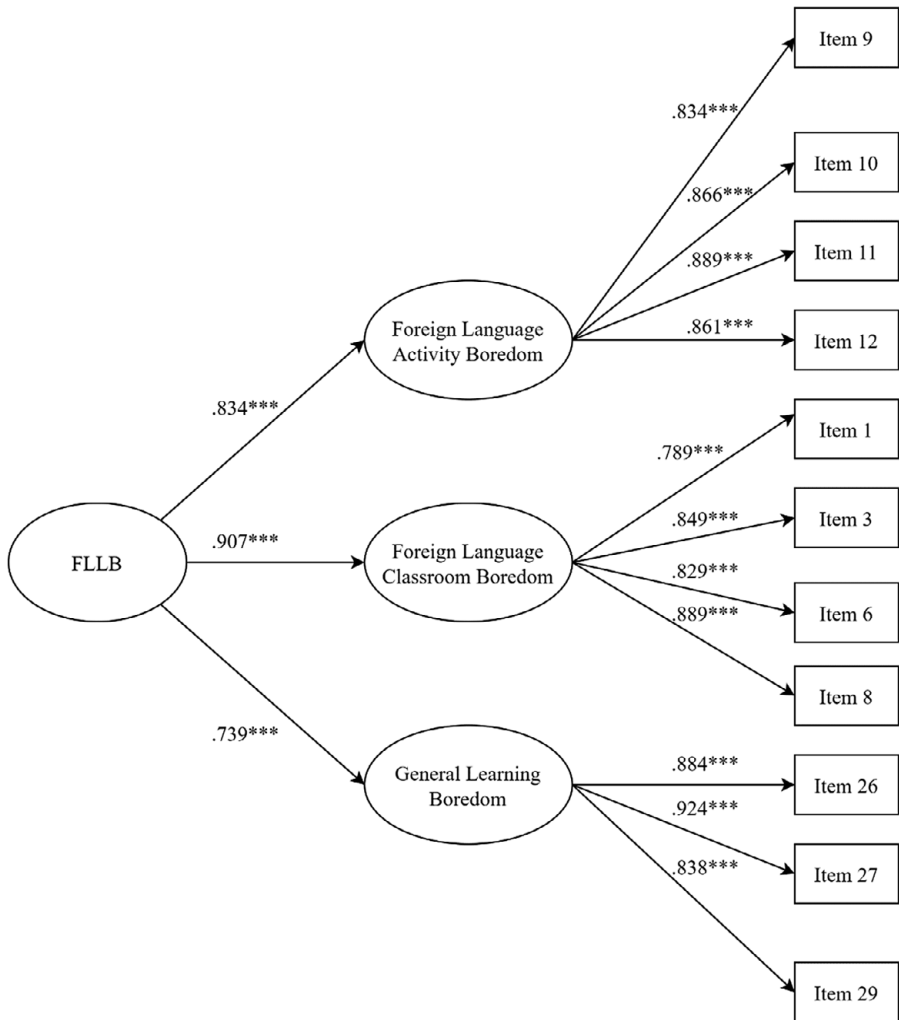

Figure 6. CFA results of the *FLLBS–SF*.
Note. All are standardized coefficients, ***p <. 001.

Phase 3: Validating the *FLLBS-SF*

The newly developed 11-item *FLLBS–SF* was subjected to a series of psychometric validations, including reliability assessment, validity evaluation, and measurement invariance tests.

Reliability assessment

The reliability of *FLLBS–SF* was evaluated by assessing its internal consistency and test–retest reliability. We calculated Cronbach's α, McDonald's ω, and Guttman's split-half coefficient as indicators of internal consistency using Datasets 1, 2, and 3 (> .70; Hair et al., 2019). The test–retest reliability was evaluated with two-way random effects intraclass correlation coefficients (ICC) based on Dataset 4: poor reliability (ICC < .40), moderate reliability (.40 ≤ ICC < .60), good reliability (.60 ≤ ICC < .75), and excellent reliability (ICC ≥. 75) (Cicchetti, 1994).

The results show excellent internal consistency for the overall scale (α = .94, ω = .93), and three subscales (Foreign Language Activity Boredom: α = .92, ω = .92; Foreign Language Classroom Boredom: α = .90, ω = .91; General Learning Boredom: α = .91, ω = .91). The Guttman's split-half reliability of the *FLLBS–SF* was also good (.87). Its test–retest reliability was acceptable (ICC$_{T1–T2}$ = .44, ICC$_{T2–T3}$ = .46). Taking together, the *FLLBS–SF* demonstrated satisfactory reliability.

Validity assessment

We further assessed convergent validity, discriminant validity, criterion validity, and predictive validity of the *FLLBS–SF*. Convergent validity was first evaluated to reflect whether items share a high proportion of variance in common (Hair et al., 2019). Specifically, we calculated (1) the correlations between the scores on the *FLLBS–SF* and the original *FLLBS* and (2) the average variance extracted (*AVE*) and composite reliability (*CR*) based on Datasets 1, 2, and 3.

The large positive correlation between the scores on *FLLBS–SF* and *FLLBS* (r = .97, p < .001) indicated good convergent validity (Hair et al., 2019). The high values of *AVE*s (>. 50) and *CR*s (>. 70) of the three factors also indicated good convergent validity (Hair et al., 2019).

