BILINGUAL PUBLISHING GROUP

Forum for Linguistic Studies
https://journals.bilpubgroup.com/index.php/fls

ARTICLE

Validation of Adapted Scales Assessing Affective Filter Hypothesis Among Second Language Chinese Learners

Wenting Bao [1, 2] , ***Farrah Dina Yusop*** [1]*

[1] *Faculty of Education, Universiti Malaya, Kuala Lumpur 50603, Malaysia*

[2] *School of Humanities and International Education Exchange, Anhui University of Chinese Medicine, Hefei 230012, China*

ABSTRACT

Affective factors, including anxiety, motivation, and self-confidence, significantly influence second language acquisition (SLA), yet existing measurement tools often lack cultural and linguistic specificity for non-English contexts. This study adapted and validated three scales, which are Foreign Language Speaking Anxiety Scale, Attitude and Motivation Test Battery (AMTB) Scale, and Foreign Language Self-Esteem Scale, under Krashen's Affective Filter Hypothesis for international students learning Chinese speaking in China. Through a multi-stage process, including scale modification, back-translation, expert opinions, and pilot testing, the revised instruments demonstrated strong psychometric properties. Expert content validity assessments yielded CVIs above 0.7, and reliability analyses run in IBM SPSS Statistics 26 produced Cronbach's α values above 0.8. Key modifications included removing redundant statements and localizing linguistic references to align with Chinese speaking contexts. The results validate the applicability of the Affective Filter Hypothesis to Chinese SLA, highlighting the necessity of culturally adapted tools to assess affective barriers and facilitators. The findings confirm that the adapted questionnaires effectively capture the key affective factors based on the Affective Filter Hypothesis in Chinese speaking settings. Researchers can employ these scales to further investigate affective influences on SLA, and educators can use the findings to tailor Chinese speaking instruction, foster higher motivation, bolster self-confidence, and reduce anxiety in the classroom.

Keywords: Affective Factor; Chinese Language Learning; Language Anxiety; Scale Validation; Second Language Acquisition

*CORRESPONDING AUTHOR:
Farrah Dina Yusop, Faculty of Education, Universiti Malaya, Kuala Lumpur 50603, Malaysia; Email: farah@um.edu.my

ARTICLE INFO
Received: 10 May 2025 | Revised: 12 June 2025 | Accepted: 17 June 2025 | Published Online: 10 July 2025
DOI: https://doi.org/10.30564/fls.v7i7.9948

CITATION
Bao, W., Yusop, F.D., 2025. Validation of Adapted Scales Assessing Affective Filter Hypothesis among Second Language Chinese Learners. Forum for Linguistic Studies. 7(7): 158–173. DOI: https://doi.org/10.30564/fls.v7i7.9948

1. Introduction

In the field of Second Language Acquisition (SLA), affective factors are widely recognized for their significant impact on learners' language learning. Positive emotions, such as enjoyment and selfconfidence, can boost motivation and classroom participation, prompting learners to devote more energy to practicing and using language skills [1]. Meanwhile, negative emotions such as anxiety and unease can reduce the efficiency of input and output, thereby discouraging learners from expressing themselves or participating in communicative activities [2,3]. It is evident that affective factors have a crucial role in SLA, whether through their motivating presence or their potential to hinder learning, and they merit further attention in future research and instructional design.

Among the theories concerning the influence of affective factors on SLA, Krashen's Affective Filter Hypothesis is regarded as particularly influential [4]. This hypothesis identifies motivation, anxiety, and selfconfidence as central affective variables, positing that their level directly affects the degree to which learners filter incoming language input, ultimately influencing learning outcomes. Despite the explanatory power of the Affective Filter Hypothesis in SLA research, accurately measuring these three affective variables remains challenging in actual studies and teaching environments [5].

Current empirical work grounded in this hypothesis has focused predominantly on English as a second language (L2). Related classroom interventions typically aim to optimize the learning environment, enhance motivation, and manage anxiety so as to lower learners' affective filter [6]. By contrast, applications of the Affective Filter Hypothesis in teaching Chinese as an L2 are still at an early stage, with only a handful of exploratory studies [7]. One likely reason is the paucity of standardized instruments for assessing motivation, anxiety, and selfconfidence in Chinese learning contexts, a gap that may stem from the hypothesis' emphasis on macrolevel affective influences rather than contextspecific measurement [8].

Some researchers have employed established instruments to assess motivation, anxiety, and selfconfidence. However, differences in cultural backgrounds, language features, and educational settings may affect the applicability, reliability, and validity of these original instruments. Therefore, adapting and validating such scales for specific populations and learning contexts is especially important. In particular, for studies on Chinese speaking among international students in China, there is currently no localized scale that closely aligns with the affective dimensions of the Affective Filter Hypothesis.

To address this gap, the present study drew Krashen's Affective Filter Hypothesis to adapt and localizes existing instruments that measure motivation, anxiety, and selfconfidence. Three established scales—the Foreign Language Speaking Anxiety Scale, the Attitude and Motivation Test Battery (AMTB), and the Foreign Language SelfEsteem Scale—were adapted to create a new, integrated instrument for international students enrolled in Chinese speaking classes in mainland China. By gathering expert opinions and conducting a pilot study, the reliability and validity of the revised scales were examined. The resulting tool aimed to provide SLA researchers with a contextspecific means of measuring affective factors in Chinese speaking environments and to support further investigation into how these factors shape Chinese speaking learning outcomes.

2. Literature Review

2.1. The Affective Filter Hypothesis

The Affective Filter Hypothesis situates affectiveladen variables at the gateway between comprehensible input and successful acquisition. In Krashen's model (**Figure 1**), language first appears as Input, yet its uptake is mediated by the Filter, which is a composite of motivation, selfconfidence, and anxiety that can either obstruct or facilitate processing. Input that passes this psychological filter is routed to the innate Language Acquisition Device (LAD), where unconscious linguistic computation converts it into Acquired Competence, the durable, implicit knowledge manifested in fluent performance [4]. A low filter—characterized by high motivation, high selfconfidence, and manageable low anxiety—maximizes both the quantity and quality of intake, accelerating the LAD's work. Conversely, a heightened filter restricts intake, starving the LAD and slowing growth in competence. By foregrounding these four constructs—Input, Filter, LAD, and Acquired Competence—this hypothesis offers a concise explanation for the uneven outcomes often observed among learners who

receive comparable instruction. Pedagogically, it mandates interventions that foster positive affect or mitigate anxiety to lower the filter and, in turn, amplify the impact of meaningful input on SLA.

