Malaysian Journal of Social Sciences and Humanities (MJSSH) (e-ISSN : 2504-8562)
2025, Volume 10, Issue 6, e003430
DOI: https://doi.org/10.47405/mjssh.v10i6.3430

RESEARCH ARTICLE

IELTS Writing Task 2 Challenges Inventory: Reliability and Validity Assessments

Chang Bao Suo[1], Boon Yih Mah[2*], Nur Ilianis Adnan[3], Hui Zanne Seng[4]

[1]Academy of Language Studies, Universiti Teknologi MARA, Cawangan Pulau Pinang, 10350 Permatang Pauh, Pulau Pinang, Malaysia.
Email: baoge003@163.com
[2]Academy of Language Studies, Universiti Teknologi MARA, Cawangan Pulau Pinang, 10350 Permatang Pauh, Pulau Pinang, Malaysia.
Email: mahboonyih@gmail.com
[3]Academy of Language Studies, Universiti Teknologi MARA, Cawangan Pulau Pinang, 10350 Permatang Pauh, Pulau Pinang, Malaysia.
Email: nurilianis@uitm.edu.my
[4]Academy of Language Studies, Universiti Teknologi MARA, Cawangan Pulau Pinang, 10350 Permatang Pauh, Pulau Pinang, Malaysia.
Email: huizanne@uitm.edu.my

CORRESPONDING AUTHOR (*):
Boon Yih Mah
(mahboonyih@gmail.com)

KEYWORDS:
EFL
Challenges
IELTS Writing Task 2
Reliability
Validity

CITATION:
Suo, C. B., Mah, B. Y., Nur Ilianis Adnan, & Seng, H. Z. (2025). IELTS Writing Task 2 Challenges Inventory: Reliability and Validity Assessments. *Malaysian Journal of Social Sciences and Humanities (MJSSH), 10*(6), e003430. https://doi.org/10.47405/mjssh.v10i6.3430

ABSTRACT

English as a Foreign Language (EFL) learners, especially Chinese undergraduates preparing for IELTS, face notable challenges in Writing Task 2, which demands advanced critical thinking and alignment with IELTS band descriptors (Task Response, Coherence/Cohesion, Lexical Resources, Grammatical Range/Accuracy), and this study aimed to validate the IELTS Writing Task 2 Challenges Inventory (IWTCI) by assessing its reliability and validity via a quantitative survey of 121 Chinese EFL students with IELTS training/testing experience who completed a 20-item Likert-scale questionnaire mapped to the four criteria. Psychometric analyses showed strong internal consistency (Cronbach's Alpha: 0.815–0.890 for subscales, 0.890 overall) and suitable factorability (KMO ≥ 0.734, Bartlett's $p < 0.001$), with descriptive statistics revealing lower mean scores in Lexical Resources ($M = 1.34–1.95$) and Coherence/Cohesion ($M = 1.58–2.12$) as primary challenges; the IWTCI was confirmed as a reliable and valid tool for identifying specific writing difficulties, highlighting the need for targeted instruction in lexical diversity and cohesive device use, and future research should expand samples and integrate qualitative methods to support evidence-based strategies for improving EFL learners' IELTS performance.

Contribution/Originality: The paper's primary contribution is that the findings provide robust evidence that this questionnaire is a valid and reliable tool for identifying key challenges in IELTS Writing Task 2 based on the four scoring criteria. It offers valuable guidance for improving instructional approaches and supporting learner development.

1. Introduction

English as a Foreign Language (EFL) is widely accepted and integrated into various sectors such as education, medicine, manufacturing, information technology, and trade. Furthermore, English is commonly used as a medium of instruction in higher education courses (Alhassan, 2021), including those related to the International English Language Testing System (IELTS). IELTS, which was founded by the University of Cambridge Testing Centre, the British Council, and IDP Education Australia in 1989 (Read, 2022; Peltekov, 2021), is designed to assess the English proficiency of individuals who do not speak English as their first language but wish to study, work or migrate to English-speaking countries (Kasimova & Ismonova, 2022). There has been a rise in the number of IELTS takers and training institutions in China due to the increasing demand for Chinese EFL IELTS students. This growing demand is due to a high tendency of Chinese students to study abroad and pursue employment opportunities within China. As a result, more and more students are applying for IELTS, and several training centres are expanding to meet this demand (Liu & Guo, 2021).

Despite the increasing attention on IELTS, teaching and learning EFL can still be challenging without access to authentic language input (Akbari, 2015). One of the most difficult components for EFL learners is the IELTS writing test, particularly Writing Task 2, which requires higher-order thinking skills. IELTS has four components: listening, reading, speaking, and writing. The writing exam consists of two tasks based on the latest test format in recent years (IELTS, n.d.). Task 1 presents a diagram, graph, chart, or table to be described in the candidate's own words, while Task 2 requires an argumentative essay on a given topic. This task tests candidates' ability to express their opinions and ideas clearly and logically. Chinese EFL learners have shown poorer performance in Writing Task 2 than in other skills, and their uncertainty about how to improve in this area has become a growing concern in EFL education in China (Bingol, 2022). Although previous research has addressed challenges in IELTS writing, systematic tools are still lacking to identify and measure learners' difficulties based on the IELTS writing band descriptors. This study seeks to address this gap by evaluating the reliability and validity of a newly developed inventory designed to assess the challenges faced by Chinese EFL learners in IELTS Writing Task 2.

IELTS writing proves to be one of the most challenging components for EFL candidates (Bingol, 2022). Compared to the listening, reading, and speaking sections, writing scores are relatively lower in countries such as China and Vietnam (Nguyen & Nguyen, 2022; Liu & Deng, 2019). Specifically, IELTS Writing Task 2 has been identified as a major obstacle (Liu & Deng, 2019). Unlike Task 1, which focuses on describing visual information, Task 2 demands more advanced writing skills, including critical thinking, argumentation, and analytical reasoning. This is further supported by findings from Tikupasang et al. (2022) and Nguyen and Nguyen (2022), who noted that candidates tend to score lower on Task 2 than Task 1. Moreover, IELTS Writing Task 2 carries double the weight of Task 1 in the overall writing score, as stated by Arzhadeeva and Kudinova (2020), making it even more crucial for candidates aiming to achieve higher band scores.