The discriminant validity was then assessed using Datasets 1, 2, and 3, to decide the extent to which the factors of FLLB are distinct from each other. Specifically, discriminant validity is established if all factor *AVE*s are greater than the squared correlation estimates (r^2) between any pair of factors (Hair et al., 2019). The comparison revealed that the *AVE* of each subscale was larger than the r^2 between any two subscales (Table 4), indicating good discriminant validity.

Criterion validity was examined to assess the correlation between FLLB and non-incidentally but theoretically related constructs (Devellis, 2016). Specifically, we

Table 4. Convergent Validity Indicators of *FLLBS–SF*

			r^2		
Factor	AVE	CR	1	2	3
1. Foreign language activity boredom	.744	.921	–		
2. Foreign language classroom boredom	.705	.905	.468***	–	
3. General learning boredom	.779	.914	.323***	.366***	–

*Note. AVE = average variance extracted, CR = composite reliability; r^2 = Coefficients of determination; ***p <. 001.*

correlated scores on the *FLLBS–SF* with a series of criteria variables, including foreign language enjoyment (Datasets 1, 2, and 3), anxiety (Datasets 1 and 3), academic boredom in English (Dataset 1), and boredom susceptibility (Dataset 1). The results indicate that FLLB measured by *FLLBS–SF* was positively correlated with foreign language anxiety ($r = .36$, $p < .001$), academic boredom in English ($r = .76$, $p < .001$), and boredom susceptibility ($r = .36$, $p < .001$), while negatively with foreign language enjoyment ($r = -.39$, $p < .001$). Taken together, the *FLLBS–SF* had good criterion validity.

Finally, the predictive validity concerns the extent to which the *FLLBS–SF* can predict learning outcomes. We examined the associations between *FLLBS–SF* scores and English achievement based on the merged Datasets 2 and 3 ($n = 1,438$). Linear regression results show that FLLBS significantly and negatively predicted L2 achievement ($\beta = -.48$, $p < .001$), suggesting that the *FLLBS–SF* had excellent predictive validity.

Invariance assessment

Cross-sectional invariance and longitudinal measurement invariance were assessed by employing multigroup CFAs in Mplus. Cross-sectional invariance tests were conducted across groups of gender [Male vs. Female], age [Adolescent: age < 18 vs. Adult: age ≥ 18], region [Rural vs. Urban], educational level [Secondary vs. Tertiary], and language status [English as L2 vs. L3] based on Datasets 1, 2, and 3. Dataset 4 was used for the longitudinal invariance test (i.e., time invariance).

To assess measurement invariance, we specified a set of increasingly stringent models, starting from the configural model (assuming equivalence of factor structures), the metric model (assuming equivalence of factor loadings), and finally, the scalar model (assuming equivalence of item intercepts). Then, we compared the more stringent model with the less stringent one to detect changes in model fits. Invariance is supported when $\Delta CFI \le -.010$, $\Delta RMSEA \le .015$, and $\Delta SRMR \le .030$ (for metric invariance), or $\Delta SRMR \le .015$ (for scalar invariance) (Chen, 2007).

The results of the invariance test are summarized in Table 5. As displayed, no significant difference was detected between a series of increasingly strict CFA models, indicating that the *FLLBS–SF* was interpreted and responded to invariantly across distinctive gender, age, regional, educational, and language status groups and three-time points spanning over a year.

Study 2: The links between FLLB and overall/skill-specific L2 achievement and proficiency

Study 2 was designed to reveal the predictive effect of FLLB on overall/skill-specific L2 achievement (vocabulary and grammar, listening, reading, and writing) and proficiency (listening, reading, and writing).

Data analysis and results

Structural equation modeling (SEM) was adopted to examine the predictive effects of boredom. It involved two steps: (1) Estimating the measurement model to assess the measurement validity of indicators and (2) Fitting the structural model to reveal the relationships among variables of interest. Data used in the SEM was the cross-sectional part of Dataset 4 (T2, $n = 1,109$). Of note was that data for boredom was based on scores on the *FLLBS–SF*.

Table 5. Measurement Invariance of *FLLBS–SF*

Invariance Model	χ^2	df	CFI	TLI	RMSEA	SRMR	ΔCFI	ΔRMSEA	ΔSRMR
Gender									
Configural	538.276	82	.984	.978	.060	.022			
Metric	552.447	90	.984	.980	.057	.024	.000	−.003	.002
Scalar	580.957	98	.983	.981	.056	.024	−.001	−.001	.000
Age									
Configural	613.616	82	.984	.979	.060	.024			
Metric	673.258	90	.983	.979	.060	.028	−.001	.000	.004
Scalar	926.656	98	.976	.973	.068	.029	−.007	.008	.001
Educational level									
Configural	657.258	82	.984	.978	.062	.024			
Metric	721.063	90	.982	.978	.062	.028	−.002	.000	.004
Scalar	1016.483	98	.974	.971	.072	.030	−.008	.010	.002
Language status									
Configural	642.952	82	.982	.975	.061	.024			
Metric	667.960	90	.981	.977	.059	.025	−.001	−.002	.001
Scalar	755.613	98	.979	.976	.061	.029	−.002	.002	.004
Region									
Configural	656.928	82	.983	.978	.062	.023			
Metric	683.949	90	.983	.979	.060	.024	.000	−.002	.001
Scalar	911.955	98	.976	.974	.067	.028	−.007	.007	.004
Time									
Configural	886.201	123	.979	.972	.081	.016			
Metric	934.730	149	.978	.974	.077	.026	−.001	−.004	.010
Scalar	1078.035	155	.975	.973	.079	.026	−.003	.002	.000

Note. All *p* <. 001.

