Studies in Second Language Acquisition (2025), **47**, 488–511
doi:10.1017/S0272263125000142

RESEARCH ARTICLE

The relationship between boredom and second language achievement: A multilevel meta-analysis

Fangwei Huang[1] and Haijing Zhang[2]

[1]School of Chinese as a Second Language, Peking University, Beijing, China and [2]Department of Chinese Language Studies, The Education University of Hong Kong, Hong Kong, Hong Kong
Corresponding author: Haijing Zhang; Email: s1142529@s.eduhk.hk

(Received 04 April 2024; Revised 08 January 2025; Accepted 27 January 2025)

Abstract

There has been a growing emphasis on researching foreign language boredom in second language acquisition in recent years. However, existing research has yet to reach a consensus regarding the effect of foreign language boredom on learners' learning achievement. To address this gap, the present study employs multilevel meta-analysis to analyze 47 effect sizes from 33 empirical studies involving a total sample size of 27,838 participants. The findings reveal that foreign language boredom illustrates a small negative effect (r = -.24, p < .001) on language achievement. Furthermore, the moderation analysis reveals that the magnitude of the effect size varies crossing educational stages, achievement measurements, domain-specific language skills, foreign language boredom measurements, teaching modes, and learning contexts. This study provides robust evidence to support the detrimental role of foreign language boredom in language acquisition and identified substantive gaps in this research field, offering valuable directions for future research.

Keywords: foreign language boredom; language achievement; multilevel meta-analysis; moderation effect

Introduction

Integrating positive psychology into applied linguistics and language acquisition marked a significant shift in examining language learning emotions (MacIntyre & Mercer, 2014). Positive psychology holds a holistic view that values both positive and negative emotions, recognizing that managing negative emotions is essential for language learning (Li & Wei, 2023; Tsang & Lee, 2023). Extensive studies have implied the feasibility of addressing foreign language boredom (FLB), a negative emotion that hinders language achievement through reducing motivation and engagement, within the framework of positive psychology (Borgonovi et al., 2023; Dewaele et al., 2023; Dewaele et al., 2024). Notably, there has been a remarkable surge of interest in researching FLB in recent years, positioning it as a prominent focal point in foreign language emotion research (Li et al., 2023; Pawlak et al., 2020). Boredom constitutes a

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0272263125000142 Published online by Cambridge University Press

prevalent experience during FL learning and potentially detrimentally influences learners' learning behavior and academic achievement (Dewaele et al., 2023). Investigating how boredom affects FL achievement can provide valuable insights into the specific mechanisms where FLB impedes language learning progress, which is crucial for educators, informing the design of engaging and effective language learning environments.

Numerous studies have identified different effect sizes of FLB on language achievement (Dewaele et al., 2023; Pawlak et al., 2022), calling for synthesizing analysis across different individual studies. One method to deepen the understanding of this issue is to conduct a meta-analysis, an approach that aggregates data from multiple studies, providing a comprehensive overview and identifying patterns or inconsistencies that individual studies might miss (Borenstein et al., 2009). By systematically synthesizing existing research, meta-analysis enhances statistical power, increases the generalizability of findings, and helps uncover underlying factors influencing the relationship between FLB and language achievement (Jak, 2015). This study intends to explore the magnitude and direction of the effect of FLB on learning achievement, as well as potential moderating factors through multilevel meta-analysis. By critically reviewing relevant research in this domain, this study provides an empirical foundation for understanding and enhancing emotional experiences throughout the foreign language learning process while offering practical suggestions for future research endeavors.

Literature review

Foreign language boredom

Boredom is a multifaceted psychological experience encompassing feelings of disinterest and lack of excitement toward the current activity or environment, indicating a state of mental disengagement or resistance (Mikulas & Vodanovich, 1993). In foreign language learning, boredom can be a commonly encountered subjective experience characterized by a lack of interest, engagement, or stimulation for FL learning tasks and activities (Pawlak et al., 2020). This phenomenon, known as foreign language boredom (FLB), represents a negative emotional state that extends from conventional boredom and manifests as a diminished drive toward learning goals and a negative emotional experience while acquiring a foreign language (Pawlak et al., 2022). Positive psychology emphasizes enhancing positive emotions and reducing negative ones to improve the overall learning experience, which fosters a balanced emotional state essential for effective learning (Dewaele et al., 2023; MacIntyre & Mercer, 2014). FLB is found to correlate with levels of enjoyment and anxiety experienced during FL study (Dewaele et al., 2023; Botes et al., 2024). As evidenced by the research of Dewaele et al. (2023), FLB is intricately linked to teacher-related factors, exerting a significant influence on learners' FL attitudes and achievement due to its influence on motivation, engagement, and psychological states (Pawlak et al., 2023; Tsang & Dewaele, 2023). FLB led to decreased motivation, loss of attention, and lack of attention to foreign language learning, thus affecting the depth and breadth of learning and negatively predicting foreign language achievement (Li et al., 2023). Understanding and mitigating boredom is vital for improving learning experiences and outcomes, aligning with the goals of positive psychology to promote overall student well-being and academic achievement.

Measurements for assessing FLB consist of laboratory-based physiological indices and self-report measurements. The former employs electrothermal activity, heart rate variability, eye-tracking techniques, and neuroimaging techniques to capture boredom-

https://doi.org/10.1017/S0272263125000142 Published online by Cambridge University Press

related physiological responses during foreign language learning tasks (Betella et al., 2014; Gawda et al., 2017). The artificiality of controlled conditions may restrict the ecological validity of the findings, and the reliance on specialized devices for data collection in the laboratory poses practical challenges, impeding the feasibility of conducting large-scale studies (Pavlenko, 2005). Self-report measurements, including questionnaires and surveys, have been extensively utilized to assess FLB (Zhao et al., 2023; Zhao & Wang, 2023). The Achievement Emotions Questionnaire (AEQ) was developed to evaluate various achievement emotions experienced by students in academic contexts (Pekrun et al., 2011). The AEQ comprises 24 scales designed to measure nine emotions experienced during class, studying, and exams including enjoyment, hope, pride, relief, anger, anxiety, hopelessness, shame, and boredom. However, AEQ needs more contextualization and conceptualization tailored explicitly to foreign language learning (Li et al., 2023).

Currently, the development of FLB scales has been addressed by Pawlak et al. (2020) and Li et al. (2023). The former devised the Boredom in Practical English Language Classes Questionnaire (BPELCQ) based on data from 111 Polish English majors, while the latter developed the Foreign Language Learning Boredom Scale (FLLBS) using data from over 3,000 non-English major college students in China. The reliability and validity of the former questionnaire require further verification, whereas the total scale and subscales of the latter have attained desirable levels of reliability and validity. Subsequently, researchers have embarked on studies centered around FLB using these scales to investigate boredom and its relationship with antecedent and sequent variables (Apridayani & Waluyo, 2022; Kruk et al., 2022). While some studies on antecedents focus on the influence of individual cognitive styles, teaching modes, and teacher's enthusiasm on FLB (Apridayani & Waluyo, 2022; Dewaele & Li, 2021), other empirical studies have explored the relationship between FLB and language achievement (Kruk, 2022; Wang et al., 2023). Among them, a significant portion has focused on the extent to which boredom can influence and predict foreign language achievement (Li & Wei, 2023; Wang et al., 2023; Zhao & Wang, 2023).

Foreign language achievement

Foreign language achievement encompasses an individual's proficiency and success in acquiring and using a non-native language, involving an evaluation of the learner's language knowledge (e.g., grammar, vocabulary, and pragmatics), abilities in comprehending, speaking, reading, and writing and skills in cross-cultural communication in the target language (Cizek, 1996; Durán, 2008; Neufeld, 1979). Many factors influence foreign language achievement, including learner's motivation, learning emotions, and access to the target language (Madigan & Curran, 2021; Peng & Kievit, 2020). It is crucial to recognize that foreign language acquisition is a dynamic process wherein learners can consistently develop and enhance their language abilities through ongoing practice and exposure (Anderman & Patrick, 2012).

Foreign language achievement measurements not only include objective scores such as course test scores and standardized language test scores but also reflect the language level and mastery of students in foreign language learning through students' self-reported scores and self-perceived ability (Botes et al., 2020; Li, 2020; Zhao & Yang, 2023). Quantifiable and standardized measurements provide consistent evaluations of a learner's language proficiency and concrete feedback on the mastery of language skills and content knowledge within a structured educational setting, offering a reliable means of comparison among students and across different contexts (Bachman & Adrian, 2022;

https://doi.org/10.1017/S0272263125000142 Published online by Cambridge University Press

Spolsky, 2000). In addition, learners' perceived language achievement, which refers to the subjective perception or belief an individual holds regarding their level of proficiency and success in acquiring a foreign/second language, has been increasingly used to assess learners' learning achievement (Dewaele et al., 2018; de Saint Léger, 2009). It is based on the learner's assessment and subjective evaluation of their language learning progress, skills, and knowledge. Several researchers have emphasized that there may be a lack of perfect alignment between perceived language learning achievement and actual proficiency levels (Babaii et al., 2016; Yoon & Lee, 2013). It becomes imperative for research studies to incorporate measurements of both perceived learning achievement and objective test scores to gain a comprehensive understanding of learners' language abilities and progress.

