Examining the Validity of FLCAS and Predictive Factors in Chinese Language Learning Anxiety

Yiping Zhang[†*]
[†]*Feinstein College of Education*

M. Shane Tutwiler[†]
University of Rhode Island

Abstract
This study validates the Foreign Language Classroom Anxiety Scale (FLCAS) for Chinese language learning among U.S. college students and examines key socio-cultural elements of Foreign Language Anxiety (FLA). These elements include demographic information (heritage status, gender, and year in college) and predictors of language learning (duration of language study, overseas experience, and self-efficacy). Our study involved 135 participants from 8 Chinese Flagship Programs in the U.S, employed Bayesian statistical methods to reassess FLCAS to determine its applicability to Chinese language learners, confirming the scale's foundational constructs and revealing its multidimensional nature. Bayesian

multilevel modeling reveals that the sociocultural factors included collectively accounted for 50% of the observed anxiety variance. Specifically, male students with overseas experience and freshmen exhibited lower levels of FLA, and self-efficacy showed a strong negative correlation with FLA. These results guide tailored educational strategies and interventions to reduce FLA and enhance language learning outcomes.

Keywords: Foreign Language Anxiety, Chinese Language Learning, FLCAS, Bayesian Statistics, Higher Education, Individual Difference

Introduction

Foreign language learning anxiety (FLA) is common in classrooms (Dewaele & MacIntyre, 2014; MacIntyre, 2017). Studies in a variety of language learning contexts have found that approximately one-third of students studying a foreign language experience at least a moderate level of FLA (Aida, 1994; Horwitz, 2001; MacIntyre & Gardner, 1989). More importantly, research shows students have experienced higher levels of FLA since the start of the COVID-19 pandemic (Liu & Yuan, 2021). This anxiety can significantly hinder language acquisition, retention, and performance, necessitating effective measures for identification and intervention (MacIntyre & Gardner, 1991). Horwitz et al. (1986) defined FLA as "a distinct complex construct" reflecting learners' self-perceptions, beliefs, feelings, and behaviors during language use (p. 128), suggesting it is not merely a transfer of general anxiety. It includes cognitive aspects (e.g., beliefs about language competencies), emotional responses (e.g., fear

of negative evaluation), and behavioral tendencies (e.g., avoiding participation) (Luo, 2014; Horwitz et al., 1986; MacIntyre, 2017; Mihai et al., 2022), emphasizing the importance of supportive learning environments.

Horwitz et al. (1986) also developed the Foreign Language Classroom Anxiety Scale (FLCAS) to gauge FLA levels. Although the FLCAS has been widely used in a range of language-learning contexts (Aida, 1994; Liu, 2006; MacIntyre & Gardner, 1989; Mak, 2011), much of the research utilizing FLCAS focuses on commonly taught European languages (Yao et al., 2022), leaving its validity underexplored among Chinese language learners—particularly those enrolled in intensive Chinese Flagship Programs (CFP) (Zhang, 2025). Chinese is classified by the Foreign Service Institute (Foreign Language Training, 2018, para.5) as a "super-hard language" for native English speakers, potentially contributing to high dropout rates (Luo, 2014). Moreover, a significant correlation between high FLA levels and dropout has been noted (Gardner et al., 1987;

Mihai et al., 2022). By reassessing the validity of FLCAS and examining how FLA varies across different sociocultural factors, this study aims to enhance student learning experiences and outcomes in these high-stakes U.S. educational settings.

Literature Review

FLCAS Constructs

Horwitz et al. (1986) identified three core constructs underlying the FLCAS– Communication Apprehension, Fear of Negative Evaluation, and Test Anxiety–as key contributors to students' high levels of FLA. Communication Apprehension refers to anxiety about speaking (McCroskey, 1977); Fear of Negative Evaluation involves distress over perceived judgments by teachers and peers (Watson & Friend, 1969, p.449) and Test Anxiety stems from fear of poor performance on exams (Sarason & Stoops, 1978). Despite being widely used, the conceptual clarity and cross language applicability of the three constructs of FLCAS has been

challenged. In particular, the validity of the original factor structures has been questioned by research studies conducted in various languages and cultural contexts. In the Japanese-learning context, Aida (1994) found FLCAS yielded four factors rather than the original three, highlighting the importance of linguistic and cultural variables. While many have relied on exploratory factor analysis (EFA) —which can be subjective in terms of factor extraction and labeling (McDonald, 2014; Park, 2014; Thompson, 2004) —relatively few have employed confirmatory factor analysis (CFA) to investigate whether prior hypothetical models of the FLCAS are supported by the data (Park, 2012, 2014).

Given the importance of contextual factors, sociocultural elements play a significant role in the manifestation and measurement of FLA. For example, demographic variables such as age, gender, prior language exposure, and cultural background—significantly shape FLA (Dewaele, 2010; Jiang & Dewaele, 2020; Park & French, 2013). Validating FLCAS across diverse sociocultural contexts

is thus crucial. Despite this need, Chinese FLA research often relies on a three-factor or unidimensional model without always confirming its suitability (Ting & Sunarti, 2022; Xiao & Wong, 2014). This issue is particularly salient because Chinese differs markedly from many European languages: its tonal system, character-based script, and culturally specific classroom practices (e.g., emphasis on error avoidance and hierarchical teacher–student interactions) may uniquely heighten or shape FLA (Luo, 2013; Ferrer & Li, 2021; Yao et al., 2022). This practice may overlook critical linguistic and cultural nuances unique to Chinese language acquisition. Therefore, reexamining FLCAS's validity is crucial, both to ensure it reflects diverse populations and to enhance the effectiveness of anxiety-reduction strategies for U.S. Chinese language learners. To address this, we employ Bayesian Confirmatory Factor Analysis (BCFA) to provide a more tailored and theoretically robust validation of the FLCAS by incorporating sociocultural factors, recognizing their

significant role in the manifestation and measurement of FLA.

FLA and Associated Variables

FLA has been found to be linked to a wide range of independent variables, including student demographic factors such as age (Dewaele, 2007; MacIntyre, 2002), gender (Dewaele, 2007; Matsuda & Gobel, 2004), year in college (Williams & Andrade, 2008), heritage status (Xiao & Wong, 2014), and language factors related to FLA, such as students' first language (Saito et al., 1999; Zhang, 2019), study abroad experience (Allen & Herron, 2003; Bensalem & Trevethan, 2024), self-efficacy toward learning foreign language (Wang et al., 2022; Zhou et al., 2023). While the studies have reflected differing dimensions of socio-cultural factors that affect students FLA, their findings may not consistently apply to the unique challenges faced by students in intensive Chinese programs, particularly in U.S. higher education contexts.

Gender and FLA.

Gender differences in FLA show mixed findings across contexts (Campbell & Shaw, 1994; Matsuda & Gobel, 2004; Sung & Li, 2019). Sung and Li (2019) found that female U.S. learners of Chinese reported higher levels of FLA compared to their male counterparts, whereas no significant gender differences were reported among Japanese English learners (Matsuda & Gobel, 2004). Such conflicting results may be influenced by sociocultural factors, as highlighted by Park and French (2013), who noted that in Korea, traditionally male-dominated societal norms could lead females to experience greater FLA, potentially affecting their social interactions during language learning. Furthermore, the gender difference in FLA might be exacerbated by the frequent assessments and high expectations inherent in the CFP (Santana & Eccius-Wellmann, 2018). Thus, these inconsistencies, shaped by sociocultural norms, underscore the importance of including gender as a variable in CFP-based FLA research (Baker et al., 2020).

Heritage Status and FLA.
Several studies have examined the effect of heritage status on FLA by comparing the levels of heritage learners' FLA with the levels of non-heritage students' FLA (e.g. Spanish, Tallon, 2011; Chinese, Xiao & Wong, 2014; Korean, Jee, 2016) because of their prior exposure and cultural connections to the language. Although heritage learners often report lower FLA, they may face unique stressors from heightened expectations, highlighting the need for targeted support (Tallon, 2011; Xiao & Wong, 2014). Yet many studies treat heritage status as a uniform demographic category without considering the variation in learners' experiences. Differences in program structure, instructional approach, or institutional culture may influence how heritage learners experience anxiety, highlighting the importance of more contextually grounded research.

Year in College and FLA.
Although many language educators intuitively believe that lower-level students may experience higher FLA due to their

unfamiliarity with the curriculum and learning community, research on the relationship between years in college and FLA is limited. Some studies found higher FLA in senior students (Onwuegbuzie et al., 1999), while others found consistent levels across academic stages (Luo, 2014). In CFPs, "year in college" is not just time-based but tied to proficiency benchmarks and study-abroad requirements, shaping learners' sociocultural exposure. These factors are critical in understanding the distinct challenges and experiences of students at different academic stages, thereby enabling us to offer more effective, culturally sensitive support tailored to their specific needs and backgrounds.

Study Abroad and FLA.

Because overseas experience is built into CFPs, examining its role is essential. Prior studies suggest that such experiences tend to lower FLA levels (Matsuda & Gobel, 2004; Thompson & Lee, 2014). For instance, Allen and Herron (2003) found students' FLA decreased significantly after a

short-term international program, and similar outcomes were reported for extended stays abroad (Matsuda & Gobel, 2004; Thompson & Lee, 2014). Regarding learning Chinese in the United States, research revealed that students who had been to China had lower levels of FLA than those who had not (Zhao & Whitchurch, 2011). Yet this research examined students' travel experience rather than their study abroad experience. This distinction is critical, as immersive academic programs offer structured opportunities for language use that casual travel does not. Because CFPs mandate both short- and long-term overseas experiences, understanding how these varying immersions impact FLA is essential for effective program design.

Language Learning Time and FLA.

Students' language learning time could potentially affect their overseas experience. As students progress through their college years, they accumulate more hours of language instruction and practice (Dewaele, 2007), boosting proficiency and confidence—essential for adapting to

intensive programs like the CFP. Furthermore, they develop effective study habits and coping mechanisms over time (Williams & Andrade, 2008; Schunk, 1989), enabling them to manage rigorous academic demands and achieve greater success both in the CFP and abroad (Allen & Herron, 2003; Cubillos et al., 2008). Hence, language learning time merits attention in FLA research, as it shapes adaptability in high-pressure language environments and deepens students' linguistic and cultural understanding.

Self-Efficacy and FLA.

Language learning time and students' overseas experience can impact self-efficacy, which is pivotal in managing FLA. Rooted in Bandura's theory of motivation, self-efficacy means students' beliefs and judgments about their capabilities to successfully perform tasks (Bandura, 1997). The more confident individuals feel in their language abilities, the less FLA they experience when faced with linguistic challenges (Bandura, 1988). Cultural factors heavily influence this

confidence. For instance, students from cultures that emphasize effort (like Japan) may develop self-efficacy based on persistence, while those from cultures that value inherent ability (typical in the West) might see their self-efficacy fluctuate with perceived natural talent. This cultural variation in self-efficacy is critical in language programs like the CFP, where high demands and intensive curricula require robust psychological resilience (Xu et al., 2022). Understanding how cultural backgrounds shape self-efficacy can help tailor educational practices to better support students, reducing FLA and enhancing success in high-pressure environments.

In this study, we include demographic information factors (e.g. gender, heritage status, and year in college) as well as language-related information (e.g. language learning time, overseas experience, as well as self-efficacy toward language class). These variables not only reflect important demographic and linguistic dimensions but also embody various sociocultural aspects that are crucial to understanding Chinese language learning. For instance, heritage learners may

experience unique forms of anxiety tied to expectations from family or community to already "know" the language, leading to greater fear of negative evaluation or feelings of inadequacy when their skills fall short (Luo, 2015). Given the inconsistent results in the literature, examining these factors in the context of intensive language programs like CFP can provide deeper insights into the sociocultural dynamics influencing language acquisition.