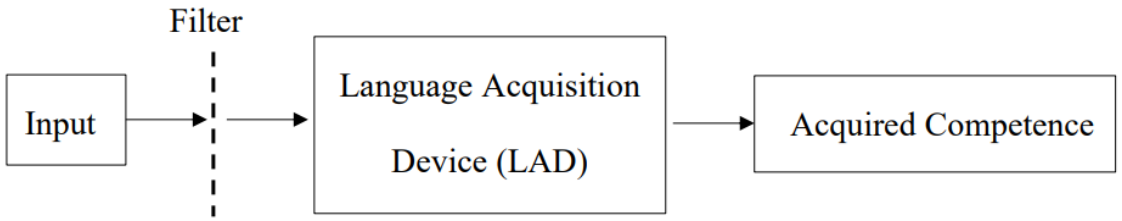

Figure 1. The Processing Process of The Affective Filter Hypothesis.

2.2. Existing Scales and Their Limitations

According to the Affective Filter Hypothesis, primary affective variables include motivation, anxiety, and self-confidence. Extensive empirical studies have shown that these factors influence learners' attention distribution, classroom participation, and language performance [3,9]. Nevertheless, the measurement tools employed vary widely, and many scales originally focus on English or are designed primarily for Western contexts, thus limiting their direct applicability to Chinese speaking class settings [7].

Currently, the most commonly used scales for affective factors based on the Affective Filter Hypothesis are included the following:

Foreign Language Classroom Anxiety Scale (FLCAS) is widely applied across various contexts in SLA. However, this scale combines anxiety related to multiple language skills (listening, speaking, reading, and writing) [10]. Although Öztürk and Gürbüz develop an adaptation focusing on speaking anxiety [11], it still primarily targets English as an L2, with no adaptation for Chinese-language teaching.

Attitude/Motivation Test Battery (AMTB) Scale is originally devised to measure motivation and attitudes toward English-language learning, covering multiple dimensions [12]. For Chinese speaking instruction and cross-cultural contexts, further verification and revision are needed.

Foreign Language Self-Esteem Scale assesses learners' positive appraisal of their foreign-language proficiency and classroom performance [13]. The original scale, however, may require structural and contextual adjustments to fit the Chinese speaking setting.

Given that international students in China differ significantly from English learners in both linguistic and cultural backgrounds, directly applying these original instruments could compromise content validity and structural validity. Therefore, employing scientifically sound approaches for cross-cultural adaptation of such scales becomes significant [14]. Accordingly, this study focused on measuring international students' learning Chinese speaking skill in China and adapted relevant scales under the framework of the Affective Filter Hypothesis. By validating the adapted instruments for anxiety, motivation, and self-confidence, it aimed to furnish researchers with a measurement tool suited to Chinese speaking learning contexts. Therefore, the goal of this study was to adapt and examine the reliability and validity of scales assessing the three affective variables—anxiety, motivation, and self-confidence—in line with the Affective Filter Hypothesis, providing a practical resource for subsequent research on international students' Chinese speaking learning.

3. Research Methods

This study targeted international students learning Chinese speaking in China and employed a comprehensive process consisting of initial scale modification, expert opinion, pilot testing, and reliability and validity analysis (**Figure 2**). First, statements related to speaking skill were selected from original scales, and any references to English or foreign language were replaced with Chinese. Next, a back-translation was conducted on the expert opinion questionnaire and all the scales statements. Experts were then invited to give feedback of the revised scale's content, and a content validity test was carried out. Based on these steps, a pilot study was conducted to verify that international students could understand and complete the scale. Finally, SPSS26 software was used to analyze the

pilot data for reliability and validity, including Cronbach's α internal consistency testing and statement analysis. The scales were subsequently refined based on the results to form the final version.

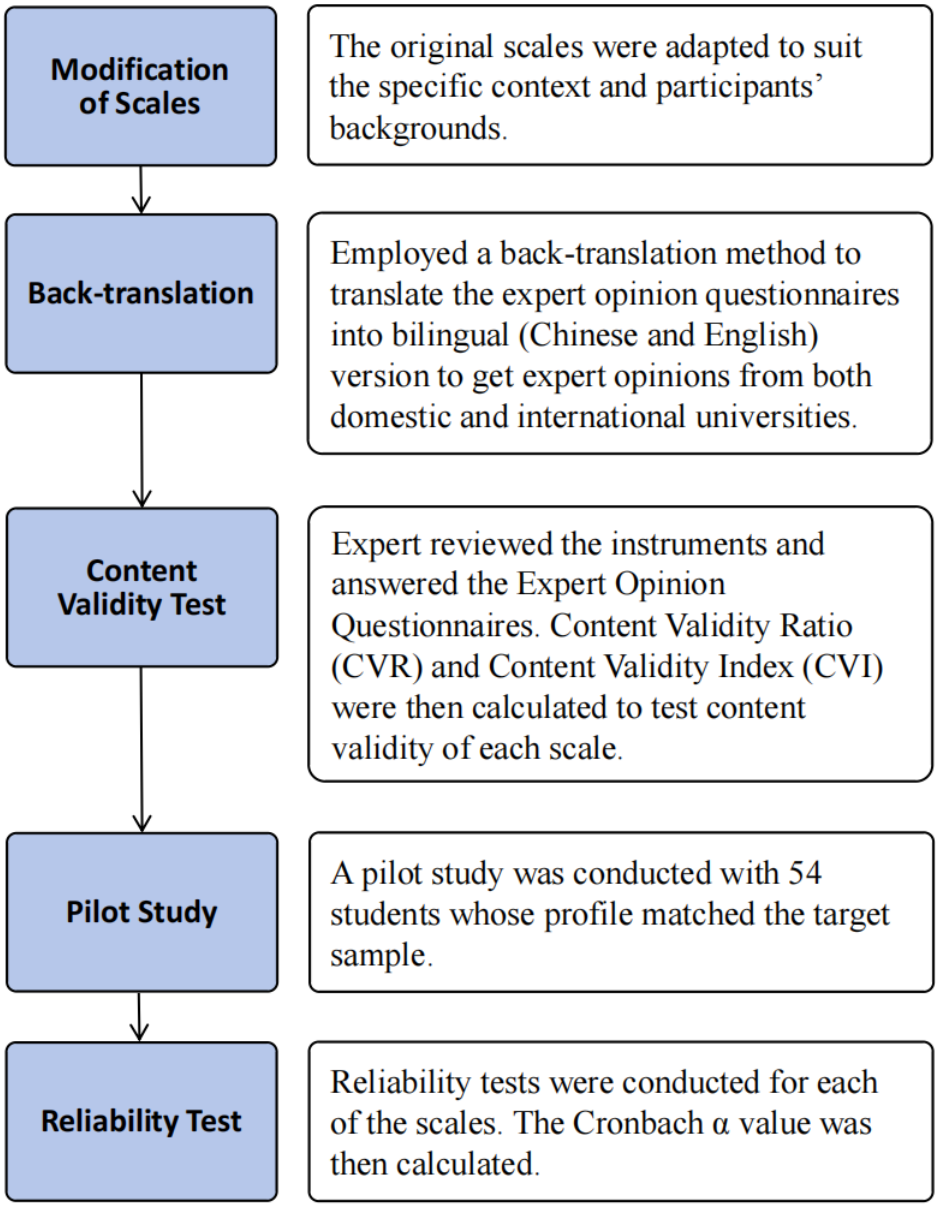

Figure 2. Stages of Instrument Adaptation, Validation, and Reliability Testing.

