

A Research on Content Validity of CET-4 Reading Comprehension ---Based on Test Task Characteristics Framework

Xie Shuangmei¹, Charanjit Kaur Swaran Singh^{2*}, Wong Wei Lun²

¹Faculty of Languages and Communications, Sultan Idris Education University, Malaysia & Panzhihua University, China

²Faculty of Languages and Communications, Sultan Idris Education University, Malaysia

***Corresponding Author:** Charanjit Kaur Swaran Singh, Faculty of Languages and Communications, Sultan Idris Education University, Malaysia

Abstract: The importance of content validity in assessing the validity of a test is crucial. Based on corresponding teaching and testing syllabuses, this study analyzes the content validity of College English Test (CET-4) reading comprehension sub-test from the perspective of the characteristic of the input and the expected response using the revised test task characteristics framework. The research subjects are 24 passages of CET-4 original materials in 2021 and 2022. The aim is to evaluate the content validity of CET-4 in these two years. The results unveil that while the CET-4 reading comprehension in 2021 and 2022 generally adheres to the requirements of the corresponding teaching and testing syllabuses in terms of text length and new word ratio, there are still certain limitations in the variety of genres and topics covered, as well as an uneven distribution of evaluated reading skills and other aspects. These findings provide valuable insights for improving the quality of the reading sub-test task.

Keywords: Content Validity, CET-4 Reading Comprehension, Test Task Characteristics Framework

1. INTRODUCTION

Reading comprehension plays a crucial role in language learning, particularly in testing, teaching, and second language acquisition research. Exams like IELTS, TOEFL, TOEIC, TEM, and CET emphasize reading comprehension. Content validity in these tests hinges on two main factors: the selection of reading materials (input characteristics) and the design of test questions, which aim to elicit specific responses. Researchers have found that Bachman and Palmer's (1999) test task characteristics framework is effective for assessing test quality. Content validity, as emphasized by Alderson (2011), is determined by how well test content aligns with curricular standards and testing syllabi. This study aims to validate the content of CET-4 reading comprehension tests from 2021 to 2022 using a revised version of Bachman and Palmer's framework. By applying this framework, test designers and educators can evaluate test quality, identify areas for improvement, and propose solutions.

2. THE COLLEGE ENGLISH TEST (CET-4)

The College English Test (CET) in China is a high-stakes, national standardized exam administered by the Ministry of Education's Examination Centre since 1987. It evaluates college students' English proficiency through listening, reading, translation, and writing components. Held twice a year in June and December, each session features three sets of test papers. While the CET has improved English teaching and learning in Chinese colleges by providing valuable feedback, it faces criticism for its limited and sometimes inappropriate item types, sparking debates on its impact on English education in China.

3. DEFINITION OF CONTENT VALIDITY

Content validity is crucial in assessment as it determines the accuracy of a test in measuring what it is intended to measure. Scholars such as Henning (2001), Heaton (2004), and the Standards (2014) have all defined what is content validity. This study aligns with Alderson's (2011) definition, which

emphasizes that content validity is assessed by comparing the test content with its specifications. Test specifications outline what the test aims to assess and how it will do so. These specifications should clearly and precisely define the language skills and areas to be covered in the test, and should be established early in the test development process. They serve as a guide for test and item writers and are essential in ensuring the test's construct validity.

4. RESEARCH OBJECTIVES AND RESEARCH QUESTIONS

The study aims to examine the content validity of reading comprehension in CET-4 from 2021 to 2022 in relation to the Curriculum Requirement (2015) and Testing Syllabus (2016). Three research questions are posed:

- 1) How do the input characteristics of CET-4's reading comprehension sub-test in 2021 and 2022 align with the Curriculum Requirement (2015) and Testing Syllabus (2016)?
- 2) How do the expected response characteristics of CET-4's reading comprehension sub-test in 2021 and 2022 correspond with the Curriculum Requirement (2015) and Testing Syllabus (2016)?
- 3) What are the implications for future test propositions in the CET-4 reading comprehension section from 2021 to 2022?

The objectives of the study are:

- 1) To investigate whether the characteristics of input of CET-4's reading comprehension sub-test in 2021 and 2022 correspond with the Curriculum Requirement (2015) and Testing Syllabus (2016).
- 2) To discover whether the characteristics of expected response of CET-4's reading comprehension sub-test in 2021 and 2022 in line with the Curriculum Requirement and Testing Syllabus.
- 3) To explore what the implications for the future test propositions in the CET-4 reading comprehension section.

5. LITERATURE REVIEW

CET-4 is designed to evaluate English proficiency among non-English major college students in China, rather than being an international assessment. Consequently, research on CET-4 is predominantly conducted by Chinese scholars, with limited involvement from foreign researchers. A number of studies in the field of language testing have focused on evaluating the validity of English language tests, particularly within the context of test task characteristics in China. Notable researchers in this area include Yang Huizhong and Weir (1998), Hou Yanping (2004), Dong Manxia (2007), Lu Qiaoling (2008), Gu Xiangdong and Li Yaguo (2010), Chen Qingxia (2014), Peng Yingying and others (2017), Gao Weixin (2023), among others. The author presents a selection of representative and relevant studies in the following table:

