Psychometric Evaluation of the Foreign Language Learner Self-Efficacy Questionnaire for English Argumentative Writing

Lei Zhang[1], Samah Ali Mohsen Mofreh[1,*], Qian Sun[2], and Sultan Salem[3]

[1]School of Educational Studies, Universiti Sains Malaysia, Penang, Malaysia
[2]College of Foreign Languages, North China University of Science and Technology, Tangshan, China
[3]Department of Economics, Birmingham Business School, College of Social Sciences, University of Birmingham, Birmingham, UK
Email: zhanglei@student.usm.my (L.Z.); samahmofreh@usm.my (S.A.M.M.); sunny8094@163.com (Q.S.); s.salem@bham.ac.uk (S.S.)
*Corresponding author

Manuscript received September 10, 2025; revised September 29, 2025; accepted November 24, 2025; published April 10, 2026

***Abstract*—Recent academic work has highlighted the need for a valid and trustworthy tool to assess Chinese undergraduates' self-efficacy in English argumentative writing. To tackle the challenge posed by the shortage of relevant tools, this study evaluated the psychometric properties of a newly adapted questionnaire, called Foreign Language Learner Self-Efficacy Questionnaire for English Argumentative Writing (FLLSEQEAW), using the Rasch model. The research included 563 undergraduate students from a key provincial university in northern China. The data were analyzed to check the questionnaire's psychometric quality. The findings confirm that the FLLSEQEAW is a unidimensional and psychometrically sound instrument. The study also acknowledges limitations and proposes future studies to gather further evidence of its validity.**

***Keywords*—psychometric evaluation, self-efficacy, English argumentative writing**

I. Introduction

Writing is a fundamental component of foreign language acquisition, yet it is frequently regarded as more challenging than other linguistic skills such as listening, speaking, and reading [1]. This complexity stems from the fact that writing is a demanding cognitive and affective activity.

This is especially true for the genre of argumentative writing [2, 3], which necessitates that authors articulate a definitive claim, substantiate it with relevant evidence, and engage with alternative perspectives [4]. Within China's educational landscape, this form of writing has become a prominent feature of large-scale national assessments, specifically the College English Test (CET-4 and CET-6), which are critical exams that evaluate the English proficiency of millions of Chinese university students.

The writing component of the CET-4 and CET-6 examinations requires test-takers to compose an essay within a 30-minute time limit. Historically, the assessed genres have varied, but recent data (from 2021 to the present) shows a strong emphasis on argumentative writing. This genre now constitutes approximately 60% of CET-4 and 90% of CET-6 writing assessments. Consequently, success demands that students cultivate the ability to critically evaluate different perspectives and effectively articulate and defend their own reasoned opinions.

The argumentative essay is a genre of academic writing that obliges the writer to comprehensively explore a topic, critically analyze evidence, and assert a well-defined position [5]. During this writing process, students learn to move beyond passive reception of information to active critique and synthesis [6] which is a skill needed in any professional or academic pursuit. Moreover, English argumentative writing demands precision in language and clarity in thought [7]. It forces students to eliminate vagueness and convey complex ideas with exactness. This discipline translates directly into professional communication, enabling students to establish credibility and professionalism in their future careers [8].

Although argumentative writing is crucial, the skills of Chinese college students in this area are significantly underdeveloped. Students frequently have trouble making clear claims [9] and supporting them with adequate, high-quality evidence [10–13]. Furthermore, they often overlook key elements of argumentation, such as providing logical justifications (warrants) or addressing and refuting opposing arguments (rebuttals) [13–17].

In fact, argumentative writing is a meta-skill that embodies and drives critical thinking [18, 19]. Unlike descriptive or narrative writing, which can often rely on observation, argumentation forces the writer to engage in higher-order cognitive processes. Additionally, it also forces students' engagement with opposing views [20], combating innate cognitive biases. Research across multiple fields demonstrates that argumentative writing practice cultivates the sophisticated habits of mind that are the ultimate goal of higher education. The enhancement of writing performance is influenced by various factors, including writing anxiety, writing strategies, and writing self-efficacy [21].

Self-efficacy is a person's belief in their own ability to plan and execute the tasks needed to achieve specific goals [22]. This belief is vital in education, as it affects a student's choice of tasks, learning strategies, motivation, effort, and ultimately, their academic achievement [23–29]. Students with high self-efficacy are typically more confident, motivated, and resilient in their learning [23–27].

In writing, self-efficacy refers to a learner's judgment of their own ability to perform writing tasks successfully [30]. It is widely recognized that self-efficacy becomes particularly critical when writing tasks are demanding and motivational conditions are low, a situation often encountered by those who learn English as second or foreign language [31]. This phenomenon is especially relevant for Chinese English learners, who face challenging contexts that necessitate proficient English writing skills for success in high-stakes examinations [32].

In the context of second-language (L2) writing in English,

students with high self-efficacy are more inclined to undertake challenging writing tasks and invest greater effort [33], as well as exhibit greater perseverance throughout the writing process [34]. Prior research has demonstrated that self-efficacy is a significant predictor of language proficiency in writing achievement [33, 35]. Moreover, research shows a strong positive link between self-efficacy and writing quality, both for native English speakers [36] and foreign language learners [32, 37, 38].

Additionally, self-efficacy is significantly associated with Chinese college students' English argumentative writing. For instance, Shen and Bai [39] assert that writing self-efficacy, as a motivational construct, reflects learners' perceived capabilities to successfully undertake English argumentative writing tasks. Teng *et al.* [32] found that self-efficacy demonstrated small to moderate correlations with performance in English argumentative writing. Zhang and Zhang's findings [38] indicated the combined effects of writing self-efficacy and Self-Regulated Learning (SRL) writing strategies on English argumentative writing achievement.

Therefore, given the importance of English argumentative writing, creating a reliable tool to measure a student's self-efficacy in this area is essential. This tool should assess their confidence in their language skills, self-regulation, and behavioral control throughout the entire writing process.

Developed specifically for EFL learners, the current proposed instrument builds on Teng *et al.*'s [32] concept of L2 writing self-efficacy to measure essential argumentative writing skills and the distinct challenges of the writing process. This study employed the Rasch model to assess the psychometric properties of the FLLSEQEAW among Chinese higher education students. The investigation was designed to answer the following key research questions:

RQ1: Does the modified self-efficacy instrument possess sufficient item psychometrics to validly measure Chinese undergraduates' English argumentative writing self-efficacy as a unidimensional trait?

RQ2: Does the 7-point Likert response instrument demonstrate proper psychometric function?

RQ3: Do all 36 items demonstrate statistically appropriate model fit indices?

RQ4: Are the person- and item-level separation and reliability acceptable?

RQ5: Do the 36 items demonstrated well between item endorsability and person proficiency?

RQ6: Does the instrument maintain invariance between different demographic groups, specifically regarding gender differences (male versus female students)?

II. Literature Review

A. The Assessment of L2 Writing Self-Efficacy

It is widely recognized that self-efficacy could be accurately assessed through the use of questionnaires or scales [40]. Meanwhile, the development of self-efficacy measures involves aligning designing items to fit specific contexts and the unique expectations of different tasks [40].

Sun *et al.* [41] introduced the Questionnaire of English Writing Self-Efficacy (QEWSE), which was based on the Self-Efficacy for Writing Scale (SEWS) by Bruning *et al.* [42] and Wang and Bai's [43] Questionnaire of English Self-Efficacy (QESE). Sun *et al.* [41] emphasized that a valid measure of writing self-efficacy must evaluate learners' confidence in their linguistic skills as well as their self-regulation. The development of QEWSE was grounded in theories of linguistics and writing self-regulation. QEWSE comprised four dimensions: ideation, skills, use of English writing, and self-regulation.