4. Scale Adaptation Process

4.1. Scale Selection and Modification

This study drew on three established scales corresponding to the three affective factors—motivation, anxiety, and self-confidence—identified in Affective Filter Hypothesis. As these instruments originated in Western contexts, they were subjected to an initial round of modification.

4.1.1. Foreign Language Speaking Anxiety Scale

Building on Horwitz's scale and the subsequent speaking-focused revision by Öztürk and Gürbüz [10,11], the term "English" in Foreign Language Speaking Anxiety Scale was replaced with "Chinese" to align with the linguistic setting of international students in China, and some scenario descriptions were localized.

4.1.2. Attitude and Motivation Test Battery (AMTB) Scale

Derived from the work of Gardner [12], the original AMTB included statements assessing attitudes toward English culture and classroom-related anxiety. Since this study focused solely on learners' motivation and attitudes toward a Chinese speaking context, relevant statements were screened or removed, and references to "English" were changed to "Chinese".

4.1.3. Foreign Language Self-Esteem Scale

Based on the work of Hassan [13], this comprehensive instrument measured self-esteem in foreign language learning. Statements related to reading, writing, listening, or "Foreign Language" were replaced with descriptions suitable for a Chinese speaking context, and the total number of statements was capped at a manageable level.

4.2. Back-Translation

This study was conducted among international students, who were learning Chinese at universities in China, which necessitated the inclusion of experts who were native Chinese language teachers with extensive experience in L2 teaching in China. As this study involved translating materials between Chinese and English in next section about the Expert Opinion Questionnaire and all the statements in the three scales, a back-translation method was employed to ensure the accuracy and reliability of the translations.

Backtranslation is a widely recognized method in translation studies used to validate the accuracy of translated documents [14]. It involves translating a document from the source language to the target language and then independently translating it back to the original language. This process allows for a comparison between the original document and the backtranslated version, helping to identify and rectify discrepancies, ensuring the translation faithfully conveys the original content's meaning and intent [15].

The back translation process in this study consisted of the following steps:

4.2.1. Initial Translation

A Translation Studies PhD lecturer (10 years' experience) converted the English questionnaire to Chinese, aligned with research objectives.

4.2.2. Blind Back-Translation

An SLA PhD lecturer (13 years' experience), blinded to the original, independently re-translated the Chinese version to English.

4.2.3. Discrepancy Resolution

The researcher revised minor lexical or syntactic variations between original and back-translated texts through iterative revisions.

4.2.4. Expert Review and Finalization

Two independent experts (25-year Translation Professor and 17-year Linguistics Associate Professor) verified semantic equivalence, finalizing the bilingual questionnaire.

After the four steps of translation, this back-translation process ensured the translations' validity, making the materials suitable for use in this cross-cultural research setting.

4.3. Content Validity Test

Content validity assesses how well an instrument measures the intended construct. This is done by obtaining experts' judgments on each statement of the three scales and by calculating the Content Validity Ratio (CVR) for individual statements and the Content Validity Index (CVI) for the overall scale [16].

Content validity was established through a panel of domain experts selected by 4 criteria: (1) a doctoral degree in applied linguistics, SLA, or educational technology; (2) at least five years of university-level teaching or research in their specialty; (3) a minimum of three peer-reviewed publications on language assessment or technology-enhanced learning within the last five years; and (4) documented experience in developing or validating research instruments. According to Gilbert and Prion [17], 5–10 experts were aimed but 15 qualified scholars were invited (8 in SLA, 7 in educational technology) to maximize disciplinary breadth. A total of 11 researchers accepted (response rate =

73.3%). **Table 1** summarizes their profiles: 7 are scholars with L2 teaching experience, and 4 are researchers recognized for significant contributions to educational-technology design and evaluation.

Table 1. Background of Experts.

Expert	Gender	Academic Position	Research Field	Experience
Expert 1	Female	Professor	SLA	22 Years
Expert 2	Female	Professor	SLA	18 Years
Expert 3	Male	Associated Professor	SLA	12 Years
Expert 4	Female	Lecturer	SLA	11 Years
Expert 5	Male	Lecturer	SLA	8 Years
Expert 6	Female	Lecturer	SLA	8 Years
Expert 7	Female	Lecturer	SLA	5 Years
Expert 8	Male	Associated Professor	Educational technology	20 Years
Expert 9	Male	Lecturer	Educational technology	15 Years
Expert 10	Female	Lecturer	Educational technology	11 Years
Expert 11	Female	Lecturer	Educational technology	5 Years

The expert opinion questionnaire asked experts to rate each statement as "essential," "useful but not essential," or "not necessary." The experts evaluated each statement in the 3 scales using the provided questionnaire. Their assessments were used to calculate the CVR for each statement and the CVI for the overall scale.

4.3.1. Analysing CVR and CVI for Each Scale

Content Validity Ratio (CVR) is a widely recognized statistical measure used to decide whether individual statements in a questionnaire should be retained or rejected, thereby confirming their content validity [18]. It is calculated using the formula: $CVR = (Ne - N/2)/(N/2)$, where "Ne" is the number of experts who rate a statement as "essential" and "N" is the total number of experts consulted. A CVR of 1.00, which is commonly adjusted to 0.99 for practical calculations, indicates unanimous agreement among the experts that a statement is essential [19]. Conversely, a CVR of 0 means that no experts consider the statement essential. statements with a CVR between 0 and 0.99 indicate varying levels of agreement that the statement is essential. statements on which is perceived to be "essential" by more than half of the experts, has some degree of content validity. The more experts (beyond 50%), perceiving a statement as "essential", the greater the extent or degree of its content validity [17].

Content Validity Index (CVI) is an average of the CVRs for all statements retained in the final instrument, assessing the overall content validity of the tool [20]. It provides a comprehensive measure of how well the statements collectively represent the domain of interest. Established guidelines suggest a CVI value of over 0.70 is acceptable for most research purposes [21], while a CVI of 0.80 or higher is preferred for higherstakes assessments [22].