Table1.*The literature review of reading comprehension task on the test task characteristics*

| No. | Author | Main findings |
|------------|----------------------------------|---|
| 1 | Hou Yanping (2004) | The characteristics of the TEM-4 reading comprehension section from 2000 to 2002 need a more acceptable content validity. |
| 2 | Dong Manxia (2007) | The reading comprehension part of the National Matriculation English Tests (NMET) still some flaws that should be taken into consideration, even with high content validity. |
| 3 | Lu Qiaoling (2008) | There is a need for future improvement in topics and genre balance, topic diversity, and assessment skill design of the original CET-4 reading comprehension part from 1997 to 2003 . |
| 4 | Gu Xiangdong and Li Yaguo (2010) | It provides some suggestions for improving CET listening comprehension test questions from the perspectives of discourse input and expected response tasks. |
| 5 | Chen Qingxia (2014) | It showed that from the characteristics of the input and expected response, the content validity of CET-4 (2010--2013) reading comprehension meet the |

| | | |
|---|---|---|
| | | Testing Syllabus (2006) and the College English Curriculum Requirements (2007). |
| 6 | Meng Fanna (2016) | She compared the reading sub-test between NMET and PETS-2 from 2011 to 2015. It provided some references to NMET's reform. |
| 7 | Peng Yingying, Gu Xiangdong & Huang Juan (2017) | In terms of discourse input and expected responses of the NMET reading sub-test in 2016, it provided recommendations to both test designers and high school English teachers to enhance the quality of the test. |
| 8 | Peng Yingying (2018) | She examined the two sets of NMET listening sub-tests in 2017 and provided valuable insights into the development of the NMET listening sub-test and English listening skills instruction in secondary schools. |
| 9 | Gao Weixin (2023) | From the analysis of reading materials from the New NMET-I, National A-roll, and National B-roll from 2021 to 2022, it showed that the content validity is not ideal and provided some suggestions on the propositions in the reading sub-test. |

The literature review discussed in Table 1 offers a thorough examination of the content validity and quality of English language tests in China, with a specific focus on Bachman & Palmer's (1999) model of test task characteristics. These studies aimed to identify shortcomings, propose enhancements, and offer insights for the improvement of language testing and teaching practices in China. The outcomes likely played a role in refining language tests and instructional methods in the country. However, recent research on the content validity of reading comprehension tests in the context of CET-4 in China is lacking. Therefore, it is crucial to conduct a validation study on CET-4 reading comprehension tests from 2021 to 2022, considering the test task characteristics from input and expected response dimensions, to provide recommendations for the improvement of CET-4 test proposition.

6. THEORETICAL FRAMEWORK

6.1. Bachman & Palmer's Model of Test Task Characteristics

By summarizing the relevant literature on test tasks, Bachman and Palmer (1999) proposed a model of test task characteristics, including five dimensions: test environment, test instructions, discourse input, expected response, and the relationship between discourse input and expected response. Specifically, the framework of test task characteristics is to be presented as follows.

| | |
|---|--|
| 1. Characteristics of the setting | |
| Physical characteristics, Participants, Time of task | |
| 2. Characteristics of the test rubric | |
| Instructions | |
| Language (native, target), Channel (aural, visual), Specification of procedures and tasks | |
| Structure | |
| Number of parts/tasks, Salience of parts/tasks, Sequence of parts/tasks, Relative importance of parts/tasks, Number of tasks/items per part | |
| Time allotment | |
| Scoring method | |
| Criteria for correctness, Procedures for scoring the response, Explicitness of criteria and procedures | |
| 3. Characteristics of the input | |
| Format | |
| Channel (aural, visual), Form (Language, non-language, both), Language (native, target, both), Length, Type (item, prompt), Degree of speededness, Vehicle ("live", "reproduced", both) | |
| Language of input | |
| Language Characteristics | |
| Organizational characteristics, Grammatical (vocabulary, syntax, phonology, graphology) | |
| Textual (cohesion, rhetorical/conversational organization) | |
| Pragmatic Characteristics | |
| Functional (ideational, manipulative, heuristic, imaginative) | |
| Sociolinguistic (dialect/variety, register, naturalness, cultural figurative language) | |
| Topical Characteristics | |
| 4. Characteristics of the expected response | |
| Format | |
| Channel (aural, visual), Form (Language, non-language, both), Language (native, target, both), Length, Type (selected, limited production, extended production), Degree of speededness | |
| Language of expected response | |
| Language Characteristics | |
| Organizational characteristics, Grammatical (vocabulary, syntax, phonology, graphology) | |
| Textual (cohesion, rhetorical/conversational organization) | |
| Pragmatic Characteristics | |
| Functional (ideational, manipulative, heuristic, imaginative) | |
| Sociolinguistic (dialect/variety, register, naturalness, cultural figurative language) | |
| Topical Characteristics | |
| 5. Relationship between input and response | |
| Reactivity (reciprocal, non-reciprocal, adaptive) | |
| Scope of relationship (broad, narrow) | |
| Directness of relationship (direct, indirect) | |

(adapted from Bachman & Palmer, 1999: 49-50)

Figure 1. Characteristics of test task

Bachman and Palmer (1999) suggest that this framework can be utilized in various ways, such as serving as a template for delineating and contrasting the characteristics of the target language use and the test task to assess their alignment. This model has been extensively utilized in developing language tests and serves as a fundamental basis for research on test content validity. Additionally, the framework of test task characteristics demonstrates practicality and applicability in evaluating the quality of test items.

6.2. Modified Framework for the Research

Therefore, in the context of a specific test with its distinct curriculum requirements and testing syllabus, adjustments must be made to certain aspects of Bachman and Palmer's framework to improve the feasibility of test assessment. Given that the CET-4 is a large-scale, standardized test administered nationwide, the characteristics of setting, test rubric, and the relationship between input and response are consistent for all test-takers, thus requiring no modifications in these areas. This study focuses solely on modifying the characteristics of input and expected response within the CET-4 reading comprehension sub-test framework.