The QEWSE proved reliable in detecting students exhibiting weaker writing self-efficacy beliefs [41]. Results from the confirmatory factor analysis supported adequate model fit for the four-factor solution (ideation, skills, use, and self-regulation), providing evidence for the internal structure of the instrument. Significant correlations between self-efficacy ratings, writing proficiency, and writing self-regulated learning strategies further supported the validity of the instrument based on its relation with other variables. Item Response Theory analyses revealed that measurement precision was highest at the lower range of the latent trait scale, justifying the primary use of the scale for detecting students exhibiting weaker writing self-efficacy beliefs.

Zhang *et al.* [40] introduced the Genre-based Second Language Writing Self-Efficacy Scale (GL2WSS). The instrument organized writing self-efficacy into four distinct dimensions: genre-based performance self-efficacy, linguistic self-efficacy, self-regulatory self-efficacy, and classroom performance self-efficacy. The development of the GL2WSS scale incorporated the specific textual characteristics of both argumentative and narrative compositions [40].

Statistical analyses indicated that GL2WSS possesses robust psychometric properties in terms of high reliability, appropriate factorial structure, and satisfactory evidence of both convergent and discriminant validity. Overall, strong empirical evidence for the GL2WSS's validity was provided as a reliable measure of writing self-efficacy [40].

Teng *et al.* [32] proposed a multidimensional framework for writing self-efficacy within an English as a L2 context, grounded in social cognitive theory and self-regulated learning, naming Second Language Writer Self-Efficacy Scale (L2WSS). Teng *et al.* [32] operationalized self-efficacy as students' judgments of their ability to utilize various linguistic or rhetorical knowledge, as well as their confidence in learning performance and metacognitive regulation. This innovative instrument encompasses three dimensions: linguistic self-efficacy, self-regulatory efficacy, and performance self-efficacy, reflecting learners' linguistic, behavioral, and self-regulatory demands throughout the writing acquisition process [32]. Overall, the framework of L2 writing self-efficacy proposed by Teng *et al.* [32] incorporated key aspects of L2 writing processes, the obstacles experienced by L2 writers, and their surrounding learning environments.

Confirmatory factor analyses conducted through structural equation modeling validated Teng *et al.* [32]'s three-dimensional structure of writing self-efficacy. Assessments of internal and composite reliability, as well as convergent and discriminant validity were evaluated, which indicated satisfactory psychometric properties of the scale.

Sun *et al.*' [41], Zhang *et al.*'s [40] as well as

Teng *et al.*'s [32] studies provided integrated perspectives to measure self-efficacy of L2 writing. However, a significant research gap remains in the assessment of self-efficacy related to the whole acquisition process of English argumentative writing. While existing scales measure confidence in general skills (e.g., linguistic, self-regulation, classroom performance), they do not typically capture a writer's judgment of their capability to execute the demands of a specific English argumentative writing task, such as formulating a thesis for a given prompt or structuring an argument for a particular audience. Addressing this gap requires instruments that move beyond generalized traits to evaluate situated, task-specific self-efficacy beliefs.

B. Linguistic Self-Efficacy, Self-Regulatory Efficacy, and Performance Self-Efficacy

According to Teng *et al.* [32], linguistic self-efficacy encompasses an individual's perceived ability to access lexical knowledge from memory, employ syntactically appropriate structures for idea expression, and adhere to discourse conventions in written composition. Findings from this dimension are expected to provide valuable insights for second language researchers and educators about how learners' self-assessments of language proficiency influence their cognitive development, behavioral transformation, and motivational engagement in writing tasks [32].

Self-regulatory efficacy pertains to learners' self-assessment of their ability to employ metacognitive strategies (including monitoring, evaluation, and goal establishment) during writing skill acquisition [32]. To operationalize the construct of self-efficacy within the framework of Self-Regulated Learning (SRL), Teng *et al*. [32] focused on two critical aspects: metacognitive control and goal orientation. Scholars generally concur that Self-Regulated Learning (SRL) constitutes a metacognitive process wherein learners examine their cognitive processes to assess their behavioral outcomes and devise alternative strategies for achievement [44].

To overcome the difficulties faced by many second language writers, research suggests that effective writing development necessitates cognitive regulation, wherein learners generate substantive writing ideas, set learning objectives to guide their practice, and employ strategic monitoring and evaluation techniques to assess their performance [45, 46]. Teng *et al.* [32] posited that incorporating self-regulatory efficacy would illustrate how second language writers actively manage their learning behaviors to conquer obstacles and psychological barriers encountered during writing development.

Performance self-efficacy pertains to learners' self-assessment of their capacity to complete academic tasks within instructional settings [32]. This construct reflects writers' confidence in successfully executing writing tasks, which correlates with their extrinsic motivation, strategy implementation, and self-regulatory capabilities [47]. The incorporation of this dimension draws upon self-regulated learning theory and social cognitive approaches to writing, both emphasizing the significance of socio-behavioral influences [32].

Within social cognitive theory, performance regulation constitutes a critical component of self-regulated learning, representing individuals' capacity to govern their observable actions [48]. Pintrich's [49] research established that learners' behavioral confidence significantly influences their knowledge acquisition, motivational levels, and academic engagement. Consequently, investigating students' task-performance confidence in classroom contexts yields valuable insights for educators to cultivate autonomous and self-initiated learners in writing pedagogy [32].

Although Teng *et al.* [32] provide a robust theoretical framework for the dimensions of L2 writing self-efficacy, a significant research gap lies in understanding the dynamic interplay between these dimensions during authentic English argumentative writing tasks. It remains unclear how a deficit in linguistic self-efficacy might impede the application of self-regulatory strategies, or how performance self-efficacy in the classroom may translate to independent argumentative writing contexts.

C. Interplay of Linguistic, Metacognitive, and Performance Factors in EFL Writing

Contemporary research conceptualizes EFL writing as a complex construct, distinct from first language composition, where achievement is influenced by a dynamic interplay of linguistic proficiency, metacognitive strategy use, and affective factors, all within a specific instructional context.

Research establishes that second language writing performance is more strongly associated with linguistic proficiency and fluency than first language writing [32, 50]. This is because second language learners face with foundational challenges in linguistic competence, struggling to communicate meaning through appropriate lexical choices and grammatical constructions [50]. This inherent difficulty with the language itself forms a primary barrier to effective composition.

A key factor exacerbating these linguistic challenges is learners' psychology, particularly self-confidence. Studies indicate that many Chinese college students lacked writing self-confidence [51, 52]. This lack of confidence is not merely a feeling but has tangible adverse effects: it impairs their ability to apply existing writing knowledge, develop effective strategies, and ultimately achieve academic success [53, 54]. Thus, confidence acts as a crucial mediator between linguistic knowledge and its successful application. Regarding students' low confidence, teachers are instrumental in designing and implementing effective motivational strategies [55–57]. For instance, teachers can facilitate collaborative discussions, outlining explicit expectations for engagement, and delivering constructive feedback to foster students' argumentative skill development [58]. Moreover, the effect of low linguistic proficiency and diminished confidence manifests in a deficit in metacognitive strategy use. In terms of EFL writing, a limited grasp of metacognitive strategies presents a significant challenge for university EFL students [59, 60]. Consequently, many EFL writers report significant difficulties in regulating their learning-to-write processes [61] and express lower confidence in their classroom writing performance [32, 62]. This creates a cyclical relationship where poor performance diminishes confidence, which in turn inhibits the use of strategies that could improve performance.

This synthesis argues that the instructional environment can either mitigate or reinforce these challenges. In many university-level EFL settings, instruction predominantly emphasizes linguistic knowledge acquisition and examination preparation [32]. This narrow focus overlooks learners' varied requirements and psychological challenges, such as their limited confidence and decreased enthusiasm for writing exercises, thereby failing to address the very factors that hinder progress [32].