In this research, each of the three scales was assessed separately for content validity. Statements marked as "essential" by less than half of the experts (less than 6 experts in this study) were removed in the first round of analysis. The remaining statements were then used to calculate the CVI. An overall CVI score of 0.7 or higher was considered indicative of adequate content validity for the scale.

The CVR and CVI are essential metrics used in this study to ensure the validity of each statement and the overall instrument. They provided quantitative evidence of how well the scale captured the intended content domains. Details on the calculation and results of the CVR and CVI for each scale presented in the following part.

CVR and CVI Analysis for Foreign Language Speaking Anxiety Scale

Table 2 illustrated the results of the CVR for the Foreign Language Speaking Anxiety Scale according to experts' opinion. Each statement on the scale was marked as "essential" by a majority of experts, thereby justifying their calculation of the CRI, which was determined to be 0.77.

Table 2. Content Validity Ratio (CVR) for Foreign Language Speaking Anxiety Scale.

Statements	Total Number of Experts Opinion			CVR
To What Extent Are You Agree with Each of the Statement.	Not Relevant	Useful, But Not Essential	Essential	
Statement 1		1	10	0.82
Statement 2			11	1.00
Statement3			11	1.00
Statement 4		1	10	0.82
Statement 5		1	10	0.82
Statement 6		2	9	0.64
Statement 7			11	1.00
Statement 8		5	6	0.09 (Removed)
Statement 9		1	10	0.82
Statement 10		2	9	0.64
Statement 11	1	3	7	0.27 (Removed)
Statement 12		1	10	0.82
Statement 13		1	10	0.82
Statement 14			11	1.00
Statement 15		1	10	1.00
Statement 16		2	9	0.64
Statement 17		1	10	0.82
Statement 18		1	10	0.82

Based on experts' feedback, modifications were also made to the scale. Specifically, four experts noted redundancy between statements 4 and 8, as well as between statements 3 and 11. Given the low CVR scores associated with statements 8 and 11, the decision was made to retain statements 3 and 4 while eliminating statements 8 and 11. Consequently, the final version of the Foreign Language Speaking Anxiety Scale achieved a CVI of 0.84, indicating a high level of content validity for the revised scale.

4.3.2. CVR and CVI Analysis for AMTB Scale

Table 3 documented expert evaluations for each statement on the Attitude and AMTB Scale. Statement 1 and 2 were removed from the scale because less than half of the experts marked them "essential". The CVI was then calculated for the remaining statements, yielding an average CVR of approximately 0.81. This average suggested a strong consensus among the experts regarding the remaining statements on the scale.

Table 3. Content Validity Ratio (CVR) for AMTB Scale.

Statement	Total Number of Experts Opinion			CVR
To What Extent Are You Agree with Each of the Statement.	Not Relevant	Useful, But Not Essential	Essential	
Statement 1	3	3	5	−0.09 (Removed)
Statement 2	2	4	5	−0.09 (Removed)
Statement 3			11	1.00
Statement 4			11	1.00

Table 3. *Cont*.

Statement	Total Number of Experts Opinion			CVR
To What Extent Are You Agree with Each of the Statement.	Not Relevant	Useful, But Not Essential	Essential	
Statement 5		1	10	0.82
Statement 6	1		10	0.82
Statement 7		1	10	0.82
Statement 8		2	9	0.64
Statement 9		1	10	0.82
Statement 10	2	1	8	0.46
Statement 11	1	1	9	0.64
Statement 12	1	1	9	0.64
Statement 13	1		10	0.82
Statement 14		1	10	0.82
Statement 15			11	1.00
Statement 16			11	1.00
Statement 17	1	1	9	0.64
Statement 18			11	1.00
Statement 19		1	10	0.82
Statement 20	1		10	0.82

4.3.3. CVR and CVI Analysis for Foreign Language Self-Esteem Scale

Table 4 provided a detailed account of expert evaluations for each statement on the Foreign Language Self-Esteem Scale, where more than half of the experts deemed each statement "essential". The calculated CVI for these statements is 0.71, suggesting that the scale exhibits relatively high validity.

Table 4. Content Validity Ratio (CVR) for Foreign Language Self-Esteem Scale.

Statement	Total Number of Experts Opinion			CVR
To What Extent Are You Agree with Each of the Statement.	Not Relevant	Useful, But Not Essential	Essential	
Statement 1	1	1	9	0.64
Statement 2			11	1.00
Statement3	1		10	0.82
Statement 4		1	10	0.82
Statement 5		1	10	0.82
Statement 6	1	2	8	0.46
Statement 7		1	10	0.82
Statement 8			11	1.00
Statement 9	1		10	0.82
Statement 10		1	10	0.82
Statement 11		2	9	0.64

Table 4. *Cont*.

Statement	Total Number of Experts Opinion			CVR
To What Extent Are You Agree with Each of the Statement.	**Not Relevant**	**Useful, But Not Essential**	**Essential**	
Statement 12	2	3	6	0.09
Statement 13			11	1.00
Statement 14	3	2	6	0.09
Statement 15	1	1	9	0.64
Statement 16		1	10	0.82

After the evaluation process for the three scales, CVR and CVI calculations were conducted, and appropriate modifications were made based on experts' feedback. These modifications resulted in scales with high content validity.

4.4. Pilot Study

This study involved adult university students who completed self-report questionnaires. Ethical approval was granted by University Research Ethics Committee (Approval No. UM.TNC2.UMREC_4566). Before data collection, the researcher provided a plain language statement outlining the study's aims, procedures, confidentiality safeguards, and the voluntary nature of participation, including the right to withdraw at any time without penalty. Participants then signed a printed consent form on site; none chose to withdraw, and no incentives or compensation were offered.

Following expert review, preliminary versions of the 5 adapted scales were uploaded to an online survey platform. Using purposive sampling, the researcher approached course instructors at a comprehensive university in Anhui Province, China, and invited international students who had studied Chinese for at least one semester and possessed beginner-level proficiency to take part. The invitation was delivered in class and reiterated via an email that contained the survey link. A total of 54 volunteers from 3 intact classes completed the pilot survey in May 2025, typically requiring about 20 minutes on their personal devices.

Since this was a pilot study, each scale was examined separately using principal-axis exploratory factor analysis (EFA) with oblimin rotation, supplemented by Horn' s parallel analysis. KMO values ranged from 0.76 to 0.81, and Bartlett's tests were significant ($p < 0.001$), confirming sampling adequacy. For all three scales, both the scree plot and parallel analysis converged on a single-factor solution. Item loadings ranged from $|\lambda| = 0.37$ to 0.93. Although Item 1 on the Foreign Language Speaking Anxiety Scale loaded more modestly ($|\lambda| = 0.37$), its corrected item–total correlation (0.32) exceeded the .30 retention benchmark, so the item was retained for theoretical completeness. These findings offered exploratory evidence of unidimensionality and subsequent sections presented internal-consistency indices Cronbach's α as additional reliability evidence.