Theoretical Framework

Horwitz et al. (1986) built on Gardner's (1985) suggestion that FLA is unique to language learning, operationalizing it through the 33-item FLCAS. Moreover, they proposed that FLCAS is related to three components: communication apprehension, test anxiety, and fear of negative evaluation, all of which can negatively affect students' foreign language performance (see Appendix A Figure 1). MacIntyre and Gardner (1991) further developed the idea and hypothesized that FLA could affect the different stages of the learning

process: the input stage (i.e., when learners encounter material for the first time), the processing stage (i.e. when learners integrate material for the new information and existing knowledge) and the output stage (i.e. when they demonstrate the new knowledge) when learners process new materials in relation to their prior knowledge. Nonetheless, Horwitz et al.'s theory has also been challenged, with critiques including the unclear causal direction between FLA and performance (Sparks & Ganschow, 1995) and the unspecified link between FLCAS items and its three constructs (Aida, 1994; Park, 2014).This led to researchers reporting different results when using FLCAS (MacIntyre & Gardner, 1989; Toyama & Yamazaki, 2018, 2021) across different contexts.

Despite the above challenges, Horwitz et al. (1986) were the first to conceptualize FLA as a unique type of anxiety specific to foreign language learning. Their theory has played a vital role in language anxiety research, with a large number of studies using it as a theoretical framework (e.g., Aida, 1994; Onwuegbuzie et al., 1999). These studies

acknowledged the uniqueness of FLA and provided evidence that the FLCAS is a reliable tool for measuring FLA (Tran, 2012).

The current study aims to reassess the validity of the FLCAS within American collegiate intensive Chinese language programs. By focusing on a population of college Chinese language learners, this research contributes to the scarce literature on Chinese as a Foreign Language (CFL), a field that is rapidly gaining academic interest due to the increasing global significance of China. This study also expands upon the traditional scope of FLCAS by incorporating a Bayesian statistical framework to evaluate anxiety constructs, and predict sociocultural variables, thereby offering a more detailed understanding of anxiety's multifaceted nature in language learning.

To date, no research has specifically investigated the validity of FLCAS in the field of Chinese language learning, nor has it examined the predictors of FLA in an intensive

Chinese language learning environment. The results of this study are poised to offer valuable empirical evidence for the operationalization of FLA in the specific context of Chinese language learning. By identifying the predictive factors of FLA with a lens on individual differences, the findings can provide insights into personalized educational strategies and interventions. This is particularly pertinent as the learning dynamics for a tonal language like Chinese, with its unique linguistic and cultural challenges, may shape FLA in ways that differ from learners of Indo-European languages. These challenges include its logographic writing system, which requires extensive character memorization; the lack of phoneme-grapheme correspondence; diglossic features such as distinctions between spoken vernacular and formal written forms; and complex sociopragmatic norms involving politeness, indirectness, and age-based honorifics.

To address this gap, our second research question employs Bayesian multilevel modeling to explore how these factors contribute to FLA among students in intensive

Chinese language programs. We will include as the predictors demographic variables: gender, heritage status, year in college; and language learning related factors: language learning time, overseas experience, as well as self-efficacy toward language learning—and examine how they are associated with FLA in the context of learning Chinese in the United States.

Specifically, the current study seeks to answer the following research questions:

RQ1: Are the three constructs of FLCAS proposed by Horwitz et al. (1986) (communication apprehension, test anxiety, and fear of negative evaluation) valid for college Chinese language learners?

RQ2: How does FLA vary across individual differences including demographic information (gender, heritage status, year in college) and language learning factors (language learning duration, overseas experience, self-efficacy)?

Methods

Setting and Participants

This study was carried out among students enrolled in the CFPs in the U.S. The Language Flagship was launched in 2002 as an important national innovation in the U.S. higher education system for language teaching and learning (Nugent & Slater, 2016). It sponsors thirteen CFPs across the United States. The CFP provides undergraduate students with pathways to professional-level proficiency in Chinese alongside the academic major of their choice. During the four or five years of study, students are required to participate in at least one overseas summer program. They are also required to spend a full year, called the Capstone Year, studying and interning at a center in either Mainland China or Taiwan. See Appendix A Table 1, which describes the outline of the CFP's four-year plan.

A total of 135 participants from the CFPs nationwide participated in this study voluntarily. Their majors spanned from liberal arts to technical and professional fields, including

but not limited to Business, Computer Science, Education, Engineering, and Political Science (see Table 1 for a summary of distribution of participants in terms of gender, heritage status, academic year, language background, as well as overseas experience).

Table 1. A Summary of Distribution of Participants in terms of Gender, Academic Year, Language Background

Category	**Subcategory**	**Count**	**Percentage (%)**
Gender	Male	41	30.37%
	Female	86	63.70%
	Other/Prefer not to disclose	8	5.93%
Academic Year	Freshmen	30	22.22%
	Sophomores	41	30.37%
	Juniors	30	22.22%
	Seniors	34	25.19%
Language Background	Mandarin/Dialects	17	12.59%
	Other	118	87.41%
International Exposure	No Overseas Experience	63	46.67%
	Visited Chinese Speaking Country (Non-academic)	24	17.78%
	Studied Abroad in Chinese Speaking Areas	48	35.56%

Instrument

This study attempts to verify the underlying constructs that represent the FLCAS developed by Horwitz et al. (1986) in the context of CFL. A questionnaire is used to collect students' demographic information, language learning competence, and FLA level. The questionnaire is based on FLCAS developed by Horwitz et al. (1986) with additional items on demographic information (see Appendix B). The first portion of the questionnaire includes 13 items, including participants' demographic information such as age, gender, year in college, native language, prior exposure to Chinese, study abroad experience, etc. The second portion includes five items of language learning self-efficacy. This self-efficacy scale is adopted from and previously validated by Martin et al. (2020). A sample item is, "*I am confident about participating in course activities*", rated on a five-point Likert scale, ranging from strongly agree (5 points) to strongly disagree (1 point). The third portion includes the 33-item self-report measure of FLCAS (Horwitz et al., 1986) which is also rated on a

five-point Likert scale. Items in FLCAS are anxiety-provoking situations, and participants will be asked to rate how they felt. For example: "*I feel nervous when I know that I'm going to be called on in language class*". This anxiety scale was confirmed by Horwitz (1986) to have high reliability and validity (Cronbach's alpha = .93; test-retest reliability: r = .83. Horwitz et al., 1986; Tóth, 2010) and has been widely used in previous research (Aida, 1994; Elkhafaifi, 2005). FLCAS was modified to measure anxiety specific to Chinese language learning in the U.S. The word "foreign language" in the original version of the scale was replaced with "Chinese language".

Data Collection

This study adhered to ethical guidelines for research with human subjects, receiving approval from the Institutional Review Board. Chinese Flagship Program directors, coordinators, and teachers across the U.S. were contacted via email for permission to survey their students. Upon approval,

students received emails with research details, consent forms, and a survey link. The 10-month data collection involved a 15-minute survey completed on Qualtrics. Of 160 contacted students, 135 completed the survey. Participants' rights to voluntary participation, confidentiality, and withdrawal were emphasized. Confidentiality measures were strictly enforced, and participants were offered a $10 gift card. Data analysis was conducted using R Studio 2024.

Analytic Plan

We adopted a Bayesian approach due to its flexibility and capacity to integrate prior knowledge, making it ideal for analyzing small sample sizes and complex models necessary for validating the FLCAS and identifying anxiety predictors among Chinese language learners. To address our first research question, we assessed multiple forms of validity for the FLCAS. Expert instructors from CLFP were invited to review the scale's items for face and content validity. For construct validity, we conducted BCFA using the 'blavaan' package in R (Merkle & Rosseel, 2018), hypothesizing a

two-level model where a higher-order 'Foreign Language Anxiety' factor influences three first-order latent constructs: Test Anxiety, Fear of Negative Evaluation, and Communication Apprehension. Each FLCAS item was assigned to its corresponding factor based on theoretical expectation, and priors were adjusted using prior predictive checks. We further examined convergent validity by correlating these anxiety constructs with self-efficacy measures, ensuring that the FLCAS aligns with established self-efficacy metrics.

For our second research question, we utilized a Bayesian multilevel model to explore how demographic variables such as heritage status, gender, year in college, and self-efficacy influence FLA across different Flagship Programs. Convergence was evaluated using standard Bayesian diagnostics, including R-hat statistics, effective sample size, and trace plots. Model fit was assessed with Bayesian R^2 and leave-one-out cross-validation (LOO-CV),

and parameter estimates were interpreted using 89% credible intervals (CIs) and the probability of direction (pd). This modeling included standardizing predictors to facilitate efficient model fitting and interpretation, capturing intra-program similarities and inter-program variations. Detailed statistical methodologies, including detailed description of our Bayesian method, selection of informative priors and the specifics of the Markov Chain Monte Carlo sampling, and other methodological details, are provided in Appendix C.

Results

RQ1 Result

Face and Content Validity

After the items were developed, the instrument was sent to a Chinese language expert to identify face and content validity. The Validation Rubric for Expert Panel was designed to measure face validity, construct validity, and content validity and was used to receive feedback from experts (Simon &

White, 2013). This rubric included items on clarity, wordiness, negative wording, overlapping responses, balance, use of jargon, appropriateness of responses listed, use of technical language, application to practice, and relationship to the problem. The instrument and review rubric were sent to two Chinese language experts to identify face and content validity. The experts agreed on the categories.

Reliability Analysis

To evaluate the internal consistency of the scale, we conducted a Bayesian reliability check to examine the interrelationship among its sub-scales. Reliability was assessed using both Cronbach's alpha and McDonald's omega. Reporting both indices provides a more comprehensive assessment of scale reliability: Cronbach's alpha remains the most widely used coefficient in language research, while omega is often considered a more robust estimate because it does not assume tau-equivalence among items. As shown in Table 2, the three constructs—Test Anxiety (TsA), Fear of

Negative Evaluation (NgE), and Communication Apprehension (CmA)—revealed high levels of internal consistency. Specifically, TsA exhibited a Cronbach's alpha of 0.90 and omega value of 0.91, indicating robust internal consistency. Similarly, NgE showed an alpha of 0.87 and omega values of 0.88, while CmA reported an alpha of 0.90 and omega value around 0.90. The high internal consistency implies that the items within each construct cohesively measure the same underlying concept, thereby providing confidence in the reliability of these constructs as tools for assessing various dimensions of anxiety.