Despite its importance, there is currently a lack of tools to systematically assess EFL learners' difficulties in IELTS Writing Task 2, particularly regarding the IELTS band descriptors. Most learners and instructors cannot identify which subskills—task response, coherence and cohesion, lexical resource, and grammatical range and accuracy—contribute most to underperformance. In addition, most past studies have

employed qualitative methods to explore these difficulties, lacking a standardised instrument that could offer structured and reliable insights into learners' self-reported challenges. Therefore, there is a need for a valid and reliable instrument that measures the specific difficulties Chinese EFL learners face concerning the IELTS Writing Task 2 band descriptors. Without such a tool, it becomes difficult for educators to develop targeted interventions to improve writing outcomes. The IELTS Writing Task 2 Challenges Inventory (IWTCI) has been developed for this purpose, but its reliability and validity remain unverified. Investigating these psychometric properties is essential to determine the instrument's usefulness in identifying the most common challenges and informing effective instructional strategies for IELTS preparation.

The main objective of this study is to examine the reliability and validity of the IELTS Writing Task 2 Challenges Inventory (IWTCI), which is designed based on the IELTS Writing Task 2 key assessment and its band descriptors. This inventory aims to identify the common challenges Chinese EFL learners face when preparing for IELTS Writing Task 2. Specifically, the study seeks to determine whether the IWTCI is a reliable and valid for measuring learners' self-reported difficulties across the four key assessment criteria: Task Response, Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy. Through a detailed analysis of the inventory's psychometric properties, the study intends to support the development of more effective teaching strategies and learner support mechanisms.

To achieve the research objective, the study addresses the following research questions:

i. What is the reliability of the IWTCI in measuring challenges across all four IELTS Writing Task 2 band descriptors?
ii. What is the validity of the IWTCI in measuring challenges across all four IELTS Writing Task 2 band descriptors?

2. Literature Review

2.1. IELTS Writing Task 2

According to IELTS (n.d.), the IELTS Writing section includes two tasks. Specifically, Task 1 asks test-takers to describe a provided diagram in over 150 words within 20 minutes. In contrast, Task 2 involves writing an article to express opinions, discuss views, analyse advantages and disadvantages, or explore causes, effects, and solutions, which should be completed in approximately 40 minutes with a minimum of 250 words. This task evaluates the test-taker's ability to respond to an issue across various academic writing topics. As stated by Bingol (2022) and Liu and Deng (2019), Task 2 is regarded as more difficult than Task 1 and needs improvement. Nguyen and Nguyen (2022) indicate that the IELTS Writing test consists of two tasks. Both the General Training and Academic IELTS writing tests feature a topic-based question, requiring test-takers to address the topic and formally present their relevant thoughts and ideas on the subject (Hyland, 2019). It is important to note that Writing Task 2 is considered the most challenging. Unlike Writing Task 1, Academic Writing Task 2 presents the biggest challenge for IELTS learners worldwide (Nguyen & Nguyen, 2022).

2.2. IELTS Writing Task 2's Scoring Criteria/Rubrics

The assessment criteria for Task 2 include four components: Task Response (TR), Coherence and Cohesion (CC), Lexical Resources (LR), and Grammatical Range and

Accuracy (GRA). These four elements span nine bands (Suo, 2019; IELTS, n.d.). As Pearson (2018) noted, the criteria from Band 0 to Band 9 become more challenging as the band level increases, reflecting that IELTS test-takers' English proficiency improves, and their errors decrease. Considering these four criteria components for IELTS Writing Band Scores, the results of IELTS writing tasks, such as those of Task 2, can assess and reflect test-takers' academic writing skills and abilities (Ariani & Arham, 2020). Furthermore, as illustrated in Figure 1 and Figure 2, the content of the IELTS Writing Key (Task 2) assessment criteria and IELTS Writing (Task 2) Band Descriptors encompasses Task Response (TR), Coherence and Cohesion (CC), Lexical Resources (LR), and Grammatical Range and Accuracy (GRA), with each of these four factors covering nine bands. The difficulty of the criteria increases from Band 0 to Band 9 for students aiming for high scores, which reflects test-takers' English proficiency levels and the frequency of their linguistic issues.

Figure 1: The IELTS Writing Key (Task 2) Assessment Criteria

The Content of IELTS Writing Key(Task 2) Assessment Criteria			
TR (Task Response)	**CC (Coherence and Cohesion)**	**LR (Lexical Resources)**	**GRA (Grammatical Range &Accuracy)**
▪ how fully the IELTS task-2 taker makes a response to the task 2. ▪ how adequately the main thoughts are supported and extended. ▪ how relevant the IELTS task-2 taker thoughts are to the task 2. ▪ how clearly the IELTS task-2 taker opens the discourse, builds up their position and make a conclusion. ▪ how appropriate the formula of making a response is to the task 2.	▪ the coherence of responses through the logical framework of information and/or thoughts (the logical movement of argument). ▪ the proper utilisation of paragraphing for topic display and organisation. ▪ the logical sequences of thoughts and/or information inner and inter sections. ▪ the flexible utilisation of substitution and references (e.g. pronouns, definite articles). ▪ the proper utilisation of discourse markers to clearly signal the steps in a response, e.g. [To begin with| In conclusion], and to mark the ties between thoughts and/or information, e.g. [as a result | so also].	▪ the range of common words utilised (e.g. the utilisation of synonymous words to prevent repetitiveness). ▪ the appropriacy & adequency of the words (e.g. topic-specific things, indicators of candidates' attitude). ▪ the accuracy of word selection and expression. ▪ the utilisation and control of collocations, idioms, sophisticated phrases. ▪ the density & communicative impact of mistakes in spelling. ▪ the density and communicative impact of mistakes in the form of word.	▪the range & appropriacy of structures utilized in an offered response (e.g. sample, compound & complex sentences). ▪ the accuracy of simple, compound and complex sentences. ▪ the density & communicative impact of linguistic blunders. ▪ the precise & suitable utilisation of punctuation.

Source: IELTS (n.d.)

