

HOW DOES THE 45-MINUTE SUMMATIVE TEST FOR NINTH GRADERS IN BIEN HOA, DONG NAI PROVINCE ALIGN WITH KEY ASSESSMENT CRITERIA SUCH AS VALIDITY, RELIABILITY, PRACTICALITY, AND RELEVANCE?

Nguyen Thi Thu Lan

Faculty of Foreign Languages

Dong Nai University, Vietnam

ABSTRACT

Testing and assessment play a crucial role in English language teaching, particularly when standardized tests determine students' final results. This article examines a 45-minute summative test (end-of-term 1) administered to ninth graders in a secondary school in Bien Hoa, Dong Nai Province, Vietnam. The study provides a detailed description of the test, its teaching context, and its intended purpose. Key assessment criteria, including validity, reliability, practicality, and relevance, are analyzed to identify the test's strengths and limitations. The analysis highlights how well the test aligns with the curriculum, measures student proficiency, and ensures fairness and efficiency. Additionally, the paper discusses the test's backwash effect on teaching and learning. Based on the findings, several recommendations for improvement are proposed to enhance test effectiveness and better support students' language development. The study aims to contribute to the refinement of assessment practices in secondary English education.

Key words: *assessment, summative assessment, standardized test, validity, reliability, practicality, relevance*

1. INTRODUCTION

Testing, from an educational perspective, is an essential component of the teaching and learning process. Without assessment, neither teachers nor students are effectively motivated, and schooling would lose its fundamental purpose. Tests serve as a tool to measure learners' progress, identify knowledge gaps, and guide teachers in adjusting their instructional methods. Moreover, testing ensures that certificates or degrees truly reflect a learner's competence rather than simply marking their attendance in an educational institution.

In many educational contexts, including the one under discussion, testing plays a crucial role in determining students' readiness to advance to the next academic level. Final exam results are not only used for awarding grades but also for making important decisions regarding students' future learning paths.

This paper aims to analyze an end-of-term test administered to ninth graders in a local secondary school, focusing on its validity, reliability, and practicality. By examining how well the test measures what it is intended to assess, how consistently it performs, and how feasible it is to administer and score, this analysis will highlight its strengths and weaknesses. Additionally, recommendations will be provided to improve the test's effectiveness, ensuring better alignment with the nationwide curriculum. Enhancing the test design can lead to a more positive backwash effect, fostering meaningful teaching and learning experiences for both educators and students.

2. DESCRIPTION OF THE TEST

2.1. The teaching context

The students taking this test are ninth graders, aged 14, who have been studying English for seven years. They are currently using the newly revised curriculum developed by the Ministry of Education and Training, designed to enhance their foreign language proficiency. As English is a compulsory subject in secondary schools, students receive structured instruction through three lessons per week.

The revised materials aim to develop students' language skills, including listening, speaking, reading, and writing, while also fostering their ability to use English effectively in real-life situations. This test is part of the assessment framework used to measure their progress and proficiency. By evaluating their comprehension and language application, the test plays a crucial role in tracking their academic development and ensuring they meet the learning objectives set by the curriculum.

2.2. Purpose of the Test

The assessment system in secondary schools includes both formative and summative evaluations. Formative assessments occur throughout the learning process and include oral tests, daily lesson reviews, mid-term tests, and 45-minute tests.

This particular test is a summative assessment, specifically an achievement test administered at the end of the first semester. Its primary purpose is to evaluate students' overall learning progress and language proficiency based on the curriculum objectives. The test measures their comprehension, language skills, and ability to apply knowledge acquired over the semester.

Additionally, this written test serves as a classification tool, as students are graded based on their performance. The scores they receive are officially recorded in their academic transcripts, contributing to their overall academic evaluation. By assessing students' achievements, the test provides valuable feedback for both teachers and learners, helping to identify strengths and areas for improvement in English language learning.

2.3 Description of the Test

The test is structured into two main sections: a multiple-choice section and a writing section. These sections aim to evaluate students' receptive and productive language skills, including listening, reading, grammar, vocabulary, pronunciation, and writing. The test aligns with the curriculum by assessing language knowledge and skills covered in the first six units of the prescribed coursebook.

Section I: Multiple-Choice Section

The first section consists of multiple-choice questions that assess receptive language skills and language knowledge. It includes three key components: listening, reading, and language proficiency (grammar, vocabulary, and pronunciation).

- **Listening Component:** This section includes two tasks. The first task consists of five True-False items based on a short story titled *Sam's Journey*, requiring students to determine the accuracy of given statements based

on the listening passage. The second task presents five multiple-choice questions related to *Freddie's Visit to Hardin Castle*, testing students' ability to comprehend spoken English, extract key details, and infer meaning.