Table 2. Internal Consistency Test Results

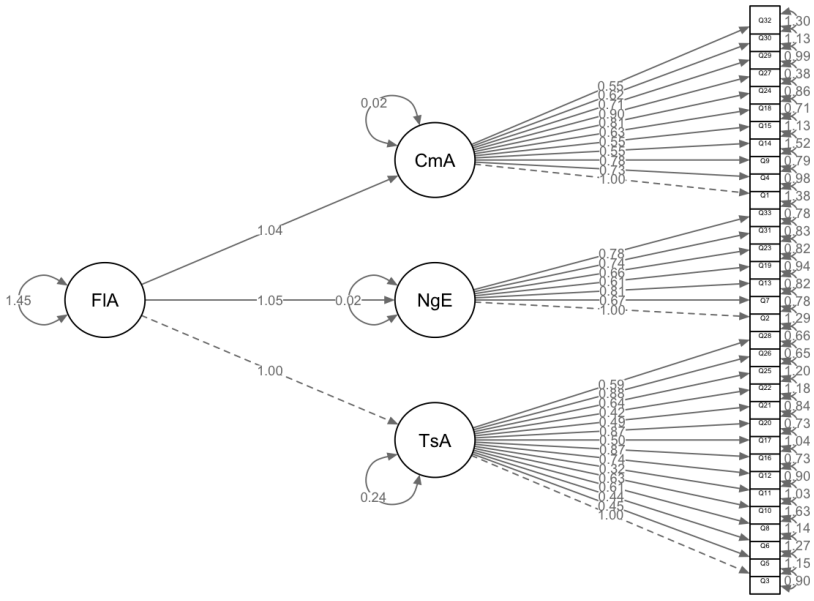
Construct	**Cronbach's Alpha**	**Omega**
Test Anxiety (TsA)	0.896	0.905
Fear of Negative Eval (NgE)	0.873	0.876
Communication Apprehension (CmA)	0.895	0.894

Construct Validity

We examined the construct validity through a two-level, three-factor model. We adjusted our priors to reflect

substantial expertise and model expectations through a sequence of prior predictive checks. We fitted BCFA model variants to assess the validity of the FLCAS constructs. Path analysis was then used to examine the relationship among variables through factor loadings, to understand the strength and significance of the connections. Additionally, the model's structure is depicted in Figure 1. The diagram presents a BCFA path diagram with observed variables loading onto three first-order factors (CmA; NgE; TsA), which are in turn explained by a single higher-order factor (FlA). Factor loadings, variances, and inter-factor correlations are depicted.

Figure 1. Path Diagram of the BCFA on Actual Data



Then we used fit indices to examine the model fit. We used several fit indexes to confirm the model fit, including Bayesian root mean square error of approximation (BRMSEA), and Bayesian Gamma Hat (BGammaHat). This model presents acceptable evidence of model fit. Considering the substantial factor loadings, excellent convergence diagnostics (Rhat =1), the BRMSEA value (0.09), and

BGammaHat (0.79) that indicates an acceptable fit, we consider that our model achieves a moderate level of fit to the data.

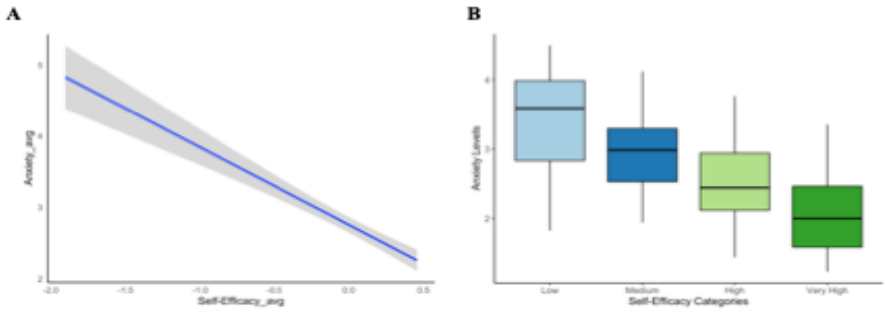
Convergent Validity

To evaluate the convergent validity of the FLCAS within the context of college Chinese language learners, we examined the correlations between the FLCAS factors, and five items designed to measure self-efficacy toward language learning. Factor scores for FLA and self-efficacy were extracted for further analysis. The results of this analysis provide critical insights into the degree to which FLCAS accurately captures the constructs of FLA and their alignment with self-efficacy.

Our Bayesian regression analysis illustrates a strong negative relationship between FLA and self-efficacy. Specifically, one-unit difference in self-efficacy was associated with a 1.49-unit difference in FLA levels, with a standard error of 0.17 and an 89% CI ranging from -1.82 to -1.16. This negative correlation emphasizes the strong convergent

validity of the FLCAS for our study population. Figure 2, panel A illustrates that students with lower self-efficacy tend to have higher FLA levels.

Figure 2. Correlation between FLA and Self-Efficacy



RQ2 Result

To answer question 2, we started by plotting the raw anxiety data against raw self-efficacy score. The self-efficacy scores are divided into quartiles, with each bin—Low, Medium, High, and Very High—representing a quarter of the data based on self-efficacy. As shown in Figure 3, panel B, the median FLA level appears to decrease as the self-efficacy category increases, with the 'Very High' self-efficacy group

exhibiting the lowest median FLA level, suggesting a potential inverse relationship between self-efficacy and FLA levels.

We then fitted the Bayesian multilevel model[1]. The baseline model included only the program effect. In the next step, we added students' demographic variables (gender, heritage status, year in college) into the model. Finally, we expanded the model further by incorporating language learning factors (language learning time, overseas experience, and self-efficacy) into the model. All models converged successfully, with no divergent transitions observed after warm-up. They demonstrated good evidence of convergence via visual examination of trace-plots. All of the estimated parameters had effective sample sizes greater than 10% of the total sample size; Monte Carlo standard errors that were less

[1] To support readers who may be less familiar with Bayesian statistics, we offer the following brief interpretive guide. An 89% credible interval (CI) represents the range within which the true value of a parameter is likely to fall with 89% certainty, given the observed data and model assumptions. If this interval does not include zero, the effect is considered more likely to be meaningful. The probability of direction (pd) reflects the certainty regarding the direction (positive or negative) of an effect. For instance, a pd of 99% indicates a 99% probability that the effect is truly in the estimated direction. Together, these values help translate model outputs into meaningful implications for instructional practice—for example, a negative estimate with high pd suggests that increases in the corresponding predictor (e.g., self-efficacy) are associated with lower FLA.

than 10% of the posterior standard deviations; and R-hat statistics are all close to 1. Visual inspection of predicted probability plots showed strong congruence with our actual, observed data. Findings from the final model were robust against a range of specified priors for the coefficient of interest.

Next, we evaluated the fit of several models using R^2 values, which quantify the proportion of variance explained by each model. The models varied significantly in their effectiveness, as evidenced by the estimated R^2 values and their corresponding 89% CIs. We used the LOO-CV to compare the predictive power and accuracy of models. The elpd_diff being 0 indicates that the model has the highest estimated predictive accuracy among those compared, since the differences in expected log predictive density (elpd) are calculated relative to it. An se_diff of 0 indicates that the standard error of this difference in elpd is also the lowest, suggesting high certainty in this model's predictive power.

As shown in Table 3, Model 0 (m_0) was a baseline model aiming to establish a foundational understanding of the constructs. It showed a minimal R^2 estimate of 2%, with an 89% CI ranging from approximately 0 to 0.08. This suggests a low explanatory power, which was further supported by a substantial negative elpd_diff in the LOO-CV scores, indicating poor predictive performance relative to other models.

Table 3. Model Progression Matrix

Model	(Intercept)	Demograp	Language Learning information			Est of R²	Est. Error	89% CI	LOO-CV	
		Linear	Linear	Quadratic	Random				elpd_diff	se_diff
m_0	✓					0.02	0	[0.00, 0.08]	-33	8.1
m_1	✓	✓				0.15	0.1	[0.06, 0.29]	-32	7.3
m_2	✓	✓	✓			0.47	0.1	[0.10,0.53]	-2.1	2.7
m_2a	✓	✓	✓	✓		0.47	0	[0.10, 0.53]	-3.1	2.9
m_2b	✓	✓	✓	✓		0.48	0	[0.10, 0.54]	-0.8	1.5
m_2c	✓	✓	✓	✓	✓	0.5	0	[0.10, 0.56]	0	0

Note. In m_2a, the quadratic effect is for language learning time, in m_2b, the quadratic effect is for self-efficacy.

Model 1 (m_1) introduced the relationship between demographic variables and FLA. It exhibited a moderate increase in explanatory power with an R^2 estimate of 15% and an 89% CI from 0.06 to 0.29. It showed better model fit and predictive accuracy, with a reduced elpd_diff compared to m_1.

Model 2 (m_2) expanded model 1 by adding language learning information into the model. It demonstrated significant improvements in explanatory power and predictive performance over the earlier models. The R^2 estimate for m_2 was 47%, with an 89% CI ranging from 0.10 to 0.53. This showcases a robust ability of m_2 to account for a large proportion of the variance within the dataset. To account for the non-normal distribution and heavy-tailed nature of language learning time, we added a non-linear term to m_2a. This adjustment does not improve in model fit and predictive accuracy with R^2 but a higher elpd_diff compared to m_2.

Therefore, we chose to drop the non-linear term of language learning time.

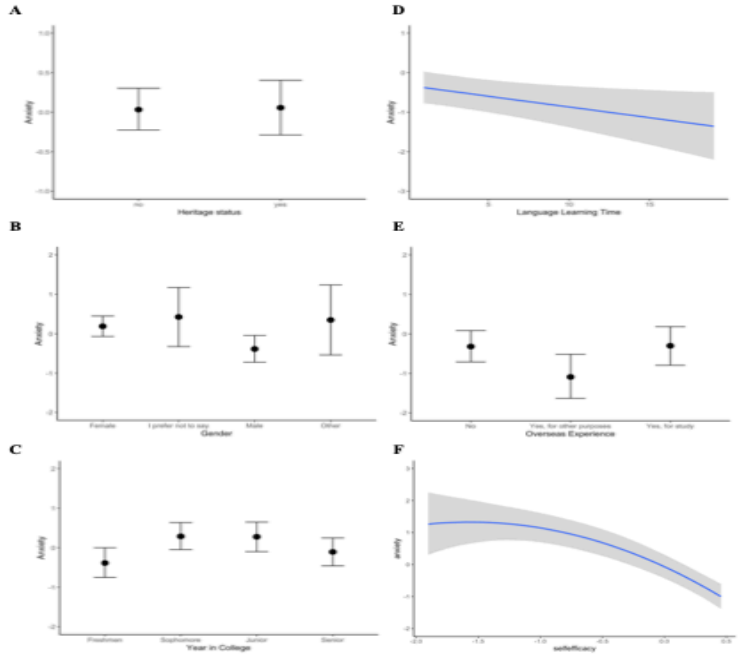
The advanced models, m_2b and m_2c, included non-linear effects for self-efficacy to address a ceiling effect, where increases in language learning self-efficacy no longer corresponded to proportional changes in FLA, potentially capturing a more complex relationship. Model m_2b integrated these non-linear features, model m_2c included random slopes for self-efficacy, reflecting program-specific variations. Model m_2c demonstrated the highest R^2 of 50% and with the narrowest 89% CI ranging from 0.10 to 0.56. Its predictive accuracy was also supported by the most favorable LOO-CV results.

Table 4. FLA Level by Demographic variables: Heritage Status, Gender and Year in College

	Parameter	Median	89% CI	pd
Heritage Status	(Intercept: No)	0.03	[-0.23, 0.30]	58.29%
	Yes	0.02	[-0.33, 0.37]	54.14%
Gender	(Intercept: Female)	0.19	[-0.09, 0.46]	87.63%
	I prefer not to say	0.24	[-0.52, 0.98]	69.74%
	Male	-0.57	[-0.91, -0.24]	99.69%
	Other	0.16	[-0.72, 1.04]	61.11%
Year in College	(Intercept: Freshmen)	-0.39	[-0.75, 0.00]	94.50%
	Sophomore	0.67	[0.26, 1.09]	99.52%
	Junior	0.66	[0.21, 1.10]	98.88%
	Senior	0.28	[-0.17, 0.72]	83.91%

Note. Controlled year in college for learning time; Controlled YC and heritage for overseas experience; controlled gender, overseas, learning time, and heritage status for self-efficacy.