In contrast to traditional, mechanically focused instruction, contemporary EFL writing research advocates for a broader view. It reconceptualizes composition not as the mere replication of linguistic and cultural norms, but as a dynamic and active interaction between writers and their social environment. This perspective necessitates an instructional approach that moves beyond grammar and exams to also nurture confidence, explicitly teach metacognitive and self-regulatory strategies, and engage students in authentic social interactions through writing.

Thus, EFL writing proficiency is not solely a product of linguistic knowledge. It is an outcome shaped by a framework where linguistic competence, metacognitive strategy use, and self-confidence are deeply intertwined. Effective argumentative writing instruction must therefore adopt a holistic approach that addresses all three components—developing language skills, building strategic competence, and fostering the psychological confidence needed to engage in the dynamic process of writing.

Grounded in the L2WSS framework by Teng *et al.* [32], the present research developed the Foreign Language Learner Self-Efficacy Questionnaire for English Argumentative Writing (FLLSEQEAW). This instrument assesses Chinese college students' self-efficacy across three core dimensions, linguistic self-efficacy, self-regulatory efficacy, and performance self-efficacy, throughout the process of learning English argumentative writing. Its application is intended to facilitate investigation into the dynamic relationships between these dimensions and to inform targeted teaching strategies that concurrently develop all three, thereby cultivating a more comprehensive sense of self-efficacy in EFL writers.

III. Materials and Methods

A. Participants

The study initially recruited 650 undergraduates from a key provincial university in Hebei Province, China. To reduce bias, participation was anonymous, with only basic demographic data (gender, age, grade) collected. Prior to any analysis of the assembled data, data screening and cleaning were conducted [63]. After screening, 87 responses were removed for insufficient effort or inaccuracies, resulting in a final sample of 563 participants.

Linacre states that scale targeting influences the minimum sample size for Rasch analysis [64]. While Wright and Stone successfully conducted an analysis with only 35 participants and 18 items [65], Linacre contends that a sample of 243 is sufficient for precise estimation of both item and person parameters, independent of how well the scale is targeted [64]. Our final sample of 563 participants comfortably exceeds the benchmark, ensuring the reliability of the psychometric properties examined.

As shown in Table 1, the final group consisted of 231 males (41.0%) and 332 females (59.0%). By academic year, 280 were first-years (49.7%), 240 were second-years (42.6%), 36 were third-years (6.4%), and 7 were fourth-years (1.2%).

Table 1. Demographic information of the participants

Variables	Level	Frequency	Percentage
Gender	Male	231	41.0%
	Female	332	59.0%
Grade	Grade One	280	49.7%
	Grade Two	240	42.6%
	Grade Three	36	6.4%
	Grade Four	7	1.2%

B. Instrument

This study used an adapted version of the Foreign Language Learner Self-Efficacy Questionnaire for English Argumentative Writing (FLLSEQEAW). The 36-item instrument measures three dimensions: linguistic self-efficacy, self-regulatory efficacy, and performance self-efficacy, addressing the language, self-regulation, and behavioral demands of the writing process.

To ensure clarity, the items were professionally translated into Chinese. Five experts validated the questionnaire, confirming its high content validity for the study. They suggested minor wording improvements to two items (Q3 and Q9) to emphasize practical application and conciseness, which were incorporated. A pilot study with five respondents confirmed that all items were clearly understood.

The final FLLSEQEAW consists of 36 items concerning three aspects: linguistic (1–11), self-regulatory (12–30), and performance (31–36). Responses were recorded on a 7-point Likert scale from "Not at all true of me" to "Quite true of me" (see Supplementary).

1) Linguistic self-efficacy scale

Building on Teng *et al.*'s Second Language Writer Self-Efficacy Scale (L2WSS) [32], which measured linguistic self-efficacy through 7 items on general writing competence, the present study developed a more specialized assessment. The FLLSEQEAW's linguistic self-efficacy dimension was expanded to 11 items to fully capture the language demands of argumentative writing. This shift moved from evaluating broad skills (e.g., "I can write compound and complex sentences with grammatical structure".) to genre-specific competencies (e.g., "I can check the cohesiveness as well as clear logic among sentences"), thereby providing a more contextualized and actionable measurement tool (see Supplementary).

2) Self-regulatory efficacy scale

The self-regulatory efficacy dimension was significantly expanded and specialized in the FLLSEQEAW. While Teng *et al.*'s L2WSS [32] conceptualized it broadly with 6 items covering general metacognitive strategies like goal-setting and planning, the FLLSEQEAW developed it into a comprehensive 19-item dimension. This expanded scale addressed the specific self-regulatory challenges of argumentative writing, including detailed aspects of goal setting, progress monitoring, focus maintenance, frustration management, and self-evaluation throughout the entire writing process (see Supplementary).

3) Performance self-efficacy scale

The performance self-efficacy dimension was refined and focused in the FLLSEQEAW. While Teng *et al.*'s L2WSS [32] conceptualized it with 7 items assessing general classroom task completion and course performance, the FLLSEQEAW condensed it to 6 items that specifically targeted key aspects of argumentative writing performance. These items directly assessed students' confidence in mastering writing knowledge, formulating ideas, and demonstrating competence through assignments and tests, providing a more targeted measurement for the specific domain (see Supplementary).

C. Data Collection

Data were collected online using the Wenjuanxing platform[1] with a convenience sampling method. Participants received instructions in Chinese that guaranteed the confidentiality of their responses and personal information. They were assured the data was for academic purposes only to encourage honest answers based on their genuine learning experiences.

D. Data Analytical Procedure

This study used Rasch model analysis to evaluate the psychometric quality of the adapted FLLSEQEAW, a questionnaire designed to measure Chinese college students' self-efficacy in English argumentative writing. The analysis specifically assessed six key properties: (a) unidimensionality, (b) rating scale function, (c) item fit (using infit/outfit MNSQ statistics and polarity indices), (d) reliability and separation indices for both items and persons, (e) Wright Map interpretation, and (f) differential item functioning. The specific criteria for each of these assessments are detailed in the following section.

Initial verification focused on assessing the unidimensionality assumption fundamental to Rasch model applications. The concept of unidimensionality implies the existence of a dominant latent trait that accounts for the variance demonstrated in the assessment responses. The explained variance represents the Rasch dimension, with the remaining variance attributable to all other dimensions and random noise. The research employed Principal Component Analysis of Residuals (PCAR) to assess scale unidimensionality, following Smith's [66] methodological approach. PCAR systematically analyzes the unexplained variance to identify potential secondary dimensions in the measurement construct [67]. The study applied established psychometric criteria requiring: (1) minimum 40% of variance explained by the Rasch dimension [67]. (2) the first contrast must explain under 15% of total variance after extracting the Rasch dimension. (3) first contrast eigenvalue <3.0 [68, 69], with ≥3:1 ratio between Rasch dimension and first contrast variance [70].

The second analysis focused on testing monotonicity, a foundational requirement of the Rasch model. The monotonicity principle asserts that response probability monotonically increases with higher levels of the latent trait being assessed. In the present research, verification of this assumption was evaluated based on the following analytical standards [71, 72]: (1) category average measures should demonstrate consistent increase in accordance with the monotonicity principle [71], and (2) the difficulty of endorsement for rating scale structure requires sequential category thresholds increasing by 1.4–5.0 logits [73].

Next, Rasch fit statistics assess the consistency of empirical data with Rasch model requirements [60] and identify misfitting items. The present research evaluated item-level psychometrics through analysis of both Infit and Outfit Mean Square statistics, along with the polarity of item. Infit MNSQ that particularly detects misfit when respondents of similar competence to the item difficulty exhibit unexpectedness. As an unweighted mean square statistic, the Outfit MNSQ demonstrates heightened sensitivity to unexpected responses [74].