Figure 2: IELTS Writing (Task 2) Band Descriptors

IELTS **WRITING TASK 2: Band Descriptors (public version)**

Band	Task response	Coherence and cohesion	Lexical resource	Grammatical range and accuracy
9	• fully addresses all parts of the task • presents a fully developed position in answer to the question with relevant, fully extended and well supported ideas	• uses cohesion in such a way that it attracts no attention • skilfully manages paragraphing	• uses a wide range of vocabulary with very natural and sophisticated control of lexical features; rare minor errors occur only as 'slips'	• uses a wide range of structures with full flexibility and accuracy; rare minor errors occur only as 'slips'
8	• sufficiently addresses all parts of the task • presents a well-developed response to the question with relevant, extended and supported ideas	• sequences information and ideas logically • manages all aspects of cohesion well • uses paragraphing sufficiently and appropriately	• uses a wide range of vocabulary fluently and flexibly to convey precise meanings • skilfully uses uncommon lexical items but there may be occasional inaccuracies in word choice and collocation • produces rare errors in spelling and/or word formation	• uses a wide range of structures • the majority of sentences are error-free • makes only very occasional errors or inappropriacies
7	• addresses all parts of the task • presents a clear position throughout the response • presents, extends and supports main ideas, but there may be a tendency to over-generalise and/or supporting ideas may lack focus	• logically organises information and ideas; there is clear progression throughout • uses a range of cohesive devices appropriately although there may be some under-/over-use • presents a clear central topic within each paragraph	• uses a sufficient range of vocabulary to allow some flexibility and precision • uses less common lexical items with some awareness of style and collocation • may produce occasional errors in word choice, spelling and/or word formation	• uses a variety of complex structures • produces frequent error-free sentences • has good control of grammar and punctuation but may make a few errors
6	• addresses all parts of the task although some parts may be more fully covered than others • presents a relevant position although the conclusions may become unclear or repetitive • presents relevant main ideas but some may be inadequately developed/unclear	• arranges information and ideas coherently and there is a clear overall progression • uses cohesive devices effectively, but cohesion within and/or between sentences may be faulty or mechanical • may not always use referencing clearly or appropriately • uses paragraphing, but not always logically	• uses an adequate range of vocabulary for the task • attempts to use less common vocabulary but with some inaccuracy • makes some errors in spelling and/or word formation, but they do not impede communication	• uses a mix of simple and complex sentence forms • makes some errors in grammar and punctuation but they rarely reduce communication
5	• addresses the task only partially; the format may be inappropriate in places • expresses a position but the development is not always clear and there may be no conclusions drawn • presents some main ideas but these are limited and not sufficiently developed; there may be irrelevant detail	• presents information with some organisation but there may be a lack of overall progression • makes inadequate, inaccurate or over-use of cohesive devices • may be repetitive because of lack of referencing and substitution • may not write in paragraphs, or paragraphing may be inadequate	• uses a limited range of vocabulary, but this is minimally adequate for the task • may make noticeable errors in spelling and/or word formation that may cause some difficulty for the reader	• uses only a limited range of structures • attempts complex sentences but these tend to be less accurate than simple sentences • may make frequent grammatical errors and punctuation may be faulty; errors can cause some difficulty for the reader
4	• responds to the task only in a minimal way or the answer is tangential; the format may be inappropriate • presents a position but this is unclear • presents some main ideas but these are difficult to identify and may be repetitive, irrelevant or not well supported	• presents information and ideas but these are not arranged coherently and there is no clear progression in the response • uses some basic cohesive devices but these may be inaccurate or repetitive • may not write in paragraphs or their use may be confusing	• uses only basic vocabulary which may be used repetitively or which may be inappropriate for the task • has limited control of word formation and/or spelling; errors may cause strain for the reader	• uses only a very limited range of structures with only rare use of subordinate clauses • some structures are accurate but errors predominate, and punctuation is often faulty
3	• does not adequately address any part of the task • does not express a clear position • presents few ideas, which are largely undeveloped or irrelevant	• does not organise ideas logically • may use a very limited range of cohesive devices, and those used may not indicate a logical relationship between ideas	• uses only a very limited range of words and expressions with very limited control of word formation and/or spelling • errors may severely distort the message	• attempts sentence forms but errors in grammar and punctuation predominate and distort the meaning
2	• barely responds to the task • does not express a position • may attempt to present one or two ideas but there is no development	• has very little control of organisational features	• uses an extremely limited range of vocabulary; essentially no control of word formation and/or spelling	• cannot use sentence forms except in memorised phrases
1	• answer is completely unrelated to the task	• fails to communicate any message	• can only use a few isolated words	• cannot use sentence forms at all
0	• does not attend • does not attempt the task in any way • writes a totally memorised response			

Source: IELTS Excellence (2025)

2.3. IELTS Writing Task 2's Challenges

As indicated by Bagheri and Riasati (2016), the writing challenges in each component of Task 2 can be identified based on Figure 1 and Figure 2. First, in terms of Task Response, some students lack focus, resulting in unbalanced coverage of issues. Some aspects may be covered more comprehensively, while others are insufficiently addressed. Second, regarding Cohesion and Coherence, there is evidence of overusing or underusing references and substitutions. Cohesive devices are sometimes misused, leading to flawed or mechanical connections. In addition, problems such as the absence of clear topic sentences and logical paragraphing have been noted.

Third, concerning Lexical Resources, students cannot use natural collocations and idiomatic expressions. Spelling and word-formation errors, as well as inappropriate and inaccurate vocabulary usage, are observed. Fourth, for Grammatical Range and Accuracy: errors in grammar and punctuation exist, along with the limited or incorrect use of complex sentences. Moreover, more existing literature is still needed to illustrate the challenges in IELTS Writing Task 2 performance, especially in proficiency.

Task Response is the cornerstone of a high-scoring IELTS Writing Task 2 essay. Panahi and Mohammaditabar (2015) conducted in-depth research, revealing that it accounts for the largest variation in total scores. Their findings demonstrate the importance of this aspect. Liang (2024) further explored the relationship between task response and scoring bands. High-band scores (7 and above) are associated with well-structured and on-topic responses. On the contrary, off-topic or illogical answers affect lower scores (5 and below). Despite its importance, Wicaksono et al. (2023) found that students often struggle with task response and completion. They frequently fail to meet the requirements and organise information effectively.