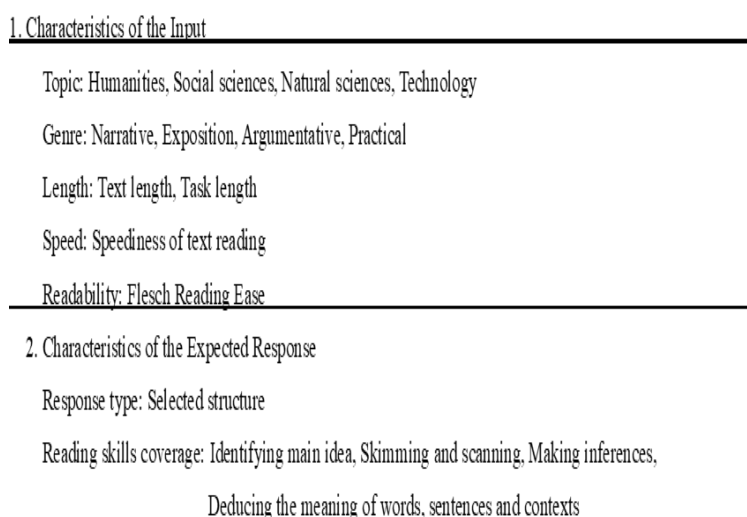


Figure2. *The modified framework for reading comprehension test*

7. METHODOLOGY

This study selected the reading comprehension components from 24 original CET-4 sub-test materials from 2021 to 2022 as research subjects. Following the Curriculum Requirements and Examination Syllabus, the aim is to investigate the content validity of reading comprehension in CET-4 by analyzing the input characteristics and expected responses. Utilizing a revised framework of test task characteristics, the study seeks to identify patterns in test paper design and offer recommendations for creating more effective and scientifically sound test papers.

8. INSTRUMENTS

In the analysis of the present study, the Microsoft Office Word 2019 and Excel 2019 in the computer are utilized undertaken a revised framework based on Bachman's Test Task Characteristics.

9. RESEARCH RESULTS

This study references the Testing Syllabus (2016 edition) and the Teaching Syllabus (2015 edition), and incorporates the revised modifications of test task characteristics suggested by Bachman and Palmer (1999) to examine the content validity of CET-4 reading subtests. The study considers two dimensions: characteristics of input and expected response. The research findings are outlined below.

Q1: How do the input characteristics of CET-4's reading comprehension sub-test in 2021 and 2022 align with the Curriculum Requirement (2015) and Testing Syllabus (2016)?

9.1. Characteristics of the Input

The characteristics of input of the revised modification framework includes topic, genre, length, readability, ratio of new words, and speed of reading materials.

9.1.1. Topic

Topic means the particular subject that an article discusses or writes about. Bachman (1990) emphasizes that the choice of topic significantly influences test takers’ performance on reading comprehension tests. Weir (1993) adds that topics should not be culturally biased or favor specific test populations. Additionally, Alderson (2000) asserts that reading comprehension tests should cover a wide range of topics to effectively assess test takers’ abilities. The Curriculum Requirements of CET-4 (2015) mandate that college students have a thorough understanding of English materials related to their major and familiar topics. Similarly, the Testing Syllabus of CET-4 (2016) highlights the importance of selecting reading materials across various topic areas, such as humanities, social science, nature science, and technology. An analysis of test papers from June 2021 to December 2022, based on the Testing Syllabus of CET-4 (2016), reveals a careful exploration of diverse topic areas.

Table2. Topics of the reading passages in CET-4 reading materials

| Years | Topics |
|-------------|--|
| 2021.06 (1) | mental health services; the decoy effect. |
| 2021.06 (2) | boredom; forest growth |
| 2021.06 (3) | electric bikes (e-bikes); the distinction between global warming and climate change |
| 2021.12 (1) | remote work on cities; curiosity |
| 2021.12 (2) | the debate of obesity; the gender disparity in the early childhood education field |
| 2021.12 (3) | cultural norms and variations in physical contact; a growth mindset of interest and its potential benefits |
| 2022.06 (1) | online classes; public debates on science |
| 2022.06 (2) | social media on employee; the distance learning and traditional classroom environment |
| 2022.06 (3) | online classes; public debates on science |
| 2022.12 (1) | healthy eating habits and lifestyles; high emotional intelligence, |
| 2022.12 (2) | the approach to improve the living conditions and well-being of seniors in nursing homes; the relationship between socioeconomic status, dietary quality, and health outcomes |
| 2022.12 (3) | the increasing global consumption of fish and the shift towards aquaculture; the importance of addressing global food instability and poverty |

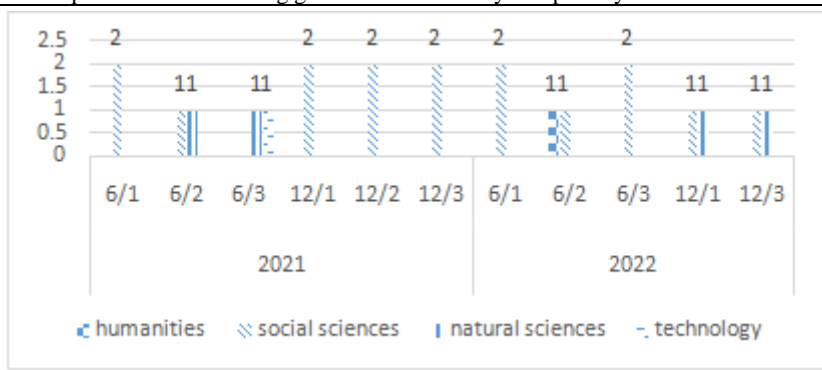


Figure3. Analysis of the topics of the texts in CET-4 reading materials

Note: CET-4 has three parallel sets of test papers are available for each test session (one test, multiple papers), 2021-6-1 refers to the first set of test paper in June, 2021.

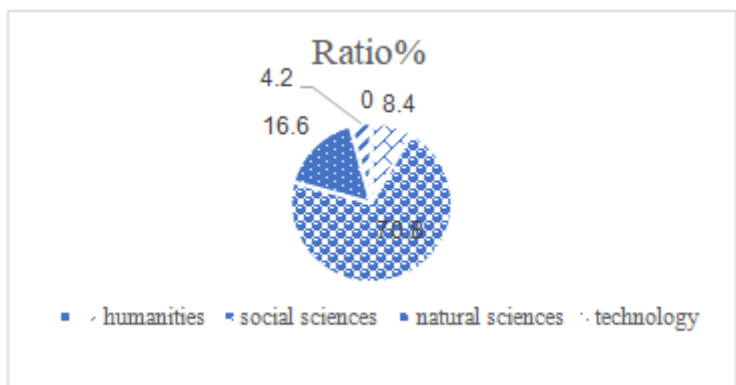


Figure4. Analysis of the topics ratio of texts in CET-4 reading materials

Analysis of Figures 3 and 4 shows that testing passages cover all four topic areas in the syllabus, with Social Science topics dominating at 70.8%, followed by Natural Science at 16.6%. These familiar topics align with Yang Huizhong and Weir’s (1998) view that topics should be accessible to test-takers. However, Humanities and Technology are underrepresented, accounting for only 8.4% and 4.2% respectively. Despite comprehensive coverage, there is a clear imbalance, particularly between Social Sciences and Technology. Most test sets focus solely on Social Sciences, highlighting the need for CET-4 reading comprehension to diversify topics and reduce the emphasis on Social Sciences to better align with syllabus and curricular standards.