The measurement model fits were good for overall L2 achievement model ($\chi^2/df =$ 296.677/51, CFI = .976, TLI = .969, SRMR = .019, RMSEA [90% CI] = .069 [.062, .077]) and skill-specific L2 achievement model ($\chi^2/df =$ 327.028/81, CFI = .982, TLI = .976, SRMR = .019, RMSEA [90% CI] = .054 [.048, .060]) (Hu & Bentler, 1999), hence allowing for the assessment of structural model. Results of the structural model show that the overall L2 achievement model was a good fit ($\chi^2/df =$ 296.677/51, CFI = .976, TLI = .969, SRMR = .019, RMSEA [90% CI] = .069 [.062, .077]), and so was the skill-specific L2 achievement model ($\chi^2/df =$ 327.028/81, CFI = .982, TLI = .976, SRMR = .019, RMSEA [90% CI] = .054 [.048, .060]). The results (Table 6) indicate that FLLB significantly (all p_s < .001) and negatively predicted learners' overall L2 achievement ($\beta = -.19$), vocabulary and grammar achievement ($\beta = -.14$), listening achievement ($\beta = -.12$), reading achievement ($\beta = -.16$) and writing achievement ($\beta = -.13$), with modest effect sizes (Cohen et al., 2017).

Results also indicate that the fit indices were satisfactory for the measurement model of the overall L2 proficiency model ($\chi^2/df =$ 293.637/51, CFI = .976, TLI = .970, SRMR = .019, RMSEA [90% CI] = .068 [.060, .075]), and skill-specific L2 proficiency model ($\chi^2/df =$ 306.810/71, CFI = .979, TLI = .973, SRMR = .019, RMSEA [90% CI] = .057 [.050,. 063]). The structural model demonstrated satisfactory fits for overall L2 proficiency model ($\chi^2/df =$ 293.637/51, CFI = .976, TLI = .970, SRMR = .019, RMSEA [90% CI] = .068 [.060, .075]) and skill-specific L2 proficiency model ($\chi^2/df =$ 306.810/71, CFI = .979, TLI = .973, SRMR = .019, RMSEA [90% CI] = .057 [.050, .063]). As indicated in Table 6, FLLB significantly (all p_s < .001) and negatively predicted overall L2 proficiency ($\beta = -.23$), listening proficiency ($\beta = -.19$), reading proficiency ($\beta = -.19$), and writing proficiency ($\beta = -.20$), with modest effect sizes (Cohen et al., 2017).

Table 6. The Predictive Effects of FLLB on Overall/Skill-Specific L2 Achievement and Proficiency

Dependent variable		β	SE	R^2
Overall				
	achievement	−.19	.03	.04
	proficiency	−.23	.03	.05
Listening				
	achievement	−.12	.03	.02
	proficiency	−.19	.03	.03
Reading				
	achievement	−.16	.03	.02
	proficiency	−.19	.03	.04
Writing				
	achievement	−.13	.03	.02
	proficiency	−.20	.03	.04
Vocabulary and grammar				
	achievement	−.14	.03	.02

Note. All p <. 001.

Study 3: The longitudinal relationships between foreign language learning boredom and achievement

Study 3 sought to reveal the longitudinal relationships between FLLB and L2 achievement.

Data analysis

Their associations were examined with cross-lagged panel modeling using Dataset 4 (T1–T3, n = 1,109) following the two steps for SEM in Study 2. The model fit was assessed based on the same recommended cutoff values in the CFA of Studies 1 and 2 (Hu & Bentler, 1999).

Results

The measurement model of the cross-lagged panel model was good (χ^2/df = 1777.487 /573, CFI = .970, TLI = .967, SRMR = .034, RMSEA [90% CI] = .044 [.041,. 046]), allowing for the structural modeling. The fit of structural model was also satisfactory (χ^2/df = 1851.368/577, CFI = .968, TLI = .965, SRMR = .063, RMSEA [90% CI] = .045 [.043, .047]) (see Figure 7).

As the figure indicates, FLLB and L2 achievement were negatively correlated at Time Point 1 (r = −.20, p < .001), but their correlations diminished to nonsignificant at Time Point 2 (r = −.07), and the effect size further decreased at Time Point 3 (r = −.05). Moreover, FLLB at Time Points 1 and 2 did not predict L2 achievement at Time Points 2 and 3, respectively. By contrast, English achievement consistently and negatively predicted subsequent FLLB (both β_s = −.12, p < .001) with modest effect sizes (Cohen et al., 2017).