Regarding Coherence and Cohesion, following the importance of task response, coherence and cohesion are equally essential for a successful IELTS Writing Task 2 essay. Arzhadeeva and Kudinova (2020) and Wicaksono et al. (2023) defined coherence as the logical flow of ideas, while cohesion involves using devices to link sentences and paragraphs. Awwalia and Suhardi (2020) and Al-Dulaimi and Al-Nuaimi (2022) further elaborated on the role of grammatical and lexical connections in preventing repetition and enriching the writing texture. As Marashi (2021) emphasised, these elements account for 25% of the Task 2 score, significantly influencing candidates' overall performance and band scores in writing and speaking. Javahery (2025) compared essays across different scoring bands and found that Band 7 essays exhibit stronger coherence and more diverse cohesive devices than Band 6 essays. Sun et al. (2024) explored the impact of theme types and thematic progression on coherence and cohesion. They found that higher-level writing features fewer dominant topical and more textual themes.

In Lexical Resources, while coherence and cohesion focus on the structure and flow of an essay, lexical resources are equally crucial for effectively conveying ideas. Lexical resources, including vocabulary range and accuracy, are crucial in IELTS Writing Task 2. Marashi (2021) and Estaji and Hashemi (2022) significantly influenced overall scores. Key components include lexical diversity, academic vocabulary usage, and content word selection. Truong (2024) found that essays rich in diverse nouns and adjectives score better. Topic familiarity, such as prompts related to urban living and health, can also enhance lexical diversity. However, as reported by Truong (2024) and Panahi and

Mohammaditabar (2015), candidates often struggle with the spelling and correct application of academic vocabulary, which may potentially lead to score reductions.

Grammatical Range and Accuracy play a crucial role in IELTS Writing Task 2. These elements in Grammatical Range and Accuracy are fundamental for ensuring that ideas are conveyed clearly and accurately. Panahi and Mohammaditabar (2015) reported that grammatical range and accuracy contribute 22.6% to the total score. Pham (2020) pointed out that a diverse grammatical range is essential for clearly expressing complex ideas in academic writing. However, candidates often face challenges in this area. Panahi and Mohammaditabar (2015) noted that common punctuation and spelling errors reduce grammatical accuracy. Wicaksono et al. (2023) reported that students find grammar and vocabulary in IELTS writing difficult, significantly affecting their performance.

2.4. Construct Validity

Validity is a central concern in developing and evaluating survey instruments, ensuring that the instrument measures the constructs it purports to measure (Beaumont et al., 2016). Validity is often established through theoretical grounding and empirical testing in educational and social research. Theoretically, an instrument must align with established models and definitions of the constructs it aims to capture (Abu Sabra, 2023). Empirically, factor analysis is one of the most widely used techniques to assess a questionnaire's underlying structure and verify whether the observed data correspond to the theorized dimensions. Specifically, exploratory and confirmatory factor analyses help researchers determine the number of latent constructs, the interrelationships among items, and whether each item loads appropriately onto its intended factor.

In factor analytic procedures, two preliminary tests are typically conducted to assess the dataset's suitability for factor extraction: the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity (Zhang et al., 2010). The KMO statistic gauges the proportion of shared variance among variables, indicating how well the items can be grouped into underlying constructs (Madsen & Tabanera, 2003). Values above 0.7 suggest that factor analysis is appropriate, while values above 0.8 or 0.9 typically indicate strong to excellent suitability (Pallant, 2016). On the other hand, Bartlett's sphericity test evaluates whether correlations within the dataset deviate significantly from zero. A statistically significant result ($p < .05$) implies that sufficient correlations exist among items, validating further factor analytic exploration (Madsen & Tabanera, 2003).

2.5. Reliability Measurement

Reliability testing is a fundamental step in ensuring that a questionnaire or survey instrument consistently measures what it is intended to measure. In other words, a reliable questionnaire produces stable and consistent results across different contexts and populations. Before interpreting data from a questionnaire, researchers must analyse the instrument's reliability to confirm that the observed responses accurately reflect the constructs under investigation rather than being artefacts of measurement error.

When discussing reliability, it is essential to distinguish it from validity. While validity pertains to whether or not a questionnaire measures what it claims to measure (i.e., correctness and relevance of the content), reliability is concerned explicitly with the consistency and stability of the measurement. A questionnaire can, in theory, be reliable but not valid; however, for robust and meaningful research, an instrument should ideally

be both valid and reliable. Reliability thus serves as a foundational checkpoint before more in-depth statistical analyses are carried out.

One of the most common statistical tools for assessing internal consistency reliability is Cronbach's Alpha. This coefficient provides an overall measure of how closely related items are within a single construct or dimension. The questionnaire items should exhibit strong intercorrelations if they measure the same underlying concept. Cronbach's Alpha ranges from 0 to 1, with higher values indicating greater internal consistency among the items:

i. Above 0.9: Excellent or very high reliability.
ii. 0.8 to 0.9: Good level of reliability.
iii. 0.7 to 0.8: Acceptable level of reliability (though slightly lower, it can still be considered reasonable in many fields).
iv. 0.6 to 0.7: Questionable or borderline reliability.
v. Below 0.6: Low reliability, indicating that the items may not measure the same concept effectively.

A Cronbach's Alpha value above 0.8 is generally considered strong in many fields. Some researchers also accept values above 0.7 as acceptable, depending on the study's purpose and the constructs' complexity. Alternatively, it is often stated that if the coefficient falls below 0.5, the instrument's internal consistency is questionable, and the questionnaire's results may not be trustworthy for further analysis.

3. Research Methodology

3.1. Research Design

In response to the identified research gap, the study aims to investigate the reliability and validity of a developed inventory - IELTS Writing Task 2 Challenges Inventory (IWTCI) to identify the challenges Chinese EFL undergraduates face in IELTS Writing Task 2. First and foremost, a quantitative survey using the IELTS Writing Task 2 Challenges Inventory (IWTCI) was conducted to identify the challenges in IELTS writing task 2 among 121 target participants who are Chinese EFL undergraduates in a public university in Beijing, China. According to Kirakowski (2000), a survey based on a questionnaire is defined as a method for eliciting, recording, and collecting information. Additionally, Brehob (2001) defines a questionnaire as a structured form that people fill out to obtain demographic information and the views and interests of those questioned. As a research tool, a questionnaire is also used to measure the validity of metacognitive strategies in writing and their impacts on EFL writing performance (Qin et al., 2022).