Foreign language boredom and achievement

Prolonged experiences of boredom during language learning can generate negative attitudes that lead to disengagement in language learning, which may impact learners' language performance and achievements (Wang & Li, 2022; Dewaele et al., 2023). A growing body of research has investigated the relationship between FLB and language learning achievement, consistently revealing a negative correlation (Li & Wei, 2023; Tsang & Dewaele, 2023; Zhao & Wang, 2023). This correlation extends beyond overall language performance, as FLB has been shown to negatively predict learners' achievement in various contexts. FLB was found to negatively predict learners' performance of English as a foreign language for specific purposes research based on aviation courses (N = 198) (Dinçer & Atay, 2022). FLB negatively affected foreign language writing test scores and self-perceived writing ability (Jiang, 2023; Li et al., 2023). Such a negative correlation may be consistent across national and educational contexts (Li & Wei, 2023; Tsang & Dewaele, 2023).

However, some studies also point out that FLB may not directly predict foreign language achievement, and it is not the only factor that directly leads to low performance; thus, the interference of other factors should be considered (Li & Lu, 2022). Based on the study of 954 English learners at a rural middle school, Li and Wei (2023) found that FLB can independently predict foreign language scores in three time points, but after entering the combined emotion model (enjoyment, anxiety, and boredom), boredom loses its adequate predictive power for academic scores. Researchers also found differences in the effect of FLB while using different measurements for learning achievement under different teaching modes (Apridayani & Waluyo, 2022; Wang & Li, 2022). FLB could significantly predict learners' self-perceived academic achievement in online classrooms for English as a foreign language, but it does not significantly predict learners' actual reading and writing performance (Wang & Li, 2022; Huang & Zhang, 2024).

Previous studies have identified several issues that warrant attention. First, various FLB measurements (such as AEQ and FLLBS) have been used in previous studies. Second, different language achievement measurements were employed, including course grades, language tests, and self-rated scores. Third, although most studies support the negative relationship between FLB and learning achievement, the exact strength of this relationship still needs to be resolved. Finally, a meta-analysis that considers potential moderators is needed to provide insight into the relationship between boredom and achievement. Therefore, these problems may bring potential risks, leading to the need for more necessary systematic and comprehensive research on FLB and the inability to draw relatively accurate, generalizable conclusions.

https://doi.org/10.1017/S0272263125000142 Published online by Cambridge University Press

Moderators

Meta-analysis provides a unique opportunity to delve into potential moderating factors, such as research characteristics, which could elucidate systematic variations in effect sizes observed across studies, offering insights into the intricate relationship between FLB and foreign language achievement. The moderation analysis within meta-analysis enhances the robustness of the analysis and facilitates a deeper understanding of the complexities underlying the impact of FLB on language achievement.

Given the multifaceted nature of FLB, different scales can provide standardized and systematic ways to measure boredom in foreign language learning, capturing its varying intensity and performance (Dewaele & MacIntyre, 2016). By distinguishing different levels of boredom, ranging from mild disinterest to profound disengagement, researchers can examine the relationship between boredom and learning achievement, including the potential threshold at which boredom becomes detrimental to foreign language learning performance (Li & Wei, 2023). Incorporating these scales enables researchers to better understand the relationship between boredom and learning achievement and identify moderating factors.

It is essential to acknowledge that the specific types of foreign language achievement measures employed in studies may moderate the relationship between boredom and learning achievement. Utilizing various measurement approaches allows for a more comprehensive evaluation of learning outcomes, as different measures may yield distinct outcomes and capture different dimensions of language achievement (Bachman & Adrian, 2022; Spolsky, 2000). By employing a range of measurements, researchers can delve deeper into the intricate dynamics between boredom and learning achievement, identifying specific areas of vulnerability to the detrimental effects of boredom and exploring the factors that influence this relationship (Bachman & Adrian, 2022). This multifaceted approach to measurement contributes to a more nuanced understanding of the complexities underlying the interplay between boredom and foreign language learning outcomes.

Specific language skills may also moderate the relationship between boredom and achievement in foreign language learning (Jiang, 2023; Liao & Dong, 2018). The impact of boredom may vary across different language skills based on their cognitive demands and nature (Li et al., 2023). Specific skills require continuous attention, concentration, and engagement, rendering them more susceptible to the negative effects of boredom (Jiang, 2023; Liao & Dong, 2018). Analyzing the relationship between boredom and specific language skills allows researchers to identify the skills most prone to boredom and ascertain their influence on academic performance. Different language skills may elicit varying levels of learner participation and motivation, resulting in differential effects of boredom on learning achievement.

Furthermore, educational stages correspond to distinct developmental milestones and cognitive abilities (Gajda, 2016; Lee et al., 2021), influencing the relationship between boredom and foreign language performance. Boredom may affect academic performance differently across different stages of education due to variations in attention span, engagement levels, and cognitive and socioemotional development (Lee et al., 2021; MacIntyre, 1992). Younger learners, such as those in secondary school, may exhibit different attention spans and levels of engagement compared to older learners in higher education (Bell & Wolfe, 2004). The effects of boredom on academic performance may vary depending on the educational stage due to these developmental disparities.

Additionally, different instructional modes, online and offline modes, diverge in terms of interactivity and engagement (Derakhshan et al., 2021), influencing the

https://doi.org/10.1017/S0272263125000142 Published online by Cambridge University Press

relationship between boredom and language learning performance (Derakhshan, Kruk et al., 2022). Online teaching typically incorporates digital platforms, multimedia resources, and interactive activities, while offline teaching relies on face-to-face interactions, physical materials, and classroom instruction (Kruk et al., 2022). The level of interaction and engagement in each mode affects learners' interest, motivation, and susceptibility to boredom (Shimray & Wangdi, 2023). Online teaching modes, characterized by greater flexibility and learner control over the learning process, can enhance learner autonomy and motivation, reducing the likelihood of boredom (Shimray & Wangdi, 2023). Conversely, offline teaching with more structured and rigid schedules may lead to higher levels of boredom (Derakhshan et al., 2022). The degree of flexibility and control embedded within each teaching mode can moderate the relationship between boredom and learning performance.

Current study

To address the limitations of individual studies, as illustrated in the literature review, a meta-analysis is crucial for gaining deep insights into how boredom affects FL learning achievement. The present study aims to advance and integrate existing empirical research and previous reviews by addressing three primary research objectives. First, a comprehensive and systematic literature review will examine the relationship between boredom and learning achievement in foreign language learning. Second, the direction (positive or negative) and the effect size of FLB on foreign language achievement should be examined as documented in the existing literature. Last, to broaden the scope of the meta-analysis by investigating potential moderators of the relationship between FLB and learning achievement, including contextual factors and language-related variables.

Based on these research objectives, the study seeks to address the following two research questions (RQs):

RQ1: How does FLB impact foreign language achievement? What is the direction and magnitude of this impact?

RQ2: How do different language achievement measurements, FLB measurements, educational stages, and teaching modes influence the relationship between FLB and language learning achievement?

Methodology

This study utilized a multilevel meta-analysis, an analytical approach designed to account for dependencies between effect sizes and effectively model the hierarchical structure of the data (López-López et al., 2018). By capturing the complexity inherent in nested data, the multilevel meta-analysis enhances the accuracy of estimates and mitigates issues related to data dependencies, allowing for a clearer interpretation of results, especially in fields where multiple effect sizes may be derived from a single study, thereby reflecting the true dynamics of the research context. (Assink, & Wibbelink 2016; Cheung, 2019; Norouzian & Bui, 2024).

Literature retrieval

Multiple methods have been employed in this study to retrieve relevant literature on FLB (Vuogan & Li, 2023). Both published and unpublished studies were considered

https://doi.org/10.1017/S0272263125000142 Published online by Cambridge University Press

eligible for inclusion, minimizing the potential impact of publication bias on the statistical outcomes (Vuogan & Li, 2023).

Several online databases were initially searched, including Web of Science, Google Scholar, Education Resources Information Clearinghouse (ERIC), China National Knowledge Infrastructure (CNKI), and Linguistics and Language Behavior Abstracts (LLBA). The search utilized various keywords, including boredom, language, foreign language, second language, classroom, emotion, language competence, language ability, language proficiency, and foreign language learning achievement. Building on references from previous research, manual searches were conducted. Selection criteria were based on frequently cited journals in previous studies, ensuring consistent and comparable literature reviews (Teimouri et al., 2019; Vuogan & Li, 2023). These influential and authoritative journals have published many groundbreaking studies in SLA and applied linguistics and are widely used in different reviews. These journals include *Studies in Second Language Acquisition, Applied Linguistics, Journal of Applied Linguistics, Foreign Language Annals, Language Learning, Canadian Modern Language Review, International Review of Applied Linguistics, Language Teaching Research, Language Testing, the Modern Language Journal, Second Language Research, System*, and *TESOL Quarterly*.

Furthermore, the backward and forward citation searches were conducted based on influential and highly cited literature. As a result, 529 relevant studies on FLB were retrieved.