9.1.2. Genre

The genre of texts in reading comprehension plays a crucial role in impacting test takers’ performance. According to Testing Syllabus (2016), texts can be classified into four main types: narrative essays, exposition articles, argumentative essays, and practical articles. It is essential for reading comprehension texts to incorporate a variety of genres in order to assess students’ skills comprehensively. Upon analyzing the selected texts, it can be observed that the frequency of these genres can be categorized as follows.

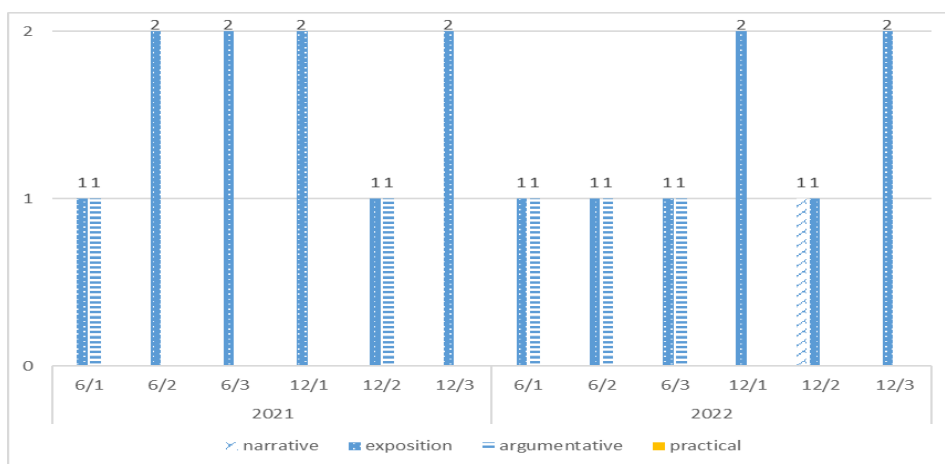


Figure5. Analysis of the genres of the texts in CET-4 reading materials

Note: CET-4 has three parallel sets of test papers are available for each test session (one test, multiple papers), 2021-6-1 refers to the first set of test paper in June, 2021.

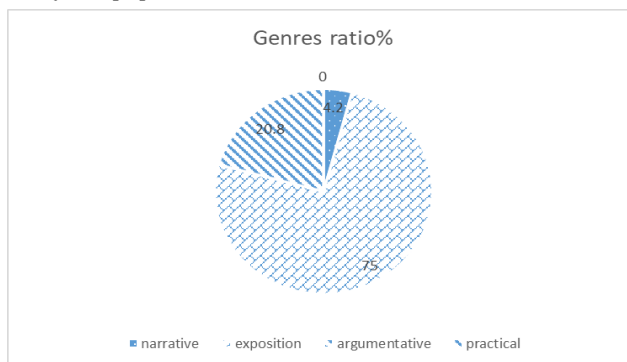


Figure6. Analysis of the genre ratio of the texts in CET-4 reading materials

The data analysis reveals that the exposition genre dominates the CET-4 reading test at 75.0%, followed by argumentative essays at 20.8%. In 2021, only two genres appeared across 12 test papers, while in 2022, three genres were identified, with a notable focus on argumentative essays. Of the 24 test papers from 2021 to 2022, only one included a narrative passage (4.2%), and none featured practical reading passages, reflecting a suboptimal test design.

Figure 6 highlights the limited range of genres in the current CET-4 reading test, which lacks practical writing forms like advertisements, directions, diagrams, and charts—materials commonly encountered in daily life. To better prepare students, it is recommended that test designers incorporate practical writing genres, ensuring a more balanced distribution of genres in the CET-4 reading comprehension subtest.

9.1.3. Readability

When selecting reading materials, it is of significance to take difficulty level into full account based on the student’s reading proficiency. This study will calculate the readability using Flesch readability formula which was developed in 1948 and is still utilized today by Microsoft office Word 2019 software. The table below presents the Readability of Reading Materials in CET-4 from 2021 to 2022. The Readability Yardstick provides a measure of readability, with lower scores indicating more challenging texts.

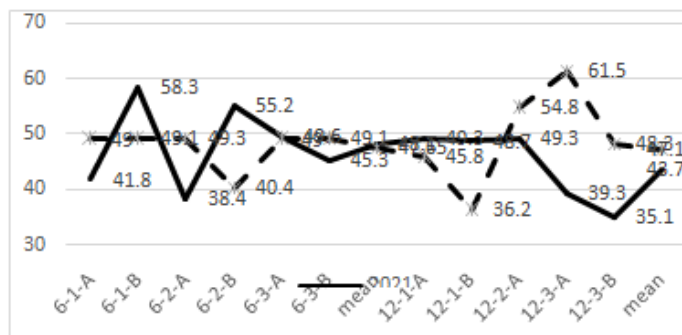


Figure7. Analysis of the readability of CET-4 reading materials

Note: CET-4 has three parallel sets of test papers are available for each test session (one test, multiple papers), 2021-6-1-A refers to the reading passage A of first set of test paper in June, 2021.

The analysis of CET-4 reading passages from June 2021 to December 2022 shows that their readability generally falls within the 40-50 range, meeting the syllabus requirements and indicating a steady trend. The December 2021 test was the most challenging, with an average readability score of 43.7, while the tests in June 2021, June 2022, and December 2022 had similar difficulty levels, with scores of 48.1, 47.65, and 47.1, respectively. The average readability disparity decreased from 4.4 in 2021 to 0.55 in 2022, reflecting greater scientific rigor. The June 2022 test was the most consistent in readability, whereas the December 2022 test was the least ideal, with significant variations and a gap of 25.2. Notably, the December 2022 test included the two most difficult passages, both with a readability index of 36.2, the lowest among all passages.

