WHIOCE PUBLISHING PTE. LTD.
PROVIDING FIRST-CLASS SCIENTIFIC INFORMATION FOR TOP SCHOLARS

2025 Volume 3, Issue 7
Lecture Notes in Education, Arts, Management and Social Science
ISSN(Online): 2705-053X

Test and Item Analysis Based on English Major Grammar: An Exploration of Twenty Multiple-Choice Questions

Nanxuan Li, Yongting Lan*
Faculty of arts and humanities, University of Southampton, Southampton, England

**Corresponding author: yongting.lan@hotmail.com*

Abstract: Grammar constitutes a vital component of language proficiency, and mastering it is crucial for English majors. This study designed and developed a set of twenty multiple-choice grammar questions to assess undergraduate English majors. Test analysis revealed that students demonstrated generally sound grammar skills with a solid foundation. However, the reliability of the test items was found to be low. Additionally, an item analysis was conducted. Future research can build upon this test to refine it, thereby more effectively evaluating the English grammar proficiency of English majors.

Keywords: English grammar test; Test analysis; Item analysis; Undergraduate English majors

Online publication: August 26, 2025

1. Introduction

For English majors, grammatical competence may constitute a vital component of their overall language proficiency. Grammar is defined as the structural and semantic framework underlying sentences and discourse[1]. It enables learners to organize and arrange information more effectively[2]. Grammatical proficiency represents a crucial skill for Chinese students[3]. This study developed a set of grammar test questions. The test comprises 20 items designed to help students achieve the B1 language proficiency level according to Common European Framework of Reference (CEFR) for Languages. The test exclusively employs multiple-choice questions, requiring examinees to select the correct answer from provided options. Multiple-choice scoring is objective; such questions are valid when they elicit instances of execution that yield inferences useful for testing purposes[1]. This paper is structured into six sections: Introduction, Test Specifications, Administration, Test Analysis, Item Analysis, and Conclusions. These analyses aim to provide guidance for subsequent test design.

2. Test specification

2.1. Test Design and item description

The concrete details about test design and item design are shown in **Table 1**.

Table 1. Test design and item descriptions

Section	Details
Test purpose	This test is designed to evaluate learners' grammatical competence, focusing on their ability to recognize and apply correct grammatical forms in practical contexts, including: Morphosyntax (e.g., word order, tense, articles); Contextual language use (e.g., coherence, pragmatic tone); Error recognition and correction.
Test format	Type: Multiple-choice format. Number of Items: 20 questions (10 selection Items + 10 error identification Items). Duration: 30 minutes. Delivery Mode: Online Test.
Test content	Word order, coherence, prepositions, verb tenses, modal verbs. Tone and negotiation in dialogue. Relative pronouns, measure words, phrasal verbs.
Item descriptions	Multiple Choice Questions: This section contains ten sentences or dialogues. Below each sentence or dialogue are four options labeled A, B, C, and D. Select the option that best completes the sentence or answers the question. Error Identification Questions: This section contains ten sentences. Below each sentence are four options labeled A, B, C, and D. Select the option containing a grammatical error.
Item specifications	Question Type: Multiple-choice questions with four options. Question Length: Each question consists of one to two sentences. Distractors: Options that appear plausible but are incorrect, designed to test common learner errors. Suitable Level: Intermediate to upper-intermediate (CEFR levels B1–B2).

2.2. Scoring criteria

The test answers are objective questions in a multiple-choice format. Scoring is based on a pass/fail system, as each question has only one correct answer. Incorrect answers receive a score of "0," while correct answers receive a score of "1"[4]. Therefore, each correct answer is worth 1 point, with a maximum total score of 20 points.

2.3. Rubrics

We designed a rubric for assessing students' test performance and the details are shown in **Table 2**.

Table 2. The rubric for assessing students' test performance

Score Range	Category	Description
18–20	Excellent	Demonstrating near-perfect understanding of grammatical concepts, with few to no errors.
15–17	Good	Demonstrating a solid grasp of grammar, though occasional minor errors indicate there is still room for improvement.
10–14	Pass	Mastering basic grammar, but with some difficulty applying it in complex contexts.
Below10	Failed	Lacking mastery over key grammatical concepts, frequent errors, and limited ability to apply rules.