According to Bond and Fox [72], the MNSQ values for both Infit and Outfit should ideally fall between 0.5 and 1.5 to confirm the suitability of the items for measuring the constructs. Additionally, item polarity was assessed and displayed the point measure correlation (PTMEA CORR) coefficient. Higher PTMEA CORR values indicate items that more effectively differentiate participants' abilities. The analysis employed an inclusion threshold ranging from +0.2 to +1.0 logits ($0.2 < x < 1.0$) for this aspect [75].

Fourth, the analysis evaluated both person and item separation indices along with their corresponding reliability coefficients. Person reliability indicates how reproducibly individuals would be ranked using equivalent test forms, while the item reliability demonstrates the internal consistency reliability of multi-item scales. Person- and item-level reliability estimates are deemed as acceptable with a threshold of 0.70, satisfactory at 0.80, and excellent at 0.90 [72]. The person separation index quantifies the discernible ability strata within the sample population [72], whereas item separation verifies the hierarchical arrangement of items. According to Malec *et al.* [76], acceptable psychometric thresholds require person separation values exceeding 2.0, while item separation should surpass 4.0.

Fifth, visual analysis of the Wright Map was utilized to determine whether items aligned with respondents' capacities. Mistargeting is operationally defined as a divergence surpassing 1.0 logit between item endorsability and person ability estimates [77].

Ultimately, the investigation of scale generalizability was operationalized through Differential Item Functioning (DIF) analysis among key demographic subgroups. Differential Item Functioning (DIF) is a statistical phenomenon indicating item bias, where an item performs differently for distinct demographic groups (e.g., gender, age) despite members of those groups having equivalent underlying ability levels [78]. According to Rasch measurement theory [74], individuals with equal ability should have the same probability of item endorsement across all demographic subgroups. DIF indicates potential measurement bias. The present analysis employed a stringent DIF detection threshold of ≥0.5 logit difference in item functioning across groups [79]. Items demonstrating significant DIF may require various corrective actions to maintain the instrument's validity and fairness across populations, including the deletion of biased items, the addition of new items, or the creation of distinct measurement tools for different demographic groups. In the present research, DIF

[1]https://www.wjx.cn/

was assessed with respect to gender differences among respondents, comparing male learners with their female counterparts.

IV. Result and Discussion

This chapter presents the findings and discusses their significance within the broader context of this research.

A. Result

1) Unidimensionality

Table 2. Standardized residual variance (in Eigenvalue units)

Item	Empirical		Modeled
Total raw variance in observations	78.0	100.0%	100.0%
Raw variance explained by measures	42.0	53.9%	54.0%
Raw variance explained by persons	24.1	30.9%	31.0%
Raw Variance explained by items	17.9	23.0%	23.1%
Raw unexplained variance (total)	36.0	46.1% 100.0%	46.0%
Unexplned variance in 1st contrast	3.3	4.3% 9.2%	
Unexplned variance in 2nd contrast	2.8	3.6% 7.9%	
Unexplned variance in 3rd contrast	2.6	3.3% 7.2%	
Unexplned variance in 4th contrast	2.0	2.5% 5.4%	
Unexplned variance in 5th contrast	1.8	2.3% 5.1%	

As shown in Table 2, the Principal Component Analysis of Residuals (PCAR) supported the scale's unidimensionality. The Rasch dimension explained 53.9% of the overall variance, surpassing the 40% benchmark [67]. Although the eigenvalue of the first contrast (3.3) slightly exceeded the 3.0 cutoff [68, 69], the variance it explained (4.3%) was well below the 15% threshold. Crucially, the ratio of variance explained by the measurement dimension to that of the first contrast was 12.5 (53.9%/4.3% = 12.5), significantly exceeding the recommended 3:1 standard [70]. These results collectively indicated that the FLLSEQEAW data adequately fit the Rasch model's unidimensional expectations.

2) Rating scale functioning

The psychometric functioning of the 7-point Likert scale was evaluated and found to be appropriate, as evidenced in Table 3. The data followed the anticipated standard trend, demonstrating a systematic progression from a negative to a positive state. Specifically, the average measure advanced monotonically from –0.48 logit to +3.14 logits, confirming a normal distribution of participant responses. The structure calibration also progressed orderly with the category levels. Supplementary to this, Fig. 1 provided a visual representation of the categorical structure, detailing the threshold progression and the interval magnitudes between each response category.

Table 3. Calibration scaling analysis

Category Label	Observed Count %	Observed Avg.	Sample Expect	Infit MNSQ	Outfit MNSQ	Structure Calibration	Category Measure
1	16 3%	–0.48	–1.81	3.46	4.09	None	(–3.42)
2	48 9%	–0.20	–0.50	1.92	2.49	–2.89	–1.62
3	71 13%	–0.01	0.14	0.99	1.16	–1.24	–0.51
4	159 28%	0.60	0.65	1.06	1.08	–1.11	0.44
5	139 25%	1.21	1.26	0.97	0.95	0.38	1.56
6	111 20%	2.00	2.07	1.18	1.14	1.17	3.24
7	19 3%	3.14	3.27	1.33	1.14	3.69	(5.55)

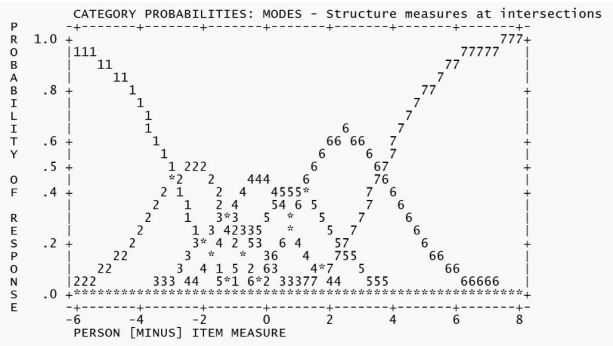

Fig. 1. Category probabilities: Modes–structure measures at intersections.

3) Item fit and item polarity

Evaluation of item fit statistics, presented in Table 4, confirmed that 35 of the 36 items conformed well to Rasch model expectations. The Infit MNSQ values for items Q2 through Q36 ranged from 0.80 to 1.24, and their Outfit MNSQ values ranged from 0.79 to 1.30, all falling within the acceptable bounds for productive measurement. The PTMEA CORR coefficients (0.2 to 1.0) for these items further verified their effective contribution to the measurement of self-efficacy.

In contrast, Item Q1 was flagged for misfit. Its Infit MNSQ was 1.40 and its Outfit MNSQ was 1.51, the latter slightly exceeding the recommended maximum of 1.5. The researchers concluded that the phrasing of Q1, concerning the "core elements of a good composition", was unclear and posed interpretative challenges for respondents. On this basis, Item Q1 is deemed non-contributory and is recommended for deletion from the final instrument.