3.2. Participants and Sampling

Based on the criterion-sampling method, this survey focuses on 121 Chinese undergraduates from one of Beijing's public universities, who previously took IELTS training courses and/ or once participated in a real IELTS test. The participants must answer the questions in the off-line questionnaires provided to them. In this research, the criterion sampling method is used as one of the purposeful sampling methods involving those who have been trained in the IELTS program or took the IELTS exam formerly, rather than randomly selecting all Chinese EFL undergraduate students. The university is identified as the research site, while the students have been chosen as the research population for several reasons. Firstly, the university's international cooperation

programs include IELTS training in their curriculum. Secondly, the students studying at the university are diligent and have good learning readiness, profiles, and language proficiency, as they have to score highly to be accepted into this prestigious university. It has many Chinese EFL undergraduates who are learning or have completed their IELTS learning and have even taken the IELTS exam. Thus, 121 participants were selected based on a criterion sampling method, consisting of males and females of different ages and genders.

3.3. Instruments

The research used 121 target participants to measure the challenges faced in the IELTS writing task 2, following the main challenges found in previous literature. A Likert scale, which contains five segments, tests the extent to which participants strongly agree, agree, neutral, disagree, and strongly disagree. Based on previous literature and official IELTS writing criteria, the questionnaire content captures the challenges of academic IELTS Writing 2 among Chinese undergraduates. As shown in Table 1, the questionnaire consists of two sections. The first section consists of three items on a personal profile. The second section contains four subsections of common challenges related to the four writing criteria in IELTS – Task Response (TR), Coherence and Cohesion (CC), Lexical Resources (LR) and Grammatical Range and Accuracy (GRA). The first subsection (TR) comprises five items. The second subsection (CC) includes five items. The third subsection (LR) consists of six items. The fourth subsection (GRA) includes four items.

Table 1: Distribution of the Items in the Questionnaire

Section	Variable	Writing Criteria	Items (Per Criteria)	Total Items (Per Section)
1	Personal Profile			3
2	Common Challenges	Task Response (TR)	5	20
		Coherence and Cohesion (CC)	5	
		Lexical Resources (LR)	6	
		Grammatical Range and Accuracy (GRA)	4	
Total				23

3.4. Data Analysis

The data analysis process was carried out using SPSS version 21.0 to ensure statistical rigour and reliability of the results. Initially, the collected data were screened, cleaned, and standardised to ensure uniformity in the dataset and manage any inconsistencies. The reliability of the instrument was assessed through Cronbach's Alpha, which measured the internal consistency across different dimensions of the questionnaire. To evaluate construct validity, factor analysis was employed, preceded by the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy and Bartlett's Test of Sphericity, which confirmed the dataset's suitability for such analysis. Additionally, descriptive statistics were generated to capture the mean, standard deviation, and response patterns for each questionnaire item. These analyses collectively verified that the questionnaire was both

reliable and valid for measuring the intended constructs regarding IELTS Writing Task 2 challenges.

4. Results

4.1. Reliability Measurement

The data were first screened and cleaned based on the questionnaire design. Next, they were standardised using SPSS 21.0 to ensure comparability across variables, which is particularly useful for handling differences in scoring methods or response scales. Cronbach's Alpha was then computed for each dimension (or subscale) and for the entire questionnaire, with the results shown in Table 2.

Table 2: Alpha Analysis

Dimension	Cronbach's Alpha	Cronbach's Alpha (Standardised Items)	Number of Items
Task Response	0.826	0.83	5
Cohesion and Coherence	0.866	0.871	5
Lexical Resources	0.872	0.869	6
Grammatical Range and Accuracy	0.815	0.819	4
Overall	0.890	0.913	20

As displayed, Cronbach's Alpha values for each dimension—including Task Response, Cohesion and Coherence, Lexical Resources, Grammatical Range and Accuracy, and Challenge-Based Suggestive Solutions—are comfortably above 0.8. This indicates that the items within each dimension are highly interrelated and consistently measure the same underlying concept. Moreover, the overall reliability score for the questionnaire is 0.890, and when standardised items are used, the coefficient rises slightly to 0.913. Both values are well above the threshold of 0.8, signifying excellent internal consistency and suggesting that, across the entire scale, respondents' answers align coherently.

The implications of these results are considerable. High internal consistency suggests that participants' responses were not arbitrary or unduly influenced by chance. Instead, the items collectively capture a stable trait or viewpoint. Thus, researchers and practitioners can be more confident that this questionnaire's findings and subsequent interpretations rest on a solid measurement foundation in applied contexts, such as educational assessments, language proficiency evaluations, or usability tests.

Despite the positive reliability outcomes, it is prudent to note that reliability alone cannot guarantee a measurement instrument's quality. A questionnaire could still lack construct validity (i.e., it may consistently measure the wrong or incomplete construct). Therefore, additional validation procedures, such as factor analysis to confirm the dimensional structure, convergent and discriminant validity tests, and content validity assessments with expert review, should be considered. This comprehensive approach ensures that a questionnaire consistently and accurately captures the intended constructs.

The reliability test, mainly using Cronbach's Alpha, demonstrates that the results possess strong internal consistency across all dimensions. These findings underpin the credibility of the subsequent analyses using data derived from this instrument. By confirming that each subset of items consistently measures its designated aspect, the questionnaire is a

robust tool for investigating the constructs of interest. As such, researchers and practitioners can proceed with greater certainty regarding the stability and coherence of the responses, enabling them to draw valid, evidence-based conclusions from the results.

4.2. Validity Analysis

Establishing the validity of a questionnaire is crucial for ensuring that the data collected accurately reflects the phenomenon or constructs under investigation. In other words, validity indicates the extent to which the results derived from a questionnaire align with real-world conditions or theoretical expectations. In predictive research and many different scientific fields, validity is a foundational measure of quality: a higher validity score suggests that the instrument genuinely captures the underlying constructs it is designed to measure. Factor analysis is one of the most common techniques for assessing this aspect of an instrument.

According to Kaiser's criteria, the suitability of data for factor analysis can be evaluated using the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity. The KMO value indicates the degree to which items share common variance and whether the dataset will likely yield meaningful factors. Specifically:

i. KMO > 0.9: The questionnaire is highly suitable for factor analysis.
ii. 0.8 ≤ KMO ≤ 0.9: The data is appropriate for factor analysis.
iii. 0.7 ≤ KMO < 0.8: The data can be analysed using factor analysis with reasonable confidence.
iv. 0.6 ≤ KMO < 0.7: The data is only marginally adequate for factor analysis.
v. KMO < 0.6: The data is unsuitable for factor analysis, suggesting that the instrument may not measure a cohesive set of underlying factors.

