HiLL PUBLISHING

Translation and Foreign Language Learning, 2026, 2(2), 370-374
https://www.hillpublisher.com/Journals/tfll/
ISSN Online: 3069-0315
ISSN Print: 3070-3077

A Comparative Study on the Content Validity of English Reading Comprehension Test Items in the 2026 College Entrance Examination and Provincial Quality Assessment: A Case Study of Fujian Province

Zhangjie Lin

College of Foreign Languages, Fujian Normal University, Fuzhou 350007, Fujian, China.

How to cite this paper: Zhangjie Lin. (2026) A Comparative Study on the Content Validity of English Reading Comprehension Test Items in the 2026 College Entrance Examination and Provincial Quality Assessment: A Case Study of Fujian Province. *Translation and Foreign Language Learning*, 2(2), 370-374.
DOI: 10.26855/tfll.2026.02.027

Received: December 20, 2025
Accepted: January 6, 2026
Published: February 28, 2026

***Corresponding author:** Zhangjie Lin, College of Foreign Languages, Fujian Normal University, Fuzhou 350007, Fujian, China.

© 2026 by the author(s).
This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) license, which permits non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and is not modified or adapted.
https://creativecommons.org/licenses/by-nc-nd/4.0/

Abstract

With the implementation of the English Curriculum Standards for Senior High Schools (2017 Edition, 2020 Revision), the college entrance examination has a growing impact on regional testing and teaching. Based on the task characteristics of Bachman and Palmer and the core literacy requirements of the Curriculum Standards, this study compares the reading comprehension items of the 2026 Fujian Quality Assessment Test (QA) and the National College Entrance Examination English Test Paper I (used in Fujian) from the two dimensions of discourse input (type, length, difficulty) and expected response (language ability, thinking quality, cultural awareness). The results show that the overall content validity of the two tests is high, and they are comparable in terms of total reading volume, overall difficulty and distribution of core question types, indicating that QA has good consistency with NCEE in these aspects. However, there are notable differences in discourse type coverage, difficulty gradient, depth of higher-order thinking assessment and cultural orientation. NCEE has a clear guiding role in the development of local tests, while QA shows regional diagnostic characteristics. Based on these findings, this paper puts forward suggestions from aspects of text selection, difficulty gradient optimization, thinking skill evaluation, cultural presentation, accuracy of test construction, teaching backwash effect and so on. These suggestions are intended to provide empirical reference for the improvement of local examination development and English reading teaching in senior high schools.

Keywords

Content validity; NCEE English; reading comprehension; quality assessment test; discourse input; expected response

1. Introduction

Language testing plays an important guiding role in senior high school English teaching, and reading comprehension is the main tool to evaluate language ability. Content validity is the fundamental criterion to judge whether the test content is consistent with the curriculum objectives (Ministry of Education, 2020). Curriculum Standards emphasizes four core qualities namely language ability, cultural awareness, thinking quality and learning ability. It also requires

reading evaluation to have diverse discourse types and real contexts.

Existing research on the validity of English reading test mainly centers on selective standardized tests like NCEE (Wang, 2021; Huang et al., 2024), while the local quality assessment test has received less research attention, despite its direct and frequent backwash effect on daily classroom teaching. Furthermore, there is a lack of systematic comparison between QA test and NCEE. Bachman and Palmer's (1996) "Discourse Input + Anticipatory Response" model provides a robust framework and is widely used in the study of NCEE reading (Zhang, 2024; Peng et al, 2017). Research shows that NCEE articles are mainly expository and narrative, with few new media types (Zhang, 2024); as for article length, a standard of 200-500 words is generally adopted (Liu & Han, 2003); the assessment of cultural awareness remains superficial, overemphasizing Western content (Xu & Liu, 2021).

To make up for these gaps, this study compares the national test I used in the 2026 Fujian college entrance examination with the Fujian QA test in April 2026. The study focuses on two main questions: (1) What is the content validity of the two tests? (2) What are their similarities, differences and implications for teaching and local test construction?

2. Research Design

This study takes the English test paper I of Fujian college entrance examination in 2026 and the English QA test paper of Fujian senior high school in April 2026 as the research objects. Each test paper contains 5 reading materials and 20 multiple-choice questions, all of which aim to test students' reading comprehension ability.

Based on Bachman and Palmer's framework and curriculum standards, this research constructs an analytical framework covering discourse input and expected response. For discourse input, each article was classified into narrative, expository, argumentative, practical, news reports, new media, etc. The actual word count (excluding instructions and options) was manually counted, then the Flesch readability score was calculated using SPSS, and the difficulty level was assigned following existing standard. For expected answers, all questions are divided into six skill types (detail understanding, main idea, inferential judgement, word guessing, purpose / attitude, discourse structure), and the share of each type was calculated; each item is also mapped to low-order thinking (LOTS) and high-order thinking (HOTS). Using Kachru's three concentric circle models (inner circle, outer circle, extended circle and global / world circle), the cultural content is classified and the frequency is recorded.