Table 4. Item fit statistics

Item	Measure	INFIT MNSQ	OUTFIT MNSQ	PTMEA CORR
1	0.70	1.40	1.51	0.63
3	0.13	1.21	1.30	0.65
2	0.76	1.24	1.25	0.65
21	–0.42	1.24	1.22	0.65
23	–0.83	1.16	1.20	0.66
20	0.19	1.19	1.19	0.67
30	0.08	1.17	1.19	0.66
11	0.46	1.12	1.13	0.69
5	–0.50	1.08	1.11	0.67
29	–0.04	1.04	1.07	0.70
25	–0.35	1.05	1.06	0.69
6	0.23	1.04	1.05	0.70
12	–0.45	1.05	1.03	0.70
24	–0.21	1.04	1.02	0.70
14	0.14	1.03	1.02	0.71
19	–0.14	0.98	1.00	0.72
7	0.11	0.97	0.99	0.72
4	0.24	0.98	0.99	0.71
8	–0.18	0.98	0.97	0.72
34	0.60	0.97	0.97	0.74
13	–0.08	0.95	0.95	0.74
9	0.18	0.92	0.92	0.74
22	–0.32	0.91	0.91	0.74
17	–0.40	0.91	0.91	0.74
36	0.21	0.91	0.90	0.75
26	–0.22	0.90	0.88	0.74
35	0.38	0.87	0.88	0.76
31	–0.32	0.87	0.88	0.75
32	0.15	0.86	0.86	0.77

28	0.08	0.84	0.85	0.76
10	–0.02	0.84	0.84	0.76
33	0.10	0.82	0.79	0.78
27	–0.13	0.81	0.82	0.77
18	0.04	0.81	0.80	0.78
16	–0.07	0.80	0.79	0.78
15	–0.09	0.80	0.80	0.78

4) Reliability and separation

As shown in Table 5, the FLLSEQEAW demonstrated high reliability, with Rasch item and person reliability indices of 0.98 and 0.97, respectively. The item separation index (6.25) and person separation index (5.36) were both satisfactory, significantly exceeding the recommended thresholds of 4 and 2. These results confirmed that the scale was highly sensitive in distinguishing between respondents with varying levels of self-efficacy in English argumentative writing.

Table 5. Item- and person-level reliability and separation

Index	Item	Person
Separation	6.25	5.36
Reliability	0.98	0.97

5) Wright map

The distribution of items and persons on the Wright Map (Fig. 2) identified Q2 as the most challenging item, followed by Q1 and Q34. The high difficulty of Q2 was posited to stem from students' fundamental lack of confidence in their lexical resource for argumentation. The difficulty of Q1 was attributed to its ambiguous phrasing, while Q34's difficulty likely reflected students' general assessment anxiety.

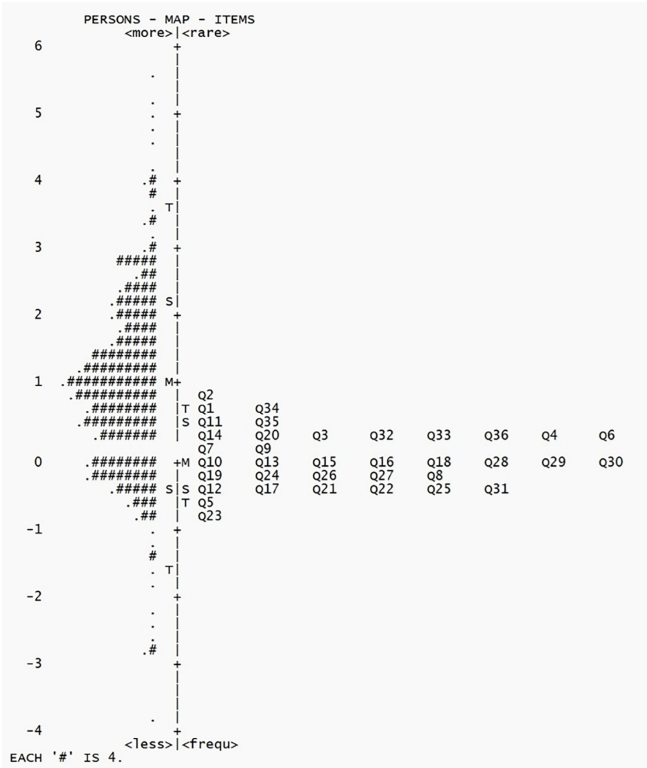

Fig. 2. Wright map.

Conversely, Item Q23 was identified as the least difficult. It was likely because timely submission was perceived as a manageable, baseline academic requirement. Furthermore, the mean person ability was closely aligned with the mean item difficulty, exhibiting 0.94 logit difference. This close match indicated that the FLLSEQEAW items were well-targeted to the ability levels of the respondents in this sample.

6) DIF

The researchers conducted a Differential Item Functioning (DIF) analysis to evaluate the consistency of item performance between male and female participants. The objective was to test for measurement invariance, ensuring that the scale operated equivalently across genders. A ~~a~~ conservative cut-off DIF contrast value of ≥0.5 logit difference was utilized [79]. The results, presented in Table 6, confirmed that all 36 items demonstrated invariant functioning. The DIF contrast values for every item fell below the critical threshold, indicating that both male and female students responded to the items in a statistically similar manner. Therefore, it was concluded that the FLLSEQEAW was free of gender bias.

Table 6. DIF across gender

Item	Male Students		Female Students		DIF contrast
	DIF measure	DIF S.E.	DIF measure	DIF S.E.	
Q1	0.75	0.06	0.63	0.07	0.12
Q2	0.84	0.06	0.65	0.08	0.19
Q3	0.13	0.06	0.11	0.08	0.02
Q4	0.30	0.07	0.15	0.08	0.15
Q5	−0.47	0.07	−0.55	0.08	0.07
Q6	0.26	0.07	0.18	0.08	0.08
Q7	0.13	0.07	0.06	0.08	0.07
Q8	−0.22	0.07	−0.12	0.08	−0.10
Q9	0.14	0.07	0.23	0.08	−0.10
Q10	−0.02	0.07	−0.02	0.08	0.00
Q11	0.51	0.07	0.38	0.08	0.13
Q12	−0.61	0.07	−0.24	0.08	−0.37
Q13	−0.12	0.07	−0.03	0.08	−0.09
Q14	0.17	0.07	0.10	0.08	0.07
Q15	−0.14	0.07	−0.02	0.08	−0.12
Q16	−0.12	0.07	0.01	0.08	−0.13
Q17	−0.53	0.07	−0.23	0.08	−0.29
Q18	0.08	0.07	−0.02	0.08	−0.10
Q19	−0.18	0.07	−0.09	0.08	−0.09
Q20	0.16	0.07	0.23	0.08	−0.08
Q21	−0.30	0.07	−0.59	0.08	0.29
Q22	−0.24	0.07	−0.45	0.08	0.22
Q23	−1.00	0.07	−0.61	0.08	−0.39
Q24	−0.10	0.07	−0.37	0.08	0.27
Q25	−0.31	0.07	−0.41	0.08	0.09
Q26	−0.29	0.07	−0.13	0.08	−0.16
Q27	−0.15	0.07	−0.11	0.08	−0.04
Q28	0.04	0.07	0.11	0.08	−0.07
Q29	−0.04	0.07	−0.02	0.08	−0.02
Q30	0.04	0.07	0.13	0.08	−0.08
Q31	−0.41	0.07	−0.21	0.08	−0.20
Q32	0.15	0.07	0.15	0.08	0.00
Q33	0.07	0.07	0.14	0.08	-0.07
Q34	0.68	0.07	0.48	0.08	0.21
Q35	0.46	0.07	0.26	0.08	0.20
Q36	0.21	0.07	0.19	0.08	0.03

B. Discussion

This section provided a comprehensive discussion of the research by delineating the principal findings, examining their broader implications, addressing the study's constraints, and suggesting productive paths for subsequent investigation.

1) Findings

Through structural equation modeling, the confirmatory factor analyses investigated by Teng *et al.* [32] empirically supported a three-factor structure for writing self-efficacy. Statistical comparisons of models upheld the conception of

writing self-efficacy as a multidimensional construct consisting of three interrelated components: linguistic, self-regulatory, and performance self-efficacy with well-established reliability and validity. Moreover, the results affirmed that the three dimensions of writing self-efficacy are interrelated components of a broader, overarching self-efficacy construct.

With the guidance of Teng *et al.*'s [32] structural dimensions, the present study investigated further and offered robust psychometric evidence for the FLLSEQEAW, with the exception of Item Q1. The results reinforced FLLSEQEAW's unidimensional structure, confirming that a single construct predominantly explains response patterns. The rating scale demonstrated monotonic increases, and analyses yielded strong separation and reliability indices for both items and persons, confirming the instrument's precision in discriminating between ability levels [69].