Discussion

RQ 1 concerned the development and validation of the *Foreign Language Learning Boredom Scale–Short Form* (*FLLBS–SF*). We reduced the original 32-item *FLLBS* to a
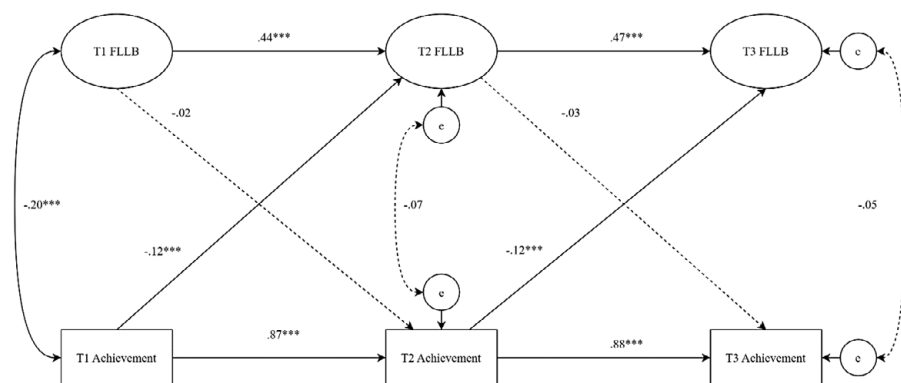
Figure 7. The longitudinal relationships between FLLB and EFL achievement.
Note. ***p < .001, all are standardized coefficients, and the dotted lines indicate nonsignificant results.

parsimonious 11-item *FLLBS–SF* measuring three factors, namely Foreign Language Activity Boredom, Foreign Language Classroom Boredom, and General Learning Boredom. The three-factor structure not only echoes the multidimensional nature of FLLB revealed in prior studies (e.g., Li, Dewaele et al., 2023; Pawlak et al., 2020), but also reflects its multilevel framework from language activities/tasks (in- and out-of-class) at the micro level, to language classrooms at the meso level, and to learning in general at the macro level. Informed by the items and the factor structure of the final *FLLBS–SF*, we further define FLLB as an individual's proneness or inclination to feel bored in relation to L2 learning in general, which is relatively stable (although evolving) and persists over a relatively long period as a combination of boredom accumulated in foreign language classrooms, (in- and out-of-class) foreign language learning activities and general learning.

Regarding the psychometric properties of the *FLLBS–SF*, we conducted a series of rigorous validation and assessment using a large, merged dataset with participants of diverse L2 backgrounds. The *FLLBS–SF* was found to have satisfactory (1) validity (i.e., face validity, construct validity, convergent/discriminant validity, criterion validity, and predictive validity), (2) reliability (i.e., internal consistency, test–retest reliability, and split–half reliability) and (3) measurement invariance (across time and across multiple heterogenous groups including age, gender, region, educational level, and language status). To sum up, the *FLLBS–SF* is a conceptually sound, psychometrically robust, stable, and parsimonious instrument to measure L2 learners' general and long-term (trait-like) boredom in relation to L2 learning.

Results for RQ2 show that FLLB had consistent negative and modest predictive effects on overall L2 achievement/proficiency and skill-specific L2 achievement/proficiency, supporting our first and second hypotheses. In other words, general FLLB was found to be a negative predictor of academic success in L2 learning across measures (curriculum-based course exams and the proficiency test) and skills or subsystems (i.e., vocabulary and grammar achievement, listening, reading, and writing). The results indicate that participants with higher levels of boredom in English learning tended to have lower scores in overall end-term English exams and the global English proficiency test, and their subsections on different skills. Resultantly, our findings extend the argument of the skill-specificity of L2 learner emotions (Li et al., 2023; Li et al., 2024) by highlighting the skill generality simultaneously. That is, although L2 learners'

emotional experiences (e.g., intensity, frequency, and causes) may vary from skill to skill, the boredom–achievement/proficiency link is a consistent pattern across skills.

The FLLB–overall L2 achievement link obtained among Chinese secondary EFL students in the current study dovetail with prior significant results obtained from EFL learners from China, Turkey, and Morocco (e.g., Dewaele, Botes, & Meftah, 2023; Li & Li, 2023; Özsaray & Eren, 2018; Zhao & Wang, 2023). The significant links differ from the nonsignificant correlations obtained from international students (Dewaele, Botes, & Greiff, 2023). The difference might be attributed to the variations in the measures of boredom and groups of participants.

Our findings also extend the literature on the role of FLLB by including overall L2 proficiency and skill-specific L2 proficiency/achievement as the outcome variables of FLLB. As reviewed, although there were no directly relevant theoretical assumptions or prior empirical evidence for the role of emotions in affecting L2 proficiency and skill-specific L2 proficiency/achievement, their links could be explained by drawing on relevant theories and literature (e.g., the control–value theory, Pekrun, 2006). FLLB, a negative achievement emotion with a low activation and an activity-related focus, is linked to a frequent lack of interest in ongoing L2 learning activities, constant demotivation, recurrent disengagement or pseudo-/superficial/procedural engagement in L2 learning, habitual distractions, short attention span, superficial information processing in L2 task completion, less self-regulated L2 learning, and fewer motivated learning behaviors (Li, Dewaele et al., 2023; Li et al., 2024, Li & Li, 2024; Pawlak et al., 2020; Pekrun, 2006). These adverse effects may further converge and impede L2 achievement/proficiency. The findings also show that the predictive effects of the general FLLB on L2 achievement/proficiency were both skill-general and skill-specific. However, boredom itself is skill-specific in terms of the way it is organized and instigated or the way it influences various learning outcomes (Li, Dewaele et al., 2023; Li & Li, 2024).