9.1.4. Length

The length includes two parts: text length and task length. Text length refers to the quantity of words in a given text. As stated by CET-4 Testing Syllabus in 2016, the length of reading passages is approximately 300-350 words. Task length refers to the quantity of words in the question items, which will also affect processing difficulty. Alderson (2000) argues that test developers face challenges when determining the appropriate length of texts for different proficiency levels in relation to the correlation between reading comprehension ability and text length. To ensure objectivity in data analysis, Microsoft Office Word 2019 is used to count the words in the passages. The quantity of words in each passage is shown in the table as follows.

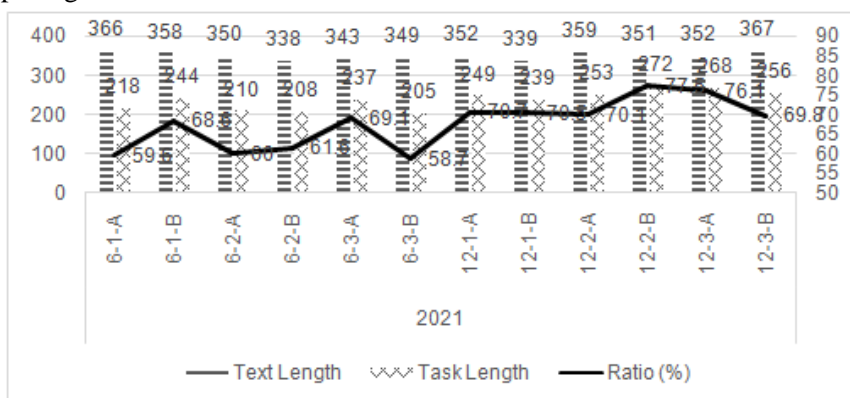


Figure8. Analysis of the length of passages and ratio in CET-4 reading materials in 2021

Note: CET-4 has three parallel sets of test papers are available for each test session (one test, multiple papers), 2021-6-1-A refers to the reading passage A of first set of test paper in June, 2021; Ratio (%)=Task length/Text length

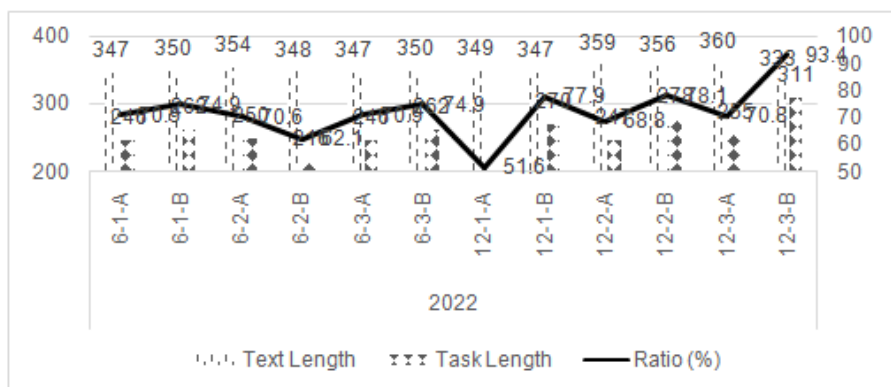


Figure9. Analysis of the length of passages and ratio in CET-4 reading materials in 2022

Note: CET-4 has three parallel sets of test papers are available for each test session (one test, multiple papers), 2021-6-1-A refers to the reading passage A of first set of test paper in June, 2021; Ratio (%)=Task length/Text length

The analysis of text passages in the CET-4 exams from June 2021 to December 2022 reveals minimal variation in length, ranging from 333 to 367 words. While some passages exceed the recommended length of 300-350 words, this deviation aligns with the overall consistency of the testing syllabus. It is crucial for the distribution of text length to follow the progression from shorter and simpler to longer and more complex sections. However, certain passages, such as those in the first and second test sessions in June and December 2021, and the third session in December 2022, do not adhere to this principle, deviating from this well-designed approach.

The data also show that the ratios between task length and text length for reading passages in these years are generally high, with the majority of ratios exceeding 70%. Particularly concerning is the second passage of the third test paper in December 2022, which represents 93.4% of the total text length, potentially overwhelming students and limiting time for critical thinking. The ratios in the December 2022 test paper fluctuate significantly, ranging from 51.6% to 93.4%. Among the twenty-four reading passages, the June 2021 test demonstrates a consistent level of difficulty, with ratios between 60% to 70% of the corresponding text length.

9.1.5. Ratio New Words

Nuttle (1982) suggests that in intensive reading, a high number of new words is acceptable, defining “quite a lot of” as equivalent to “two or three percent”. Other similar tests also recommend that the ratio of new words in testing should not exceed 3%. The Testing Syllabus (2016) theoretically states that the proportion of new words in testing passages should remain below 3%. For the purpose of this study, new words are defined as those exceeding the Curriculum Requirements, excluding proper nouns, derivative or compound words, and words with meanings provided in Chinese brackets. The results of the new words’ ratios in reading passages from 2021 to 2022 are then presented.

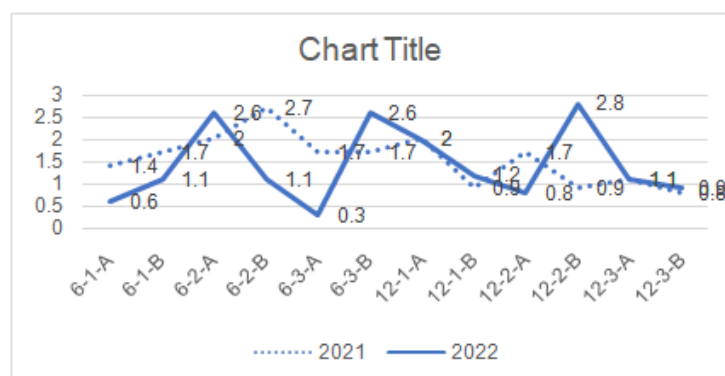


Figure10. Analysis of the ratio of new words (%) in CET-4 reading materials

Note: CET-4 has three parallel sets of test papers are available for each test session (one test, multiple papers), 2021-6-1-A refers to the reading passage A of first set of test paper in June, 2021.