- **Reading Component:** The reading section contains ten questions centered around a passage about *Simmonds and Different Signs*. These items assess students' comprehension skills, including their ability to recognize key ideas, infer meaning, and apply reading strategies. The inclusion of varied text types aims to reflect real-world reading scenarios, enhancing students' ability to interpret different forms of written communication.
- **Language Knowledge Component:** This component consists of three subcategories: grammar, vocabulary, and pronunciation. The grammar section includes five multiple-choice questions that evaluate students' understanding of key grammatical structures covered in the syllabus. The vocabulary section consists of two questions that assess word meaning and usage in context. The pronunciation section contains two items where students must identify words with distinct phonetic features, testing their awareness of English pronunciation patterns.

Section II: Writing Section

The writing section requires students to demonstrate their ability to produce written English through three distinct tasks.

- **Reordering Task:** In this task, students are given jumbled words and must arrange them into grammatically correct and meaningful sentences. This exercise evaluates students' knowledge of English syntax, sentence structure, and word order.
- **Sentence Transformation Task:** The second task requires students to rewrite three given sentences while maintaining their original meaning but changing their grammatical structure. This tests students' ability to apply alternative grammatical patterns and demonstrate flexibility in language use.
- **Guided Paragraph Writing Task:** In the final task, students are required to write a paragraph of 100–120 words about a personal experience. Guiding questions are provided to assist them in structuring their response. This task assesses students' ability to organize ideas coherently, apply appropriate vocabulary and grammar, and produce a well-structured written text.

3. ANALYSIS OF THE TEST

3.1. Analysis of Test Validity

An initial review of the test under analysis suggests that it covers a broad range of language skills and structures from the first six units of the secondary school coursebook. This comprehensive coverage contributes significantly to its overall validity, as it aligns with the content students have studied. However, when examining the effectiveness of this test in truly measuring students' attainment, several limitations emerge.

One notable issue pertains to the pronunciation section. For instance, Question 17 is designed to assess students' pronunciation skills. On the one hand, this inclusion is beneficial, as it encourages learners to develop a foundational understanding of English pronunciation, which is crucial for the refinement of productive language skills. On the other hand, this test component fails to fully achieve its intended objective. The primary concern is that the assessment method is indirect; rather than requiring students to demonstrate their pronunciation abilities through actual speech, they are merely asked to identify words that are pronounced differently from others. This approach diminishes the communicative value of the language and does not reflect real-world pronunciation usage. As Hughes (2003, p. 31) asserts, "if people try to measure an ability indirectly, they can no longer take for granted what they are doing." Moreover, the coursebook's pronunciation focus emphasizes the contrast between /h/ and /r/ sounds, yet the test assesses knowledge of silent /h/, thereby failing to align with the intended learning outcomes.

Another critical aspect of the test's validity concerns its treatment of communication skills. The textbook includes sections specifically designed to equip learners with functional language for real-life interactions, such as expressing gratitude, apologizing, and making promises. Despite the significance of these communicative functions, they are underrepresented in the test. The only question addressing this area is Question 11, which focuses on apologies. Given

the importance of communicative competence in language learning, the test would benefit from a more balanced inclusion of pragmatic functions to better reflect students' ability to use English in authentic contexts.

The reading section, in contrast, is more communicative, as it integrates grammar and vocabulary within contextualized passages. Additionally, it employs an integrative approach by assessing not only reading comprehension but also students' mastery of various linguistic elements. A detailed examination of this section (Questions 21–27) shows that it includes a cloze test incorporating diverse grammatical structures, such as verb tenses (simple present and present perfect), conjunctions (so, because, although, however), prepositions, and lexical items (adjectives, nouns). While this design supports language acquisition, concerns arise regarding the second reading task. This task comprises three questions that assess students' comprehension of brief real-life texts, including notices, emails, signs, or labels. The brevity of these texts contrasts starkly with the reading passages that students typically encounter in class, which adhere to the assessment requirements set by the Ministry of Education and Training (2014), stipulating that reading texts should be between 140 and 160 words in length. Moreover, the topics of these texts are not closely aligned with the thematic content outlined in the syllabus, further undermining the test's validity.

An additional issue pertains to the separation of language skills in the test. The coursebook emphasizes an integrated skills approach, wherein reading is combined with speaking, and listening is linked to writing. However, the test largely isolates these skills, deviating from the instructional approach students are accustomed to.

The writing section consists of three tasks: reordering words to form correct sentences, sentence transformation, and paragraph writing. The first two tasks primarily assess grammatical competence, as students are required to manipulate isolated sentences to demonstrate their understanding of specific grammatical structures, including conditional sentences, the present perfect tense, and the use of "wish" constructions. While these tasks effectively gauge grammatical accuracy, they lack contextualization, which limits their ability to assess students' overall writing proficiency.