Inclusion/exclusion criteria

This study applied carefully defined inclusion and exclusion criteria to determine the final research sample. The following eligibility criteria were considered.

First, the included studies must be empirical, with a focus on quantitative research. Studies employing mixed methods (both quantitative and qualitative) are also eligible for inclusion. However, systematic reviews, scoping reviews, and purely qualitative studies are not considered for inclusion. A meta-analysis that provides effect sizes can be included. Second, the selected studies must include at least one measure of FLB and one measure of foreign language achievement.

Third, the selected studies should report a relationship between FLB and foreign language achievement. This relationship should be expressed as a correlation coefficient or other statistics that can be translated into a correlation (e.g., t or F). Studies that analyze correlation matrices using multiple regression or structural equation model analysis can also be included.

In addition, the current study would include all forms of study design if a correlation coefficient can be retrieved from it. In the current meta-analysis, pretest data would be used when experimental studies with pretest and posttest data are included to avoid confounding effects and enhance the accuracy and reliability of effect estimates. This approach ensures comparability of participants' baseline characteristics, controlling for initial differences and minimizing external variables. The uniform collection of pretest data enhances consistency and reduces the impact of heterogeneity factors—such as variations in intervention implementation, assessment tools, and time spans—on the meta-analysis results (Vuogan & Li, 2023). In cases where the study examines differences between groups, different groups will be entered into the database and labeled separately.

Last, several potential moderating variables, such as teaching modes and educational stage, were identified. However, the presence or absence of a moderating variable did

https://doi.org/10.1017/S0272263125000142 Published online by Cambridge University Press

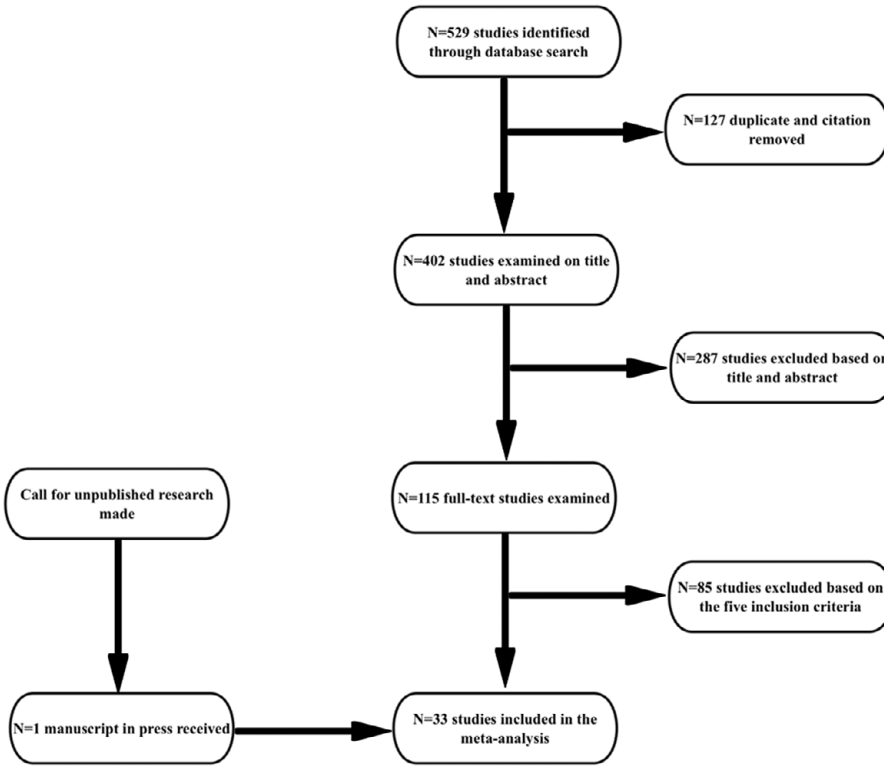


Figure 1. Database search flowchart.

not exclude any studies. Figure 1 illustrates the workflow for document inclusion and exclusion in this study.

The Newcastle-Ottawa Scale (NOS) was adapted to check methodological transparency and quality of individual studies included in the meta-analysis. By assessing selection, comparability, and outcome, the NOS standardizes study ratings, which facilitates the inclusion of high-quality research and improves the robustness of meta-analysis conclusions (Stang, 2010). The scoring index of methodological transparency is presented in Table 1 (See the elaboration of the index in Appendix). Each item is scored on a scale of 0 and 1, with selection scoring up to 3 points, comparability up to 2 points, and outcome up to 4 points, totaling a maximum score of 9 points. Studies scoring over 5 are suitable for meta-analyses and over 7 are considered high quality (Stang, 2010). After scoring, the studies included in the study are suitable for meta-analysis (score above 5), with the mean score equal to 7.06 and SD equal to 1.51.

Coding procedure and coding system

To comprehensively capture various aspects of each study in the sample, this study integrated a comprehensive literature coding system tailored to the research purpose, research questions, and previous studies in the field (Vuogan, 2023). It consists of five

https://doi.org/10.1017/S0272263125000142 Published online by Cambridge University Press

Table 1. Scoring index of methodological transparency

Dimension	Assessment item	Scoring criteria (point)
Selection	Clarity and selection of study population	1: Clear and adequate definition 0: Unclear or inadequate definition
	Representativeness of the sample	1: Sample representative of the target population 0: Sample not representative
	Participation rate	1: High participation rate and reported 0: Low participation rate or not reported
Comparability	Control of major confounding factors	1: Controlled major confounding factors 0: Not controlled
	Use of multivariate analysis	1: Conducted multivariate analysis 0: Not conducted
Outcome	Measurement of boredom	1: Used reliable and validated measurement tools 0: Unreliable or invalid tools
	Measurement of achievement	1: Consistent and validated assessment methods 0: Inconsistent or invalid methods
	Completeness of data collection	1: Reasonable handling of missing and outlined data 0: Unreasonable handling of missing and outlined data
	Appropriateness of statistical analysis	1: Used appropriate statistical method 0: Inappropriate statistical methods

parts, including 39 coding indices: background information, research methods, participants/sample information, instrument of boredom, and instrument of achievement.

The coding scheme was adaptively revised to characterize the diverse research features in this area, as shown in Table 2. This study also explores some of these features as potential moderators of the overall correlation effect size. Two coders, both doctoral students in applied linguistics, participated in the coding process. The intercoder agreement reached 98%, indicating a high level of agreement and ensuring the reliability of subsequent analyses (Marsden et al., 2015).

Data analysis

The data analysis process adheres to the steps outlined in Figure 2 and is conducted in accordance with established guidelines and recommendations for multilevel meta-analysis (Assink & Wibbelink, 2016; Cheung, 2019; Cui et al. 2024; Norouzian & Bui, 2024), comprising the following six steps, effect size calculation, overall effect size estimation, outlier detection and influence analysis, heterogeneity analysis, moderator analysis, and publication bias assessment. All statistical analyses were conducted using the "metafor" package (Viechtbauer, 2010) in R (Version 4.1.2; R Core Team, 2021).

Effect size calculation

The effect size was calculated using Fisher's z transform of the correlation coefficients. Fisher's z was chosen because it has the properties of stable variance and producing a normal distribution indicator, which facilitates more reliable meta-analysis inferences (Jak, 2015). The conversion to Fisher's z helps to handle studies with different sample sizes and different effect sizes, while ensuring that the variance of the effect size is

https://doi.org/10.1017/S0272263125000142 Published online by Cambridge University Press

Table 2. Coding system

Index	Description
Background Information	
Author's name	Scholars who conducted and published the study
Title of publication	Title of paper
Year of publication	Year in which the study was published
Type of publication	Classification of a document such as a journal article, book, or conference paper
Research Methods	
Methodology	Quantitative or mixed methodology
Study context	Foreign language, second language, not specific
Country	Country where the research was conducted
Participants/sample information	
Participant age	Mean age or age range of the participants
Participants' nationality	Nationality of the participants
Participants' LI	Native language of the participants
Participants' L2	The target language of the participants
Participants' academic status	Participants' educational stage (elementary school, middle school, high school, university, and mixed)
Participant sample size	Total number of participants
Male	Number/rate of male
Female	Number/rate of female
Participants' proficiency level	Proficiency level as illustrated in individual studies
Participants' language learning experience	Average/range of learning time that participants had been learning the target language
Instrument of boredom	
Instrument source	Adapted or adopted sources
Instrument type	Questionnaire or another
Creator of instrument	Origin study of the instrument
Items	Number of items
Likert scale	Likert scale form
Pilot	Yes/No
Mean of boredom	Mean score of boredom measurement
Standard deviation (SD)	SD of boredom measurement
Reliability index	As illustrated in individual study
Reliability estimate	Value of reliability index
Instrument of achievement	
Achievement type	The type of achievement measure used in the study (course score, self-perceived achievement, self-report course score, standardized language test, not available)
Measurement source	Whether the achievement measure was adopted/adapted or newly developed
Pilot	Yes/No
Measurement level	Range of scores
Mean	Mean score of participants
Standard deviation (SD)	SD of achievement score
Analysis type	Type of inferential statistical test used
Reliability	Whether measures of reliability were reported for the achievement tool
Reliability index	If reliability was reported, what reliability index was used
Reliability estimate	Reliability coefficient as reported
Reported effect size (r):	Observed correlation between anxiety and achievement
Teaching modes	Offline, online, and not available

properly estimated (Jak, 2015). The variance of Fisher's z transform effect size is calculated according to the respective formulas for consistency and accuracy. The calculation of the effect size was initially performed in Excel, and the results were then converted back to the correlation coefficient (r) for easy interpretation.