To ensure coding consistency and research reliability, all 40 reading items were independently coded by two raters majoring in English teaching. The initial inter-rater reliability coefficient reached 0.89, and all divergent coding results were discussed and adjusted until full consensus was achieved, ensuring the credibility of subsequent data analysis.

3. Results and Analysis

3.1 Discourse Type

From Table 1, the NCEE paper covers four text types (expository, narrative, linear practical writing, and practical writing with non-linear features), showing more variety; the QA test, in contrast, is mostly made up of expository (60%) and argumentative (40%), with narrative, practical writing (including linear and non-linear), and news types all missing. New media discourse is not included in either test, and this lack of non-linear and new media texts points to a traditional choice of text types; it falls behind the curriculum's call for real and varied reading materials. The QA test depends heavily on expository texts; this is suitable for factual understanding, but it restricts the test's ability to measure how well students grasp story details or pull information from everyday documents.

Table 1. Distribution of Discourse Types

Test	Discourse Type					
	Expository	Argumentative	Narrative	Practical writing	Practical (non-linear)	News / New media
QA	3 (60%)	2 (40%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
NCEE	2 (40%)	0 (0%)	1 (20%)	1 (20%)	1 (20%)	0 (0%)

3.2 Article Length

As shown in Table 2, every reading text stays within 200–500 words; the QA has a total of 1,362 words (average ~272), and the NCEE totals 1,256 (average ~251), both close to the suggested 300-word average. The way length

varies is similar—QA passages range from 242 to 302 (a range of 60), NCEE passages from 217 to 278 (a range of 61). Neither test orders passages strictly from short to long, and this is acceptable because their purposes differ (the QA for diagnosis and review, the NCEE for standardized selection). But the QA places its longest text (302 words) at position B, not later, which may disturb the reading flow of students and add to their mental effort early on. In contrast, the NCEE's first four passages grow in length more smoothly, giving a gentler start and perhaps helping with time management.

Table 2. Passage Lengths (words)

Test	Passage				
	A	B	C	D	E
QA	242	302	272	295	248
NCEE	217	268	272	278	221

3.3 Text Difficulty (Flesch Readability)

The QA test's average Flesch readability score is 56.8 (relatively difficult), while the NCEE's is 60.7 (standard); the NCEE covers four difficulty levels and the QA covers three, both using a multi-tiered design to separate students effectively. However, the QA has 40% difficult passages, against 20% for the NCEE, so the QA is a bit harder overall; the difficulty order in both is B→A→E→C→D, which does not strictly go from easy to hard. This order may keep the test from building student confidence step by step, since weaker test-takers could face hard texts before easy ones; also, the QA's higher share of difficult passages might make the test seem more rigorous, but it could also be less fair for average students, who make up most of the test-takers.

Table 3. Flesch Scores and Difficulty Levels

Readability Metrics	Passage				
	A	B	C	D	E
QA	61.5 (Standard)	69.7 (Standard)	47.4 (Difficult)	47.2 (Difficult)	58.0 (Relatively difficult)
NCEE	66.6 (Standard)	75.8 (Relatively easy)	54.3 (Relatively difficult)	49.8 (Difficult)	57.0 (Relatively difficult)

3.4 Language Ability

According to table 4, detail understanding gets 40% and text structure 25% from both tests, making up 65% together and matching the main testing points; the share for main idea (10%) and word guessing (5%) stays the same across the two. The NCEE includes 20% inference items (the QA only 10%), while the QA has 10% purpose/attitude questions not found in the NCEE. This redistribution means both tests maintain a similar overall proportion of higher-order skill items, though they emphasize different sub-skills; all six skill types are covered by the QA, but the 2026 NCEE paper lacks purpose/attitude items, which suggests that advanced critical reading skills are only partly covered in this test format.

The NCEE's push toward higher-level logical thinking is reflected in its stronger emphasis on inference, which fits the curriculum's aim of building deep understanding; by including purpose/attitude items, the QA can test how well students pick up an author's stance, a skill that is being valued more and more in academic reading, but the NCEE's decision to leave out this question type may narrow the range of reading ability it measures.

Table 4. Language Ability Distribution

Test	Skill Type					
	Detail Understanding	Main Idea	Inferential Judgment	Word Meaning Guessing	Purpose & Attitude	Text Structure
QA	40%	10%	10%	5%	10%	25%
NCEE	40%	10%	20%	5%	0%	25%

3.5 Thinking Quality

Table 5 shows that both tests keep a nearly 1:1 balance between LOTS and HOTS, with the QA at 50% LOTS and 50% HOTS, and the NCEE at 45% LOTS and 55% HOTS; all HOTS items in the NCEE are analytical (55%), while the QA adds 10% evaluation-type questions. Neither test includes memory or creative questions, which is a design choice often seen in objective multiple-choice tests; the NCEE's larger share of analytical items puts more weight on logical reasoning and text deconstruction. However, no evaluation items appear in the NCEE, so students are not required to make a critical judgment about the strength of arguments or the author's viewpoint, and these are core parts of critical thinking. The QA gives a small look into this side by including a few evaluation questions, but its overall analytical share is still lower than the NCEE's; this suggests both tests could be improved by a more even distribution of HOTS sub-categories, so that the range of cognitive skills set by the curriculum can be more fully captured.