4.5. Reliability Test

The next process was reliability test, which reported the reliability analysis of the three scales using data from a pilot study. SPSS 26 was used for data cleaning and for calculating Cronbach's α to assess internal consistency. Cronbach's α values range from 0 to 1, with higher scores indicating better reliability [23].

Data cleaning included three steps: identifying missing values, verifying data formatting, and detecting outliers. Any incomplete questionnaires were deemed invalid. To address outliers in Likerttype data, which generally deviate from normal distribution, this study employed the Extreme Response Style (ERS) approach [24]. Respondents who consistently chose extreme answers were considered inattentive or possibly misunderstanding the statements [25]. Their responses were treated as invalid.

Cronbach's α represents how consistently a set of statements measures a single construct. While thresholds can vary, $\alpha \geq 0.9$ is often viewed as excellent, 0.8–0.9 as good, 0.7–0.8 as acceptable, 0.6–0.7 as questionable, 0.5–0.6 as poor, and < 0.5 as unacceptable [26]. A statement Deletion Analysis was also performed; if removing a particular statement significantly increased α or if a statement's

total correlation fell below 0.3, that statement was considered for removal.

4.5.1. Reliability Analysis for the Foreign Language Speaking Anxiety Scale

Comprising 16 statements, the Foreign Language Speaking Anxiety Scale displayed no missing values or outliers during data cleaning, resulted in all 54 responses valid. SPSS26 generated a Cronbach's α of 0.959, signifying high reliability. According to the statement deletion analysis, all statement-total correlations exceeded 0.3, and although removing statement 1 produced a slight increase in Cronbach's α, the change was negligible. Consequently, no modifications were needed, and the scale demonstrated robust internal consistency.

4.5.2. Reliability Analysis for AMTB Scale

Originally comprising 18 statements, the AMTB scale measures students' speaking motivation. After data cleaning, all 54 responses were complete and valid. The initial Cronbach's α was 0.887, indicating good reliability. However, statement 8 had a low correlation with the total score ($0.23 < 0.3$), suggesting removal. With statement 8 excluded, the 17-statement scale's Cronbach's α rose to 0.892, demonstrating improved internal consistency. Accordingly, the revised scale consists of 17 statements with high reliability.

4.5.3. Reliability Analysis for Foreign Language Self-Esteem Scale

This 16-statement scale measures students' self-esteem, and the overall Cronbach's α was 0.92, indicating excellent reliability. Although it suggests statements 6 and 14 could marginally increase Cronbach's α if removed, their statement-total correlations are already above 0.3, and the potential gains in reliability are minimal. Consequently, the scale demonstrates strong internal consistency, and no modifications are required.

After the reliability test, all the three scales demonstrated satisfactory Cronbach's α values, indicating strong internal consistency. The final Cronbach's α values are presented in **Table 5**. Notably, only one statement (statement 8 of the AMTB scale) was removed to enhance reliability.

Table 5. Cronbach's α of all the Independent Variables in This Study.

Scale	Cronbach's α	Number of Statements
Foreign Language Speaking Anxiety Scale	0.959	16
AMTB Scale	0.892	17
Foreign Language Self-Esteem Scale	0.92	16

Through validity (expert opinion) and reliability (pilot study) processes, the finalized scales (Appendix A) for anxiety, motivation, and self-esteem display appropriate psychometric properties and are ready for Chinese speaking class settings.

5. Discussion

5.1. Scale Interpretation

The results of this study indicated that the adapted scales demonstrated excellent reliability and validity for international students learning Chinese speaking in China. First, the scales exhibited high internal consistency, with high Cronbach's α values exceeding 0.8. These values were comparable to those reported for the original scales in Western contexts. Despite statement reduction and cultural adjustments, the adapted scales reliably measured the intended constructs. A high Cronbach's α indicates strong coherence among statements, reflecting consistent responses from international students and ensuring dependable internal reliability. Additionally, content validity tests showed that the adapted scales covered the critical aspects of Chinese speaking. Both the CVR for individual statements and the overall CVI exceeded established thresholds, with each statement's CVR surpasses Lawshe's critical value. These robust psychometric indicators confirmed that the adapted scales effectively and consistently assessed relevant psychological traits in the target population.

Notably, items removed during adaptation were discarded chiefly because they were either tangential to the

study focus—spoken Chinese—or culturally incongruent with learners' experiences. In contrast, the retained items address more universal facets of language education and were thus well suited to a variety of instructional settings and national contexts. Future scale development and adaptation should keep these considerations in mind, rigorously vetting items for topical relevance and cultural fit to maximise both precision and cross-context applicability.

5.2. Advantages of Adapted Scales

Compared with the original scales, the adapted scales offered several advantages. First, statement optimization was achieved by carefully screening and revising the original statements to remove those unsuitable for the Chinese context or redundant in content. As a result, the scale was more concise, reducing the response burden on participants while retaining statements that focused on key dimensions of Chinese as a L2. This statement reduction did not compromise reliability; instead, it enhanced internal consistency. For example, after eliminating certain low-correlated statements, the Cronbach's α for the AMTB Scale increased, reflecting a higher degree of coherence among the remaining statements, which indicated that appropriate statement deletion and integration improve overall consistency. Second, in terms of cultural adaptation, the revised scales were more aligned with the cultural context of learning Chinese as a L2. Learning Chinese uniquely requires mastering Chinese characters and navigating the critical role of tone in semantic transmission, posing special challenges for learners whose native languages are non-tonal. Moreover, learning Chinese involves deep integration with Chinese traditional culture, history, and social norms [27]. Through careful cross-cultural adaptation, the scales retained consistency with the original constructs while incorporating local cultural characteristics. As previous research noted, the lack of rigorous cross-cultural adaptation when applying foreign scales might lead to discrepancies and reduced validity [28]. In this study, the adaptation strictly followed principles of cross-cultural measurement equivalence, with translations modified to align with Chinese linguistic habits and cultural backgrounds. Consequently, the adapted scales more accurately reflected the real experiences of international students learning Chinese speaking, offering greater cultural appropriateness, enhanced usability, and stronger explanatory power in the L2 Chinese context.