https://doi.org/10.1017/S0272263125000142 Published online by Cambridge University Press

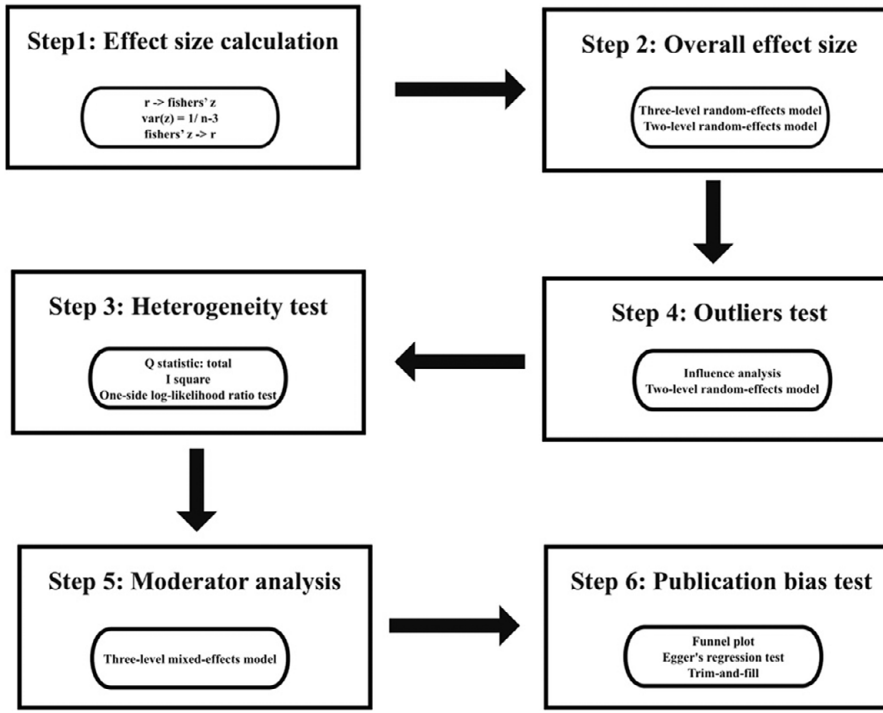


Figure 2. Flow chart of data analysis.

Overall effect size estimation

A multilevel random model implemented using the "metafor" package in R (Viechtbauer, 2010) was employed to estimate the overall effect size, considering the variance at three levels: within-study variance, between-studies variance, and overall variance. This modeling strategy allows the variance to be partitioned into different levels, thereby addressing the dependency of effect sizes within studies. By incorporating random effects at both the study and effect size levels, the model accurately reflects the nested structure of the data, providing a more robust estimate of the overall effect size (Assink, & Wibbelink 2016; Cheung, 2019).

Outlier and influence analysis

Outlier and influence analyses involving calculating Cook's distance, DFBETAS, and leverage (Hat) values for each study were conducted to identify studies or effect sizes that exerted disproportionate influence on the overall results (Viechtbauer & Cheung, 2010). The influence of each effect size was evaluated within the multilevel model, allowing for a more nuanced understanding of how individual studies contribute to the overall model. Outliers and influential studies were further investigated using influence index plots and leave-one-out analyses to assess their impact on the overall effect size estimate.

Heterogeneity analysis

Heterogeneity was quantified using the total Q-statistic and I^2 indices, which describe the proportion of variance attributable to different levels. A one-sided log-likelihood

https://doi.org/10.1017/S0272263125000142 Published online by Cambridge University Press

ratio test was performed to examine the significance of variance components at each level. The results from these tests provide evidence on whether substantial heterogeneity exists within or between studies, validating the necessity of the multilevel model (Huedo-Medina et al., 2006). Low heterogeneity suggests that the model has successfully accounted for the variability in effect sizes.

Moderator analysis

To explore potential sources of heterogeneity, moderator analysis was performed using a multilevel mixed-effects model, allowing for the inclusion of moderator variables to explain variability in effect sizes across and within studies (Norouzian & Bui, 2024). The mixed-effects model considers both fixed effects (moderator variables) and random effects (study and effect size levels), providing a comprehensive understanding of how moderators influence the overall effect (Cui et al., 2024). The inclusion of moderators enables the examination of contextual or methodological factors that may impact the observed effect sizes.

Publication bias assessment

Studies reporting relatively high effect sizes were more likely to be published than those reporting low effects, and published papers were more likely to be included in the meta-analysis, so the combined effect sizes might be overestimated, known as publication bias (Borenstein et al., 2009). This study adopted the Funnel Plot, Egger's regression test, and the trim-and-fill method to assess publication bias. The Funnel Plot graphically shows the presence or absence of publication bias and a simple scatter plot for each study and effect size. When the funnel plot appears roughly symmetrical, the study may not have serious publication bias (Sterne & Egger, 2001). Egger's regression test was employed to statistically test for funnel plot asymmetry at both the two-level and multilevel model structures (Egger et al., 1997). Additionally, the trim-and-fill method was applied to estimate the number of missing studies due to potential publication bias (Duval & Tweedie, 2000).

Results

Results of the methodological transparency analysis

After conducting a methodological transparency analysis using the NOS scale and considering the background of independent studies, all included studies achieved an acceptable level, with total scores above five. However, some issues were identified, examining the three dimensions of methodological transparency. In the selection dimension, while most studies had good sample representativeness, 42.4% (K = 14) did not report response rates and questionnaire return rates. In the comparability dimension, only 24.2% (K = 8) of the independent studies controlled for confounding variables, whereas the majority did not. Only 51.2% (K = 17) of the studies in the outcome assessment dimension reported handling missing and omitted values. In addition, this study identified a lack of transparent reporting on language proficiency across independent studies. Specifically, 54.5% of the studies (K = 18) did not report participants' language proficiency, while two included participants with mixed proficiency levels. Among the 13 studies that did provide proficiency information, there was significant variability and a lack of clarity with classifications ranging from beginner to intermediate, intermediate to advanced, and others that were often based on differing or unspecified criteria.

https://doi.org/10.1017/S0272263125000142 Published online by Cambridge University Press

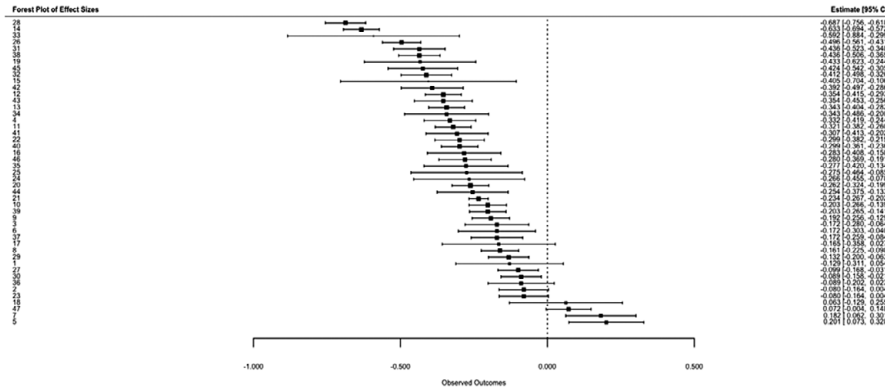


Figure 3. Effect sizes and confidence intervals for each sample.

Results of the overall effect and outlier analysis

This study included 33 independent studies with a total of 47 effect sizes. The effect sizes and confidence intervals for each sample are displayed in Figure 3.

The multilevel random model showed that the overall effect size was significant, with an effect size of r = -.240, 95% confidence interval (CI) = [-.303, -.187], df = 46, t = -8.504, $p < .001$. The two-level random model illustrated that the overall effect size was significant, with r = -.250, 95% CI = [-.310, -.202], df = 46, t = -9.507, $p < .001$. The Heterogeneity analysis indicated significant heterogeneity, Q (46) = 553757.301, $p <$ .001. To assess the influence of individual studies on the outcomes, an influence analysis was conducted. A study is considered influential if its removal significantly alters the model's results (Viechtbauer & Cheung, 2010). The outlier detection process revealed no significant outlier distribution, as illustrated in Figure 4.