RQ3 was concerned with the association between FLLB and L2 achievement across the three time points that spanned over a year. Our results only partially support the third hypothesis that FLLB and English achievement would reciprocally and negatively predict each other over time. First, the cross-lagged panel modeling results showed cross-sectional links between FLLB and L2 achievement at three different time points. Specifically, FLLB and L2 achievement were significantly negatively correlated at Time Point 1. However, this relationship was no longer significant at Time Point 2 and was weakened further by Time Point 3. This counterintuitive finding suggests that the effects of learners' emotions fluctuate and may interact with each other and with motivational constructs over time (Dewaele & Meftah, 2023). The finding that FLLB no longer predicted L2 achievement at Time 2 and 3 could mean that more positive emotions and attitudes had neutralized its effect as the year progressed.

Secondly, the cross-lagged panel modeling results also show that L2 achievement consistently predicted subsequent FLLB negatively. This result extends prior cross-sectional studies on their associations (e.g., Li, Dewaele et al., 2023) by specifying the effect of directionality over time. Cross-sectional studies typically assume that the effect goes from FLLB to achievement; however, such a design cannot decide the accurate directionality. Our study adopted a longitudinal design, revealing that the impact went from L2 achievement to FLLB rather than the opposite. The current findings dovetail with Alamer and Lee's (2021) findings that the effect went from L2 achievement to anxiety rather than the reverse. The unidirectional negative impact of L2 achievement on FLLB might result from a change in the emotional dynamics. Inspired by the control–value theory (Pekrun, 2006), previous learning experience is

an essential environmental factor that precedes emotions. Good scores in L2 exams may have initiated a positive feedback loop, resulting in reduced FLLB and vice versa. That is, in response to prior language achievement indicated as high course exam scores, learners may also have felt increased subjective control over the learning process, which further alleviates boredom and vice versa.

Another interesting finding was that FLLB did not predict subsequent L2 achievement across time. Similar patterns emerged in the study by Li and Li (2023). The local educational context may help to explain this pattern. Specifically, our participants in Substudy 3 were all Chinese secondary students in a competitive educational context where the government, the educational sector, school administrators and instructors, and parents generally place a high value on education and the subject of English of interest. English has been emphasized as a crucial asset for success and personal development (e.g., university admission and job opportunities) (Hu, 2005). The participants' extrinsic motivation to achieve higher in English is expected to grow as their upgrade from middle school to high school gets closer (temporally in parallel with our research project). We could thus expect that the debilitative effect of perceived boredom on L2 achievement was declining as a result of the increasing extrinsic motivation to achieve higher in English. In other words, participants may have decided that despite their FLLB, they would work hard to get good results, which would explain why FLLB did not predict any subsequent L2 achievement. In addition, the strong autoregressive stability among FLLB across time (i.e., the effect from FLLB at a prior time point to FLLB at a subsequent time point) could have reduced the power of FLLB in explaining the variance of L2 achievement (Adachi & Willoughby, 2015).

Limitations, Implications, and Suggestions for Future Research

The study has the following limitations. First, our participants were all Chinese EFL learners, which limits the generalizability of our findings. Future studies could cross-validate the *FLLBS–SF* in more heterogeneous L2 contexts. Second, since English proficiency was only collected at one time point, it was not possible to establish causality between FLLB and English proficiency. Future studies could adopt a longitudinal design to explore their causal relationships and their developmental trajectories (Alamer & Lee, 2021; Kenny, 1979).

Some pedagogical implications can be drawn from this study. Firstly, L2 educators should raise awareness about the importance of L2 learner emotions and help learners boost positive emotions and avoid or reduce FLLB. It is worth noting that how students feel matters not only for their well-being but also for their overall/skill-specific L2 performance. That is, an intervention on FLLB is not only an intervention on students' emotional well-being in L2 contexts but also could be an indirect intervention to boost their L2 achievement/proficiency. Secondly, the factors identified in Study 1 have some clear pedagogical implications for L2 teachers. For example, they could try to reduce boredom by diversifying task/exercise format, avoiding excessive repetition in tasks, reducing the length of the same task (see the items for Foreign Language Activity Boredom), establishing a positive L2 classroom atmosphere (see the items for Foreign Language Classroom Boredom), and collaborating with teachers for other subjects to improve students' emotional experiences in learning in general (see the items for General Learning Boredom). The other implication for L2 educators is the need to

identify, remove, or loosen the effect from prior poor L2 achievement to subsequent higher levels of boredom.