6. Implication

These findings carried important implications for SLA research, especially in less commonly studied contexts like Chinese as an L2. First, the successful validation of the scales offered researchers a new set of localized instruments for examining affective factors in Chinese language learning. Prior SLA research has often focused on English as a L2, this study extended this work by providing tools tailored to Chinese L2 learners. Therefore, future studies can confidently include measures of motivation, anxiety, and self-confidence when investigating Chinese learning, knowing that these scales were culturally and linguistically appropriate. Moreover, these results reinforced theoretical frameworks that highlight affective factors in SLA. The clear operability of three affective factors in the Chinese learning environment supported the generality of the Affective Filter Hypothesis across different languages.

Also, researchers can build on this by exploring how these measured affective factors correlate with language outcomes. For instance, do students with higher motivation scores indeed achieve better speaking proficiency over time? Does anxiety as measured by our scale predict slower development of speaking skills? Having reliable instruments enables such questions to be tested empirically. Additionally, this study encourages comparative research: scholars might compare affective profiles of Chinese learners with those of learners of other languages. As a result, this study bridges a gap in SLA research by providing validated instruments for affective factors in Chinese L2 contexts, paving the way for more nuanced and data-driven investigations into how affective factors impact language acquisition.

7. Limitations and Future Directions

While this study provided valuable insights, several limitations must be acknowledged. First, the sample size was relatively small and drawn exclusively from a single university offering an international student program. This limited sample may affect the generalizability of our findings. Also, although the EFA produced single-factor

solution with satisfactory item loadings, and reliability analyses supplied further evidence of internal consistency, a rigorous examination of convergent and discriminant validity as well as a full confirmatory factor analysis (CFA) will require a much larger dataset. Accordingly, future studies should recruit at least 200 participants to obtain stable CFA parameter estimates and support these additional validity checks.

Additionally, the primary aim of this study was to adapt existing scales for empirical research on Chinese speaking skills, focusing specifically on this sub-skill. Future research could adapt similar scales to assess listening, reading, and writing skills in Chinese L2 contexts.

Furthermore, since the current study utilized well-established scales that have been widely applied in various L2 instructional settings, our focus was primarily on content validity and internal consistency reliability. Future research could expand the sample size to conduct factor analyses in order to examine the underlying structure of each scale. Such analyses would confirm whether all statements load onto their intended constructs and provide opportunities to further refine the scales if necessary.

8. Conclusions

This study successfully adapted and validated three affective scales under Krashen's Affective Filter Hypothesis for international students learning Chinese speaking. The revised instruments—Foreign Language Speaking Anxiety Scale, AMTB Scale, and Foreign Language Self-Esteem Scale demonstrated high validity and reliability, providing SLA researchers with context-specific tools for measuring anxiety, motivation, and self-confidence in Chinese-speaking environments.

The findings underscored the critical role of localized scale adaptation in addressing cultural and linguistic disparities, ensuring accurate affective factor assessment. By bridging the gap in Chinese SLA research, this study not only advanced theoretical understanding of affective filters but also offered practical tools for educators to design targeted interventions.

Author Contributions

Conceptualization, W.B. and F.D.Y.; methodology, W.B. and F.D.Y.; software, W.B.; validation, W.B.; formal analysis, W.B.; investigation, W.B.; resources, W.B.; data curation, W.B.; writing—original draft preparation, W.B.; writing—review and editing, F.D.Y.; visualization, W.B.; supervision, F.D.Y.; project administration, F.D.Y. All authors have read and agreed to the published version of the manuscript.

Funding

This work received no external funding.

Institutional Review Board Statement

The study was conducted in accordance with the Declaration of Helsinki, and approved by the Research Ethics Committee of the Universiti Malaya (protocol code UM.TNC2.UMREC_4566 and the date of approval is March 30th, 2025).

Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

Data Availability Statement

The data supporting the findings are not publicly available due to ethical and privacy considerations. However, researchers who wish to access the data for academic or research purposes may request it by contacting the corresponding author via email.

Acknowledgments

The authors extend their sincere gratitude to the anonymous reviewers for dedicating their valuable time and providing insightful and constructive comments. We are also deeply thankful to the editors for their expert guidance and efficient handling of the manuscript throughout the review process.

Conflicts of Interest

The authors declare no conflict of interest.

Appendix A

1. Finalized Scales

Foreign Language Speaking Anxiety Scale 外语口语焦虑量表

Directions: This scale is prepared to collect information about your level of Chinese language speaking anxiety that you experience. After reading each statement, please circle the number which appeals to you most, which indicates the extent you agree with the statement. There are no right or wrong answers for the statements in this questionnaire. Thanks for your contribution.

说明：本量表旨在收集您在汉语口语交流中所体验到的焦虑程度。在阅读每个陈述后，请圈选最符合您情况的数字，以表明您对该陈述的认同程度。本问卷中的各项均没有对错之分。感谢您的参与。

1 = Strongly disagree 非常不同意

2 = Disagree 不同意

3 = Neutral 中立

4 = Agree 同意

5 = Strongly agree 非常同意

Statements 问题	1	2	3	4	5
1. I am never quite sure of myself when I am speaking in Chinese. 在说汉语时，我从未对自己感到十分自信。	1	2	3	4	5
2. I am afraid of making mistakes in Chinese classes. 我害怕在汉语课上犯错误。	1	2	3	4	5
3. I tremble when I know that I am going to be called on in Chinese classes. 当我知道自己将在汉语课上被点名时，我会紧张得发抖。	1	2	3	4	5
4. I get frightened when I don't understand what the teacher is saying in Chinese. 当我不理解汉语老师在说什么时，我会感到恐惧。	1	2	3	4	5
5. I start to panic when I have to speak without preparation in Chinese classes. 当我要在汉语课上需要即兴发言时，我会感到慌张。	1	2	3	4	5
6. I get embarrassed to volunteer read texts or answer questions in Chinese classes. 在汉语课上自愿阅读课文或者回答问题会让我感到尴尬。	1	2	3	4	5
7. I feel nervous while speaking Chinese with native speakers. 当与母语为汉语的人交谈时，我会感到紧张。	1	2	3	4	5
8. I don't feel confident when I speak Chinese in classes. 当我在课堂上说汉语时，我感到很不自信。	1	2	3	4	5
9. I am afraid that my Chinese teacher is ready to correct every mistake I make. 我担心我的汉语老师会纠正我每一个错误。	1	2	3	4	5
10. I always feel that the other students speak Chinese better than I do. 我总觉得其他学生说汉语比我说得好。	1	2	3	4	5
11. I feel very self-conscious about speaking Chinese in front of other students. 当我在其他学生面前说汉语时，我感到很不自在。	1	2	3	4	5
12. I get nervous and confused when I am speaking in Chinese classes. 当我在汉语课上说话时，我感到紧张和困惑。	1	2	3	4	5
13. I get nervous when I don't understand every word my Chinese teacher says. 当我不理解我汉语老师说的每个单词时，我会感到紧张。	1	2	3	4	5
14. I feel overwhelmed by the number of rules I have to learn to speak Chinese. 我觉得我必须学习很多规则才能说好汉语，这让我感到不知所措。	1	2	3	4	5
15. I am afraid that others will laugh at me when I speak Chinese. 当我说汉语时，我害怕别人会嘲笑我。	1	2	3	4	5
16. I get nervous when the Chinese teacher asks questions which I haven't prepared in advance. 当汉语老师问我一些我没有提前准备的问题时，我会感到紧张。	1	2	3	4	5

2. AMTB Scale 态度和动机量表

Directions: This questionnaire is prepared to collect information on your motivation for learning Chinese. After reading each statement, please circle the number which appeals to you most, which indicates the extent you agree with statement. There are no right or wrong answers for the statements in this questionnaire. Thanks for your contribution.