Figure 10 illustrates that the ratios of new words are effectively controlled in the testing papers from June 2021 to December 2022. Across all twenty-four papers, the proportion of new words remains below the recommended 3%, with the ratio varying between 0% and 2.7% in each passage. This suggests that the testing papers are not overly challenging for test takers to comprehend. Therefore, it aligns well with both the Curriculum Requirement (2015) and Testing Syllabus (2016).

9.1.6. Reading Speed

Yang and Weir (1998) suggest that when engaging in reading comprehension, the time spent on reading should be carefully considered, with the ratio of reading to answering questions being approximately 1:0.75. Given that 15 minutes is allocated for this section, it is logical to allocate 8.6 minutes for reading the text and 6.4 minutes for answering the questions. According to the Testing Syllabus (2016), the recommended reading speed for CET-4 is 70 words per minute. Therefore, it is advisable for the text to contain around 602 words ($8.6 * 70$) for optimal reading quantity. The speed of reading passages from 2021 to 2022 is presented in the following results.

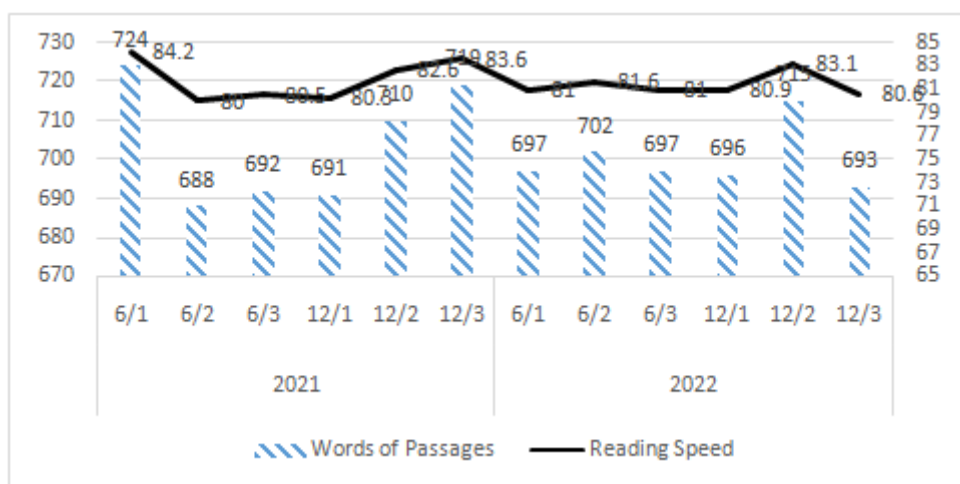


Figure11. Analysis of the reading speed of CET-4 reading materials

Note: CET-4 has three parallel sets of test papers are available for each test session (one test, multiple papers), 2021-6-1 refers to the first set of test paper in June 2021.

Figure 11 shows that all reading passages exceed 602 words, except for the first June 2021 test paper, which has 584 words. This excessive word count results in a reading speed higher than the curriculum requires. The data reveals reading speeds ranging from 80.0 wpm in June 2021 to 83.6 wpm in December 2021, with the only passage meeting the standard recorded at 67.9 wpm in the first June 2021 test paper. Given the curriculum's minimum requirement of 70 wpm, the actual speeds of 80.0-83.6 wpm significantly exceed this threshold.

In summary, the analysis of 24 test papers focused on topic, genre, text length, reading speed, new word ratio, and readability. The results show that, aside from the higher-than-expected reading speed, limited topics, and unbalanced genres, the other aspects align with the testing syllabus and curriculum requirements.

Q2: To what extent do the characteristics of expected response of CET-4’s reading comprehension sub-test in 2021 and 2022 correspond with the Curriculum Requirement (2015) and Testing Syllabus (2016)?

9.2. Characteristics of Expected Responses

This section examines the characteristics of expected response from the response type and reading skills coverage to explore the content validity of the reading comprehension sub-test of CET-4 from 2021 to 2022, based on the revised framework of test task characteristics.

9.2.1. Response Types

Response types, as categorized by Bachman and Palmer (1999), include selected response, limited response, and extended production response. Heaton (2007) suggests that certain item types are better suited for assessing reading comprehension in specific tests. Anderson (2011) notes that no single

method is universally optimal for evaluating reading abilities; different techniques may be required based on test objectives, difficulty, or scope. To accurately measure learners' language abilities, examiners should minimize the influence of response formats.

The CET-4 reading comprehension sub-test relies solely on a selected response format, which accounts for 35% of the total score. While this format ensures high reliability, convenience, and cost efficiency, it may compromise content validity. It lacks real-life relevance, can lead to guessing, and fails to capture the authenticity and interactivity essential in communicative testing.

9.2.2. Reading Skills Coverage

The reading comprehension sub-test aims to evaluate students' ability to extract information from written materials. The Testing Syllabus (2016) explicitly states the nine language skills evaluate in reading comprehension sub-test. These skills are categorized into four distinct categories: identifying the main idea, skimming and scanning, deducing the meaning of words, sentences, and contexts, and making inferences. For convenience and brevity, these categories will be referred to as Sub-skill I, Sub-skill II, Sub-skill III, and Sub-skill IV. The specific reading skills are outlined in the following figures.