3. Administration

This project invited 29 undergraduate English majors from a university in southern China to participate in a 30-minute online test. Prior to the test, it was confirmed that they had recently studied grammar or taken grammar courses. If they felt unwell during the test, they could stop at any time. The testing platform displayed a timer, and upon time expiration, the system automatically locked the submission function. The system issued a reminder five minutes before the test ended and automatically recorded and saved all responses when time ran out. All participants completed the test within the allotted

time, and all test results were valid.

4. Test analysis

4.1. Mean, median, and mode

The average score of 14.4 indicates a solid grasp of grammatical concepts. The median score of 15 suggests that half of the students achieved 15 points or higher. The close proximity of this value to the mean indicates a relatively symmetrical distribution, with no significant skew caused by extreme values. The most frequent score was 17, indicating that a significant portion of students performed above average.

4.2. Reliability

Reliability refers to the measurement of internal consistency[5]. We employed SPSS 30.0 to examine the internal consistency of the test items. The Cronbach's alpha coefficient for this test was 0.55, below the conventional threshold for high-risk assessments. However, given the exploratory nature of this study and the limited number of items (20), this value remains a viable reference for item-level analysis. Consequently, subsequent analyses of item validity, difficulty, discrimination, and interference were conducted to identify problematic items and propose revisions.

4.3. Validity

4.3.1. Construct validity

Construct validity refers to "the vertical correspondence between a construct which is at an unobservable, conceptual level and a purported measure of it which is at an operational level"[6]. First, the test is explicitly designed to assess syntactic competence by focusing on three core domains: morphological syntax, contextual application, and error recognition. The first domain covers word order, tense, and articles; the second includes coherence, relative pronouns, and pragmatic tone; while the third area emphasizes identifying and correcting common grammatical errors. The alignment between test design and construct ensures precise measurement of grammatical knowledge. Second, construct validity is further demonstrated through diverse test tasks. The 20 multiple-choice questions are divided into two types, ensuring the test evaluates both recognition ability and error-correction ability, thereby capturing different dimensions of grammatical competence. Third, the test comprehensively covers key grammatical domains: word order (Task 14), tense (Task 16), articles (Task 15), modal verbs (Task 18), and coherence/pragmatics (Task 2). The scope of grammatical structures aligns closely with the theoretical framework of grammatical competence, effectively enhancing construct validity. This grammar test demonstrates high construct validity by precisely measuring grammatical proficiency through diverse tasks aligned with both theoretical and practical dimensions.

4.3.2. Content validity

Content validity refers to the extent to which a measurement tool "covers" the target concept[5]. First, the explicit purpose of this test is to assess learners' grammatical competence, with particular emphasis on identifying and applying correct grammatical forms in practical contexts. This objective is achieved through the following dimensions: morphological syntax (including word order, tense usage, articles, and agreement); contextual grammar (including pragmatic tone, coherence, and relative pronouns); and error recognition (focusing on common grammatical pitfalls). This comprehensive focus ensures high alignment between the test and its objectives, encompassing both theoretical knowledge and practical application. Second, the test demonstrates robust content validity, with all tasks closely aligned to the testing objectives and covering multiple grammatical dimensions, such as morphosyntax and pragmatic intonation. While comprehensively covering a broad range of grammatical areas, it provides clear and actionable feedback. The test content is divided into two sections: multiple-choice questions and error identification questions. The first section emphasizes morphosyntax (e.g., word order and adverb placement) and contextual grammar (e.g., relative pronouns and measure words), as demonstrated

in Task 1 and Task 8. The second section focuses on common error types, such as preposition usage, subject-verb agreement, and tense issues, as seen in Task 11 and Task 16.

5. Item analysis

The item analysis in this study (including item facility analysis, item discrimination analysis, and item distractor analysis) was conducted using Excel's built-in formulas rather than statistical software such as SPSS. This method yields results equivalent to traditional CTT analysis methods while facilitating item-by-item inspection and transparent presentation of findings.

5.1. Item difficulty/facility analysis

The item difficulty/facility assessment system is used to determine whether a project is simple or difficult[7]. In this study, the difficulty index p is calculated using an Excel formula, specifically the ratio of the number of correct answers to the total number of participants. The results are shown in **Table 3**.