The final task, paragraph writing, instructs students to compose a 100–120 word narrative about a personal experience, following a set of guiding questions. This guided writing approach provides students with essential scaffolding to support their writing process, thereby enhancing their ability to express ideas coherently. However, the test would benefit from a more integrated approach that aligns writing tasks with reading passages. Hughes (2003, p. 84) emphasizes that an ideal writing task should require students to engage in real writing, suggesting that meaningful, content-rich prompts would contribute to a more authentic assessment of students' writing abilities.

In short, while the test demonstrates considerable validity in terms of content coverage, several shortcomings prevent it from fully meeting its intended purpose. The pronunciation section lacks direct assessment, reducing its communicative relevance. The test's limited emphasis on functional language further detracts from its ability to measure students' real-world communicative competence. Additionally, the reading section includes texts that do not align with syllabus requirements, and the test structure fails to reflect the integrated skills approach promoted in the coursebook. Finally, the writing section would benefit from tasks that emphasize meaningful, contextualized writing rather than isolated sentence manipulation. Addressing these concerns through revisions and refinements would enhance the test's validity and improve its overall effectiveness as a tool for measuring students' language proficiency.

3.2. Analysis of Test Reliability

Reliability is a crucial criterion in test evaluation, ensuring that an assessment consistently measures what it is intended to assess. Various scholars have provided different definitions of reliability. Bachman and Palmer (1996) define reliability as "the consistency of measurement," emphasizing that a test is reliable when it yields stable and consistent results across different administrations. Similarly, Bynom (n.d.) highlights that reliability and validity are fundamental principles in testing, although they may sometimes appear to be in conflict.

Brown (2001) further refines this concept, asserting that reliability is closely linked to the consistency of test scoring. In this regard, the summative test under analysis demonstrates a relatively high degree of reliability, particularly in terms of scoring and administration. A significant proportion of the test (approximately 85%) consists of objective multiple-choice questions, which facilitate quick and consistent marking. Furthermore, these items can be graded by

less experienced teachers with minimal subjectivity, ensuring a standardized evaluation process. The test also employs discrete-point testing, covering a wide range of grammatical structures included in the textbook. This ensures that students are assessed on their comprehensive understanding of the course content rather than relying on rote memorization of a limited number of lessons. As Heaton (1988) states, a well-constructed test should contain a representative sample of the course material.

The format of the test further enhances its reliability. It includes four distinct task types: multiple-choice questions, sentence reordering, sentence transformation, and paragraph writing. The familiarity of these question types to learners contributes to the reliability of the test, as Heaton (1990) suggests that a good test should reflect the instructional methods employed in the classroom. Additionally, the test layout is well-organized, with clear and explicit instructions that minimize ambiguity. According to Hughes (2003), a well-presented test with unambiguous instructions contributes to higher reliability by reducing misinterpretations and enhancing student performance consistency.

While the test format strengthens its reliability, certain limitations should be acknowledged. The heavy reliance on multiple-choice questions (30 out of 36 items) may restrict the overall validity of the test by not fully assessing students' productive language skills. However, the multiple-choice items themselves are carefully constructed, each with only one correct answer. The inclusion of plausible distractors further enhances reliability, as these distractors are of similar length and structure to the correct answer, making the test more challenging for students who are unsure. Additionally, the test is free of structural errors, further ensuring that all students are assessed under the same conditions.

Despite these strengths, some concerns remain. One major issue is the possibility of academic dishonesty, a common challenge in standardized testing. The multiple-choice format, while efficient for scoring, can make it easier for students to engage in cheating. To further improve reliability, measures such as randomized question orders, different test versions, or stricter invigilation protocols could be implemented.

In conclusion, the summative test demonstrates a strong level of reliability due to its structured format, objective scoring, and alignment with the syllabus. However, potential threats such as the risk of cheating and the limited range of question formats should be addressed to further enhance the overall reliability of the assessment.

3.2 Practicality and Relevance of the Test

3.2.1. Practicality

According to Bachman and Palmer (1996), practicality refers to the human resources, material resources, and time required to develop, administer, and score a test. Hughes (2003) further defines a practical test as one that is easy and cost-effective to construct, administer, and interpret. A highly practical test should minimize costs and logistical challenges while maintaining efficiency in scoring and evaluation.

One of the main reasons for the existence of this particular test is its economic feasibility. With a large number of students—approximately 400 ninth graders—the test developers and administrators must consider cost-effectiveness in its design. The test consists of four tasks printed on two sheets of paper, which keeps printing costs low. Additionally, since the test contains a significant proportion of objective questions, it reduces the time and effort required for both test writing and grading.

From an administrative perspective, the test is highly efficient. It is designed to be easily managed in a large-scale setting, with multiple-choice and other objective-format questions enabling quick and reliable scoring. The test requires minimal resources in terms of grading personnel and time. Since a large percentage of the test consists of objective items, each paper can be scored quickly, often in less than six minutes if the scorer is experienced and employs an effective mnemonic device. This efficiency is crucial when handling a large volume of test papers within a limited grading period.