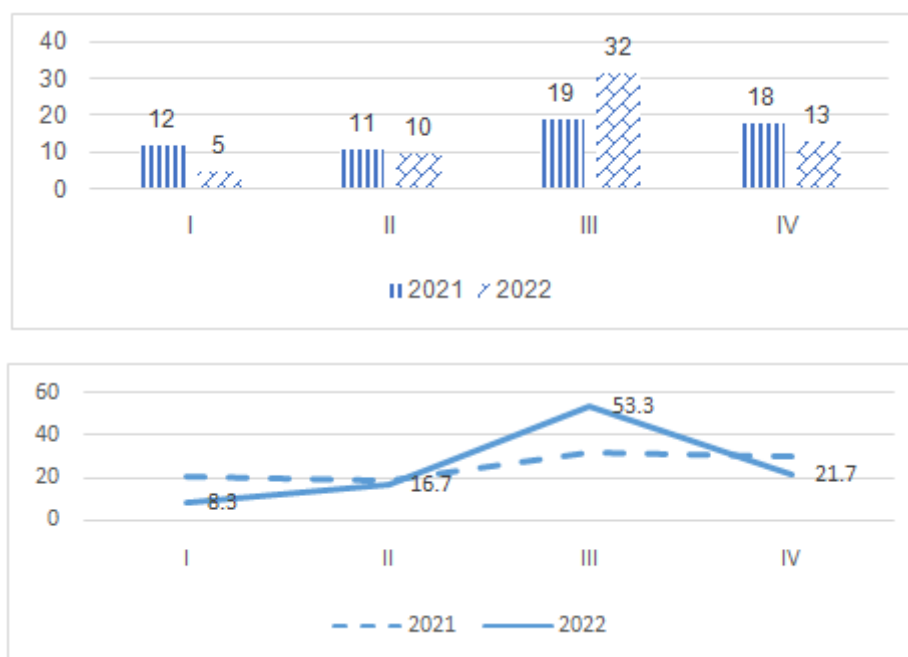


Figure 12. Analysis of distribution and ratio of reading skills of CET-4 reading materials

Note: Sub-skill I refers to the ability of identifying main idea, Sub-skill II is the ability of skimming and scanning, Sub-skill III is the ability of deducing the meaning of words, sentences and contexts and Sub-skill IV is the ability of making inferences.

The analysis of the data from 2021 reveals that the skill of deducing meaning from words, sentences, and contexts was the most prominent, comprising 31.7% of the total with 19 items. Following closely behind was the ability to make inferences, which accounted for 30.0% with 18 items. These two skills collectively accounted for 61.7% of the total sixty items. Similarly, in 2022, the trend persisted with higher-order skills dominating. The data showed that deducing meaning continued to be the primary sub-skill, representing 53.3% with 32 items. Making inferences followed with 21.7% and 13 items. Together, these two skills comprised 75.0% of the total sixty items.

The analysis of reading skills across twenty-four test papers reveals a varied distribution, with only some papers covering all four skills, while approximately half focus solely on three reading skills. The third set of test sessions in June 2022 is even worse, containing only two reading skills.

Overall, the study suggests that to align with testing syllabus and curriculum requirements, a more balanced coverage of reading skills is necessary. This should involve consideration of various test formats, adoption of different response types, and inclusion of text titles.

Q3:By doing this research, what are the implications for the future test propositions in the CET-4 reading comprehension section from 2021 to 2022?

To ensure comprehensive reading passages, test designers should select a diverse range of topics, rather than focusing solely on social sciences, and maintain a balanced distribution of genres, emphasizing practical passages. Incorporating various test formats, such as True-or-False and Short-answer questions, is crucial for assessing overall language competence. Tasks should be concise, and passages should be logically arranged from short to long and simple to complex, reducing test-takers' anxiety. Each passage should cover all four reading skill categories. Additionally, the absence of titles in passages is problematic, as titles provide essential context. Adding titles is necessary to reflect real-life reading scenarios.

10. DISCUSSIONS

The analysis reveals that the distribution of topic areas in these testing papers is limited, with a focus on social science, natural science, humanities, and technology. While natural and social sciences are suitable for CET-4 reading materials, a mix of familiar and unfamiliar topics is needed to engage test-takers and assess different proficiency levels effectively.

The testing syllabus does not adequately reflect the diverse range of genres it specifies, undermining the communicative assessment's authenticity. The limited variety of reading materials leads students to focus on test-taking strategies over genuine skill development. To address this, greater emphasis should be placed on analyzing stylistic characteristics in English teaching and testing, enabling students to improve their language proficiency effectively.

The readability of testing passages is generally adequate, but their organization needs improvement to optimize test-takers' psychological comfort. Arranging passages from easy to difficult is beneficial, and adjusting test paper order based on readability is justified. Deviations from this order affect the test's apparent validity. The current task length design is suboptimal, imposing undue stress on test-takers. Further research is needed to explore the relationship between text length and task length, particularly within time-constrained settings.

The new word ratio aligns with the 2016 syllabus, providing a balanced challenge. However, the reading speed exceeds syllabus limits. To align with curriculum standards and accurately assess reading abilities, a reduction to around 70 words per minute is recommended.

In summary, apart from the issues of length and new word ratio, other aspects such as readability, reading speed, topic area, and genre do not effectively match the criteria set forth in the Testing Syllabus of CET-4 (2016) and the Curriculum Requirements (2015).

Regarding Research Question 2, reading comprehension is a complex skill, and using a single item format may not fully capture test-takers' abilities. To address this, incorporating a variety of response types in assessments is recommended. The current test papers disproportionately emphasize higher-order reading skills, while neglecting lower-order skills, which are foundational. Since CET-4 test-takers are non-native English speakers and non-English majors, future test papers should place greater emphasis on lower-level reading skills. To align with the syllabus, reading comprehension tests should comprehensively evaluate all reading skills.

Regarding Research Question 3, test designers must thoroughly analyze test papers to develop more scientifically rigorous reading comprehension assessments that accurately reflect students' abilities. While the CET-4 reading comprehension test exhibits high content validity, some minor issues persist. To better meet educational demands, it is crucial to refine the test by diversifying topics, balancing genres, ensuring readability, organizing reading skill distribution, and incorporating constructed response questions.