Figure 3. FLA Level Based on Students Demographic and Language Learning Information



Note. All panels show 89% CIs. Panel A: FLA levels by heritage status, with similar medians between heritage and non-heritage learners; Panel B: FLA levels across genders; males exhibit the lowest anxiety; Panel C: Distribution of FLA by academic year, from freshmen to seniors; Panel D: Trend of FLA levels relative to language learning time; Panel E: FLA variations by overseas experience type: none, for non-academic purposes, and for study; Panel F: Trend showing decreasing FLA with increasing self-efficacy.

Next, we introduce how each factor predicts FLA one by one. As shown in Table 4 and Figure 3, panel A, the data shows that heritage students exhibit a slightly higher level of FLA,

with a wide 89% CI (-0.33, 0.37), and a pd of 54.14%, reflecting strong uncertainty that this effect is positive.

As shown in Table 4 and Figure 3, panel B, the data shows that male students are associated with a notably lower level of FLA, with a clear and negative effect (median: -0.57, 89% CI: [-0.91, -0.24]), and a high pd of 99.64%, reflecting strong certainty that this effect is negative. Students who identified as 'I prefer not to say' and 'other' showed a higher level of FLA, 0.22 and 0.16, respectively. However, the pd values for 'I prefer not to say' (68.71%) and 'Other' (61.11%) suggest that the predictions for these gender categories are less certain.

As shown in Table 5 and Figure 3, panel C, as students' progress through their academic careers, the FLA levels fluctuate. Notably, sophomore (median: 0.67, 89% CI: [0.26, 1.09]) and junior (median: 0.66, 89% CI: [0.21, 1.10]) students show a rise in FLA, and with level of confidence (99.50% and 99.17% respectively). This upward trend in FLA

appears less distinct among seniors (median: 0.28, 89% CI: [-0.17, 0.72]), with the respective pd values (84.3%) indicating reduced confidence compared to sophomores.

Table 5. FLA Level by Language Learning Factors: Learning Time; Overseas Experience; Self-Efficacy

	Parameter	Median	89% CI	pd
Learning time	(Intercept)	-0.23	[-0.73, 0.09]	89.89%
	Learning time	-0.05	[-0.10, -0.01]	96.88%
Overseas	(Intercept: No)	-0.32	[-0.71, 0.08]	90.17%
	Oversea: Yes, for other purposes	-0.77	[-1.24, -0.28]	99.53%
	Oversea: Yes, for study	0.01	[-0.38, 0.41]	51.93%
Self-efficacy	(Intercept)	0.07	[-0.27, 0.42]	63.44%
	Self-efficacy	-1.79	[-2.18, -1.38]	100%
	Self-efficacy squared	-0.57	[-0.96, -0.18]	98.92%

Note. Controlled year in college for learning time; Controlled YC and heritage for overseas experience; controlled gender, overseas, learning time, and heritage status for self-efficacy.

As shown in Table 5 and Figure 3, panel D, the data revealed a negative relationship between learning time and FLA. The magnitude of the effect is small with a median of -0.05, but

the 89% CI [-0.10, -0.01] and 96.88% pd indicate a substantial certainty of the reverse relationship, further supported by the results presented in Table 8 and Figure 9.

Overseas experience seems to buffer FLA for students traveling for non-study purposes (Median: -0.77, 89% CI: [-1.24, -0.28]), with a pd of 99.53%, indicating a substantial certainty of a negative effect on FLA. By contrast, studying abroad shows minimal effect on FLA (Median: 0.01, 89% CI: [-0.38, 0.41]) with a pd of 51.93%, suggesting near-equal probabilities of increase or decrease in FLA, which warrants cautious interpretation (See Table 4 and Figure 3, panel E).

For self-efficacy toward language learning, the data showed a strong negative relationship with FLA (Table 4 and Figure 3, panel F), suggesting that as self-efficacy increases, anxiety decreases substantially (Median: -1.79, 89% CI: [-2.18, -1.38]), and a pd of 100%. The non-linear effects of self-efficacy (Median: -0.57, 89% CI: [-0.96, -0.18]) also reveal

a substantial pd of 98.92%, indicating a high probability that increases in self-efficacy led to diminishing returns in FLA. We interpret the quadratic term in Figure 3, panel F.

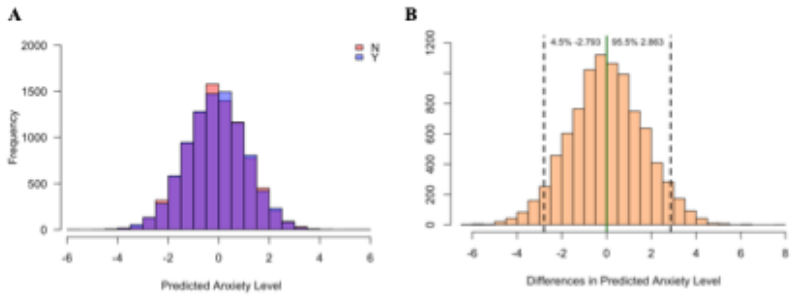
To explain these patterns further, we introduced posterior predictive cases. This method involves simulating the FLA predictions for students who embody typical characteristics identified in the model—yet were not part of the original dataset. It represents the probability that a future observation will exceed the existing data, given the model (Kruschke, 2014). By doing this, we can visualize how specific variables, like gender, influence outcomes when all other variables are held constant. This is crucial for confirming the model's applicability and reliability, particularly when extended to hypothetical scenarios that reflect the real world.

Predicted FLA Levels of Posterior Predictive Cases

As shown in Figure 4, our posterior predictive case analysis indicated that the median predicted FLA level for heritage students was, on average -0.9 [89% CI: -2.13,1.86], while for non-heritage students, it was -0.1 (89% CI: -1.93 to 2.89).

The model predicts that approximately 50.2% of cases showed higher FLA for non-heritage students. The comparisons between the two categories showed minor differences.

Figure 4. Posterior Predictive Case of Different Heritage Status

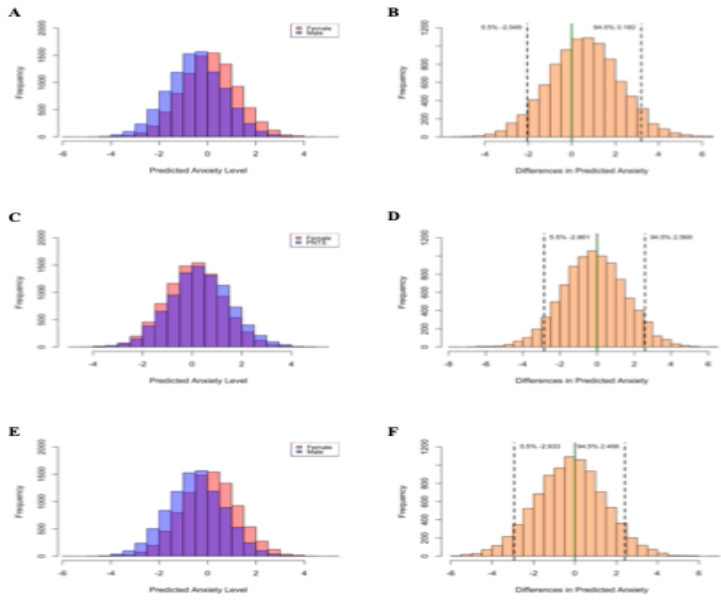


Note. Panel A predicted FLA levels among heritage and non-heritage students. Panel B shows the distribution of differences in predicted FLA level between these two groups of students. This histogram details the spread and the central tendency of difference.

As shown in Figure 5, the posterior predictive case analysis indicated that the predicted FLA level for female students was, on average 0.07, while for male students, it was 0.58 (89% CI: -1.93 to 2.89). The model predicts that approximately 64.04% of cases showed higher anxiety for

females. Yet this pattern does not extend to other categories, where differences are minimal.

Figure 5. Posterior Predictive Case of Different Genders

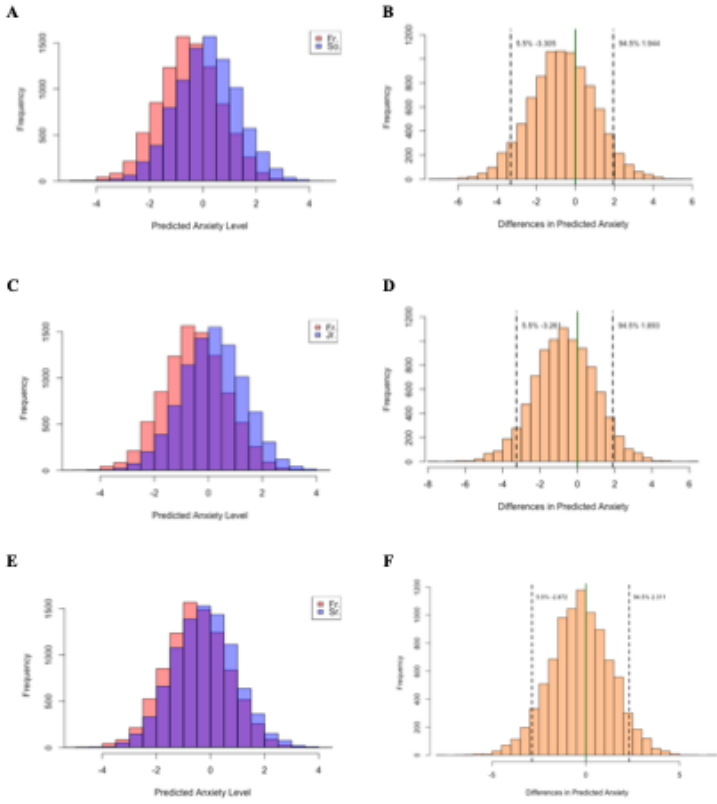


Note. Panel A shows the predicted FLA levels among female and male students, while Panel B illustrates the distribution of differences in predicted FLA between these groups. Panel C and D depict the same comparisons between female students and those who prefer not to disclose their gender. Finally, Panel E and F present the predicted FLA levels and differences between female students and those identifying as "other".

When examining academic year in college, freshmen are predicted to have lower FLA with a posterior median of

-0.49, compared to sophomore (0.17) and junior (0.16). The difference in FLA is -0.69 (89% CI: -3.27 to 2.0) for sophomore and -0.69 (89% CI: -3.34 to 2.0) for junior, with 33.69% and 33.63% probabilities, respectively, of observing higher anxiety for freshmen. Compared to seniors, freshmen show a median predicted anxiety difference of -0.31 (89% CI: -2.92 to 2.30), with a 42.18% probability of increased anxiety in freshmen. This suggests a lower probability of increased FLA among freshmen compared to other academic years. See Figure 6.

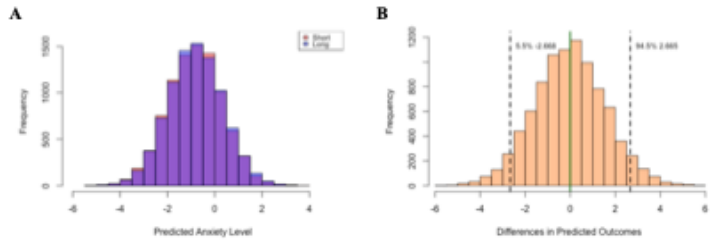
Figure 6. Posterior Predictive Case of Different Year in College



Note. Panel A and B display the predicted FLA levels for freshmen and sophomores, with Panel B showing the difference in predicted FLA between these two groups. Panel C and D present the predicted FLA levels for freshmen and juniors. Panel E and F present FLA levels for freshmen and seniors. These plots represent the distribution of FLA scores for each class year, emphasizing the range and central tendencies within these student groups, and they compare the predicted FLA level distributions between freshmen and sophomores/juniors/seniors.