Results of heterogeneity analysis

The heterogeneity test for the multilevel model did not yield significant results with Q (46) = 768.013, $p < .001$). Then, a comparison of model fit between the two-level and

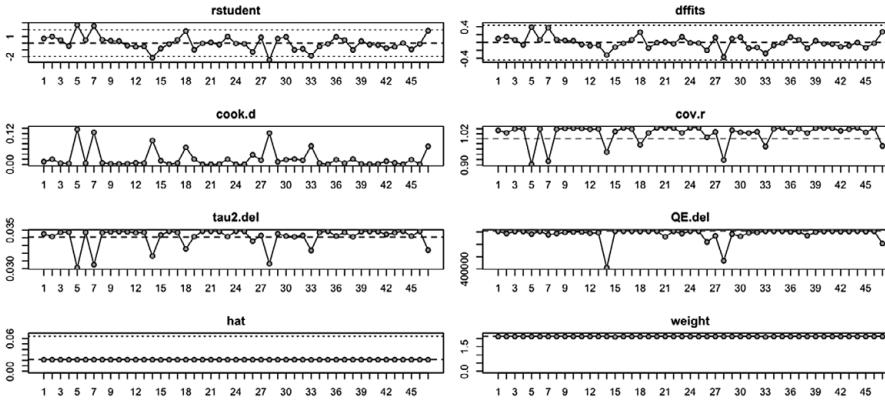


Figure 4. Results of outlier detection.

https://doi.org/10.1017/S0272263125000142 Published online by Cambridge University Press

Table 3. Model comparison.

	df	AIC	BIC	AICc	LogLink	LRT	*p*-value
Full	3	−19.552	−14.066	−18.980	12.776		
Without level 2	2	169.208	172.865	169.487	−82.604	190.760	< .001
Without level 3	2	−21.260	−17.602	−20.981	12.630	.292	.590

multilevel models was conducted, as presented in Table 3. When the within-study variance was constrained to zero, the model fit significantly worsened, as indicated by the log-likelihood ratio test, indicating heterogeneity at the within-study level. In contrast, fixing the between-study variance at zero did not result in a significant decline in model fit, indicating that between-study variance has a minimal impact on the observed heterogeneity.

Results of moderator analysis

To address RQ2, this study conducted a multigroup analysis on seven moderate factors: educational stage, language achievement measurements, domain-specific skills, foreign language boredom measurements, teaching modes, learning contexts, and methodology transparency level.

First, the included literature was divided into subgroups based on educational stages: elementary school, middle school, high school, university, and a mixed group (referring to studies that did not specify the educational stage of the language learners or included learners from different educational stages). Second, to investigate the differential impact of FLB on foreign language achievement measurements, the included literature was categorized into four groups: course score, self-perceived achievement, self-reported score, and standardized test score (e.g., International English Language Testing System [IELTS]/Test of English as a Foreign Language [TOEFL], College English Test [CET]4/6, and Hanyu Shuiping Kashi [HSK]). One study that did not report the assessment form was labeled as NA. Third, domain-specific skills were categorized into five groups: overall language proficiency, reading, writing, speaking, and reading and writing. Studies that combined reading and writing scores without distinguishing between the two were classified into the reading and writing skills group.

Additionally, two studies did not provide specific descriptions of language proficiency measurement and were marked as NA. Fourth, this study included four scales—AEQ, BPELC, FLLBS, and PSBEC—to explore the effects of FLB measurements on language achievement. Fifth, as the impact of FLB on foreign language achievement may vary under different instructional modes (in-person and online), this study conducted a moderation analysis of instructional modes. Finally, the moderator analysis of the FL/SL learning background and methodology transparency level was conducted, respectively. Table 4 illustrates the combined effect sizes, confidence intervals, and *p*-values for each subgroup.

Results of publication bias

Funnel plot, Egger's regression test, and trim-and-fill were employed to assess publication bias. The funnel plot is a graphical tool used to detect publication bias in research. It is depicted as a scatter plot where the horizontal axis represents the effect size of each study, and the vertical axis denotes the standard error. Each circle on the

https://doi.org/10.1017/S0272263125000142 Published online by Cambridge University Press

Table 4. The moderator analysis.

Moderator analysis group	*K*	Effect size and 95% confidence interval			*f*	*df*	*p*-value
		r	Lower limit	Upper limit			
Educational stage					.564	42	.690
Elementary	6	−.273	−.488	−.122			.008***
Middle school	13	−.286	−.399	−.189			***
High school	3	−.243	−.459	−.037			.023**
University	21	−.210	−.290	−.136			***
Achievement measurements					3.166	42	.023**
Course score	21	−.236	−.318	−.162			***
Self-perceived achievement	5	−.415	−.584	−.298			***
Self-report course score	5	−.155	−.315	.004			.056
Standardized language test	15	−.175	−.272	−.082			***
NA	1	−.230	−.569	.101			.166
Domain-specific skills					1.156	41	.347
Overall language proficiency	29	−.245	−.320	−.186			***
Reading	5	−.130	−.300	.038			.125
Reading and writing	3	−.292	−.521	−.082			.008**
Speaking	3	−.246	−.471	−.032			.026**
Writing	5	−.370	−.547	−.229			***
NA	2	−.164	−.416	.086			.191
Boredom measurements					1.008	43	.398
AEQ	8	−.213	−.346	−.086			.002***
BPELC	2	−.003	−.298	.292			.981
FLLBS	36	−.250	−.322	−.188			***
PSBEC	1	−.211	−.561	−.133			.220
Teaching modes					.022	45	.883
Offline	35	−.234	−.310	−.173			***
Online	12	−.246	−.358	−.143			***
Learning context					1.508	44	.233
Foreign language	40	−.245	−.313	−.187			***
Second language	5	−.206	−.383	−.034			.020**
Method level					.148	45	.702
Pass		−.255	−.368	−.155			***
Good		−.234	−.305	−.170			***

Note: **= $p < .05$; *** = $p < .001$

plot corresponds to specific data. The funnel plot was asymmetrical (Figure 5), suggesting that publication bias is not a significant concern (Sterne & Egger, 2001).

The two-level Egger's regression test showed a low risk of publication bias, z = .057, $p = .955$. The multilevel Egger test with the modified predictor was nonsignificant, z = 1.206, p = .996. The trim-and-fill showed that 0 effect sizes were missing to the right side of the mean effect size (SE = 1.414). Based on the above results, this meta-analysis appears to have a low risk of publication bias.

Discussion

Methodological transparency and overall effect size

The results of methodological transparency analysis highlight several weaknesses in the research field. First, the omission of reporting response rate can lead to nonresponse bias, where the opinions of non-respondents may significantly differ from those of respondents, thereby skewing the results; second, the absence of controlling confounding variables may

https://doi.org/10.1017/S0272263125000142 Published online by Cambridge University Press

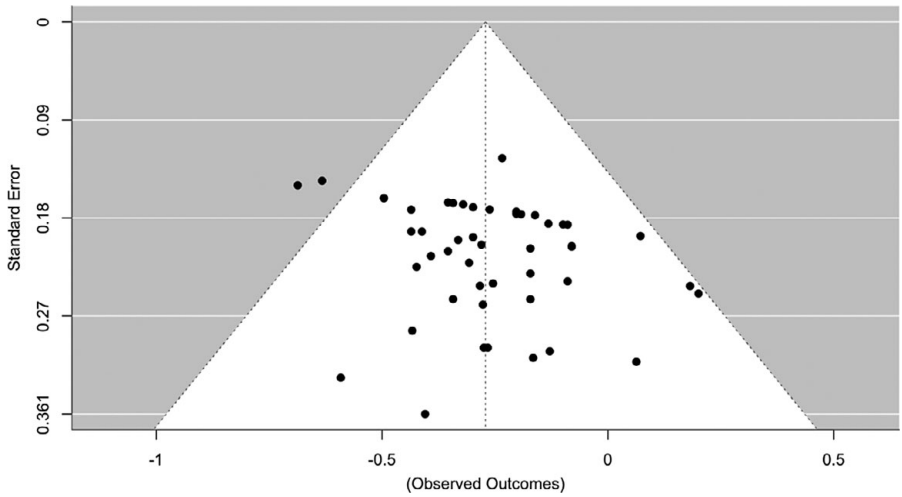


Figure 5. Funnel plot of effect size.

raise the research risk, as extraneous variables may influence the results, leading to erroneous conclusions about the relationships under investigation; and third, inadequately addressing missing data can result in biased estimates and reduced statistical power, compromising the validity of the findings in this field (DeKeyser, 2007; Little & Rubin, 2019). These deficiencies urge rigorous methodological practices with improved reporting standards, enhanced control of confounding variables, and robust handling of missing data to advance the quality of emotion research in SLA. Furthermore, methodological, and interpretive challenges were aroused due to the lack of clear reporting on language proficiency across independent studies, which impedes the ability to draw meaningful conclusions, as proficiency is a critical covariate in language learning research (DeKeyser, 2007). Additionally, using differing and ambiguous criteria for language proficiency may create inconsistencies and complicate comparisons across studies (DeKeyser, 2007). The transparency issues identified in this study are consistent with recent research in applied linguistics, which emphasizes the importance of transparent practice as part of a broader quality framework (Isbell et al., 2022; Plonsky et al., 2020). These issues are not isolated; rather, many of the challenges observed align with the broader research agenda in SLA (De Costa, 2016). As proposed by Plonsky (2024), transparency is a foundational element within the study quality framework, which is essential for achieving a more robust evaluation of research practices and ensuring that findings can be synthesized more reliably (Isbell et al., 2022). These results yearn for transparent measurements to enhance the comparability of studies, facilitate synthesizing research, and improve the overall quality and reliability of research in language learning.