Additionally, Bartlett's test of sphericity examines whether the correlation matrix differs significantly from the identity matrix. If the p-value is less than 0.05, the test indicates enough correlation among variables to proceed with factor analysis. A significant result further supports the appropriateness of the dataset for uncovering latent constructs.

Based on these criteria, the questionnaire was assessed for factor analysis suitability. SPSS 21.0 was used to conduct a factor analysis suitability check on the questionnaire data, focusing on KMO and Bartlett's tests. The results are presented in Table 3.

Table 3: KMO and Bartlett's Test

Dimension	KMO	Bartlett's Approx. Chi-Square	df	Sig.
Task Response	0.734	376.622	10	0
Cohesion and Coherence	0.834	312.682	10	0
Lexical Resources	0.834	403.217	15	0
Grammatical Range and Accuracy	0.754	185.325	6	0
Overall	0.865	1645.237	190	0

From these findings, each dimension exhibits a KMO value above 0.7, while the overall KMO stands at 0.865. This indicates that the data is more than adequate for factor analysis. Furthermore, Bartlett's test results are significant ($p < 0.001$), confirming sufficient interrelationships among variables. Thus, based on both KMO and Bartlett's test results,

we can conclude that the questionnaire demonstrates acceptable validity characteristics, supporting the notion that it measures the intended constructs.

Alongside factor analysis, descriptive statistics offer an initial glimpse into how respondents interact with different questionnaire dimensions (Wu & Zuo, 2008). Examining mean scores helps identify which areas participants rate most positively or negatively, while the range (minimum and maximum values) underscores the variability in responses. Standard deviations further illuminate how spread out the data are, indicating whether participants share relatively uniform perceptions or whether experiences are distributed more unevenly (Chen et al., 2025).

These descriptive results complement the earlier findings and further clarify areas of strength and concern across the dimensions. Examining Table 4, we see that:

i. Task Response (Q1–Q5) averages range roughly between 3.04 and 3.27, with standard deviations hovering between 1.40 and 1.62. The results suggest a moderate overall rating, though the data has some spread.
ii. Cohesion and Coherence (Q1–Q5) show lower means, between 1.58 and 2.12, indicating comparatively less favourable scores. However, the standard deviations (approx. 0.70–0.89) are relatively tight, signalling moderate agreement among respondents.
iii. Lexical Resources (Q1–Q6) also has lower averages (1.34–1.95) with small standard deviations, suggesting that most participants rated these items in a narrow range.
iv. Grammatical Range and Accuracy (Q1–Q4) has mean values between 3.00 and 3.30, implying a moderate to moderately positive response. Standard deviations (around 1.34–1.48) show slightly more dispersion.

Table 4: Descriptive Statistics for Each Dimension

Dimension	Item	N	Min	Max	Mean	Std. Deviation
Task Response	Q1	120	1	5	3.1	1.40467
	Q2	120	1	5	3.1583	1.58775
	Q3	120	1	5	3.2667	1.62301
	Q4	120	1	5	3.1167	1.58344
	Q5	120	1	5	3.0417	1.58986
Cohesion and Coherence	Q1	120	1	4	1.5833	0.79477
	Q2	120	1	5	1.9333	0.88625
	Q3	120	1	5	1.8833	0.77982
	Q4	120	1	3	2.1167	0.81151
	Q5	120	1	5	2.0333	0.70928
Lexical Resources	Q1	120	1	5	1.85	0.72934
	Q2	120	1	4	1.7333	0.65764
	Q3	120	1	4	1.95	0.68415
	Q4	120	1	4	1.8083	0.70169
	Q5	120	1	3	1.8	0.60252
	Q6	120	1	4	1.3417	0.57242
Grammatical Range and Accuracy	Q1	120	1	5	3	1.34101
	Q2	120	1	5	3.125	1.36931
	Q3	120	1	5	3.1083	1.48265
	Q4	120	1	5	3.3	1.39988

In the reported results, “Lexical Resources” and “Cohesion and Coherence” scored lower on average, pointing to potential areas of difficulty or perceived inadequacy that may

warrant targeted interventions or instructional support. Constructs like "Lexical Resources" and "Cohesion and Coherence" often reflect more nuanced or technical language skills, which can be influenced by participants' linguistic backgrounds, disciplinary training, and exposure to particular text genres (Wang, 2022). Lower scores in these areas might suggest that respondents have less experience or feel less competent when producing or evaluating language in contexts that demand robust vocabulary usage and coherent organization.

The findings from the reliability and validity analyses also intersect with broader discussions around test fairness and appropriateness in language-related assessments. (Ermiş & Uluçınar Sağır, 2025). When a dimension such as "Lexical Resources" exhibits lower mean scores, it could be due to genuine performance gaps or mismatches between the measured construct and participants' actual language practices. Factor analysis can help clarify whether an item set is genuinely capturing a cohesive domain (i.e., lexical variety, depth, or sophistication) or whether it overlaps with other dimensions.

The questionnaire's validity is supported by the KMO values, which all exceed 0.7, and the statistically significant Bartlett's tests, confirming the appropriateness of factor analysis. The descriptive statistics offer further insight into each dimension's performance. Specifically, Task Response stands out with relatively higher average scores within the first four dimensions, while Cohesion and Coherence and Lexical Resources show lower means, signalling potential areas for improvement. These findings underscore that the questionnaire adequately measures its intended constructs and provides meaningful data distributions for each dimension.

The convergence of robust KMO values, significant Bartlett's test results, and coherent item loadings suggests that the questionnaire is empirically sound for measuring multiple dimensions—from task response to language proficiency and solution-oriented thinking. From a pedagogical perspective, the relatively high "Task Response" scores can inform future curriculum design by highlighting where learners or respondents feel most adept. In contrast, the more moderate or lower scores in "Cohesion and Coherence" and "Lexical Resources" indicate areas for possible curricular refinement and additional support. Ultimately, combining factor analysis with descriptive statistics provides a holistic view of the instrument's structure and how respondents engage with different facets of the measured constructs, reinforcing such a questionnaire's theoretical and practical value in research and instructional settings.