When examining the influence of language learning time on FLA, there is a median difference of 0.02 between short and long learning time groups, with a 50% probability that longer learning time demonstrates slightly lower anxiety, which indicates the uncertainty of this relationship (See Figure 7.)

Figure 7. Posterior Predictive Case of Different Language Learning Time

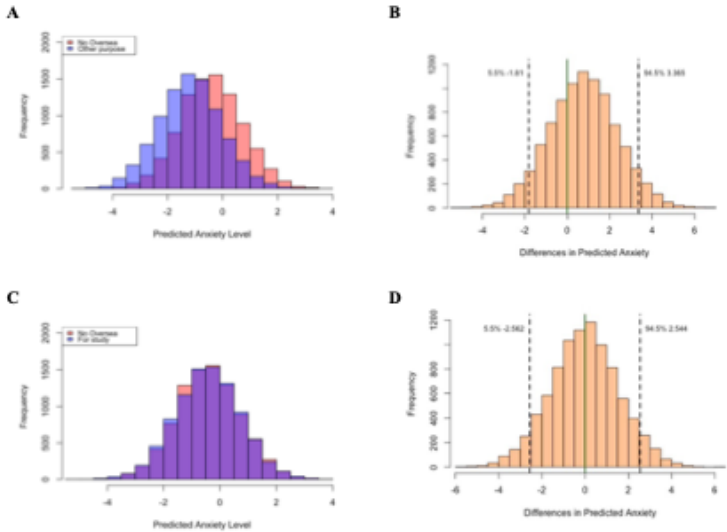


Note. Panel A shows the distribution of predicted FLA levels for different lengths of language learning time, highlighting the range and central tendencies. Panel B presents the distribution of differences in predicted FLA levels, comparing students with short and long language learning times.

In terms of overseas experience, students without such experience have a median predicted FLA level of -0.43, while those who traveled for non-academic purposes show a lower median level of -1.18. The posterior difference is 0.77 (89%

CI: -1.86 to 3.30), with 68.69% probability of higher FLA among students without overseas experience. It suggests a potential impact on FLA levels. When comparing students without overseas experience to those who traveled for study purposes, the difference is marginal at 0.02 (89% CI: -2.59 to 2.53), and with probabilities evenly split at 49.51%, indicating subtle influences of study abroad experience on student FLA. See Figure 8 for posterior predictive case of overseas experience.

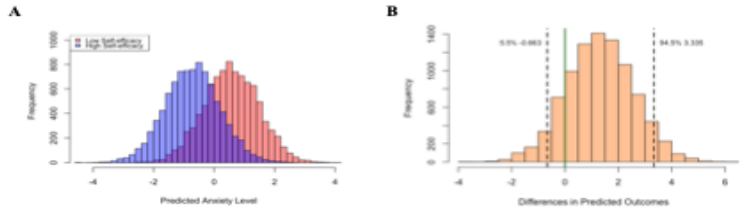
Figure 8. Posterior Predictive Case of Overseas Experience



Note. The histograms on the left depict predicted FLA levels among students with no overseas experience and overseas experience for purposes other than study (top) and study abroad (bottom). The figures on the right are the distribution of differences in predicted FLA levels based on their study abroad experience.

Lastly, the estimate of self-efficacy demonstrates a clear trend (See Figure 9). Specifically, there's a posterior median difference of 1.3 between low and high self-efficacy groups, with an 84.98% probability of observing lower FLA as self-efficacy increases (89% CI: -0.74, 3.32), indicating a strong negative association between self-efficacy and FLA.

Figure 9. Histograms of Predicted FLA Levels and Differences by Self-Efficacy



Note. Panel A shows the distribution of predicted FLA levels for students with low and high self-efficacy. Panel B presents the distribution of differences in predicted FLA outcomes between these two groups. The dotted lines indicate the median difference, highlighting the influence of self-efficacy on FLA levels.

We conducted the robustness check by fitting the models with wider priors. The direction and the magnitude of the estimates remain the same.

Discussion

Scale Validation

Our validation process–including face, content, construct, and convergent–alongside the internal consistency analyses, substantiates the FLCAS as an effective tool for measuring FLA in the U.S. CFL context. The high internal consistency reliabilities of the FLCAS across all 33 items and in each component support earlier studies where Cronbach's alpha coefficients of the FLCAS were satisfactory (Chen et al., 2024; Li et al., 2021; Liu, 2006). We also found that FLA is a multi-dimensional construct. This aligns with studies that examined the FLCAS in different contexts, such as Aida (1994), who explored FLA in learning Japanese in the U.S. as a foreign language, Park (2014) with English learners in

Korea, and Zhang and Lai (2023) who studied Chinese EFL students.

Our study makes two important contributions to the field. First, our study extends the understanding of FLA by validating the FLCAS scale and demonstrating that it is potentially a two-level, three-construct factor. The three constructs are communicative anxiety, test anxiety, fear of negative evaluation, with a higher-level structure (FlA) above these three constructs. This finding indicates that FLA is more complex than previously understood. Due to this, it is important that future researchers aim to incorporate this complexity to existing theories. Another contribution is the application of BCFA to assess construct validity and employing posterior predictive simulation to compare prototypical cases. This methodological choice not only enhances the rigor of our validity assessments but also sets a precedent for future research in the field.

This validation process provides a level of support for our theoretical model. While the fit is not perfect, the strong relationship between observed variables and their respective latent factors, alongside satisfactory model convergence, lends credence to the theoretical constructs under investigation. It should be noted that the BRMSEA (0.09) and BGamma Hat (0.79) indicate only moderate model fit, which may be partly attributable to the limited sample size (Shi et al., 2019). Future research may explore model refinements to address the moderate fit and further validate the theoretical framework, with a particular emphasis on the implications of sample size.

Predictors of FLA

Our study utilized a Bayesian approach to assess the impacts of various socio-cultural factors, including demographic factors (heritage status, gender, academic year) and language learning experiences (overseas experience, self-efficacy, language learning time) on Chinese language learning anxiety. The analysis revealed that these factors collectively account for about 50% of the variance in anxiety levels among

college-level Chinese language students, as indicated by R^2 values. This substantial explanatory and predictive power are a notable improvement over the baseline model, which only considered program-specific random effects.

Delving into the demographic factors, we observed that students' gender and academic year largely influenced FLA levels, while heritage status did not emerge as a strong predictor. Our findings indicated no significant difference in FLA between heritage and non-heritage students, contrasting with studies like Luo (2013) which suggested lower FLA among heritage learners with a Chinese language background. Typically, such reduced FLA might be attributed to these learners' greater exposure to the language and cultural context (Luo, 2014). This discrepancy may be attributed to the unique structure and objectives of CFPs, which offer intensive language training and cultural immersion, potentially equalizing anxiety levels across different heritage statuses. The rigorous curriculum and comprehensive support systems in

these programs might offset the typical advantages of heritage learners, resulting in similar FLA levels among all students. Further research could explore how the specific educational strategies and support mechanisms within Flagship programs influence FLA among diverse learner groups.

Similarly to Arnaiz and Guillen (2012) and Dewaele et al. (2016), our study found that students who identified as male typically reported lower FLA than those who identified as female. This challenges previous research that suggested a minimal impact of gender on FLA, such as Jiang and Dewaele (2020) and Zhao and Whitchurch (2011). Moreover, our research brings a new perspective into this conversation by adding more categories beyond the traditional binary of female and male. Our results show that students who identified as 'I prefer not to say' and 'other' demonstrated a higher level of FLA than female students. These findings should be interpreted with caution, though, due to the weak pd values, likely caused by limited data in these categories.

Nevertheless, they do challenge oversimplified interpretations of gender differences in FLA studies (Piniel & Zólyomi, 2022). Thus, future research may benefit from examining how societal and cultural constructs of masculinity and femininity shape learning dynamics within intensive Chinese language programs in the U.S.

We also found variability in FLA levels across different years in college. Sophomores and juniors exhibited higher levels of FLA compared to freshmen, whereas seniors tended to report lower FLA. These findings diverge from Onwuegbuzie et al.'s (1999), who noted a linear increase in FLA with academic progression, and Luo (2014), who suggested uniform FLA levels across all academic standings. The heightened FLA among sophomores and juniors could be attributed to the pressures of preparing for overseas study and proficiency tests. These academic milestones challenge students both culturally and academically. Adopting a new culture or navigating the clash between one's own culture and

a new one might further exacerbate this anxiety (Marginson, 2024). Therefore, these years are pivotal and may require additional attention to help reduce students' stress and promote a healthier and more productive academic experience.

Regarding language learning factors, our study found that overseas experience and self-efficacy greatly impacted FLA levels, while we observed only a slight difference among students with varying lengths of exposure to Chinese. Those who had longer exposure experienced marginally lower levels of FLA. This finding partially aligns with Dewaele (2007), who noted increased proficiency and confidence with more instruction and practice over time. Nevertheless, Price (1991) and MacIntyre and Gardner (1994) suggested that more time spent in learning does not necessarily equate to lower FLA or higher achievement. The discrepancy can be attributed to the necessity of high-quality, engaging instruction. Merely increasing the duration of exposure is insufficient; the effectiveness of instruction plays a crucial role. This is also

supported by Williams and Andrade (2008) and Schunk (1989), who emphasize the importance of developing effective study habits and coping mechanisms to manage rigorous programs successfully. These insights demonstrate the complex relationship between FLA and learning time, highlighting the importance of both the quantity of learning time and the quality of learning experiences in shaping students' adaptability in high-pressure environments. Therefore, it is essential for language programs to focus not only on increasing the amount of instructional time but also on enhancing the quality of education provided.

Overseas experiences appear to act as a buffer against FLA, especially for students who have traveled for non-academic purposes, leading to lower levels of FLA compared with those who have not traveled. Non-academic experiences may be particularly effective because they often involve less evaluative pressure than classroom settings, provide more opportunities for spontaneous interaction with

native speakers, and allow learners greater freedom to explore language and culture in authentic contexts. This aligns with studies by Jiang and Dewaele (2020), Matsuda and Gobel (2004), and Thompson and Lee (2014), suggesting that authentic language practice and cultural immersion are key factors in reducing FLA.

The effectiveness of study abroad programs in reducing FLA is less conclusive. One explanation may be the distinct characteristics of the Flagship programs, particularly the test anxiety associated with post-study abroad exams, which are required to measure proficiency gains. Such academic requirements could potentially reduce benefits typically associated with language and cultural immersion of overseas exposure. Additionally, different curriculum expectations in study abroad programs in Taiwan or mainland China may also contribute to this uncertainty. For instance, there might be no syllabus provided at the beginning of the program, preventing students from understanding the holistic structure of the curriculum. This lack of clarity can affect

students' confidence in navigating course requirements. Furthermore, differences in pedagogical approaches may further affect students' experiences: while instruction with a student-centered approach can encourage active engagement, teacher-centered approaches have been associated with negative emotions such as loss of self-confidence and feelings of isolation (Chen & Yu, 2019). Further research could focus on how various aspects of study abroad programs influence FLA, aiding in the design of more effective educational experiences and support systems for language learners.