Conclusion

The current study has corroborated, extended, and enriched the control–value theory in the L2 context. FLLB was found to predict L2 achievement/proficiency in a skill-general way. The skill generality should be highlighted in the control–value theory framework, in parallel with the skill (subsystem) specificity (Li, Dewaele et al., 2023; Li & Li, 2024). This underlines the need to include both curriculum-based language course exams and global proficiency tests as measures for academic achievement in L2, which is the core dependent variable in the control–value theory framework and in L2 learner emotion research. In addition, our findings specified the directionality between FLLB and L2 achievement over time. L2 achievement was found to negatively influence FLLB unidirectionally, which contradicts the reciprocity claimed by the control–value theory (Pekrun, 2006). Moreover, our findings resonate with the waxing and waning of the effects of FLLB—and other emotions—highlighted in Li and Li (2023). To conclude, the new 11-item *FLLBS–SF* was successfully developed and validated. It has sound psychometric properties and robust measurement invariance, and we recommend it for future research.

Author contribution. **Chengchen Li**: Conceptualization; Research Design; Data Collection; Data Process and Analyses; Writing–Abstract, Introduction, Literature Review, Motivation and Rationale, Discussion, and Implications; Revision & Editing; Resources & Project Administration; Funding Acquisition. **Enhao Feng:** Statistical Analyses; Writing– Methodology and Results; Drafting: Discussion; Revision. **Xian Zhao**: Drafting –Literature Review. **Jean-Marc Dewaele**: Drafting – Conclusion; Revision & Editing.

References

Adachi, P., & Willoughby, T. (2015). Interpreting effect sizes when controlling for stability effects in longitudinal autoregressive models: Implications for psychological science. *European Journal of Developmental Psychology*, *12*(1), 116–128. https://doi.org/10.1080/17405629.2014.963549

Alamer, A., & Lee, J. (2021). Language achievement predicts anxiety and not the other way around: A cross-lagged panel analysis approach. *Language Teaching Research*. https://doi.org/10.1177/13621688211033694

Apridayani, A., & Waluyo, B. (2022). Antecedents and effects of students' enjoyment and boredom in synchronous online English courses. *Journal of Multilingual and Multicultural Development*, 1–16. https://doi.org/10.1080/01434632.2022.2152457

Awang, Z. (2012). *A handbook on SEM structural equation modelling: SEM using AMOS graphic* (5th ed.). Universiti Teknologi Mara Kelantan.

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, *6*, 149. https://doi.org/10.3389/fpubh.2018.00149

Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, *71*(3), 425–440. https://doi.org/10.1007/s11336-006-1447-6

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modelling*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Chen, W., Sun, P., & Yang, Z. (2022). Understanding Chinese second language learners' foreign language learning boredom in online classes: Its conceptual structure and sources. *Journal of Multilingual and Multicultural Development*, 1–17. https://doi.org/10.1080/01434632.2022.2093887

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*(3), 309–319. https://doi.org/10.1037/14805-012

Cohen, L., Manion, L., & Morrison, K. (2017). *Research methods in education* (8th ed.). Routledge.

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, *10*(1), 1–9. https://doi.org/10.7275/jyj1-4868

Devellis, R. (2016). *Scale development: Theory and applications* (4th ed.). Sage Publications.

Dewaele, J. M., Albakistani, A., & Kamal Ahmed, I. (2024). Levels of foreign language enjoyment, anxiety and boredom in emergency remote teaching and in in-person classes. *The Language Learning Journal*, *52*(1), 117–130. https://doi.org/10.1080/09571736.2022.2110607

Dewaele, J.-M., Botes, E., & Greiff, S. (2023). Sources and effects of foreign language enjoyment, anxiety, and boredom: A structural equation modelling approach. *Studies in Second Language Acquisition*, *45*(2), 461–479. https://doi.org/10.1017/S0272263122000328

Dewaele, J.-M., Botes, E., & Meftah, R. (2023). A three-body problem: The effects of foreign language anxiety, enjoyment, and boredom on academic achievement. *Annual Review of Applied Linguistics*, *43*, 7–22. https://doi.org/10.1017/S0267190523000016

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*(2), 237–274. https://doi.org/10.14746/ssllt.2014.4.2.5

Dewaele, J.-M., MacIntyre, P. D., Albakistani, A., & Kamal Ahmed, I. (2023). Emotional, attitudinal and sociobiographical sources of flow in online and in-person EFL classrooms. *Applied Linguistics*. https://doi.org/10.1093/applin/amad071

Dewaele, J.-M., & Meftah, R. (2023). The trajectory of English foreign learners' emotions and motivation from the start to the end of their learning journey: A pseudo-longitudinal investigation. *Journal of the European Second Language Association*, *7*(1), 16–30. https://doi.org/10.22599/jesla.94

Dörnyei, Z., & Dewaele, J.-M. (2022). *Questionnaires in second language research: Construction, administration, and processing*. Routledge. https://doi.org/10.4324/9781003331926

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (5th ed.). Sage Publications.

Fowler, F. J. (2014). *Survey research methods (*5th edition). SAGE.

Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion* Quarterly,*73*(2), 349–360. https://doi.org/10.1093/poq/nfp031

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage.

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organisational Research Methods*, *7*(2), 191–205. https://doi.org/10.1177/1094428104263675

Hu, G. (2005). Contextual influences on instructional practices: A Chinese case for an ecological approach to ELT. *TESOL Quarterly*, *39*(4), 635–660.