A final validation step assessed Differential Item Functioning (DIF) across gender. The analysis revealed no significant DIF, indicating consistent response patterns between male and female respondents and supporting the scale's measurement invariance. Ultimately, the 36-item FLLSEQEAW was distinguished by its authentic relevance to the specific challenges of English argumentative writing, thereby addressing a significant gap in the assessment tools available for EFL learners.

2) Implications

A psychometrically sound instrument is essential for obtaining precise measurements. The FLLSEQEAW, validated in this study, provides educators with a reliable tool to assess Chinese college students' self-efficacy in English argumentative writing. It enables the identification of both student strengths and areas needing support, thereby informing the design of targeted instructional strategies. For students, the scale fosters self-reflection and personal development by helping them understand their own confidence levels. Furthermore, it establishes a crucial baseline for tracking the evolution of self-efficacy over time, allowing for the evaluation of teaching effectiveness and student progress. Ultimately, this reliable instrument holds significant promise for enhancing student learning, refining instructional practices, and improving educational outcomes by strengthening students' academic self-belief.

3) Limitations and future research directions

Despite its contributions, this study is subject to several limitations that point to valuable avenues for future research. Firstly, the generalizability of the findings is constrained by the convenience sampling strategy, which was restricted to undergraduates from a single key provincial university in northern China. Subsequent studies should endeavor to incorporate a more heterogeneous sample from a broader spectrum of universities and geographical regions within China to enhance the external validity of the FLLSEQEAW.

Secondly, in an effort to mitigate social desirability bias, the collection of demographic information was limited to gender. Consequently, the examination of Differential Item Functioning (DIF) was confined to this single variable. Future validation work should include a wider array of demographic factors (such as academic grade, university prestige, and geographic location) to establish the scale's invariance across more diverse subpopulations.

Thirdly, while the scale demonstrated adequate monotonicity, the observed frequencies for the extreme response categories were substantially lower than the others. This may be attributed to the homogeneity of the participant pool. To address this, future studies could investigate the performance of a simplified 5-point Likert scale and intentionally sample from universities of differing tiers (e.g., municipal versus national) to improve the distribution of responses across categories.

Finally, while the Principal Component Analysis of Residuals (PCAR) supported the scale's unidimensionality, the instrument is theorized to encompass three specific dimensions. To further probe this structural model, applying a combination of Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) on data from new samples is strongly recommended.

V. Conclusion

In summary, the newly adapted FLLSEQEAW bridges a critical gap between theory and practice. It provides researchers and educators with a valid tool to diagnose specific weaknesses in students' confidence, both in their argumentation skills and linguistic knowledge. This instrument can evaluate the impact of new teaching methods on students' self-belief and inform targeted interventions to break the cycle of low confidence and poor performance. Additionally, it opens several promising avenues for future research to deepen the current understanding. Subsequent studies could move beyond validating the instrument itself and toward exploring some dynamic factors that influence these self-efficacy dimensions, such as personal characters, gender differences, malleable environmental and instructional factors that could enhance each self-efficacy dimension.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Lei Zhang and Qian Sun conducted the research; Lei Zhang wrote the paper; Qian Sun analyzed the data; Samah Ali Mohsen Mofreh and Sultan Salem supervised the whole research and provided valuable guidance throughout the article writing process. All authors had approved the final version.

References

[1] M. Liao, and Y. Liao, "Improving Chinese college students' argumentative writing: A presentation-assimilation-discussion-exercise approach," *Frontiers in Psychology*, vol. 13, 874531, pp. 1–14, June 2022. doi: 10.3389/fpsyg.2022.874531

[2] O. Noroozi, S. K. Banihashem, H. J. Biemans, M. Smits, M. T. Vervoort, and C. L. Verbaan, "Design, implementation, and evaluation of an online supported peer feedback module to enhance students' argumentative essay quality," *Education and Information Technologies*, vol. 28, no. 10, pp. 12757–12784, 2023. doi: 10.1007/s10639-023-11683-y

[3] D. Yang and X. Zeng, "Epistemic modality and projection in argumentative writing: a corpus-driven comparison of Chinese English majors and native English speakers," *Theory and Practice in Language Studies*, vol. 11, no. 4, pp. 412–421, April 2021. doi: 10.17507/tpls.1104.11

[4] S. Latifi, O. Noroozi, and E. Talaee, "Worked example or scripting?

Fostering students' online argumentative peer feedback, essay writing and learning," *Interactive Learning Environments*, vol. 31, no. 2, pp. 655–669, 2023. doi: 10.1080/10494820.2020.1799032

[5] Y. Liunokas, "Assessing students' ability in writing argumentative essay at an Indonesian senior high school," *IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature*, vol. 8, no. 1, pp. 184–196, 2020. doi:10.24256/ideas.v8i1.1344

[6] L. Casado-Ledesma, I. Cuevas, and E. Martín, "Learning science through argumentative synthesis writing and deliberative dialogues: A comprehensive and effective methodology in secondary education," *Reading and Writing*, vol. 36, no. 4, pp. 965–996, 2023. doi:10.1007/s11145-021-10191-0

[7] E. Pohan, E. Emilia, and I. L. Damayanti, "Using task-based learning to teach argumentative writing and critical thinking in higher education," *English Review: Journal of English Education*, vol. 13, no. 2, pp. 553–562, June 2025. doi:10.25134/erjee.v13i2.11839

[8] S. Izzatullakhon, "Writing their future: How academic writing skills drive success in studies and careers for bachelor's students," *Eurasian Journal of Academic Research*, vol. 4, no. 11, pp. 110–120, November 2024. doi:10.5281/zenodo.14172590

[9] R. Yang, "An empirical study of claims and qualifiers in ESL students' argumentative writing based on Toulmin model," *Asian-Pacific Journal of Second and Foreign Language Education*, vol. 7, no. 6, pp. 1–17, 2022. doi: 10.1186/s40862-022-00133-w

[10] L. Yang, "Focus and interaction in writing conferences for EFL writers," *Sage Open*, pp. 1–13, January–March 2022. doi: 10.1177/21582440211058 2

[11] D. Liu, "Research on the assessment rubrics of English argumentative essay," *Hei Longjiang Education-Theory and Practice*, vol. 12, pp. 70–71, 2018.

[12] Y. Liu, M. Xi, and Y. Chen, "An analysis of evidence use in argumentative essays by Chinese EFL learners," *Foreign Language and Literature*, vol. 36, no. 5, pp. 141–149, 2020.

[13] Y. Zhang, "An investigation into the development of structure and evidence use in argumentative writing," *Theory and Practice in Language Studies*, vol. 8, no. 11, pp. 1441–1448, November 2018. doi: 10.17507/tpls.0811.08

[14] R. Yang, "An empirical study on the scaffolding Chinese university students' English argumentative writing based on Toulmin model," *Heliyon*, vol. 8, pp. 1–11, November 2022. doi: 10.1016/j.heliyon.2022.e12199

[15] K. Guo, Y. Zhong, D. Li, and S. K. W. Chu, "Effects of chatbot-assisted in-class debates on students' argumentation skills and task motivation," *Computers & Education*, vol. 203, 104862, pp. 1–19, June 2023. doi: 10.1016/j.compedu.2023.104862

[16] Q. T. Liu, B. W. Liu, and Y. R. Lin, "The influence of prior knowledge and collaborative online learning environment on students' argumentation in descriptive and theoretical scientific concept," *International Journal of Science Education*, vol. 41, no. 2, pp. 165–187, November 2018. doi:10.1080/09500693.2018.1545100

[17] Y. Su, K. Liu, C. Lai, and T. Jin, "The progression of collaborative argumentation among English learners: A qualitative study," *System*, vol. 98, 102471, pp. 1–25, 2021. doi: 10.1016/j.system.2021.102471