5. Conclusion

This study evaluated the reliability and validity of a questionnaire designed to explore challenges in IELTS Writing Task 2. The reliability analysis, measured through Cronbach's Alpha, indicates strong internal consistency, with all subscales scoring above 0.8. The overall reliability coefficient of 0.890, rising to 0.913 with standardised items, shows that the instrument yields stable and consistent results throughout diverse samples and contexts.

Validity was tested through the KMO measure and Bartlett's test, both of which confirmed the dataset's suitability for factor analysis. These results support the instrument's ability to capture the constructs it intends to measure accurately. Among the four assessment criteria, "Lexical Resources" and "Cohesion and Coherence" received lower mean scores,

indicating areas where learners encounter more difficulties. These findings portray the need for focused pedagogical strategies to strengthen performance in these dimensions.

Together, the reliability and validity outcomes underscore the practical utility of this questionnaire in both research and educational settings. Educationally, the results offer meaningful insights into students' strengths and areas for development, which can inform instructional planning and support measures. Higher scores in "Task Response" show learners' strengths in analytical thinking and responding effectively to prompts. In comparison, lower language use and cohesion scores lead to the need for targeted enhancement in these areas.

Ethics Approval and Consent to Participate

This study prioritises strict research ethics, with the pilot test addressing the dissertation's first research question. Academic and ethical rigour are ensured via a mandatory defence of the proposal's first three chapters (Problem Statement, Introduction, Literature Review, Methodology). The 4–8-month Research Ethics Committee (REC) approval process involves documenting the study's background, literature, methodology, and instruments, followed by departmental and committee reviews. Final data collection authorisation at the research site is granted only after the REC confirms ethical compliance, issuing approval code REC/02/2025(PG/MR/110). Participants provided informed consent, confirming voluntary involvement and study understanding. All procedures adhere to REC guidelines, ensuring anonymity, confidentiality, and safeguarding rights throughout data collection and analysis.

Acknowledgement

I deeply thank the participants for their pivotal contributions, whose engagement with the IWTCI provided invaluable data to analyse EFL writing challenges. Gratitude also goes to my academic advisors for guiding the research design and ensuring psychometric rigour, particularly their insights into IELTS pedagogy and statistical validation. Thanks to the institution supporting data collection and reviewers for refining the manuscript, enhancing its clarity and relevance. Finally, I acknowledge EFL and IELTS writing scholars, whose foundational research underpinned this study. Their work provided insights into the IWTCI's reliability and validity.

Funding

No funding.

Conflict of Interest

The authors report no potential conflict of interest regarding this study in terms of the research, authorship, or publication of this article.

References

Abu Sabra, M. A. (2023). Adapted management of early warning signs of relapse questionnaire: Validity and reliability. *Archives of Psychiatric Nursing, 45*, 16–25. https://doi.org/10.1016/j.apnu.2023.04.015

Al-Dulaimi, A. S. I., & Al-Nuaimi, H. S. S. (2022). Cohesion: A fundamental element in writing English texts. *Journal of Language Studies, 5*(3), 37-45. https://doi.org/10.25130/jls.5.3.16

Akbari, Z. (2015). Current challenges in teaching/learning English for EFL learners: The case of junior high school and high school. *Procedia-Social and Behavioral Sciences, 199*, 394-401. https://doi.org/10.1016/j.sbspro.2015.07.524

Alhassan, A., Ali, N. A., & Ali, H. I. H. (2021). EFL students' challenges in English-medium business programmes: Perspectives from students and content teachers. *Cogent Education, 8*(1), 1888671. https://doi.org/10.1080/2331186X.2021.1888671

Ariani, N. & Arham, M. (2020). An analysis of EFL test takers' problems in IELTS writing task, *Tamaddun, 19*, 132-140. https://doi.org/10.33096/tamaddun.v19i2.81

Arzhadeeva, D., & Kudinova, N. (2020). Using debate to develop writing skills for IELTS Writing Task 2 among STEM students. *Journal of Language and Education, 6* (24), 30-43. https://doi.org/10.17323/jle.2020.10424

Awwalia, L., & Suhardi, S. (2020). Cohesive devices analysis: Mind your writing texture! *Eralingua: Jurnal Pendidikan Bahasa Asing dan Sastra*, *4*(2), 174–183. https://doi.org/10.26858/eralingua.v4i2.13102

Bagheri, M. S., & Riasati, M. J. (2016). EFL graduate students' IELTS writing problems and students' and teachers' beliefs and suggestions regarding writing skill improvement. *Journal of Language Teaching and Research, 7*(1), 198. http://dx.doi.org/10.17507/jltr.0701.23

Beaumont, V., Mrzljak, L., Dijkman, U., Freije, R., Heins, M., Rassoulpour, A., Tombaugh, G., Gelman, S., Bradaia, A., Steidl, E., Gleyzes, M., Heikkinen, T., Lehtimäki, K., Puoliväli, J., Kontkanen, O., Javier, R. M., Neagoe, I., Deisemann, H., Winkler, D., … Munoz-Sanjuan, I. (2016). The novel KMO inhibitor CHDI-340246 leads to a restoration of electrophysiological alterations in mouse models of Huntington's disease. *Experimental Neurology*, *282*, 99–118. https://doi.org/10.1016/j.expneurol.2016.05.005

Bingol, M. (2022). Difficulties of IELTS writing module experienced by test-takers: A case study in Tishk International University Continuous Education Center. *International Journal of Social Sciences and Educational Studies, 9*(4), 287-297. https://doi.org/10.23918/ijsses.v9i4p287

Brehob, K. (2001). *Questionnaire.* Usability glossary. https://www.usabilityfirst.com/glossary/questionnaire/

Chen, N., Zong, Y., Yang, C., Li, L., Yi, Y., Zhao, J., Zhao, X., Xie, X., Sun, X., Li, N., & Jiang, L. (2025). KMO-driven metabolic reconfiguration and its impact on immune cell infiltration in nasopharyngeal carcinoma: A new avenue for immunotherapy. *Cancer Immunology, Immunotherapy, 74*(3), 75. https://doi.org/10.1007/s00262-024-03928-7

Ermiş, M., & Uluçınar Sağır, Ş. (2025). Validity and reliability study of metacognitive listening strategies teaching self-efficacy scale for teachers. *International Journal of Assessment Tools in Education, 12*(1), 147–164. https://doi.org/10.21449/ijate.1403112

Estaji, M., & Hashemi, M. (2022). Phraseological competence in IELTS academic writing task 2: phraseological units and test-takers' perceptions and use. *Language Testing in Asia, 12*(1), 1–25. https://doi.org/10.1186/s40468-022-00180-7

Hyland, K. (2019). *Second language writing.* Cambridge university press.