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

Isbell, D. R., & Son, Y.-A. (2022). Measurement properties of a standardized elicited imitation test: An integrative data analysis. *Studies in Second Language Acquisition*, *44*(3), 859–885. https://doi.org/10.1017/S0272263121000383

Jin, Y., Jie, W., & Wang, W. (2022). Exploring the alignment between the College English Test and language standards. *Foreign Language World* (《外语界》), *2*, 24–32.

Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology*, *30*(1), 17–24. https://doi.org/10.1037/h0057123

Kenny, D. A. (1979). *Correlation and causality*. Wiley.

Koo, T. K., & Li, M. Y. (2000). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic* Medicine,*15*(2): 155–163.

Kruk, M., & Zawodniak, J. (2017). Nuda a praktyczna nauka języka angielskiego. *Neofilolog*, *49*(1), 115–131. https://doi.org/10.14746/n.2017.49.1.07

Li, C. (2021). A control-value theory approach to boredom in English classes among university students in China. *The Modern Language Journal*, *105*(1), 317–334. https://doi.org/10.1111/modl.12693

Li, C. (2022). Foreign language learning boredom and enjoyment: The effects of learner variables and teacher variables. *Language Teaching Research*, https://doi.org/10.1177/13621688221090324

Li, C., & Dewaele, J. -M. (2020). Thepredictive effects of trait emotional intelligence and online learning achievement perceptions on foreign language class boredom among Chinese university students. *Foreign Languages and Foreign Language Teaching 5*, 33–44. https://doi.org/10.13458/j.cnki.flatt.004711

Li, C., Dewaele, J.-M., & Hu, Y. (2023). Foreign language learning boredom: Conceptualization and measurement. *Applied Linguistics Review*, *14*(2), 223–249. https://doi.org/10.1515/applirev-2020-0124

Li, C., & Han, Y. (2022). The predictive effects of foreign language anxiety, enjoyment, and boredom on learning outcomes in online English classrooms. *Modern Foreign Languages* (《现代外语》), *45*(2), 207–219.

Li, C., Jiang, G., & Dewaele, J. M. (2018). Understanding Chinese high school students' foreign language enjoyment: Validation of the Chinese version of the foreign language enjoyment scale. *System*, *76*, 183–196. https://doi.org/10.1016/j.system.2018.06.004

Li, C., & Li, W. (2023). Anxiety, enjoyment, and boredom in language learning amongst junior secondary students in rural China: How do they contribute to L2 achievement? *Studies in Second Language* Acquisition,*45*(1), 93–108. https://doi.org/10.1017/S0272263122000031

Li, C., & Li, W. (2024). The relationships of language attitude, emotions, and foreign language achievement: Urban-rural comparisons based on structural equation modelling. *Foreign Languages and Their Teaching* (《外语与外语教学》), *1*(1), 57–69. https://doi.org/10.13458/j.cnki.flatt.004711

Li, C., & Li, W. (2024). The joint contributions of cognitive abilities and emotions to second language writing proficiency. *Modern Foreign Languages* (《现代外语》). *4*, 465–477.

Li, C., Li, W., & Jiang, G. (2024). Emotions in second language learning: Retrospect and prospect. *Modern Foreign Languages* (《现代外语》), *1*, 63–75.

Li, C., Li., W., & Lu, X. (2023). Contributions of foreign language writing emotions to writing achievement. *System*, *116*, 103074. https://doi.org/10.1016/j.system.2023.103074

Liu, E., & Wang, J. (2023). The effects of student and teacher variables on anxiety, enjoyment, and boredom among Chinese high school EFL learners. *International Journal of Multilingualism*, 1–22. https://doi.org/10.1080/14790718.2023.2177653

Loiacono, E. T., Watson, R. T., & Goodhue, D. L. (2002). WebQual: A measure of website quality. *Marketing Theory and Applications*, *13*(3), 432–438.

Ma, L., Liu, L., Luo, S., & Liu, J. (2023). Exploring urban-rural differences in foreign language enjoyment, anxiety, boredom, and burnout among Chinese secondary students. *Journal of Multilingual and Multicultural Development*, 1–19. https://doi.org/10.1080/01434632.2023.2298691

Marsh, H. W., Ellis, L. A., Parada, R. H., Richards, G., & Heubeck, B. G. (2005). A short version of the self description questionnaire II: Operationalizing criteria for short-form evaluation with new applications of confirmatory factor analyses. *Psychological Assessment*, *17*(1), 81–102. https://doi.org/10.1037/1040-3590.17.1.81

Ministry of Education of the People's Republic of China. (2022). *Compulsory education curriculum programs and standards* (2022 version). https://www.gov.cn/zhengce/zhengceku/2022-04/21/content_5686535.htm

Özsaray, A. E., & Eren, A. (2018). Achievement emotions, epistemic curiosity, and graded performance of undergraduate students in English preparatory classes. *International Journal of Curriculum and Instructional Studies (IJOCIS)*, *8*(1), 39–58. https://doi.org/10.31704/ijocis.2018.003

Pawlak, M., Kruk, M., Zawodniak, J., & Pasikowski, S. (2020). Investigating factors responsible for boredom in English classes: The case of advanced learners. *System*, *91*, 102259. https://doi.org/10.1016/j.system.2020.102259

Pawlak, M., Kruk, M., Zawodniak, J., & Pasikowski, S. (2022). Examining the underlying structure of after-class boredom experienced by English majors. *System*, *106*, 102769. https://doi.org/10.1016/j.system.2022.102769

Pawlak, M., Zawodniak, J., & Kruk, M. (2023). Investigating changes in after class-boredom over time: Insights from a longitudinal study of English majors in Poland. In Li, C. & Li, W. (Eds). *Emotions in foreign language learning: Developments in theory, research, and practice in honor of Jean-Marc Dewaele*. Multilingual Matters.