Finally, the result demonstrates the essential role of self-efficacy in predicting FLA variations in the CFL setting. This finding is supported by several previous studies, such as Piniel and Zólyomi (2022), Akkakoson (2016), Dewaele et al. (2008), and Thompson & Khawaja (2015), which have consistently documented a negative correlation between higher self-efficacy and FLA. This relationship is further corroborated by recent research exploring how self-efficacy

influences emotions and attitudes in language learning (Pawlak & Csizér, 2022; Wang et al., 2022; Wu et al., 2022). Our research contributes to the field by suggesting that the relationship between FLA and self-efficacy might be non-linear, indicating that self-efficacy may not have a uniform effect across all learners and underscoring the need for tailored interventions.

This trend could be attributed to our participants, who generally exhibit high self-efficacy, likely due to their commitment to a rigorous four-to-five-year program in Chinese language studies. Moreover, the socio-cultural context, including the cultural values and expectations around education and success, plays a crucial role in shaping self-efficacy. The supportive learning environment featured by accessible faculty, peer collaboration, and community engagement opportunities provided by the programs may also help build students' confidence (DiFrancesca, 2020), and make them feel more comfortable in class (White et al., 2023). This finding emphasizes the need to explore additional

social cultural factors and interventions for managing and reducing FLA, especially among students who are already highly competent. It suggests that enhancing self-efficacy alone may not be sufficient for these learners to manage or reduce FLA. Therefore, further research should explore additional socio-cultural factors, such as family expectations for heritage learners, requirements under a different educational system, student-teacher and peer dynamics, as well as specific interventions tailored to these highly capable learners.

Despite its strengths, this study has some limitations. First, due to sample size constraints, our study included only six socio-cultural variables and did not consider interactions between different variables. Future studies could replicate this research with a larger sample size and include a broader range of socio-cultural variables. Secondly, this study is primarily focused on students from CFPs, which are characterized by rigorous curricula, selective admissions, and structured

overseas experience. Their unique features may limit how generalizable the findings are to students in non- CFPs or less intensive Chinese language programs. Moreover, the self-selecting and highly motivated nature of CFP participants may also influence levels of language anxiety and should be considered in future comparative studies. Future research could compare students in CFPs with those in regular Chinese classes to evaluate the extent to which these predictors of FLA hold across varied instructional contexts. Additionally, by implementing a longitudinal study, future research could explore the evolution of FLA over time and the effectiveness of intervention. Despite these limitations, our study provides valuable initial insights into FLA among Chinese language learners in an intensive learning program, setting a stage for more comprehensive future studies.

Conclusions

In this study, we have demonstrated that FLA is a multidimensional construct, encompassing test anxiety, fear of negative evaluation, and communication apprehension.

This finding not only supports Horwitz's (1986) original conceptualization but also extends it by validating these dimensions in the specific context of CFL. We further highlighted the importance of considering students' demographic and language learning factors such as heritage status, gender, academic year, learning time, overseas experiences, and self-efficacy in understanding FLA. Using advanced Bayesian methods, we have provided a robust analysis of the predictors of FLA and are the first to specifically explore FLA among students in the CFPs. This exploration into the intensive language learning context offers insights suggesting the need for dynamic models of FLA that account for complex interactions.

The results of this study offer important implications for practice. We suggest that researchers should incorporate socio-cultural elements into scale development to ensure comprehensive assessment tools that are sensitive to diverse learner contexts. Furthermore, we advocate for educators to

provide tailored language education, recognizing individual differences and the specific needs related to students' academic standings and overseas experiences. Our findings also point to the potential benefits of integrating non-academic overseas experiences into language curricula. At a minimum, this could involve increasing exposure to the language and culture through relaxed settings, such as cultural activities, festival celebrations, language clubs, or even interactive virtual reality environments. Administrators might consider modifying study abroad programs to include more cultural immersion activities, such as local homestays and community engagement projects. Additionally, providing pre-departure orientation sessions can sensitize students to potential challenges and reduce FLA, potentially improving language proficiency outcomes (Bensalem & Trevethan, 2024). Finally, while these implications stem from intensive Chinese language programs, they may also inform broader language learning contexts. Approaches that build self-efficacy, mitigate fear of negative evaluation, and increase

informal exposure to the target language hold promise for programs with varied structures, durations, and learner populations.

Given the evolving understanding of gender as a complex social construct, it is crucial for educators to approach gender differences with caution (Dewaele et al., 2016). Before implementing strategies aimed at reducing FLA among female students, further investigation into the magnitude of these differences is necessary. Adopting inclusive teaching strategies that accommodate a broad spectrum of gender identities can benefit all students. Moreover, the important role of self-efficacy in reducing FLA suggests targeted interventions could be beneficial. At the same time, it is important for educators and researchers to consider that the relationship between self-efficacy and FLA may be non-linear; while initial increases in self-efficacy might significantly reduce FLA, further increases may see diminishing returns. Identifying low self-efficacy beliefs can

be a first step towards promoting more positive and realistic attitudes to foreign language learning. Overall, this study provides new evidence on how demographic and learning factors shape FLA in intensive Chinese programs, offering a foundation for more nuanced interventions and pointing to the need for continued inquiry in this underexplored context.

References

Aida, Y. (1994). Examination of Horwitz, Horwitz, and Cope's construct of foreign language anxiety: The case of students of Japanese. *The Modern Language Journal*, *78*(2), 155–168. https://doi.org/10.2307/329005

Akkakoson, S. (2016). Speaking anxiety in English conversation classrooms among Thai students. *Malaysian Journal of Learning and Instruction*, *13*(1), 63–82.

Allen, H. W., & Herron, C. (2003). A mixed-methodology investigation of the linguistic and affective outcomes of summer study abroad. *Foreign Language Annals*, *36*(3), Article 3.

Bandura, A. (1988). Self-efficacy conception of anxiety. *Anxiety Research*, *1*(2), 77–98. https://doi.org/10.1080/10615808808248222

Bandura, A. (1997). *Self-efficacy and agency of change*. New York: Raven Press.

Bensalem, E., & Trevethan, R. (2024). Composition and implications of foreign-language anxiety among American

study-abroad students in Morocco. *International Journal of Applied Linguistics*, *34*(1), 46–62. https://doi.org/10.1111/ijal.12477

Chen, W., & Yu, S. (2019). Implementing collaborative writing in teacher-centered classroom contexts: Student beliefs and perceptions. *Language Awareness*, *28*(4), 247–267. https://doi.org/10.1080/09658416.2019.1675680

Cubillos, J. H., Chieffo, L., & Fan, C. (2008). The impact of short-term study abroad programs on L2 listening comprehension skills. *Foreign Language Annals*, *41*(1), 157–186. https://doi.org/10.1111/j.1944-9720.2008.tb03284.x

Dewaele, J.-M. (2007). Predicting language learners' grades in the L1, L2, L3 and L4: The effect of some psychological and sociocognitive variables. *International Journal of Multilingualism*, *4*(3), 169–197. https://doi.org/10.2167/ijm080.0

Dewaele, J.-M. (2010). *Emotions in Multiple Languages*. Palgrave Macmillan UK. https://doi.org/10.1057/9780230289505

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), Article 2. https://doi.org/10.14746/ssllt.2014.4.2.5

Dewaele, J.-M., Petrides, K. V., & Furnham, A. (2008). Effects of trait emotional intelligence and sociobiographical variables on communicative anxiety and foreign language anxiety among adult multilinguals: A review and empirical investigation. *Language Learning*, *58*(4), 911–960. https://doi.org/10.1111/j.1467-9922.2008.00482.x

DiFrancesca, H. (2020). Interactive, team-based learning environment enhances students learning and confidence. *The FASEB Journal*, *34*(S1), 1–1. https://doi.org/10.1096/fasebj.2020.34.s1.06671

Elkhafaifi, H. (2005). Listening comprehension and anxiety in the Arabic language classroom. *The Modern Language Journal*, *89*(2), 206–220. https://doi.org/10.1111/j.1540-4781.2005.00275.x

Foreign Language Training. (n.d.). United States Department of State. Retrieved April 26, 2024, from https://www.state.gov/foreign-language-training/

Gardner, R. C. (1985). Social psychology and second language learning: The role of attitudes and motivation. *(No Title)*. https://cir.nii.ac.jp/crid/1130000797160023680

Gardner, R. C., & Others, A. (1987). *The Role of Anxiety in Second Language Performance of Language Dropouts. Research Bulletin No. 657*. https://eric.ed.gov/?id=ED323774

Horwitz, E. (2001). Language anxiety and achievement. *Annual Review of Applied Linguistics*, *21*, 112–126. https://doi.org/10.1017/S0267190501000071

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, *70*(2), 125–132. https://doi.org/10.2307/327317

Jean-Marc Dewaele, Peter MacIntyre, Carmen Boudreau, & Livia Dewaele. (2016). Do girls have all the fun? Anxiety and enjoyment in the foreign language classroom. *Theory and Practice of Second Language Acquisition*, *2*(1), 41–63.

Jee, M. J. (2016). Exploring Korean heritage language learners' anxiety: 'We are not afraid of Korean!' *Journal of Multilingual and Multicultural Development*, *37*(1), 56–74. https://doi.org/10.1080/01434632.2015.1029933

Jiang, Y., & Dewaele, J.-M. (2020). The predictive power of sociobiographical and language variables on foreign language anxiety of Chinese university students. *System*, *89*, 102207. https://doi.org/10.1016/j.system.2020.102207

Kruschke, J. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan.* Academic Press.

Liu, M., & Yuan, R. (2021). Changes in and effects of foreign language classroom anxiety and listening anxiety on Chinese undergraduate students' English proficiency in the COVID-19 context. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.670824

Luo, H. (2013). Chinese language learning anxiety and its associated factors. *Journal of Chinese Language Teachers Association*, *48*(2), 109–133.

Luo, H. (2014). Foreign language speaking anxiety: A study of Chinese language learners. *Journal of the National Council of Less Commonly Taught Languages*, *15*(1), 99–117.

MacIntyre, P. D. (2002). Motivation, anxiety and emotion in second language acquisition. In P. Robinson (Ed.), *Individual Differences and Instructed Language Learning* (pp. 45–68). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.2.05mac

MacIntyre, P. D. (2017). An overview of language anxiety research and trends in its development. In C. Gkonou, M. Daubney, & J.-M. Dewaele (Eds.), *New Insights into Language Anxiety* (pp. 11–30). Multilingual Matters. https://doi.org/10.21832/9781783097722-003

MacIntyre, P. D., & Gardner, R. C. (1989). Anxiety and second-language learning: Toward a theoretical clarification. *Language Learning*, *39*(2), 251–275. https://doi.org/10.1111/j.1467-1770.1989.tb00423.x

MacIntyre, P. D., & Gardner, R. C. (1991). Language anxiety: Its relationship to other anxieties and to processing in native

and second languages. *Language Learning*, *41*(4), 513–534. https://doi.org/10.1111/j.1467-1770.1991.tb00691.x

Marginson, S. (2024). Student self-formation: An emerging paradigm in higher education. *Studies in Higher Education*, *49*(4), 748–762. https://doi.org/10.1080/03075079.2023.2252826

Martin, F., Stamper, B., & Flowers, C. (2020). Examining student perception of readiness for online learning: Importance and confidence. *Online Learning*, *24*(2), Article 2. https://doi.org/10.24059/olj.v24i2.2053

Matsuda, S., & Gobel, P. (2004). Anxiety and predictors of performance in the foreign language classroom. *System*, *32*(1), 21–36. https://doi.org/10.1016/j.system.2003.08.002

McDonald, R. P. (2014). *Factor analysis and related methods*. Psychology Press. https://doi.org/10.4324/9781315802510

Merkle, E. C., & Rosseel, Y. (2018). blavaan: Bayesian structural equation models via parameter expansion. *Journal of Statistical Software*, *85*(4). https://doi.org/10.18637/jss.v085.i04

Mihai, M., Albert, C. N., Mihai, V. C., & Dumitras, D. E. (2022). Emotional and social engagement in the English language classroom for higher education students in the COVID-19 online context. *Sustainability*, *14*(8), Article 8. https://doi.org/10.3390/su14084527

Nugent, M., & Slater, R. (2016). The Language Flagship: Creating expectations and opportunities for professional- level language learning in undergraduate education. In D. Murphy & K. E. Romaine (Eds.), *Exploring the US Language Flagship Program* (Vol. 1, pp. 9–28). Blue Ridge Summit: Multilingual Matters. https://doi.org/10.21832/9781783096107-005

Pajares, F., & Urdan, T. (2006). *Self-Efficacy Beliefs of Adolescents*.