The results revealed the negative effect of FLB on foreign language achievement (r = -.250, p < .001). According to Plonsky and Oswald (2014), an effect size close to .25 is considered small, around .40 is moderate, and above .60 is large in SLA. Therefore, the results indicate a small effect of FLB on language achievement, consistent with previous studies (Li et al., 2023; Li & Lu, 2022; Tze et al., 2016). Boredom can hinder the consolidation of new knowledge and skills, hinder language practice, and negatively affect long-term memory (Pawlak et al., 2020). When learners are bored, they may be

https://doi.org/10.1017/S0272263125000142 Published online by Cambridge University Press

less likely to actively engage in learning activities, retain information, or effectively develop language skills (Li, 2022). As a result, foreign language performance may suffer due to the reduced effort, motivation, and overall learning engagement caused by boredom (Kruk & Zawodniak, 2018).

Moderation effects

The subgroup analysis of different educational stages revealed inconsistent impact effect size of FLB on foreign language achievement. The effect size of different educational stages approximately follows the order of middle school group > the elementary school group > the high school group > university group, indicating that the influence of FLB on foreign language achievement tends to weaken as the educational stage advances, which may be attributed to learners' cognitive maturity and changes in language learning content. As the educational stage rises, learners' cognitive abilities, including attention, autonomous learning skills, memory, and critical thinking abilities, further develop, which enables learners to handle negative feedback more effectively during the learning process promptly and engage in positive emotion regulation (Bell & Wolfe, 2004), reducing the adverse impact of negative emotions on foreign language learning achievement (Ozfidan & Burlbaw, 2019). Additionally, as the educational stage progresses, the content of foreign language learning changes. In the initial stages of language learning, the emphasis is often on basic vocabulary and grammar, which can be repetitive and unstimulating (Langacker et al., 2006). As learners progress to higher stages and their proficiency in the foreign language improves, they begin to encounter more diverse and interesting language content, which may evoke enjoyment and alleviate the negative impact of FLB (Pawlak et al., 2023; Zhao & Wang, 2023).

However, the middle school group exhibited the highest effect size, indicating that the achievement of middle school learners is most affected by FLB, which may be related to the unique characteristics of the middle school stage. Middle school learners are transitioning from concrete operational thinking to abstract thinking (Gajda, 2016), which may challenge them to grasp and persevere in learning abstract and complex language aspects. Middle school often marks a shift from a more nurturing and playful learning environment in elementary school to a more structured and academically demanding environment, which may make some students feel less adapted and potentially affect their engagement and interest in foreign language learning (Madjar & Chohat, 2017; Papi & Hiver, 2020). Moreover, it was found that the effect of boredom on language achievement was slightly greater in FL contexts compared to SL contexts, which may arise from varying levels of exposure and immersion. In FL contexts, learners may experience more boredom due to structured classroom settings and limited real-life application, whereas in SL contexts, learners can benefit from practical use (Pawlak et al., 2020). Additionally, learners may face more pressure and less intrinsic motivation in FL than in the SL context due to cultural and educational differences (Chen et al., 2022).

The results of FLB's effect on different language achievement measurements illustrated significant between-group differences with inconsistent effect size. This finding is consistent with the meta-analysis results by Teimouri et al. (2019) on foreign language anxiety, which implies that the use of different foreign language achievement measurements (subjective measurement and objective measurement) requires more research attention (Li & Zhang, 2021). Both self-perceived foreign language achievement and self-reported course grades involve subjective measurement issues, especially

https://doi.org/10.1017/S0272263125000142 Published online by Cambridge University Press

self-perceived foreign achievement, which is essentially subjective and influenced by learners' emotions, attitudes, and self-concepts (Babaii et al., 2016; Ding & Zhao, 2020). On the contrary, course grades and standardized test scores are objective criteria for measuring learners' foreign language performance, aiming to evaluate specific knowledge and skills (Durán, 2008). The results of the two categories of foreign language achievement measurements suggest that future researchers must be cautious about relying solely on self-reported academic achievement indicators, as it may introduce potential research risks (Botes et al., 2020; Cizek, 1996). To reduce the risks associated with self-report data, future studies could consider using triangulation methods that combine self-report data with objective measurements (such as exam scores or grades) and other sources (such as teacher evaluations or observational data), which provides a more comprehensive and accurate reflection of foreign language learners' performance and learning experiences.

The results show that foreign language boredom has the highest impact on writing, followed by writing and reading, overall foreign language proficiency, speaking, and reading (p>.05) that did not reach statistical significance. These findings partially support previous studies (Teimouri et al., 2019), indicating that writing skills can be significantly influenced by negative emotions (boredom and anxiety), followed by speaking and reading. This pattern may be attributed to the distinctive characteristics associated with different foreign language skills. Learners may experience negative emotions when facing the challenge of production in a foreign language, which requires a deep understanding of grammar, vocabulary, and syntax, as well as the ability to organize thoughts and ideas coherently (Garner, 2016). When individuals feel bored, their cognitive resources could be more optimally utilized, making it challenging to generate original ideas, organize thoughts coherently, and employ appropriate language structures (Jiang, 2023).

Furthermore, the results illustrate that the effect size of FLB on learning achievement varies when different boredom measurements are used. When using the achievement emotions questionnaire (AEQ) and foreign language learning boredom scale (FLLBS), significant effects of cross-independent studies of FLB on learning achievement were observed, suggesting that these two scales demonstrate high measurement stability and reliability. However, AEQ is not a specialized scale focusing on foreign language learning emotions despite the good reliability, which may lead to a lack of context-specific content and requires more precise measurement dimensions of foreign language emotions (Botes et al., 2021). Therefore, more attention is needed to avoid potential risks when using AEQ in a foreign language context. The cross-language background stability of the FLLBS has been confirmed, which has been supported by studies on foreign language learning with different native language backgrounds (Derakhshan et al., 2022; Dewaele et al., 2024). Boredom in the personal English language classroom scale (BPELC) and the Perceived Second/Foreign Language Boredom Emotion Checklist (PSBEC) did not reach statistical significance. BPELC lacks rigorous reliability and validity tests and had a small sample size during initial development, resulting in decreased reliability (Li et al., 2023). The PSBEC is based on a precursor scale measuring boredom in the context of foreign language learning in Iran (Rad et al., 2022), which lacks cross-language and cross-cultural reliability and validity. Given the confirmed stability of the FLLBS across individual studies, the current study suggests that future investigations consider the FLLBS as a reliable instrument for assessing boredom in foreign language learning contexts as the reliability of the FLLBS has been validated across diverse native language backgrounds, rendering it a more accurate and context-specific measure.

https://doi.org/10.1017/S0272263125000142 Published online by Cambridge University Press

Moreover, this study found that the impact of FLB on learning achievement in the online environment is slightly greater than that in the offline environment, suggesting that there are differences in learners' emotional experiences in different classroom environments, and online classroom environments may have a more substantial influence on shaping learners' foreign language emotions (Zhang & Huang, 2023). Online learning often lacks the same level of social interaction and real-time feedback as offline learning, which can exacerbate feelings of boredom and disengagement, hindering the improvement of language performance (Derakhshan et al., 2022). Without the variety of face-to-face interactions and other offline activities, learners may quickly lose interest and become bored with language learning, resulting in lower achievement in online learning than in offline study (Chen et al., 2022; Jongsma et al., 2023).

Conclusion

This study employed the meta-analysis to examine the relationship between FLB and foreign language achievement, drawing on a comprehensive analysis of 33 empirical studies ($N = 27{,}838$) with 47 effect sizes and potential moderating factors. The findings indicate that FLB has an overall small negative effect on foreign language achievement. By integrating the findings of previous studies on FLB, this study presents conclusive evidence supporting the significance of boredom as a substantial negative predictor of foreign language achievement. The study also reveals that the negative effect sizes of FLB on achievement vary depending on educational stages, FL/SL learning contexts, domain-specific skills, measurements of boredom and achievement, and teaching modes.

Implications for future studies

This study underscores the need for future research to report on methodological practices with enhanced reporting standards, improved control of confounding variables, and effective handling of missing data to advance the quality and reliability of research in SLA. Moreover, the study emphasizes the importance of considering the reliability of FLB measurements and the distinct characteristics of both subjective and objective foreign language achievement measurements to avoid potential research risks. To mitigate the risks inherent in self-report data, future research is recommended to consider employing triangulation methods that integrate self-report data with objective measures (such as exam scores or grades) and additional sources (such as teacher evaluations or observational data). Considering the demonstrated stability of the FLLBS across multiple individual studies, the present research recommends that future investigations adopt the FLLBS, an effective instrument for accurately measuring boredom in diverse educational settings, as a reliable tool for assessing FLB. In addition, it is suggested that future researchers explore measurement corrections for these relationships and employ nonlinear statistical measurements to uncover potential nonlinear associations between the two constructs. This study also identified substantive gaps in the research field of foreign language boredom, particularly regarding boredom induced by specific foreign language skills. Previous studies primarily focus on writing, while the relationship between boredom and other foreign language skills remains insufficiently explored. Notably, research investigating the relationship between FLB and foreign language listening needs to be included, underscoring the need for further investigation.

https://doi.org/10.1017/S0272263125000142 Published online by Cambridge University Press

In addition, this study highlights the theoretical advantage of multilevel meta-analysis in adequately representing the underlying processes by which the data were generated (Gucciardi et al., 2022). Multilevel meta-analysis facilitates the accurate capture of variance across different levels and mitigates issues arising from nested data, providing a more authentic representation of the underlying research processes (Norouzian & Bui, 2024). Therefore, this study advocates for the adoption of multilevel meta-analysis in future research within the field of SLA, where scholars increasingly employ multiple measurements for exploring emotions and their relationships with various variables (performance or different psychological and cognitive variables.