Table 3. The difficulty distribution analysis results about items

Difficulty Distribution		Items
P-value ≥ 0.7	Easier items	3, 7, 8, 9, 11, 12, 14, 15, 16, 17, 18, 19, 20
0.4 ≤ P-value < 0.7	Medium difficulty items	1, 2, 5, 6, 13
P-value < 0.4	Difficult items	4, 10

The difficulty of the questions (as reflected by their P-values) indicates the complexity of the question design, but it is also influenced by the characteristics of the respondent group[8]. Thirteen questions (65% of the test) were relatively easy for students. These questions primarily assessed basic grammar knowledge, such as fundamental verb agreement or common sentence structures. While such questions can boost student confidence, they may not sufficiently challenge top-performing students. Another 5 questions (25% of the test) were classified as medium difficulty. These questions strike a balance between challenge and accessibility, focusing on intermediate grammar skills. Such questions are crucial for identifying the average student ability level and ensuring test balance. The remaining two questions (10% of the test) are considered high difficulty. These pose challenges for most students, involving complex grammatical rules (such as conditional clauses or advanced sentence structures) or obscure concepts. While effective at identifying advanced learners, these questions require careful review to ensure fairness. Among the easiest questions, nearly all students answered Questions 12, 14, 15, and 20 correctly, indicating solid mastery of fundamental grammar rules. Conversely, only about 28% of students answered Question 4 correctly, and 34% answered Question 5 correctly. These two questions involved less common grammar rules, effectively distinguishing between different proficiency levels.

5.2. Item discrimination analysis

The discrimination index in this study was calculated using the point-biserial correlation coefficient (rpb)[9], which measures the correlation between an item's score (0/1) and the total test score (excluding that item). The results are shown in **Table 4**.

Table 4. The item discrimination analysis results about items

Pt-Biserial Classification	Items
rpb > 0.3	1, 5, 6, 7, 9, 11, 12, 14, 16, 18, 19, 20
0 < rpb ≤ 0.3	2, 3, 4, 10, 13, 17
rpb < 0	8, 15

Matlock-Hetzel emphasized the importance of test item discrimination, noting that high-quality test items should effectively distinguish between high-scoring and low-scoring examinees, thereby ensuring the test accurately assesses examinees' abilities or knowledge levels[8]. Item discrimination (two-column correlation coefficient) measures a test item's ability to distinguish high-scoring students from low-scoring ones. Items 1, 5, 6, 7, 9, 11, 12, 14, 16, 18, 19, and 20 were classified as high-discrimination items, effectively differentiating high-scoring students from low-scoring ones. Items 2, 3, 4, 10, 13, and 17 were classified as having moderate discrimination. These items possess limited discrimination ability, meaning they may not effectively distinguish between students of different proficiency levels. They may involve simpler grammar rules or lower difficulty. Items 8 and 15 were categorized as having negative discrimination. For example, Item 8 exhibited slight negative discrimination, indicating that academically stronger students found it more difficult to understand than academically weaker students.

5.3. Item distractor analysis

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14	Item 15	Item 16	Item 17	Item 18	Item 19	Item 20
1	C	C	C	D	A	C	A	C	B	A	B	C	D	C	B	C	C	C	B	D
2	C	C	C	A	C	B	D	C	A	B	B	C	C	C	B	D	C	C	B	D
3	B	A	C	A	B	B	A	C	B	B	D	C	A	C	B	C	C	B	B	D
4	B	A	B	A	A	C	A	A	B	D	B	C	C	C	B	C	C	C	B	D
5	C	B	C	B	D	C	A	C	B	C	B	C	B	C	B	C	B	B	B	D
6	D	D	C	B	A	B	A	A	B	C	B	C	A	C	B	C	C	C	B	D
7	C	C	C	B	A	B	D	A	B	C	B	C	C	C	B	B	C	B	B	D
8	A	B	C	D	A	D	A	C	A	C	B	C	B	C	B	C	C	C	B	D
9	C	C	C	B	A	C	A	C	B	D	B	C	D	C	B	C	C	C	B	D
10	C	B	B	B	A	C	A	C	B	D	B	C	C	C	B	C	C	C	B	D
11	B	B	C	A	A	C	A	C	B	C	B	C	C	C	B	B	C	C	B	D
12	A	B	C	C	B	B	A	C	A	D	B	C	B	C	B	C	C	C	B	D
13	C	B	C	C	A	C	A	A	A	B	B	C	C	C	B	C	C	C	B	D
14	A	B	C	D	A	B	A	A	B	D	B	C	C	C	B	C	C	C	D	D
15	C	A	C	D	B	B	A	C	B	B	B	C	B	C	B	C	C	C	B	D
16	C	B	C	A	C	C	D	C	A	B	B	C	C	C	B	C	C	C	B	D
17	A	B	C	B	C	D	C	C	A	C	A	A	A	B	B	A	A	B	C	B
18	C	B	C	D	A	C	C	C	B	D	B	C	D	C	B	C	C	C	B	D
19	C	C	C	D	A	D	D	A	B	C	B	C	D	C	B	C	C	B	B	D
20	C	D	C	D	B	D	A	C	B	D	B	C	B	C	B	C	C	C	C	D
21	C	B	C	D	A	C	A	C	B	C	B	C	C	C	B	C	C	B	B	D
22	A	A	C	A	A	C	A	C	B	B	B	C	C	C	B	C	C	C	B	D
23	A	B	B	A	A	B	A	C	A	B	B	C	B	C	B	C	C	C	B	D
24	C	B	C	B	A	C	A	C	B	D	B	C	B	C	B	C	B	C	B	D
25	C	C	B	D	A	B	D	C	A	D	C	C	C	B	B	C	C	B	B	D
26	A	B	C	A	C	B	A	A	B	D	B	C	C	C	B	B	C	C	D	D
27	C	A	C	A	A	C	A	C	B	C	B	C	A	C	B	C	C	C	B	D
28	C	B	C	D	C	C	D	C	A	B	B	C	A	C	B	C	C	B	B	D
29	C	C	C	B	A	C	A	A	B	C	B	C	C	C	C	C	C	C	B	D
	C	B	C	B	A	C	A	C	B	C	B	C	C	C	B	C	C	C	B	D
A	7	5	0	9	19	0	21	8	9	1	1	1	5	0	0	1	1	0	0	0
B	3	15	4	8	4	10	0	0	20	8	26	0	7	2	28	3	2	8	25	1
C	18	7	25	2	5	15	2	20	0	10	1	28	13	27	1	25	26	21	2	0
D	1	2	0	10	1	4	6	0	0	10	1	0	4	0	0	1	0	0	2	28