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. https://doi.org/10.1007/s10648-006-9029-9

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(1), 36–48. https://doi.org/10.1016/j.cedpsych.2010.10.002

Reynolds, C. R., Altmann, R., & Allen, D. N. (2021). *Mastering modern psychological testing: Theory and methods* (2nd ed.). Springer. https://doi.org/10.1007/978-3-030-59455-8_1

Rolstad, S., Adler, J., & Rydén, A. (2011). Response burden and questionnaire length: is shorter better? A review and meta-analysis. *Value in Health*, 14(8), 1101–1108. https://doi.org/10.1016/j.jval.2011.06.003

Tremblay, A. (2011). Proficiency assessment standards in second language acquisition research: "Clozing" the gap. *Studies in Second Language Acquisition*, 33(3), 339–372. https://doi.org/10.1017/S0272263111000015

Tsang, A., & Dewaele, J. (2023). The relationships between young FL learners' classroom emotions (anxiety, boredom, & enjoyment), engagement, and FL proficiency. *Applied Linguistics Review*. https://doi.org/10.1515/applirev-2022-0077

Wang, X., & Li, Y. (2022). The predictive effects of foreign language enjoyment, anxiety, and boredom on general and domain-specific English achievement in online English classrooms. *Frontiers in Psychology*, 13, 1050226. https://doi.org/10.3389/fpsyg.2022.1050226

Wang, H., Wang, Y., & Li, S. (2023). Unpacking the relationships between emotions and achievement of EFL learners in China: Engagement as a mediator. *Frontiers in Psychology*, 14, 1098916. https://doi.org/10.3389/fpsyg.2023.1098916

Yeung, S. S., Tsang, A., Lam, W. W. C., & Law, T. S. T. (2023). Maternal education, classroom emotions, and literacy outcomes among grade 3/4 EFL learners. *Journal of Multilingual and Multicultural Development*, 1–15. https://doi.org/10.1080/01434632.2023.2187397

Zhao, X., Lan, G., & Chen, T. (2023). Motivational intensity and self-perceived Chinese language proficiency: A moderated mediation model of L2 enjoyment and boredom. *Language Teaching Research*, https://doi.org/10.1177/13621688231180465

Zhao, X., & Wang, D. (2023). Grit, emotions, and their effects on ethnic minority students' English language learning achievements: A structural equation modelling analysis. *System*, 113, 102979. https://doi.org/10.1016/j.system.2023.102979

Zhang, T. (2022). Unpacking boredom factors of Chinese foreign language major students in translation classes: A sequential mixed methods study. Frontiers in Psychology,13, 895223. https://doi.org/10.3389/fpsyg.2022.895223

Zuckerman, M. (1979). *Sensation seeking: Beyond the optimal level of arousal*. Hillsdale, N.J.: L. Erlbaum Associates. New York: distributed by the Halsted Press Division of Wiley. http://archive.org/details/sensationseeking0000zuck

Appendix

The Foreign Language Learning Boredom Scale–Short Form

To what extent do you agree with the following statements in relation to English learning?
  1= Strongly Disagree, 2 = Disagree, 3 = Undecided, 4 = Agree, 5 = Strongly Agree

	Chinese	English
No.	因子(分量表)和题项	Factors (Subscales) and Items
	因子**1:**外语课堂无聊	*Factor 1: Foreign Language Classroom Boredom*
1	英语课没什么意思。	The English class bores me.
2	上英语课我容易发呆。	My mind begins to wander in the English class.
3	英语课上时间过得好慢。	Time is dragging on in English class.
4	英语课上我总是想办法打发时间。	I always think about what else I might be doing to kill the time rather than sitting in this English class.
	因子**2:**外语活动无聊	*Factor 2: Foreign Language Activity Boredom*
5	长(英语)课文分析真的太枯燥了。	An analysis of a long (English) text is really dreary.
6	跟读(英语)课文好枯燥乏味。	It is really boring to repeat the (English) text after the modeling audio.
7	太多类似的(英语)练习好没意思。	So many similar types of (English) exercises make me lose interest.
8	对同一话题,太多重复的(英语)练习会让我很烦。	So much practice on the same (English–related) subject matter makes me restless.
	因子**3:**一般学习无聊	*Factor 3: General Learning Boredom*
9	我就是对学习不太感兴趣的人。	I'm somebody who is not interested in studying.
10	学习(不仅仅英语学习)是一件无聊的事情。	Studying (English) is dull in general.
11	我学习所有科目(包括英语)只是因为我不得不学。	I'm forced to learn all the subjects, including English.