[18] Y. Hu and A. Saleem, "Insight from the association between critical thinking and English argumentative writing: catering to English learners' writing ability," *PeerJ*, vol. 11, e16435, November 2023. doi: 10.7717/peerj.16435

[19] Q. Wang and G. E. Newell, "Teaching and learning argumentative writing as critical thinking in an EFL composition classroom," *Learning, Culture and Social Interaction*, vol. 51, 100891, pp. 553–562, January 2025. doi: 10.1016/j.lcsi.2025.100891

[20] S. Latifi, O. Noroozi, and E. Talaee, "Peer feedback or peer feedforward? Enhancing students' argumentative peer learning processes and outcomes," *British Journal of Educational Technology*, vol. 52, no. 2, pp. 768–784, 2021. doi:/10.1111/bjet.13054

[21] Q. Wang, "Text memorization: An effective strategy to improve Chinese EFL learners' argumentative writing proficiency," *Frontiers in Psychology*, vol. 14, 1126194, pp. 01–11, April 2023. doi: 10.3389/fpsyg.2023.1126194

[22] A. Bandura, *Social Foundations of Thought and Action: A Social Cognitive Theory*, New Jersey: Prentice Hall, 1986.

[23] A. A. Sabti, S. Md Rashid, V. Nimehchisalem, and R. Darmi, "The Impact of writing anxiety, writing achievement motivation, and writing self-efficacy on writing performance: A correlational study of Iraqi tertiary EFL Learners," *SAGE Open*, vol. 9, no. 4, pp. 1–13, October–December 2019. doi: 10.1177/2158255019894289

[24] S. U. Anam and E. Stracke, "The role of self-efficacy beliefs in learning English as a foreign language among young Indonesians," *Tesol Journal*, e440, pp. 1–12, 2019. doi: 10.1002/tesj.440

[25] L. Mendoza, T. Lehtonen, S. Lindblom-Ylänne, and H. Hyytinen, "Exploring first-year university students' learning journals: Conceptions of second language self-concept and self-efficacy for academic writing," *System*, vol. 106, pp. 1–12, February 2022. doi: 10.1016/j.system.2022.102759

[26] S. Yun, P. Hiver, and A. H. Al-Hoorie, "Academic buoyancy: Exploring learners' everyday resilience in the language classroom," *Studies in Second Language Acquisition*, vol. 40, no. 4, pp. 805–830, December 2018. doi: 10.1017/S0272263118000037

[27] M. Pawlak, "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*, vol. 105, pp. 1–14, April 2022. doi: 10.1016/j.system.2022.102735

[28] M. M. Gebauer, N. McElvany, W. Bos, O. Köller, and C. Schöber, "Determinants of academic self-efficacy in different socialization contexts: investigating the relationship between students' academic self-efficacy and its sources in different contexts," *Social Psychology of Education*, vol 23, no. 2, pp. 339–358, December 2019. doi: 10.1007/s11218-019-09535-0

[29] Ö. Özer and D. İ. Akçayoğlu, "Examining the roles of self-efficacy beliefs, self-regulated learning and foreign language anxiety in the academic achievement of tertiary EFL learners," *Participatory Educational Research*, vol. 8, no. 2, pp. 357–372, April 2021. doi: 10.17275/per.21.43.8.2

[30] F. Pajares and G. Valiante, "Gender differences in writing motivation and achievement of middle school students: A function of gender orientation?" *Contemporary Educational Psychology*, vol. 26, pp. 366–381, July 2001. doi: 10.1006/ceps.2000.1069

[31] J. Kormos, "The role of individual differences in L2 writing," *Journal of Second Language Writing*, vol. 21, no. 4, pp. 390–403, 2012. doi: 10.1016/j.jslw.2012.09.003

[32] L. S. Teng, P. P. Sun, and L. Xu, "Conceptualizing writing self-efficacy in English as a foreign language contexts: Scale validation through structural equation modeling," *TESOL Quarterly*, vol. 52, no. 4, pp. 911–942, December 2018. doi: 10.1002/tesq.432

[33] T. Sun, C. Wang, R. G. Lambert, and L. Liu, "Relationship between second language English writing self-efficacy and achievement: A meta-regression analysis," *Journal of Second Language Writing*, vol. 53, 100817, September 2021. doi: 10.1016/j.jslw.2021.100817

[34] T. Sun and C. Wang, "College students' writing self-efficacy and writing self-regulated learning strategies in learning English as a foreign language," *System*, vol. 90, no. 102221, pp. 1–17, 2020. doi: 10.1016/j.system.2020.102221

[35] K. A. Robinson, T. Perez, A. White-Levatich, and L. Linnenbrink-Garcia, "Gender differences and roles of two science self-efficacy beliefs in predicting post-college outcomes," *The Journal of Experimental Education*, vol. 90, no. 2, pp. 344–363, 2022. doi: 10.1080/00220973.2020.1808944

[36] S. Graham, K. R. Harris, E. Fishman, J. Houston, K. Wijekumar, P. W. Lei, and A. B. Ray, "Writing skills, knowledge, motivation, and strategic behavior predict students' persuasive writing performance in the context of robust writing instruction," *The Elementary School Journal*, vol. 119, no. 3, pp. 487–510, March 2019. doi: 10.1086/701720

[37] J. Chen and L. J. Zhang, "Assessing student-writers' self-efficacy beliefs about text revision in EFL writing," *Assessing Writing*, vol. 40, pp. 27–41, April 2019. doi: 10.1016/j.asw.2019.03.002

[38] J. Zhang and L. J. Zhang, "Profiling L2 students' writing self-efficacy and its relationship with their writing strategies for self-regulated learning," *System*, vol. 122, 103253, pp. 1–15, June 2024. doi: 10.1016/j.system.2024.103253

[39] B. Shen and B. Bai, "Chinese university students' self-regulated writing strategy use and EFL writing performance: Influences of self-efficacy, gender, and major," *Applied Linguistics Review*, vol. 15, no. 1, pp. 161–188, April 2022. doi:10.1515/applirev-2020-0103

[40] J. Zhang, L. J. Zhang, and Y. Zhu, "Development and validation of a genre-based second language (L2) writing self-efficacy scale," *Frontiers in Psychology*, vol. 14, 1181196, pp. 01–13, June 2023. doi: 10.3389/fpsyg.2023.1181196

[41] T. Sun, C. Wang, and S. Y. Kim, "Psychometric properties of an English writing self-efficacy scale: Aspects of construct validity," *Reading and Writing*, vol. 35, no. 3, pp. 743–766, 2022. doi: 10.1007/s11145-021-10206-w

[42] R. Bruning, M. Dempsey, D. F. Kauffman, C. McKim, and S. Zumbrunn, "Examining dimensions of self-efficacy for writing," *Journal of Educational Psychology*, vol. 105, no. 1, pp. 25–38, 2013. doi: 10.1037/a0029692

[43] C. Wang and B. Bai, "Validating the instruments to measure ESL/EFL learners' self-efficacy beliefs and self-regulated learning strategies," *TESOL Quarterly*, vol. 51, no. 4, pp. 931–947, December 2017.