Figure 1. The item distractor analysis results about items

The Distractor Efficiency (DE) of multiple-choice questions (MCQs) options is a component of psychometric analysis, through which examiners evaluate the credibility and functionality of distractors[10]. Distractors are categorized into two types: functional distractors and non-functional distractors. Functional distractors attract low-scoring students while high-scoring students avoid them. As shown in Figure 1, in item 6, option D attracted 4 students with conceptual misunderstandings, while the correct answer C dominated among high-scoring groups. Non-functional distractors are rarely or never selected, indicating a lack of plausibility. For instance, in Item 3, option A received no selections, suggesting it was either obviously incorrect or irrelevant to the question. In Question 1, 18 students selected the correct answer C, demonstrating its dominance. However, option A attracted 7 students, indicating partial validity; 3 students

chose B, suggesting its limited effectiveness. Furthermore, only 1 student selected D, necessitating revision of this option. In Question 6, 15 students selected the correct answer C, while no one chose option A, confirming its invalidity. Option D attracted 4 students, successfully testing a common misconception. Finally, in Question 15, 27 students selected the correct answer C, while only 1 chose A, suggesting this option may confuse even high-scoring students. Only 1 student selected D, indicating its limited distractor effect.

6. Conclusion

In terms of test analysis, this set of questions demonstrates strong construct validity and content validity. The mean and mode indicate that the tested group possesses strong overall learning abilities. Regarding item analysis, difficulty analysis shows that this grammar test effectively assesses students' proficiency levels, highlighting strengths while identifying areas for improvement. In terms of item discrimination analysis, this analysis highlights both the strengths and weaknesses of grammar testing for English majors in China. By revising problematic items and fully leveraging high-discrimination items, future tests can achieve higher validity and reliability. Furthermore, the analysis of distractors indicates that while most items and their distractors function effectively, there is room for improvement in matching distractors with common student errors, enhancing clarity of wording, and ensuring balanced difficulty levels. Additionally, this test has limitations in item difficulty design, such as imbalances in item difficulty. Future item design can build upon this foundation for enhancement.