Park, G.-P. (2014). Factor analysis of the foreign language classroom anxiety scale in Korean learners of English as a foreign language. *Psychological Reports*, *115*(1), Article 1. https://doi.org/10.2466/28.11.PR0.115c10z2

Pawlak, M., & Csizér, K. (2022). The impact of self-regulatory strategy use on self-efficacy beliefs and motivated learning behavior in study abroad contexts: The case of university

students in Italy, Poland and Turkey. *System*, *105*, 102735. https://doi.org/10.1016/j.system.2022.102735

Piniel, K., & Zólyomi, A. (2022). Gender differences in foreign language classroom anxiety: Results of a meta-analysis. *Studies in Second Language Learning and Teaching*, *12*(2), 173–203.

Price, M. L. (1991). *The subjective experience of foreign language anxiety: Interviews with highly anxious students [Extras]*. Prentice-Hall.

Saito, Y., Horwitz, E. K., & Garza, T. J. (1999). Foreign language reading anxiety. *The Modern Language Journal*, *83*(2), 202–218.

Santana, J. C., & Eccius-Wellmann, C. (2018). Gender Differences in Test Anxiety in High-Stakes. *Electronic Journal of Foreign Language Teaching*, *15*(1), 100–111.

Schunk, D. H. (1989). Self-efficacy and achievement behaviors. *Educational Psychology Review*, *1*(3), 173–208. https://doi.org/10.1007/BF01320134

Sparks, R. L., & Ganschow, L. (1995). A strong inference approach to causal factors in foreign language learning: A response

to MacIntyre. *The Modern Language Journal*, *79*(2), 235–244. https://doi.org/10.1111/j.1540-4781.1995.tb05436.x

Sung, K.-Y., & Li, X. (2019). Factors Influencing Chinese Language Learning Anxiety in the Classroom Setting. *New Waves-Educational Research and Development Journal*, *22*(2), 1–15.

Tallon, M. (n.d.). Heritage Speakers of Spanish and Foreign Language Anxiety: A Pilot Study. *Texas Papers in Foreign Language Education*, *15*(1).

Thompson, A., & Khawaja, A. (2015). Foreign language anxiety in Turkey: The role of multilingualism. *Journal of Multilingual and Multicultural Development*, *37*, 1–16. https://doi.org/10.1080/01434632.2015.1031136

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications* (pp. x, 195). American Psychological Association. https://doi.org/10.1037/10694-000

Ting, H.-L., & Sunarti. (2022). Language learning anxiety in Mandarin as foreign language classroom. *Asian Pendidikan*,

2(1), Article 1. https://doi.org/10.53797/aspen.v2i1.4.2022

Tóth, Z. (2010). *Foreign language anxiety and the advanced language learner: A study of Hungarian students of English as a foreign language*. Cambridge Scholars Publishing.

Toyama, M., & Yamazaki, Y. (2018). Exploring the components of the foreign language classroom anxiety scale in the context of Japanese undergraduates. *Asian-Pacific Journal of Second and Foreign Language Education*, *3*(1), 4. https://doi.org/10.1186/s40862-018-0045-3

Toyama, M., & Yamazaki, Y. (2021). Classroom interventions and foreign language anxiety: A systematic review with narrative approach. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.614184

Tran, T. T. T. (2012). A review of Horwitz, Horwitz and Cope's theory of foreign language anxiety and the challenges to the theory. *English Language Teaching*, *5*(1), 69–75.

Wang, M., Wang, H., & Shi, Y. (2022). The role of English as a foreign language learners' grit and foreign language anxiety in their willingness to communicate: Theoretical

perspectives. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.1002562

Watson, D., & Friend, R. (1969). Measurement of social-evaluative anxiety. *Journal of Consulting and Clinical Psychology*, *33*(4), 448–457. https://doi.org/10.1037/h0027806

Williams, K. E., & Andrade, M. (2008). Foreign language learning anxiety in Japanese EFL university classes: Causes, coping, and locus of control. *Electronic Journal of Foreign Language Teaching, 5*.

Wu, X., Yang, H., Liu, J., & Liu, Z. (2022). English use anxiety, motivation, self-efficacy, and their predictive effects on Chinese top university students' English achievements. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.953600

Xiao, Y., & Wong, K. F. (2014). Exploring heritage language anxiety: A study of Chinese heritage language learners. *The Modern Language Journal*, *98*(2), 589–611. https://doi.org/10.1111/modl.12085

Xu, W., Zhang, H., Sukjairungwattana, P., & Wang, T. (2022). The roles of motivation, anxiety and learning strategies in

online Chinese learning among Thai learners of Chinese as a foreign language. *Frontiers in Psychology*, *13*. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2022.962492

Yan, W., & Gaier, E. L. (1994). Causal attributions for college success and failure: An Asian-American comparison. *Journal of Cross-Cultural Psychology*, *25*(1), 146–158. https://doi.org/10.1177/0022022194251009

Zhang, S., & Lai, C. (2023). Investigation of factors underlying foreign language classroom anxiety in Chinese university English majors. *Applied Linguistics Review*. https://doi.org/10.1515/applirev-2021-0062

Zhang, X. (2019). Foreign language anxiety and foreign language performance: A meta-analysis. *The Modern Language Journal*, *103*(4), Article 4.

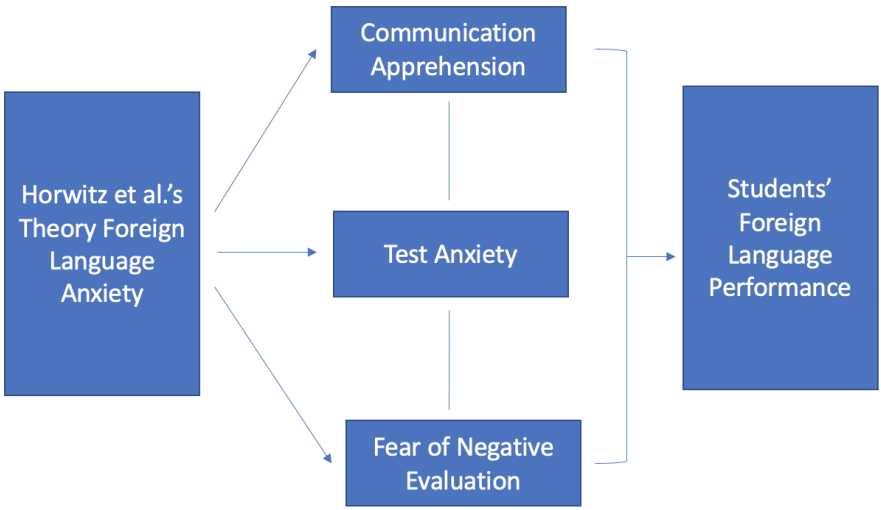
Zhang, Y. (2025). *Screening of language anxiety in learning Chinese as a foreign language: A systematic scoping review*. Manuscript submitted for publication.

Zhou, S., Chiu, M. M., Dong, Z., & Zhou, W. (2023). Foreign language anxiety and foreign language self-efficacy: A meta-analysis. *Current Psychology*, *42*(35), 31536–31550. https://doi.org/10.1007/s12144-022-04110-x

Appendix A

Supplementary Figures and Tables

Figure A1. The Theory of Foreign Language Classroom Anxiety



Note. This figure was adapted from Horwitz et al. (1986).

Table A1. The Overview of the CLFP Students' 4-5 Year Academic Plan

Year	Activity
Year 1	Chinese Class at the University (CHN 111 & CHN 112)
Year 1 Summer	Summer Domestic Immersion Program (CHN 211 & CHN 212)
Year 2	Chinese Class at the University (CHN 311 & CHN 312)
Year 2 Summer	Study Abroad: Overseas Immersion Program
Year 3	Chinese Class at the University (CHN 411 & CHN 412)
Year 4/5	Study Abroad: Capstone Year (Mainland China; Taipei)

Appendix B

Questionnaire

Part1: Background Information

1. Gender:

- Female
- Male
- Other
- I prefer not to say

2. Major:

- Business & Economics
- Education
- Social science & Political science
- Engineering
- Law
- Medicine
- Humanities
- Math & Statistics
- Public health

- Other _____ (specify)

0. I am (choose one or more):

- White
- Hispanic
- African-American
- Chinese American
- Asian American but not Chinese American
- Asian (international students)_____(specify)
- other_____(specify)

0. What is your native language (check all those apply)

- English
- Mandarin
- Cantonese
- Taiwanese
- another variety of Chinese other than Mandarin, Cantonese or Taiwanese_____ _____(please specify)
- other languages_____(plea se specify)

0. Are you currently in/taking courses in the Chinese flagship program?

- Yes
- No

0. I am taking Chinese as an(n):

- required course
- elective course

0. What is your purpose in taking Chinese classes?

0. What year are you in?

- Freshmen

- Junior
- Sophomore
- Senior (year 4 & 5 in the flagship program)

0. Are you studying at URI

- Yes, if yes, what is your students' ID
- No

0. How long have you been learning Chinese?

0. Have you been to China Mainland or Taiwan?

- Yes, for study; If so, how long?
- Yes, for other purposes
- No

0. I am confident about completing basic course requirements (e.g. daily homework, projects).

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. I am confident about navigating through the course materials, and find what I need for my Chinese class (e.g. class google folder; Brightspace; daily schedule)

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. I am confident about participating in course activities (e.g. discussion, quizzes, assignments)

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree

- Strongly disagree

0. I know where to access the gradebook for feedback on performance.

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. I feel comfortable about seeking help or assistance from others (teachers; tutors; classmates)

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

Part2: Chinese Language Classroom Anxiety

In this section, please indicate whether you agree or disagree with statements involving factors that might make you feel anxious in your foreign language class.

0. I never feel quite sure of myself when I am speaking in my Chinese language class.

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. I don't worry about making mistakes in Chinese language class.

- Strongly agree
- Somewhat agree

- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. I feel nervous when I know that I'm going to be called on in Chinese class.

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. It frightens me when I don't understand what the teacher is saying in the Chinese class.

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. It wouldn't bother me at all to take more Chinese language classes.

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. During Chinese class, I find myself thinking about things that have nothing to do with the course.

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. I keep thinking that the other students are better at Chinese than I am.
- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. I usually feel relaxed during tests in my Chinese class.
- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. I start to panic when I have to speak without preparation in Chinese class.
- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. I worry about the consequences of failing my Chinese language class.
- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. I don't understand why some people get so upset over Chinese language classes.
- Strongly agree
- Somewhat agree

- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. In language class, I can get so nervous I forget things I know.

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. It embarrasses me to volunteer answers in my Chinese class.

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. I would not be nervous speaking Chinese language with native speakers.

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. I get upset when I don't understand what the teacher is correcting.