Pedagogical implications

The meta-analysis indicates that FLB negatively predicts FL achievement, highlighting the need to address boredom in language education. It is suggested that Educators use engaging and diverse materials in various task types (e.g., discussions, group work, games) to cater to students' interests and reduce boredom. Providing timely and constructive feedback can motivate students by helping them see their progress, reducing boredom. Multiple assessment methods (e.g., formative assessment, peer assessment, self-assessment) can make the evaluation process more engaging. By integrating these strategies, educators can effectively mitigate the negative impact of boredom on foreign language achievement, enhancing learners' engagement and learning outcomes.

Limitations

However, this study relies on synthesizing existing literature, which introduces the possibility of incomplete literature searches and potential biases. The findings may be influenced by methodological variations, sample characteristics, and potential publication bias. This study did not include studies published in other languages, such as Polish, Turkish, Arabic, and Persian, which are indeed significant in boredom research, and their exclusion may limit the generalizability of our findings. Therefore, this study strongly recommends that future research endeavors strive to incorporate studies published in diverse languages to provide a more holistic and nuanced understanding of FL boredom. Furthermore, more potential moderators that may influence the relationship between boredom and achievement in foreign language learning need to be explored. Factors such as language acquisition order (L2, LX), the presence of multiple foreign languages, and changes in the language learning environment may all impact the strength of this relationship.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/S0272263125000142.

References

Anderman, E. M., & Patrick, H. (2012). Achievement goal theory, conceptualization of ability/intelligence, and classroom climate. In *Handbook of research on student engagement* (pp. 173–191). London: Springer.

Apridayani, A., & Waluyo, B. (2022). Antecedents and effects of students' enjoyment and boredom in synchronous online English courses. *Journal of Multilingual and Multicultural Development*, *45*(1), 1–16.

Assink, M., & Wibbelink, C. J. (2016). Fitting multi-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology*, *12*(3), 154–174.

https://doi.org/10.1017/S0272263125000142 Published online by Cambridge University Press

Babaii, E., Taghaddomi, S., & Pashmforoosh, R. (2016). Speaking self-assessment: Mismatches between learners' and teachers' criteria. *Language Testing*, *33*(3), 411–437.

Bachman, L., & Adrian, P. (2022). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford: Oxford University Press.

Betella, A., Zucca, R., Cetnarski, R., Greco, A., Lanatà, A., Mazzei, D., … & Verschure, P. F. (2014). Inference of human affective states from psychophysiological measurements extracted under ecologically valid conditions. *Frontiers in neuroscience*, *8*, 87270.

Bell, M. A., & Wolfe, C. D. (2004). Emotion and cognition: An intricately bound developmental process. *Child development*, *75*(2), 366–370.

Borenstein, M., Cooper, H., Hedges, L., & Valentine, J. (2009). Effect sizes for continuous data. *The handbook of research synthesis and meta-analysis*, *2*, 221–235.

Borgonovi, F., Pokropek, M., & Pokropek, A. (2023). Relations between academic boredom, academic achievement, ICT use, and teacher enthusiasm among adolescents. *Computers & Education*, *200*, 104807.

Botes, E., Dewaele, J.-M., & Greiff, S. (2020). The foreign language classroom anxiety scale and academic achievement: An overview of the prevailing literature and a meta-analysis. *Journal for the Psychology of Language Learning*, *2*(1), 26–56.

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*(4), 858–876.

Botes, E., Dewaele, J-M., Greiff, S., & Goetz, T. (2024). Can personality predict foreign language classroom emotions? The devil's in the detail. *Studies in Second Language Acquisition*, *46*(1), 51–74.

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*, *45*(8), 3291–3307.

Cheung, M. W. L. (2019). A guide to conducting a meta-analysis with non-independent effect sizes. *Neuropsychology review*, *29*(4), 387–396.

Cizek, G. J. (1996). Learning, achievement, and assessment: Constructs at a crossroads. In *Handbook of classroom assessment* (pp. 1–32). Elsevier.

Cui, X., Ding, Q., Yu, S., Zhang, S., & Li, X. (2024). The deficit in cognitive reappraisal capacity in individuals with anxiety or depressive disorders: meta-analyses of behavioral and neuroimaging studies. *Clinical Psychology Review*, *114*, 102480.

De Costa, P. I. (Ed.). (2016). *Ethics in applied linguistics research: Language researcher narratives*. New York: Routledge.

DeKeyser, R. (Ed.). (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. Cambridge: Cambridge University Press.

de Saint Léger, D. (2009). Self-assessment of speaking skills and participation in a foreign language class. *Foreign Language Annals*, *42*(1), 158–178.

Derakhshan, A., Fathi, J., Pawlak, M., & Kruk, M. (2022). Classroom social climate, growth language mindset, and student engagement: The mediating role of boredom in learning English as a foreign language. *Journal of Multilingual and Multicultural Development*, *45*(3), 1–19.

Derakhshan, A., Kruk, M., Mehdizadeh, M., & Pawlak, M. (2021). Boredom in online classes in the Iranian EFL context: Sources and solutions. *System*, *101*, 102556.

Derakhshan, A., Kruk, M., Mehdizadeh, M., & Pawlak, M. (2022). Activity-induced boredom in online EFL classes. *ELT Journal*, *76*(1), 58–68.

Dewaele, J.-M., Albakistani, A., & Ahmed, I. K. (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.

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

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.

Dewaele, J.-M., & Li, C. (2021). Teacher enthusiasm and students' social-behavioral learning engagement: The mediating role of student enjoyment and boredom in Chinese EFL classes. *Language Teaching Research*, *25*(6), 922–945.

https://doi.org/10.1017/S0272263125000142 Published online by Cambridge University Press

Dewaele, J.-M., & MacIntyre, P. D. (2016). Foreign language enjoyment and foreign language classroom anxiety: The right and left feet of the language learner. *Positive Psychology in SLA*, *215*(236).

Dewaele, J.-M., Witney, J., Saito, K., & Dewaele, L. (2018). Foreign language enjoyment and anxiety: The effect of teacher and learner variables. *Language Teaching Research*, *22*(6), 676–697.

Dinçer, R., & Atay, D. (2022). An examination of the relationship between enjoyment, boredom, classroom environment and achievement in esp context. *Boredom, Classroom Environment and Achievement in Esp Context*. [Preprint]. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4210611

Ding, Y., & Zhao, T. (2020). Emotions, engagement, and self-perceived achievement in a small private online course. *Journal of Computer Assisted Learning*, *36*(4), 449–457.

Durán, R. P. (2008). Assessing English-language learners' achievement. *Review of Research in Education*, *32*(1), 292–327.

Duval, S., & Tweedie, R. (2000). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, *95*(449), 89–98.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(7109), 629–634.

Gajda, A. (2016). The relationship between school achievement and creativity at different educational stages. *Thinking Skills and Creativity*, *19*, 246–259.

Garner, J. R. (2016). A phrase-frame approach to investigating phraseology in learner writing across proficiency levels. *International Journal of Learner Corpus Research*, *2*(1), 31–67.

Gawda, B., Szepietowska, E., Soluch, P., & Wolak, T. (2017). Valence of affective verbal fluency: fMRI studies on neural organization of emotional concepts joy and fear. *Journal of Psycholinguistic Research*, *46*, 731–746.

Gucciardi, D. F., Lines, R. L., & Ntoumanis, N. (2022). Handling effect size dependency in meta-analysis. *International Review of Sport and Exercise Psychology*, *15*(1). 152–178.

Huang, F., & Zhang, H. (2024). Explaining the penetrating role of technology in online foreign language learning achievement. *Foreign Language Annals*.

Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? *Psychological Methods*, *11*(2), 193.

Isbell, D. R., Brown, D., Chen, M., Derrick, D. J., Ghanem, R., Arvizu, M. N. G., … & Plonsky, L. (2022). Misconduct and questionable research practices: The ethics of quantitative data handling and reporting in applied linguistics. *The Modern Language Journal*, *106*(1), 172–195.

Jak, S. (2015). *Meta-analytic structural equation modelling*. Switzerland: Springer.

Jiang, C. (2023). Chinese undergraduates' English reading self-efficacy, intrinsic cognitive load, boredom, and performance: A moderated mediation model. *Frontiers in Psychology*, *14*, 1093044.

Jongsma, M. V., Scholten, D. J., van Muijlwijk-Koezen, J. E., & Meeter, M. (2023). Online versus offline peer feedback in higher education: a meta-analysis. *Journal of Educational Computing Research*, *61*(2), 329–354.