doi: 10.1002/tesq.355
[44] E. L. Usher and F. Pajares, "Self-efficacy for self-regulated learning: A validation study," *Educational and psychological measurement*, vol. 68, no. 3, pp. 443–463, May 2008. doi: 10.1177/0013164407308475
[45] R. M. Manchón, "Writing to learn the language: Issues in theory and research," in *Learning-to-Write and Writing-to-Learn in an Additional Language*, R. M. Manchón, Ed., John Benjamins Publishing Company, 2011, pp. 61–82.
[46] L. J. Zhang, "Second language writing as and for second language learning," *Journal of Second Language Writing*, vol. 22, pp. 446–447, 2013. doi: 10.1016/j.jslw.2013.08.010
[47] S. E. Golparvar and A. Khafi, "The role of L2 writing self-efficacy in integrated writing strategy use and performance," *Assessing Writing*, vol. 47, pp.1–15, January 2021. doi: 10.1016/j.asw.2020.100504
[48] B. J. Zimmerman, "From cognitive modeling to self-regulation: A social cognitive career path," *Educational Psychologist*, vol. 48, no. 3, pp. 135–147, 2013. doi: 10.1080/00461520.2013.794676
[49] P. R. Pintrich, "A conceptual framework for assessing motivation and self-regulated learning in college students," *Educational Psychology Review*, vol. 16, no. 4, pp. 385–407, December 2004. doi: 10.1007/s10648-004-0006-x
[50] I. Lee, "EFL writing in schools," in *Handbook of Second and Foreign Language Writing*, R. M. Manchón and P. K. Matsuda, Eds., Walter de Gruyter GmbH & Co KG, 2016, pp. 91–112.
[51] B. Li, "Research on correlation between English writing self-efficacy and psychological anxiety of college students," *Frontiers in psychology*, vol. 13, pp. 01–10, July 2022. doi: 10.3389/fpsyg.2022.957664
[52] Y. Zhang and H. Guo, "A study of English writing and domain-specific motivation and self-efficacy of Chinese EFL learners," *Journal of Pan-Pacific Association of Applied Linguistics*, vol. 16, no. 2, pp. 101–121, 2012.
[53] L. S. Teng, "Changes in teachers' beliefs after a professional development project for teaching writing: two Chinese cases," *Journal of Education for Teaching*, vol. 42, no. 1, pp. 106–109, 2016. doi: 10.1080/02607476.2015.1135228
[54] L. J. Woodrow, "College English writing affect: Self-efficacy and anxiety," *System*, vol. 39, no. 4, pp. 510–522, December 2011. doi: 10.1016/j.system.2011.10.017
[55] J. E. Casal and X. Lu, "'Maybe complicated is a better word': Second-language English graduate student responses to syntactic complexity in a genre-based academic writing course," *International Journal of English for Academic Purposes: Research and Practice*, pp. 95–114, 2021. doi: 10.3828/ijeap.2021.7
[56] J. Utari, B. bUlfah, and D. Surayatika, "The correlation between self confidence and speaking ability of the fourth semester students of Indo Global Mandiri University," *ULIL ALBAB: Jurnal Ilmiah Multidisiplin*, vol. 1, no. 12, pp. 4255–4261. 2022. doi: 10.56799/jim.v1i12.1063
[57] Sabarun, U. Widiati, N. Suryati, and M. N. A. Azman, "The correlation between different interventions in writing strategy and self-confidence in constructing argumentative essays at higher education level," *Asian Journal of University Education (AJUE)*, vol. 20, no. 1, pp. 221–234, March 2024. doi: 10.24191/ajue.v20i1.26029
[58] A. G. Saputra, "Investigating students' anxiety in argumentative writing," *ELP (Journal of English Language Pedagogy)*, vol. 8, no. 2, pp. 263–276, July 2023. doi: 10.36665/elp.v8i2.793
[59] C. Qin, R. Zhang, and Y. Xiao, "A questionnaire-based validation of metacognitive strategies in writing and their predictive effects on the writing performance of English as foreign language student writers," *Frontiers in Psychology*, vol. 13, pp. 01–14, December 2022. doi: 10.3389/fpsyg.2022.1071907
[60] F. Teng, "The role of metacognitive knowledge and regulation in mediating university EFL learners' writing performance," *Innovation in Language Learning and Teaching*, vol. 14, pp. 436–450, May 2019. doi: 10.1080/17501229.2019.1615493
[61] Y. V. Sasmita and L. Setyowati, "Problems faced by EFL students in learning to write," *Linguista: Jurnal Ilmiah Bahasa, Sastra, Dan Pembelajarannya*, vol. 5, no. 1, pp. 11–25, June 2021. doi: 10.25273/linguista.v5i1.9404
[62] J. D. Vleeschauwer, "An investigation of the relationships among Thai EFL learners' writing performance, self-efficacy, and anxiety," *European Journal of English Language Studies*, vol. 3, no. 1, pp. 23–36, 2023. doi: 10.12973/ejels.3.1.23
[63] Z. Dörnyei and T. Taguchi, *Questionnaires in Second Language Research: Construction, Administration, and Processing*, 2nd ed., New York: Routledge, 2010.
[64] J. Linacre, "Sample size and item calibration stability," *Rasch Mes Trans.*, vol. 7, no. 4. p. 328, 1994.
[65] B. D. Wright and M. H. Stone, *Best Test Design*, Chicago: MESA PRESS, 1979.
[66] E. V. J. Smith, "Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals," *Journal of Applied Measurement*, vol. 3, no. 2, pp. 205–231, January 2002.
[67] J. M. Linacre, "Data variance explained by Rasch measures," *Rasch Measurement Transactions*, vol. 20, no. 1, pp. 1045–1054, 2006.
[68] S. Galli, F. Chiesi, and C. Primi, "The construction of a scale to measure mathematical ability in psychology students: an application of the Rasch model," *TPM (Testing Psicometria Metodologia)*, vol. 15, no. 1, pp. 3–18, 2008.
[69] J. M. Linacre, *A User's Guide to WINSTEPS® Rasch-Model Computer Programs: Program Manual 4.4.6*, Chicago: Mesa-Press, 2019.
[70] S. E. Embretson and S. P. Reise, *Item Response Theory for Psychologists*, 1st ed., Mahwah: Lawrence Erlbaum Associates, Inc, 2000.
[71] J. M. Linacre, "Understanding Rasch measurement: optimizing rating scale category effectiveness," *Journal of Applied Measurement*, vol. 3, no. 1, pp. 85–106, February 2002.
[72] T. G. Bond and C. M. Fox, *Applying the Rasch Model Fundamental Measurement in the Human Sciences*, 3rd ed., London and New York: Routledge & T. & F. Group, 2015.
[73] J. M. Linacre, "Investigating rating scale category utility," *Journal of Outcome Measurement*, vol. 3, no. 2, pp. 103–122, 1999.
[74] A. Tennant and P. G. Conaghan, "The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper?" *Arthritis Care & Research*, vol. 57, no. 8, pp. 1358–1362, December 2007. doi: 10.1002/art.23108
[75] T. G. Bond and C. M. Fox, *Applying the RM: Fundamental Measurement in the Human Sciences*, New Jersey and London: Lawrence Erilbaum Associations, 2007.
[76] J. F. Malec, L. C. Torsher, W. F. Dunn, D. A. Wiegmann, J. J. Arnold, D. A. Brown, and V. Phatak, "The mayo high performance teamwork scale: reliability and validity for evaluating key crew resource management skills," *Simulation in Healthcare*, vol. 2, pp. 4–10, 2007. doi: 10.1097/SIH.0b013e31802b68ee
[77] M. Cantó-Cerdán, P. Cacho-Martínez, F. Lara-Lacárcel, and Á. García-Muñoz, "Rasch analysis for development and reduction of Symptom Questionnaire for Visual Dysfunctions (SQVD)," *Scientific Reports*, vol. 11, no. 14855, pp. 1–10, 2021. doi: 10.1038/s41598-021-94166-9
[78] J. F. Pallant and A. Tennant, "An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS)," *British Journal of Clinical Psychology*, vol. 46, no. 1, pp. 1–18, 2007.doi:10.1348/014466506x96931
[79] W. C. Wang, "Assessment of differential item functioning," *Journal of Applied Measurement*, vol. 9, no. 4, pp. 387–408, 2008.

Copyright © 2026 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).