11. CONCLUSIONS

Past studies have shown that analyzing test task characteristics can enhance the content validity of English tests, providing valuable insights for test design and teaching. This study finds that while the length and new word ratio of CET-4 input align with the Testing Syllabus (2016) and Curriculum Requirements (2015), other aspects—such as readability, reading speed, topic area, and genre—do

not. The test formats are limited, and the coverage of reading skills is uneven. Despite these issues, CET-4 reading comprehension demonstrates substantial content validity. To meet evolving demands, it is essential to diversify topics, balance genres, maintain readability, and incorporate constructed response types. However, this study did not assess the quality of question design, such as clarity, suitability, or complexity, highlighting areas for future research.

12. ACKNOWLEDGEMENT

I express my profound gratitude to my supervisor, Charanjit Kaur Swaran Singh, and Wong Wei Lun, for their invaluable guidance and feedback to accomplish this article.

REFERENCES

- Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.
- Alderson, J. C. (2011). *Assessing Reading*. Beijing: Foreign Language Teaching and Research Press.
- American Psychological Association (2014). *Standards for educational and psychological testing*. Washington, D. C: American Psychological Association.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1999). *Language testing in practice*. Shanghai: Shanghai Foreign Language Teaching Press.
- Chen, Q. X. (2014). *Study on the Content Validity of CET-4 (2010--2013) Reading Comprehension Subtests* (Master's dissertation). Guangxi Normal University.
- Dong, H. W. (2016). *A Study of the CET-4 Reading Comprehension from Communicative Language Testing Perspective* (Master's dissertation). Jilin University.
- Dong, M. X. (2007). *Content validity study on reading comprehension tests of NMET* (Master's dissertation). Chongqing University.
- Grabe, W., & F. L. Stoller (2005). *Teaching and researching reading*. Beijing: Foreign Language Teaching and Research Press.
- Goodman, K. S. (2014). *Reading: A psycholinguistic guessing game. Making Sense of Learners Making Sense of Written Language*. Routledge
- Gu, X. D., & Li, Y. G. (2010). Analysis of discourse input and expected response tasks in CET listening tests after reform. *Journal of Xi'an International Studies University*, (4), 71-74, 79.
- Gao, W. X. (2023). *Research on the Content Validity of the Reading Comprehension in College Entrance English Examination of 2021 and 2022*. (Master's dissertation). Hubei Normal University.
- Heaton J.B. (2004). *Writing English Language Tests*. Beijing: Foreign Language Teaching and Research Press.
- Henning, G. (2001). *A Guide to Language Testing: Development, Evaluation and Research*. Beijing: Foreign Language Teaching and Research Press.
- Hughes, A. & Hughes, J. (2020). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Hu, Y. Q., & Xiong, S. C. (2022, March). Investigating Text Complexity of Reading in National Matriculation English Test. In 2022 4th International Conference on Natural Language Processing (ICNLP). 385-389. DOI: 10.1109/ICNLP55136.2022.00071
- Hou, Y. P. (2004). *Content Validity Analysis of TEM-4 Reading Comprehension* (Doctor's dissertation). Shanghai Foreign Language University.
- Khalifa, H., & Weir, C. J. (2018). *Examining reading: Research and practice in assessing second language reading*. Shanghai: Shanghai Foreign Language Teaching Press.
- Lu, Q. L. (2008). Content validity analysis of CET-4 reading comprehension. *Higher Education Forum*. (5), 110-114.
- Meng, F. N. (2016). *A Chronological Analysis on the Task Characteristics of NMET and PETS-2 reading Tests (2011-2015)*. *Examinations Research*. (6), 72-79.
- Peng, Y. Y. (2018). Analysis of the input and expected response task characteristics in the NMET English listening test in 2017. *Educational Measurement and Evaluation*, (4), 11-16.
- Peng, Y. Y., Gu, X. D., & Huang, J. (2017). The Research on the Test-Designed Quality of the Reading Comprehension Among Test Papers of NMET National Version and Provincial Autonomous Ones in 2016---The Analysis of Test Papers Based on Bachman & Palmer's Test Task Characteristics. *Educational Measurement and Evaluation*, (03)57-63.
- Weir, C. J. (1993). *Understanding and developing language tests*. Hemel Hempstead: Prentice Hall International (LTK) Ltd.

- Yang, H., & Weir, C. (1998). *Research on the Validity of College English Tests Band 4 and 6*. Shanghai: Shanghai Foreign Language Education Press.
- Zhao, H. L., & Mu, L. (2016). A study on the quality of CATI Level 2 interpretation (consecutive interpretation)--based on Bachman's test task characteristic. *Foreign Language Testing and Teaching*, (2), 1-8.
- Zhao, C. H., & Huang, J. Y. (2020). The impact of the scoring system of a large-scale standardized EFL writing assessment on its score variability and reliability: Implications for assessment policy makers. *Studies in Educational Evaluation*, 67, 100911. <https://doi.org/10.1016/j.stueduc.2020.100911>
- Zhou, H., & Hu, P. P. (2017). A study on the content validity of English listening test questions in Zhejiang Province from 2012 to 2017. *Educational Measurement and Evaluation*, (11), 19-26.
- The department of higher education of the ministry of education. (2015). *Syllabus for College English Teaching*. Beijing: Higher Education Press.
- The department of higher education of the ministry of education. (2016). *Syllabus for College English Testing*. Beijing: Higher Education Press.

AUTHOR'S BIOGRAPHY



Miss Xie, is a Teacher Educator at the University of Panzhihua in the Department of Foreign Languages. He holds a MA in Applied Linguistics from the Southwest University, Chongqing, China. Her research interests and areas of postgraduate supervision include Language Testing, Language Teaching Theory, and Translation.

Citation: Xie Shuangmei et al. "A Research on Content Validity of CET-4 Reading Comprehension ---Based on Test Task Characteristics Framework" *International Journal on Studies in English Language and Literature (IJSELL)*, vol 12, no. 8, 2024, pp. 24-36. DOI: <https://doi.org/10.20431/2347-3134.1208004>.

Copyright: © 2024 Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.