说明：本问卷旨在收集您学习汉语动机方面的信息。在阅读每个陈述后，请圈选最符合您情况的数字，以表明您对该陈述的认同程度。本问卷中的各项均无对错之分。感谢您的参与。

1 = Strongly disagree 非常不同意

2 = Disagree 不同意

3 = Neutral 中立

4 = Agree 同意

5 = Strongly agree 非常同意

Statements 问题	1	2	3	4	5
1. When I have a problem understanding something in my Chinese class, I always ask my teacher for help. 当我在中文课上遇到问题时，我总是向我的老师寻求帮助。	1	2	3	4	5
2. I really work hard to learn to speak Chinese. 我真的很努力学习说汉语。	1	2	3	4	5
3. When I am studying Chinese, I ignore distractions and pay attention to my task. 当我学习汉语时，我会忽略干扰，专注于我的任务。	1	2	3	4	5
4. Learning Chinese is really great. 学习汉语真的很棒。	1	2	3	4	5
5. I really enjoy learning Chinese. 我真的很喜欢学习汉语。	1	2	3	4	5
6. Chinese is a very important part of the school program. 汉语是学校课程的一个非常重要的部分。	1	2	3	4	5
7. I wish I could have many native Chinese speaking friends. 我希望我能有许多说汉语的朋友。	1	2	3	4	5
8. The more I get to know native Chinese speakers, the more I like them. 我越了解中国人，我就越喜欢他们。	1	2	3	4	5
9. Studying Chinese is important because it will allow me to be more at ease with people who speak Chinese. 学习汉语很重要，因为这将使我更加轻松地与说中国人交往。	1	2	3	4	5
10. Studying Chinese is important because it will enable me to better understand and appreciate the Chinese way of life. 学习汉语很重要，因为这将使我更好地了解和欣赏中国的生活方式。	1	2	3	4	5
11. I have a strong desire to know all aspects of Chinese. 我非常渴望了解中国的各个方面。	1	2	3	4	5
12. I would like to learn as much Chinese as possible. 我想尽可能多地学习汉语。	1	2	3	4	5
13. I wish I were fluent in speaking Chinese. 我希望我能说一口流利的汉语。	1	2	3	4	5
14. I enjoy the activities of our Chinese class much more than those of my other classes. 我更喜欢我们汉语课的活动，而不是我其他课程的活动。	1	2	3	4	5
15. I look forward to the time I spend in Chinese class. 我期待着在中文课上度过的时间。	1	2	3	4	5
16. Studying Chinese is important because I will need it for my career. 学习汉语很重要，因为我将需要它来支持我的职业。	1	2	3	4	5
17. Studying Chinese is important because other people will respect me more if I know Chinese. 学习汉语很重要，因为如果我懂汉语的话，其他人会更尊重我。	1	2	3	4	5

3. Foreign Language Self-Esteem Scale 外语自尊量表

Directions: This questionnaire is to self-report on the degree of self-esteem in speaking in Chinese language. After reading each statement, please circle the number which appeals to you most, which indicates the extent you agree with statement. There are no right or wrong answers for the statements in this questionnaire. Thanks for your contribution.

说明：本问卷旨在对您中文口语表达的自尊程度进行自我报告。在阅读每个陈述后，请圈选最符合您情况的数字，以表明您对该陈述的认同程度。本问卷中的各项均无对错之分。感谢您的参与。

1 = Strongly disagree 非常不同意

2 = Disagree 不同意

3 = Neutral 中立

4 = Agree 同意

5 = Strongly agree 非常同意

Statements	1	2	3	4	5
1. My ability to learn Chinese is high. 我学习汉语的能力很高。	1	2	3	4	5
2. I can express myself in Chinese for basic conversation. 我可以用汉语进行基本的交流。	1	2	3	4	5
3. I participate effectively in Chinese discussions of the topic I've learned. 我能有效地参与我所学习主题的汉语讨论。	1	2	3	4	5
4. I can speak Chinese well after learning. 我学习后可以很好地说汉语。	1	2	3	4	5
5. I can understand what others say in Chinese for basic conversation. 我能听懂别人在汉语中进行基本的交流。	1	2	3	4	5
6. I speak Chinese with almost no foreign accent. 我说汉语几乎没有外国口音。	1	2	3	4	5
7. I feel good about myself when speaking in the Chinese classroom. 当我在汉语课堂上说话时，我觉得很自信。	1	2	3	4	5
8. I feel at ease when I talk to my Chinese instructors. 当我和我的汉语老师交谈时，我感觉很自在。	1	2	3	4	5
9. I feel relaxing talking in Chinese in front of my classmates. 当我在同学面前用汉语交谈时，我感到很放松。	1	2	3	4	5
10. I am better Chinese learners than my classmates. 我比我的同学汉语学得更好。	1	2	3	4	5
11. My Chinese instructors have high expectations of me. 我的汉语老师对我期望很高。	1	2	3	4	5
12. My Chinese schoolmates like me. 我的汉语同学很喜欢我。	1	2	3	4	5
13. I can understand Chinese very well. 我能很好地理解汉语。	1	2	3	4	5
14. I attend Chinese class sessions on time. 我准时参加汉语课程。	1	2	3	4	5
15. I volunteer myself for any Chinese classroom activities. 我自愿参加任何汉语课堂活动。	1	2	3	4	5
16. I like discussions in Chinese for basic conversation. 我喜欢进行基本的汉语交流讨论。	1	2	3	4	5

References

[1] Łodej, M., Osmoła, A., 2024. Emotions in language learning: Understanding foreign language enjoyment and anxiety in higher education. Neofilolog. 62(1), 78–104. DOI: https://doi.org/10.14746/n.2024.62.1.6

[2] Henter, R., 2014. Affective factors involved in learning a foreign language. Procedia – Social and Behavioral Sciences. 127, 373–378. DOI: https://doi.org/10.1016/j.sbspro.2014.03.274

[3] Wang, L., Wu, X., 2020. Influence of affective factors on learning ability in second language acquisition. Revista Argentina de Clínica Psicológica. 29(2), 1232. DOI: https://doi.org/10.24205/03276716.2020.365

[4] Krashen, S., 1982. Principles and Practice in Second Language Acquisition. Pergamon Press: Oxford, UK. pp. 29.