Kruk, M. (2022). Dynamicity of perceived willingness to communicate, motivation, boredom and anxiety in Second Life: The case of two advanced learners of English. *Computer Assisted Language Learning*, *35*(1–2), 190–216.

Kruk, M., Pawlak, M., Shirvan, M. E., & Shahnama, M. (2022). The emergence of boredom in an online language class: an ecological perspective. *System*, *107*, 102803.

Kruk, M., & Zawodniak, J. (2018). Boredom in practical English language classes: Insights from interview data. In Szymański *et al.* (Eds), *Interdisciplinary Views on the English Language, Literature and Culture*. Poland: Wydział Humanistyczny Uniwersytetu Zielonogórskiego.

Langacker, R. W., von Stutterheim, C., Carroll, M., Lambert, M., Behrens, B., Strauss, S., Pavlenko, A., Paribakht, T. S., Wesche, M., & Rinner, S. (2006). *Educating for advanced foreign language capacities: Constructs, curriculum, instruction, assessment*. Washington, D.C.: Georgetown University Press.

Lee, J. S., Xie, Q., & Lee, K. (2021). Informal digital learning of English and L2 willingness to communicate: roles of emotions, gender, and educational stage. *Journal of Multilingual and Multicultural Development*, *45*(2), 596–561.

Li, C. (2020). A positive psychology perspective on Chinese EFL students' trait emotional intelligence, foreign language enjoyment and EFL learning achievement. *Journal of Multilingual and Multicultural Development*, *41*(3), 246–263.

https://doi.org/10.1017/S0272263125000142 Published online by Cambridge University Press

Li, C. (2022). Foreign language learning boredom and enjoyment: The effects of learner variables and teacher variables. *Language Teaching Research*, 13621688221090324.
Li, C., Dewaele, J.-M., & Hu, Y. (2023). Foreign language learning boredom: Conceptualization and measurement. *Applied Linguistics Review*, *14*(2), 223–249.
Li, C., & Lu., X. (2022). Research on Boredom in Foreign Language Learning: A Scoping Review Approach. *Foreign Language Education*, *43*(6), 70–77.
Li, C., & Wei, L. (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.
Li, C., Wei, L., & Lu, X. (2023). Contributions of foreign language writing emotions to writing achievement. *System*, *116*, 103074.
Li, M., & Zhang, X. (2021). A meta-analysis of self-assessment and language performance in language testing and assessment. *Language Testing*, *38*(2), 189–218.
Liao, J., & Dong, J. (2018). The effects of working memory and boredom on L2 reading performance. *American Educational Research Journal*, *55*(6), 1339–1368.
Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data*. New Jersey: John Wiley & Sons.
López-López, J. A., Page, M. J., Lipsey, M. W., & Higgins, J. P. (2018). Dealing with effect size multiplicity in systematic reviews and meta-analyses. *Research Synthesis Methods*, *9*(3), 336–351.
MacIntyre, P. D. (1992). Anxiety and Language Learning from a Stages of Processing Perspective [Doctoral Dissertation, The University of Western Ontario].
MacIntyre, P. D., & Mercer, S. (2014). Introducing positive psychology to SLA. *Studies in Second Language Learning and Teaching*, *4*(2), 153–172.
Madigan, D. J., & Curran, T. (2021). Does burnout affect academic achievement? A meta-analysis of over 100,000 students. *Educational Psychology Review*, *33*, 387–405.
Madjar, N., & Chohat, R. (2017). Will I succeed in middle school? A longitudinal analysis of self-efficacy in school transitions in relation to goal structures and engagement. *Educational Psychology*, *37*(6), 680–694.
Marsden, E., Mackey, A., & Plonsky, L. (2015). The IRIS Repository: Advancing research practice and methodology. In Mackey, A &, Marsden, E (Eds), *Advancing methodology and practice*. New York: Routledge.
Mikulas, W. L., & Vodanovich, S. J. (1993). The essence of boredom. *The Psychological Record*, *43*(1), 3–12.
Neufeld, G. G. (1979). Towards a theory of language learning ability. *Language learning*, *29*(2), 227–241.
Norouzian, R., & Bui, G. (2024). Meta-analysis of second language research with complex research designs. *Studies in Second Language Acquisition*, *46*(1), 251–276.
Ozfidan, B., & Burlbaw, L. M. (2019). A literature-based approach on age factors in second language acquisition: Children, adolescents, and adults. *International Education Studies*, *12*(10), 27–36.
Papi, M., & Hiver, P. (2020). Language learning motivation as a complex dynamic system: A global perspective of truth, control, and value. *The Modern Language Journal*, *104*(1), 209–232.
Pavlenko, A. (2005). *Emotions and multilingualism*. Cambridge: Cambridge University Press.
Pawlak, M., Kruk, M., Csizer, K., & Zawodniak, J. (2023). Investigating in-class and after-class boredom among advanced learners of English: intensity, interrelationships and learner profiles. *Applied Linguistics Review*, *15*(6), 2537–2564.
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.
Pawlak, M., Kruk, M., Zawodniak, J., & Pasikowski, S. (2022). Examining the underlying structure of after-class boredom experienced by English majors. *System*, *106*, 102769.
Pawlak, M., Zawodniak, J., & Kruk, M. (2020). *Boredom in the foreign language classroom: A micro-perspective*. Switzerland: Springer Nature.
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.
Peng, P., & Kievit, R. A. (2020). The development of academic achievement and cognitive abilities: A bidirectional perspective. *Child Development Perspectives*, *14*(1), 15–20.
Plonsky, L. (2024). Study quality as an intellectual and ethical imperative: A proposed framework. *Annual Review of Applied Linguistics*, 1–15.

https://doi.org/10.1017/S0272263125000142 Published online by Cambridge University Press

Plonsky, L., Marsden, E., Crowther, D., Gass, S. M., & Spinner, P. (2020). A methodological synthesis and meta-analysis of judgment tasks in second language research. *Second Language Research*, *36*(4), 583–621.

Plonsky, L., & Oswald, F.L. (2014). How big Is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*, 878–912.

Rad, S. E. M., Roohani, A., & Mirzaei, A. (2022). Developing and validating Precursors of Students' boredom in EFL classes: an exploratory sequential mixed-methods study. *Journal of Multilingual and Multicultural Development*, *45*(8), 3010–3027.

Shimray, R., & Wangdi, T. (2023). Boredom in online foreign language classrooms: Antecedents and solutions from students' perspective. *Journal of Multilingual and Multicultural Development*, 1–16.

Spolsky, B. (2000). Language testing in The Modern Language Journal. *The Modern Language Journal*, *84*(4), 536–552.

Stang, A. (2010). Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. *European Journal of Epidemiology*, *25*, 603–605.

Sterne, J. A., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *Journal of Clinical Epidemiology*, *54*(10), 1046–1055.

Teimouri, Y., Goetze, J., & Plonsky, L. (2019). Second language anxiety and achievement: A meta-analysis. *Studies in Second Language Acquisition*, *41*(2), 363–387.

Tsang, A., & Dewaele, J. M. (2023). The relationships between young FL learners' classroom emotions (anxiety, boredom, & enjoyment), engagement, and FL proficiency. *Applied Linguistics Review*, *15*(5), 2015–2034.

Tsang, A., & Lee, J. S. (2023). The making of proficient young FL speakers: The role of emotions, speaking motivation, and spoken input beyond the classroom. *System*, *115*, 103047.

Tze, V. M., Daniels, L. M., & Klassen, R. M. (2016). Evaluating the relationship between boredom and academic outcomes: A meta-analysis. *Educational Psychology Review*, *28*(1), 119–144.

Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software*, *36*(3), 1–48.

Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, *1*(2), 112–125.

Vuogan, A., & Li, S. (2023). A systematic review of meta-analyses in second language research: Current practices, issues, and recommendations. *Applied Linguistics Review*, *15*(4), 1621–1644

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.

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.

Yoon, E., & Lee, H.-K. (2013). Do the effects of self-assessment differ by L2 language level? A case of Korean learners of English. *The Asia-Pacific Education Researcher*, *22*, 731–739.

Zhang, H., & Huang, F. (2023). Perceived teachers' enthusiasm and willingness to communicate in the online class: The mediating role of learning enjoyment and group interaction for Chinese as a second language. *Language Teaching Research*, 13621688231216199.

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*, 13621688231180465.

Zhao, X., & Wang, D. (2023). The role of enjoyment and boredom in shaping English language achievement among ethnic minority learners. *Journal of Multilingual and Multicultural Development*, 1–13.

Zhao, Y., & Yang, L. (2023). A latent profile analysis of Chinese high school students' control and value appraisals: Associations with enjoyment, boredom, and language achievement. *System*, *115*, 103068.

Cite this article: Huang, F., & Zhang, H. (2025). The relationship between boredom and second language achievement: A multilevel meta-analysis. *Studies in Second Language Acquisition,* 47: 488–511. https://doi.org/10.1017/S0272263125000142

https://doi.org/10.1017/S0272263125000142 Published online by Cambridge University Press