Appendix: Test items

Item 11
Sentence: I am good in playing chess.
Options:
(A) I am
(B) good in
(C) playing
(D) chess

Item 12
Sentence: Neither the teacher nor the students was ready for the test.
Options:
(A) Neither the teacher
(B) nor the students
(C) was
(D) ready for the test

Item 13
Sentence: Each of the players brought their own equipment.
Options:
(A) Each of
(B) the players
(C) their
(D) own equipment

Item 14
Sentence: She gave to me the keys.
Options:
(A) She
(B) gave
(C) to me
(D) the keys

Item 15
Sentence: He bought a expensive car last week.
Options:
(A) He bought
(B) a
(C) expensive car
(D) last week

Item 16
Sentence: She has went to the store already.
Options:
(A) She
(B) has
(C) went
(D) to the store

Item 17
Sentence: It's a leather brown jacket.
Options:
(A) It's
(B) a
(C) leather brown
(D) jacket

Item 18
Sentence: You mustn't to touch that switch.
Options:
(A) You
(B) mustn't
(C) to touch
(D) that switch

Item 19
Sentence: We called up the meeting because of bad weather.
Options:
(A) We
(B) called up
(C) the meeting
(D) because of

Item 20
Sentence: The homework was finished by he.
Options:
(A) The homework
(B) was
(C) finished
(D) by he

Part 1: Multiple-Choice Items (10 Items)
There are ten sentences or dialogues in this section. Beneath each sentence or dialogue there are four options marked A, B, C and D. Choose the one that best completes the sentence or answers the question.

Item 1
A: Does she know how to fix the issue?
B: She has ____ experience with this.
(a) so small an experience
(b) a such small experience
(c) such little experience
(d) a too little experience

Item 2
A: Will they accept the proposal?
B: I believe ____.
(a) it
(b) so
(c) that
(d) not

Item 3
A: I can't believe you lost your phone again!
B:
_____
_____
A: It's fine, let's try to find it.
(a) I really don't care.
(b) I look ridiculous!
(c) I'm sorry, it was an accident.
(d) What are you talking about?

Item 4
In this experiment, they are wakened several times during the night, and asked to report what they ____.
A) had been dreaming
B) have been dreaming
C) are dreaming
D) had dreamt

Item 5
An important lecture _____ tomorrow, the professor has to stay up late into the night to prepare for it.
A. to be given
B. will be given
C. is to be given
D. given

Item 6
Above the bushes are the mountains, _____ magnificence the lake faithfully reflects on the surface.
A. whom
B. which
C. whose
D. that

Item 7
A: How was the meeting?
B: The manager spoke ____ about the changes.
(a) openly
(b) open
(c) openness
(d) more open

Item 8
A: Is this the book ____ you told me about?
B: Yes, it's the one.
(a) what
(b) whom
(c) that
(d) whose

Item 9
A: Do you have ____ money left?
B: No, I've spent it all.
(a) some
(b) any
(c) few
(d) a few

Item 10
____ right now, she would not be late for the class
A) Would she leave
B) If she leave
C) Were she to leave
D) If she had left

Part 2: Multiple-Choice Error Identification Items (10 Items)
There are ten sentences in this section. Beneath each sentence there are four options marked A, B, C and D in them. Choose the one that has grammatical errors.

Disclosure statement

The author declares no conflict of interest.

References

[1] Purpura J E, 2004, Assessing grammar (Vol. 8). Cambridge University Press.

[2] Supakorn P, Feng M, Limmun W, 2018, Strategies for Better Learning of English Grammar: Chinese vs. Thais. English Language Teaching, 11(3): 24-39.

[3] Windsor R J, 2021, The effectiveness of an online grammar study scheme for Chinese undergraduate students. Smart Learning Environments, 8(1): 3.

[4] Bachman L F, Palmer A S, 1996, Language testing in practice: Designing and developing useful language tests (Vol. 1). Oxford University Press.

[5] Price P C, Jhangiani R, Chiang I C A, 2015, Reliability and validity of measurement. In Research Methods in Psychology – 2nd Canadian Edition. eCampusOntario Press.

[6] Peter J P, 1981, Construct validity: A review of basic issues and marketing practices. Journal of marketing research, 18(2): 133-145.

[7] Muslih R A A, 2023, Examining item facility and item discrimination of Multiple-Choice Questions (MCQs) created by English teachers at a Private Junior High School in Bandung (Doctoral dissertation, UIN Sunan Gunung Djati).

[8] Matlock-Hetzel S, 1997, Basic Concepts in Item and Test Analysis. Ed.gov.

[9] Kılıç A F, Uysal I, 2022, To what extent are item discrimination values realistic? A new index for two-dimensional

structures. International Journal of Assessment Tools in Education, 9(3): 728-740.

[10] Rezigalla A A, Eleragi A M E S A, Elhussein A B, et al., 2024, Item analysis: the impact of distractor efficiency on the difficulty index and discrimination power of multiple-choice items. BMC Medical Education, 24(1): 445.