Overall, the test achieves a high level of practicality by being cost-effective, easy to administer, and quick to score, making it an ideal choice for large-scale assessment scenarios.

3.2.2.Relevance

The relevance of a test refers to how well it aligns with the students' background, educational experience, and the learning objectives outlined in the curriculum. This test meets the relevance criteria satisfactorily and can be rated an 8 out of 10 in this regard.

One of the test's strengths is that it is designed for a student population that shares a similar cultural and educational background. Even though students come from different classes, they have been taught using the same curriculum and textbook. As a result, there is little risk of the test being more (or less) valid for certain groups of students, ensuring fairness in assessment.

Additionally, the test content closely aligns with the textbook syllabus. Most of the grammar and vocabulary questions are derived from the materials covered in class, ensuring that students are tested on what they have learned. Similarly, the listening and writing tasks are relevant to the themes and topics presented in the coursebook. This alignment enhances the test's validity by ensuring that it measures the knowledge and skills that students were expected to acquire.

In conclusion, this test demonstrates a high level of practicality by being cost-effective, easy to administer, and quick to grade. Its relevance is also commendable, as it aligns well with students' educational backgrounds and the syllabus content. These factors contribute to making the test a suitable and efficient tool for assessing students' language proficiency.

4. SUGGESTIONS FOR IMPROVING THE TEST

To enhance the effectiveness and fairness of the test, several modifications can be made to better align it with the learning objectives and assessment criteria. The following recommendations aim to improve the validity, reliability, and overall impact of the test on student learning.

Firstly, regarding the pronunciation section, it would be beneficial to replace the current format, which requires students to identify phonetic differences in words, with a more practical assessment method. A more effective approach could involve students pronouncing specific sets of words containing target sounds and recording their pronunciation. This could be incorporated into ongoing assessment or oral evaluations conducted throughout the term. Such an approach would provide a more direct and authentic measure of students' pronunciation abilities. Given the time and logistical constraints of a written test, an alternative solution could be to omit this section and allocate the available space to communication-based tasks, which are essential for real-life language use. Without adequate representation in the test, students may perceive communicative skills as less important, leading to insufficient emphasis in their learning process.

Secondly, improvements can be made to the reading section, which is a critical component of the test. The selected texts should meet the length requirements outlined by the Ministry of Education and Training, ensuring that they adequately challenge students. Furthermore, the texts should be thematically relevant to the syllabus to maintain coherence with classroom instruction. The comprehension questions should also vary in complexity, covering different cognitive levels as described in Bloom's Taxonomy, such as recognition, comprehension, and application. By incorporating higher-order thinking questions, the test can better assess students' analytical and inferential skills while also offering opportunities for more advanced students to challenge themselves.

In terms of writing, greater integration with reading tasks would be beneficial. A follow-up writing activity linked to the reading section could help students develop parallel writing skills, reinforcing their ability to construct well-structured responses based on textual input. Additionally, the sentence transformation section, which requires students to rewrite sentences while maintaining their original meaning, could be improved by presenting the original sentences in quotation marks. This minor adjustment would enhance both the accuracy and face validity of the test, making it clearer and more formal. For instance, question 34 could be revised as follows:

34. "The last time I saw my uncle was two weeks ago."
I haven't

By implementing these changes, the test would more effectively assess students' language proficiency while promoting meaningful learning experiences.

5. CONCLUSION

Effective language assessment should go beyond merely testing what students have memorized or are familiar with; it should challenge them to apply their knowledge in meaningful ways. The analysis of the summative test administered to ninth-grade students highlights both its strengths and areas for improvement. While the test demonstrates reliability in scoring and alignment with the syllabus, certain aspects, such as pronunciation assessment, reading text selection, and communicative competence evaluation, require refinement. By incorporating more authentic tasks, ensuring appropriate text length, and integrating writing with reading, the test can better reflect students' overall language proficiency. Additionally, a balanced approach to assessment—one that includes both objective and performance-based tasks—will enhance validity and backwash effects. Ultimately, improving the test structure will not only provide a more accurate measure of students' abilities but also encourage more effective teaching and learning practices in the English language classroom.

6. REFERENCES

- Batchman, L, and A. Palmer (1996). *Language Testing in Practice*. Oxford: Oxford University Press.
- Brown, H.D (2001). *Teaching by Principles. In Interactive Approach to Language Pedagogy*. 2nd Ed. Longman. San Fransisco University.
- Heaton, J.B (1988). *Writing English Language Tests*. London: Longman.
- Heaton, J.B (1990). *Classroom Testing*. London: Longman.
- Hughes, A. (2003). *Testing for Language Teachers*. 2nd Ed. Cambridge: Cambridge University Press.