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. Even if I am well prepared for Chinese class, I feel anxious about it.
- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. I often feel like not going to my Chinese class.
- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. I feel confident when I speak in a Chinese language class.
- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. I am afraid that my Chinese teacher is ready to correct every mistake I make.
- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. I can feel my heart pounding when I'm going to be called on in Chinese class.
- Strongly agree
- Somewhat agree

- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. The more I study for a Chinese test, the more confused I get.

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. I don't feel pressure to prepare well for Chinese class.

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. I always feel that the other students speak the Chinese language better than I do.

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. I feel very self-conscious about speaking the Chinese language in front of other students.

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. Chinese class moves so quickly I worry about getting left behind.

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. I feel more tense and nervous in my Chinese class than in my other classes.

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. I get nervous and confused when I'm speaking in my Chinese class.

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. When I'm on my way to Chinese class, I feel very sure and relaxed.

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. I get nervous when I don't understand every word the Chinese teacher says.

- Strongly agree
- Somewhat agree
- Neither agree nor disagree

- Somewhat disagree
- Strongly disagree

0. I feel overwhelmed by the number of rules you have to learn to speak Chinese.

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. I am afraid that the other students will laugh at me when I speak the Chinese language.

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. I would probably feel comfortable around native speakers of Chinese.

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

0. I get nervous when the Chinese teacher asks questions that I haven't prepared in advance.

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

Appendix C

Justification for Bayesian

We leverage the Bayesian approach to answer our two research questions for its flexibility, depth of analysis, and ability to incorporate prior knowledge. Recent literature highlights the advantages of Bayesian statistics in the field of education and language research (Yang, 2019). Scholars argue that Bayesian approaches are adept at handling the complex models, such as those needed to understand the intricacies of FLA among Chinese language learners (König & van de Schoot, 2018; Lee, 2007). Moreover, this approach is particularly valuable in small sample contexts, as it allows us to incorporate informative priors (Lee, 2007). These priors are informed by a synthesis of existing literature and expert opinion, thus enriching our understanding of FLA among Chinese language learners.

Furthermore, within the domain of second language research, Norouzian et al., (2018) suggest that Bayesian statistics bolster theory development and emphasize

parameters of substantive interest rather than merely statistical significance. This method excels in providing a dynamic quantification of evidence, as it can update the strength of evidence for or against hypotheses with each new data point (Jeffreys, 1935; Rouder, 2014). In our study, this translates to an ability to rigorously test the validity of the FLCAS in capturing the anxiety of Chinese language learners and to identify individual predictors of anxiety with greater precision.

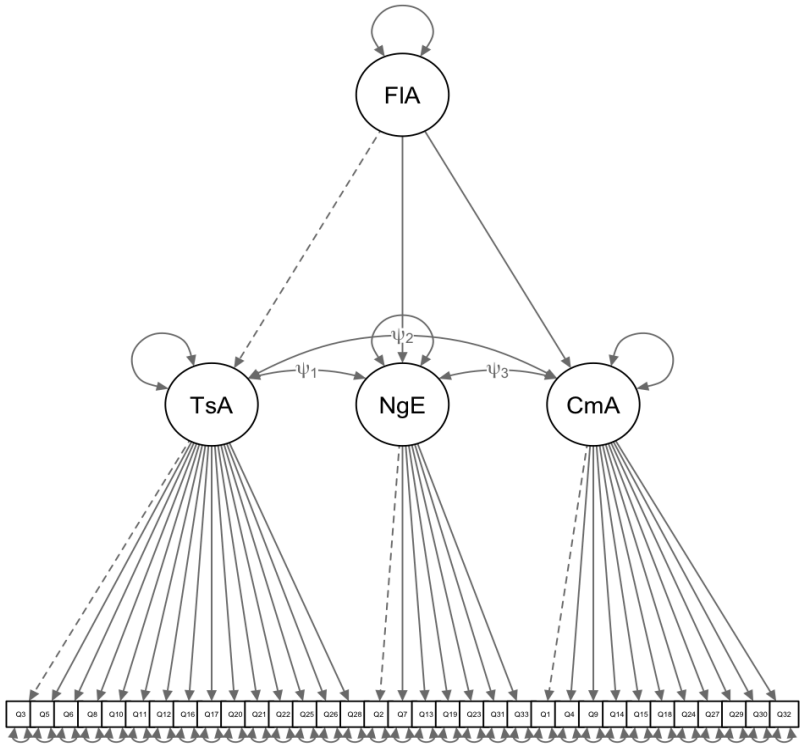
Thus, the Bayesian approach significantly contributes to the field, moving beyond the limitations of traditional frequentist methods to provide a richer picture of FLA in an educational setting. We use posterior predictive simulation to compare those prototypical cases which is not possible in a traditional statistical approach. An 89% credible interval (CI) was used for a more stable result (Kruschke, 2014). Priors were carefully selected based on historical data and expert consultations, ensuring they reflect realistic expectations for the parameters being estimated.

Detailed Data Analytic Plan

To answer RQ1, we examined various sources of validity evidence, including face validity, content validity, construct validity, and convergent validity (Cizek et al., 2008). We first invited expert instructors of CLFP to independently evaluate the questions to review the face validity and content validity of the FLCAS items. Given the small sample size and the subjective nature associated with EFA in terms of factor extraction, factor rotation, and factor labeling (Bryant & Yarnold, 1995; Thompson, 2004), our study did not utilize EFA. Instead, we established the reliability analysis of the scale by using Composite Reliability, including coefficient alpha and coefficient omega (Bacon et al., 1995). We then used BCFA with the "blavaan" package in R (Merkle & Rosseel, 2018) to test the construct validity. We proposed a two-level, three-factor model, centering on a higher-order latent factor ('Foreign language anxiety- FlA') influenced by three lower-order latent variables ('Test Anxiety-TsA', 'Fear of negative evaluation-NgE', 'communication anxiety-CmA'),

each measured through observed variables (items 1-33) (see Figure 1). The last step was to check the convergent validity by examining if those factors correlated with 5 items of self-efficacy (See appendix A, questions 12-16). Both anxiety and self-efficacy factor scores were retracted through BCFA.

Figure 1. Path Diagram of the Two-level Three-structure Scientific Model



Note. This path diagram illustrates a *two-level three-structure BCFA* model with observed variables (Q1-Q33) informing three latent factors—TsA, NgE, and CmA. These, in turn, are hypothesized to contribute to the overarching construct FlA. Factor loadings (λ), inter-factor covariances (ψ), and error terms (ε) are also represented, depicting the scientific relationships and measurement structure.

Upon establishing the validity of the Chinese learning anxiety scale, we proceeded to examine the predictive relationship between FLA and student-level measures including the demographic variables and self-efficacy. We first analyzed the raw score to obtain an unadjusted and more intuitive view of the data. Subsequently, we employed factor scores derived from our BCFA, which allowed us to have a more refined analysis that considers the underlying structure of the data and the interrelationships between the variables.

Outcome Variable (FLA)

Measured by FLCAS and standardized (mean = 0, SD = 1) to facilitate efficient model fitting and interpretation McElreath (2020). Anxiety standardized score ranges from a minimum

value of -2.04 to a maximum of 2.44, and normally distributed.

Predictor Variables

Demographic variables: dummy coded variables including heritage status, gender, year in college; Language learning variable: overseas experience: dummy coded variable; language learning time: a continuous variable; self-efficacy: a continuous variable standardized (mean = 0, SD = 1). Self-efficacy factor scores range from a minimum of -1.90 to a maximum of 0.46.

Covariate

Program participation, factor-coded to control for variations in anxiety levels across eight Flagship programs.

To answer RQ2, we employed a Bayesian multilevel model to accommodate the hierarchical structure of our data, with students nested within eight different Flagship Programs. This model structure allows us to account for the fact that observations within the same Flagship Program may

be more similar to each other than to those from different groups. It helps to correct for potential biases that could occur in estimated standard errors and inferences if such clustering is ignored. By sampling the parameter space via Markov Chin Monte Carlo sampling, the Bayesian approach allowed us to use regularizing priors (Gelman et al., 2020) to regularize our estimates and prioritize estimates that fit within the bounds of scientific plausibility (McElreath, 2020). Regularization also helps to guard against potentially inflated effect estimates that arise in scenarios where the design might be under-powered to detect some effects due to missingness or sparsity at various levels of aggregation (Gelman & Carlin, 2014).

Our hypothesized data generating model leveraged a within-program and between-program design, which allows us to jointly model and test hypotheses about the relationship within prototypical programs, and between students in

different programs. Using self-efficacy as example, our data generating model was:

$$anxiety_{ip} \sim normal(\mu, \sigma)$$

$$\mu = \alpha + \alpha_p + \beta * self\_efficacy$$

$$\alpha \sim normal(0,\ 0.5)$$

$$\beta \sim normal(0,\ 0.5)$$

$$\alpha_p \sim normal(0,\ \sigma_p)$$

$$\sigma \sim exponetial(1)$$

$$\sigma_p \sim exponetial(1)$$

This represents a measure of anxiety that is normally distributed with a mean of μ and standard deviation of σ. The mean value of anxiety is modeled as baseline value of anxiety (α) that deviates by programs, accounting for potential differences in anxiety by program (α_p). Population and program level variances are specified ($\sigma,\ \sigma_p$). We then express our priors for our model parameters; in this case we set to constrain our prior predictions to ranges of the standardized anxiety score (-3 to 3) and to regularize estimates, and which

constrain the variance estimates to be greater than zero. The posterior estimate of the slope (β) and intercept parameters (α) were used to generate simulated predictions of average values for students in different programs. We then evaluated the proportion of posterior simulated estimates greater than zero in probability of direction (pd). The last step was to use posterior distribution to generate data for prototypical scenarios to visualize and understand how our model behaves under different conditions. Similar analyses were applied to demographic variables.

Final models were fitted with three chains and with four thousand sample draws, to ensure a thorough exploration of the posterior distribution and to bolster the precision of our inferences (Bürkner, 2017; R Core Team, 2024).

More broadly, we followed the Bayesian data analytic workflow proposed by Gelman & Carpenter (2020). In general, we first specified models with standard normal

distributions. Then we conducted prior predictive checks to establish priors that would capture a realistic range of anxiety factor score. After setting our priors, we fitted our hypothesized models to simulated data to ensure they would properly converge and compute the values of interest. We then fitted these models to the actual data. To determine the model with the highest predictive accuracy, we utilized Leave-One-Out Cross-Validation (LOO-CV). This approach allowed us to select the model that not only best fit the data but also had the most robust predictive power, ensuring that our conclusions would be more reliable. Finally, we fitted and interpreted the models as described above.

In order to ensure the validity of our causal inference and avoid table II fallacy, we used backdoor criteria for individual analysis of different variables. We first identified variables that we are interested in, and this selection was guided by prior domain knowledge and a review of literature, as well as our research interest. We controlled different variables based on the Directed Acyclic Graph (DAG) that

indexed our scientific model (Shi et al., 2019). See appendix A for the scientific model and details on which variables were controlled. We then conducted a robustness check by fitting the models with wider priors, examining the direction and the magnitude of the estimates.

For demographic information (gender, heritage status, year in college) predicting anxiety: we controlled for program differences.

Overseas experience predicting FLA: we adjusted for year in college, and heritage status.

Language learning time predicting FLA: we adjusted for year in college.

Self-efficacy predicting FLA: we adjusted for gender, heritage status, language learning time, oversea, year in college, ensuring a robust assessment of how self-efficacy influences anxiety free from these confounding factors (Rohrer, 2018).