[5] Ren, L., Kasuma, S.A.A., 2025. Latent profile analysis for the investigation of emotional factors in second language acquisition. Language Related Research. 16(1), 87–113. DOI: https://doi.org/10.29252/LRR.16.1.4

[6] Zheng, Y., 2021. Review of enlightenment of emotional factors in Krashen's affective filter hypothesis on English teaching research in college. Proceedings of the 2021 4th International Conference on Humanities Education and Social Sciences (ICHESS 2021); 17–19 September 2021; Xishuangbanna, China. Atlantis Press: Paris, France. pp. 1518–1524. DOI: https://doi.org/10.2991/assehr.k.211220.256

[7] Chen, Y., 2022. Analysing the impact of the affective filter hypothesis on teaching spoken Chinese as a foreign language. Education Review USA. 6(9), 516–518. DOI: https://doi.org/10.26855/er.2022.09.012

[8] Wang, T., 2024. Analysis of the affective factors of high school struggling students' English learning based on affective filter hypothesis. Journal of Education and Educational Research. 8(2), 178–180. DOI: https://doi.org/10.54097/vep4rp21

[9] Pan, Z., Wang, Y., Derakhshan, A., et al., 2023. Unpacking Chinese EFL students' academic engagement and psychological well-being: The roles of language teachers' affective scaffolding. Journal of Psycholinguistic Research. 52(5), 1799–1819. DOI: https://doi.org/10.1007/s10936-023-09974-z

[10] Horwitz, E.K., Horwitz, M.B., Cope, J., 1986. Foreign language classroom anxiety. Modern Language Journal. 70(2), 125–132.

[11] Öztürk, G., Gürbüz, N., 2014. Speaking anxiety among Turkish EFL learners: The case at a state uni-versity. Journal of Language and Linguistic Studies. 10(1), 1–17.

[12] Gardner, R.C., 2004. Attitude/Motivation Test Battery: International AMTB Research Project. University of Western Ontario: London, ON, Canada.

[13] Hassan, B.A., 1992. Foreign language self-esteem and the oral communicative ability of university students. Mansoura Faculty of Education Journal. 19(3), 1–19.

[14] Klotz, A.C., Swider, B.W., Kwon, S.H., et al., 2023. Back-translation practices in organizational research: Avoiding loss in translation. Journal of Applied Psychology. 108(5), 699–727. DOI: https://doi.org/10.1037/apl0001050

[15] Bundgaard, K., Brøgger, M.N., 2018. Who is the back translator? An integrative literature review of back translator descriptions in cross-cultural ada-ptation of research instruments. Perspectives. 27(6), 833–845. DOI: https://doi.org/10.1080/0907676X.2018.1544649

[16] Alqahtani, T.M., Yusop, F.D., Halili, S.H., et al., 2023. Content validity of the Constructivist Learning in Higher Education Settings (CLHES) scale in the context of the flipped classroom in higher education. Humanities and Social Sciences Communications. 10(1), 1–12. DOI: https://doi.org/10.1057/s41599-023-01754-3

[17] Gilbert, G.E., Prion, S., 2016. Making sense of methods and measurement: Lawshe's content validity index. Clinical Simulation in Nursing. 12(12), 530–531. DOI: https://doi.org/10.1016/j.ecns.2016.08.002

[18] Beck, C.T., Gable, R.K., 2001. Ensuring content validity: An illustration of the process. Journal of Nursing Measurement. 9(2), 201–215.

[19] Lawshe, C.H., 1975. A quantitative approach to content validity. Personnel Psychology. 28(4), 563–575.

[20] Polit, D.F., Beck, C.T., 2006. The content validity index: Are you sure you know what's being reported? Critique and recommendations. Research in Nurs-ing & Health. 29(5), 489–497. DOI: https://doi.org/10.1002/nur.20147

[21] Lu, E.S., Katz, R., Miller, J.B., et al., 2022. Peripapillary choroidal vascularity and visual correlates in non-arteritic anterior ischemic optic neuropathy using swept-source optical coherence tomography. Frontiers in Ophthalmology. 2, 848040. DOI: https://doi.org/10.3389/fopht.2022.848040

[22] Yan, Y., Xin, T., Wang, D., et al., 2014. Application of the Geriatric Anxiety Inventory-Chinese Version (GAI-CV) to older people in Beijing communities. International psychogeriatrics. 26(3), 517–523. DOI: https://doi.org/10.1017/S1041610213002007

[23] Kennedy, I., 2022. Sample size determination in test-retest and Cronbach alpha reliability estimates. British Journal of Contemporary Education. 2(1), 17–29. DOI: https://doi.org/10.52589/BJCE-FY266HK9

[24] Rocereto, J.F., Puzakova, M., Anderson, R.E., et al., 2011. The role of response formats on extreme response style: A case of Likert-type vs semantic differential scales. In: Sarstedt, M., Schwaiger, M., Taylor, C.R. (eds.). Measurement and Research Methods in International Marketing (Advances in International Marketing, Vol. 22). Emerald Group Publishing Limited: Leeds, England, UK. pp. 53–71. DOI: https://doi.org/10.1108/S1474-7979(2011)0000022006

[25] Van Vaerenbergh, Y., Thomas, T.D., 2013. Response styles in survey research: A literature review of antecedents, consequences, and remedies. International Journal of Public Opinion Research. 25(2), 195–217. DOI: https://doi.org/10.1093/ijpor/eds021

[26] Kalkbrenner, M.T., 2023. Alpha, omega, and H internal consistency reliability estimates: Reviewing these options and when to use them. Counseling Outcome Research and Evaluation. 14(1), 77–88. DOI: https://doi.org/10.1080/21501378.2021.1940118

[27] Chu, W.H., Lin, D.Y., Chen, T.Y., et al., 2015. The relationships between ambiguity tolerance, learning strategies, and learning Chinese as a second language. System. 49, 1–16. DOI: https://doi.org/10.1016/j.system.2014.10.015

[28] Taguchi, N., Xiao, F., Li, S., 2016. Assessment of study abroad outcomes in Chinese as a second language: Gains in cross-cultural adaptability, language contact and proficiency. Intercultural Education. 27(6), 600–614. DOI: https://doi.org/10.1080/14675986.2016.1217126