IELTS. (n.d.). *IELTS Home of the English Language Test*. IELTS. https://ielts.org/organisations/ielts-for-organisations/test-types/ielts-academic-test#Writing

IELTS Excellence. (2025). *IELTS Writing Task 1 & Task 2 band descriptors.* IELTS Excellence. https://ielts-excellence.com/writing-band-descriptors

Javahery, P. (2025). *Analyzing discourse patterns in ChatGPT-generated IELTS Writing Task 2 essays: A discourse analysis approach*. Research Square. https://doi.org/10.21203/rs.3.rs-5608928/v1

Kasimova Guljakhon Mahmudovna, & Ismonova Tursunoy Zokirjonovna. (2022). The teaching methods Of English skills. *Galaxy International Interdisciplinary Research Journal, 10*(5), 178–182. https://internationaljournals.co.in/index.php/giirj/article/view/1825

Kirakowski, J. (2000). *Questionnaires in usability engineering: a list of frequently asked questions* (3rd ed.). Human Factors Research Group, Ireland.

Liang, S. (2024). Documenting the task response features of IELTS Writing Task 2, band scores 5-7 to improve teaching and learning for Chinese students who use IELTS preparation templates. *The Education Review, USA, 8*(3), 355-362. https://doi.org/10.26855/er.2024.03.003

Liu, F. L., & Deng, Z. M. (2019). Production-oriented approach to IELTS writing task 2. *Journal of Literature and Art Studies, 9*(11), 1200-1211. https://doi.org/10.17265/2159-5836/2019.11.012

Liu, J., & Guo, F. (2021). A case study on the teaching cognition of IELTS writing teachers: From the perspective of "Learning to Write" [雅思写作教师教学认知调查——“Learning to Write”视域下的案例研究]. *Overseas English Testing: Pedagogy and Research*, 3, 39. https://doi.org/10.12677/oetpr.2021.32006

Madsen, J., & Tabanera, R. (2003). Bartlett's test applied to variance component models. *Scandinavian Journal of Statistics, 30*(3), 609–616. https://doi.org/10.1111/1467-9469.00352

Marashi, S. M. (2021). Impact of teaching cohesive devices to Iranian EFL learners: Case of IELTS writing task 2 examination. *JL3T (Journal of Linguistics, Literature and Langauge Teaching), 6*(2), 115–126. https://doi.org/10.32505/JL3T.V6I2.2050

Nguyen, H. N., & Nguyen, D. K. (2022). Vietnamese learners' performance in The *IELTS Writing Task, 2*, 170-189. https://doi.org/10.54855/ijte.222111

Pallant, J. (2016). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS* (6th ed.). McGraw- Hill Education.

Panahi, R., & Mohammaditabar, M. (2015). The strengths and weaknesses of Iranian IELTS candidates in academic writing task 2. *Theory and Practice in Language Studies, 5*(5), 957. http://dx.doi.org/10.17507/tpls.0505.10

Pearson, W. S. (2018). Written corrective feedback in IELTS writing task 2: Teachers' priorities, practices, and beliefs. *The Electronic Journal for English as a Second Language, 21*(4), 314–320. https://files.eric.ed.gov/fulltext/EJ1172568.pdf

Peltekov, P. (2021). The International English Language Testing System (IELTS): A critical review. *JELTL (Journal of English Language Teaching and Linguistics), 6*(2), 395-406. https://dx.doi.org/10.21462/jeltl.v6i2.581

Pham, Q. H. P. (2022). A corpus-based approach to grammar instruction in IELTS writing classes. *RELC Journal, 53*(3), 723-730. https://doi.org/10.1177/0033688220952170

Qin, C., Zhang, R., & Xiao, Y. (2022). 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, 13,* 1071907. https://doi.org/10.3389/fpsyg.2022.1071907

Read, J. (2022). Test review: The International English Language Testing System (IELTS). *Language Testing*, *39*(4), 679-694. https://doi.org/10.1177/02655322221086211

Sun, G., Ahmad, N. K., & Nawi, N. R. C. (2024). Applying theme types and thematic progression to cohesion and coherence of IELTS Academic Writing Task 2: A qualitative descriptive study. *Forum for Linguistic Studies, 6*(6), 473–488. https://doi.org/10.30564/fls.v6i6.7367

Suo, C. (2019). *Yasi xiezuo guanfang tiku xunlian sucheng [A quick guide for the IELTS writing training with the official question bank] (2nd ed.)*. China Textile Press.

Tikupasang, O., Hardiyanti, S., & Arman, A. (2022). EFL student's difficulties toward IELTS writing task 2. *AMCA Journal of Education and Behavioral Change*, *2*(1), 28-31. https://doi.org/10.51773/ajeb.v2i1.141

Truong, T. T. L. (2024). Analyzing lexical characteristics in IELTS Writing Task 2 essays: A study of Vietnamese students from non-English majors. *International Journal of Research Publication and Reviews, 5*(12), 3233–3245. https://doi.org/10.55248/gengpi.5.1224.250102

Wang, Y. (2022). Financial crisis prediction model of listed companies based on statistics and AI. *Scientific Programming*, 2022, 1–10. https://doi.org/10.1155/2022/1118023

Wicaksono, R. R., Nurhamdah, N., & Kalsum, K. (2023). Analysis of iain parepare students' academic writing levels on IELTS preparation class. *Al-Irsyad, 2*(2), 132–146. https://doi.org/10.58917/aijes.v2i2.77

Wu, M., & Zuo, Y. (2008). Trimmed and winsorized standard deviations based on a scaled deviation. *Journal of Nonparametric Statistics, 20*(4), 319–335. https://doi.org/10.1080/10485250802036909

Zhang, L., Jiang, W., & Chen, G. (2010). A note on Bartlett's *M* Test for homogeneity of variances. *Communications in Statistics - Theory and Methods*, *39*(16), 2890–2900. https://doi.org/10.1080/03610920903147812