Table 5. Cognitive Level Distribution

Test	Level (Category)					
	Remember (LOTS)	Understand (LOTS)	Apply (LOTS)	Analyze (HOTS)	Evaluate (HOTS)	Create (HOTS)
QA	0	10	0	8	2	0
NCEE	0	9	0	11	0	0

3.6 Cultural Awareness

In both tests, inner-circle culture takes up the largest part (NCEE 60%, QA 40%); the QA includes one expanding-circle passage, while the NCEE's non-inner-circle content deals with global and universal topics (40%). The NCEE goes into deeper cultural meanings, such as urban ecology and public aesthetic education, but the QA tends to stay at the level of introducing concepts, with little exploration of the values behind them; this difference may come from the NCEE's broader cultural goal of getting students involved in contemporary global issues, while the QA's cultural content tends to offer conceptual introductions and spends less time uncovering deeper cultural values. Still, the QA's inclusion of an expanding-circle text shows an effort to diversify the cultural viewpoints, even though it still does not do enough to promote intercultural comparison or to build appreciation of Chinese local culture, things that the curriculum clearly encourages. Table 6 below gives a summary of the comparison.

Table 6. Cultural Awareness Category Distribution

Table Head	Category		
	Inner-circle	Expanding-circle	Global thematic
QA	2 (40%)	1 (20%)	2 (40%)
NCEE	3 (60%)	0 (0%)	2 (40%)

4. Conclusion and Suggestions

Both tests are high in content validity, and they match the Curriculum Standards; passage lengths are reasonable with an average near 300 words; the difficulty is moderate to high and covers four tiers, so students of different levels can be told apart effectively. The two examinations both attach equal importance to detail comprehension and text structure, and the ratio of lower-order to higher-order thinking questions hovers around 1:1, fully satisfying the core competency assessment requirements. However, differences can be identified across text type, difficulty sequencing, depth of ability evaluation and cultural presentation. The NCEE paper adopts a richer spectrum of discourse genres and sets up a more progressive difficulty curve; meanwhile its test items place greater weight on analytical thinking, and the cultural texts selected carry deeper connotations. In contrast, the local Fujian QA test relies heavily on expository articles, with an unbalanced difficulty sequence. Though it incorporates a small proportion of evaluative items absent from the NCEE, its cultural materials only cover superficial popular concepts without digging into underlying values. This adaptive adjustment of the QA test matches its positioning as a diagnostic tool rather than a

selective assessment.

These findings are instructive not only for Fujian local test construction, but also for daily senior high English reading instruction. For the development of Fujian provincial quality assessment tests, developers should expand genre diversity by adding non-linear and new media passages and arrange reading materials in ascending length and difficulty order; the difficulty gradient can be optimized to achieve balanced distribution across all tiers; test designers ought to raise the proportion of analytical and evaluative questions and explore innovative item forms; cultural material selection should integrate profound Chinese and foreign cultural themes to realize two-way intercultural comparison. For front-line senior high English teachers in Fujian, a practical priority is to diversify discourse genre training in reading classes, especially the logical analysis of argumentative texts and information extraction from non-linear materials. Teachers need to balance the training of low-order and high-order cognitive skills, and guide students to conduct in-depth comparison between Chinese and foreign cultural contexts in reading practice. By taking NCEE papers as the benchmark of ability standards and QA tests as regular diagnostic tools, teachers can build a virtuous interactive cycle between classroom teaching and formal assessment.

References

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford University Press.

Huang, R., et al. (2024). Quality evaluation of high school English test items based on the teaching-learning-assessment integration theory. *Curriculum, Teaching Material and Method*, (2), 112-118.

Kachru, B. (1985). Standards, codification and sociolinguistic realism: English in the outer circle. In R. Quirk & H. Widowson (Eds.), *English in the world*. Cambridge University Press.

Liu, R., & Han, B. (2003). *Language testing and its methods*. Foreign Language Teaching and Research Press.

Ministry of Education. (2020). *Senior high school English curriculum standards (2017 edition, revised in 2020)*. People's Education Press.

Peng, Y., Gu, X., & Huang, J. (2017). Quality analysis of NCEE English reading comprehension items based on Bachman and Palmer's framework. *Educational Measurement and Evaluation*, (3), 57-64.

Wang, Q. (2021). Diachronic analysis of content validity of reading comprehension in a specific city's NCEE. *Foreign Language Education Research*, (4), 78-85.

Xu, G., & Liu, J. (2021). Deepening reform of foreign language testing in Gaokao from the perspective of competency cultivation – A review of 40 years' evolution of reading tests. *Foreign Language Teaching Theory and Practice*, (1), 81-90.

Zhang, L. (2024). A diachronic study on content validity of NCEE English reading items based on curriculum standards – Taking Zhejiang NCEE (2018–2023). *Science and Education Guide*, (3), 41-43.