



Designing and Developing an English language Test for Secondary Certificate in Libya: A Case of Study on Libyan Teachers and Students in Two Secondary Schools in Albeida City

Haneen K. Saleh^{1*}, Machalla M. Megaiab²

¹ Center of Languages, Mohamed Bin Ali Al-Senussi University, Albeida, Libya

² English Department, Faculty of Languages, Tobruk University, Tobruk, Libya

تصميم وتطوير اختبار اللغة الإنجليزية لشهادة الثانوية العامة في ليبيا: دراسة حالة للمعلمين والطلاب الليبيين في
مدرستين ثانويتين بمدينة البيضاء

حنين صالح^{1*}، ماشالله مجياب²

¹ مركز اللغات، جامعة محمد بن علي السنوسي، البيضاء، ليبيا

² قسم اللغة الإنجليزية، كلية اللغات، جامعة طبرق، طبرق، ليبيا

*Corresponding author: hanen.k.khames@ius.edu.ly

Received: July 07, 2025

Accepted: August 20, 2025

Published: August 24, 2025

Abstract:

Designing a language test is a challenging task. This study aimed to develop an English test for third-year secondary students and verify its validity and reliability. A quantitative approach was adopted using closed-ended instruments: a proficiency test, a questionnaire for 40 students, and interviews with 10 teachers. Data were analyzed using SPSS with descriptive statistics. The results showed average student performance, with strength in grammar and weaknesses in reading and vocabulary. Students reported limited time and poor curriculum coverage. Teachers appreciated the test structure but recommended improving it to reflect the four skills more clearly. The study recommends reviewing the content, adjusting the difficulty, and involving teachers to enhance assessment quality and fairness.

Keywords: language Test, Validity and Reliability, Quantitative approach, Assessment Quality.

المخلص:

يُعد تصميم اختبار اللغة مهمةً صعبة. هدفت هذه الدراسة إلى تطوير اختبار لغة إنجليزية لطلاب الصف الثالث الثانوي والتحقق من صلاحيته وموثوقيته. تم اعتماد نهج كمي باستخدام أدوات مغلقة: اختبار الكفاءة، واستبيان لـ 40 طالباً، ومقابلات مع 10 معلمين. تم تحليل البيانات باستخدام برنامج SPSS مع الإحصاء الوصفي. أظهرت النتائج أداءً متوسطاً للطلاب، مع قوة في القواعد ونقاط ضعف في القراءة والمفردات. أفاد الطلاب بمحدودية الوقت وتغطية المنهج الدراسي الضعيفة. أعرب المعلمون عن تقديرهم لهيكل الاختبار لكنهم أوصوا بتحسينه ليعكس المهارات الأربع بشكل أوضح. توصي الدراسة بمراجعة المحتوى وتعديل الصعوبة وإشراك المعلمين لتحسين جودة التقييم وإنصافه.

الكلمات المفتاحية: اختبار اللغة، الصلاحية والموثوقية، نهج كمي، جودة التقييم.

Introduction

Education is the process through which individuals acquire knowledge, skills, and competencies necessary for personal and social development (Ellis, 2003). In the realm of education, the terms "test", "evaluation", "assessment" and "teaching" play crucial yet distinct roles, each contributing in a distinct

way to the learning process. A test refers to a formal examination designed to measure specific knowledge, skills, or abilities under controlled conditions (Bachman, 1990). For instance, a mathematics test might assess students' problem-solving skills through a series of structured questions. Evaluation, on the other hand, encompasses a broader range of activities aimed at gathering and interpreting evidence to make informed judgments about educational effectiveness (Hughes, 2003). An evaluation of a new teaching method might involve analyzing test results alongside qualitative data from student feedback and classroom observations to assess its impact on learning outcomes.

Assessment is a multifaceted concept that encompasses both formal testing and ongoing evaluation within educational settings. It involves the systematic collection of data to track student progress and comprehension across various aspects of learning. According to Smith (2018), assessment methods can extend beyond traditional tests to include continuous observations, student interviews, and portfolio assessment, offering a comprehensive view of student learning. Furthermore, Fulcher (2010) highlights that effective assessment practices not only evaluate student achievement but also significantly influence instructional approaches. For example, educators can leverage assessment data to adapt their teaching methods to meet the diverse learning needs of students, fostering a more individualized and supportive environment. This approach allows educators to pinpoint areas of student strength and those needing further assistance, ultimately improving overall learning outcomes and increasing student engagement.

Teaching, as a foundational component of education, refers to the deliberate efforts of educators to facilitate learning experiences that foster student acquisition of knowledge and skills. Teachers utilize assessment data to tailor their instructional methods. For example, a language teacher might adjust lesson plans based on assessment feedback to enhance students' speaking proficiency. This iterative process of assessment-informed teaching contributes to improving educational outcomes by addressing areas of weakness and building on strengths in both curriculum delivery and student comprehension (Bachman, 1990).

In recent years, educational assessments have become an essential component of teaching and learning processes, serving as vital tools to evaluate students' performance, inform instructional strategies, and ensure that learning outcomes align with curriculum goals. In particular, English language testing plays a crucial role in determining students' proficiency and guiding improvements in language education. This study focuses on the development and design of an English language test in Libya, aiming to create a valid and reliable tool that supports both teaching effectiveness and student achievement. According to Harmer (2002), a proficiency test offers a general overview of students' abilities and knowledge.

Students' knowledge encompasses grammar and vocabulary, while their abilities include reading ability, listening ability, writing ability, and speaking. Therefore, this research centers on grammar, vocabulary, and reading, aiming to create a well-rounded test that accurately reflects students' proficiency levels. By focusing on these areas, the study seeks to identify specific strengths and weaknesses, enabling targeted educational in Libyan secondary schools. As noted by Giraldo (2018), informed decision-making by teachers, guided by assessment results, plays a pivotal role in shaping effective teaching and learning processes.

Therefore, this research offers concrete examples and recommendations to empower teachers in enhancing language assessment practices and fostering improved learning outcomes among students. Sultana (2019) confirms that teachers need to know how to implement language tests effectively to foster better learning outcomes. Language testing is an essential part of the teaching and learning process. Weir (2005) suggested that effective language testing requires an understanding of both the trait being measured and the methods used for assessment. Language is an essential part of the teaching and learning. Brown (2004) noted that teaching and learning are interdependent, making it challenging to separate them. When teachers instruct, they require tests to determine how well their students have achieved the course objectives. Therefore, teaching and testing should be interconnected, with testing promoting effective teaching and, when necessary, addressing and improving poor teaching practices. Testing is an active process that provides valuable information about both teachers and student.

There are various types of assessments used in English language testing. According to Hughes (2003), tests can be categorized based on their purpose, including proficiency tests, diagnostic tests, achievement tests, and placement tests. Some tests designed to assess students' performance in specific English skills either during or at the end of a program. Teachers create achievement tests to evaluate specific topics within one skill. In contrast, proficiency tests, such as IELTS and TOEFL, measure students' overall proficiency in English. Placement tests are administered at the beginning of a program to group students based on a certain educational standard, while diagnostic tests are conducted to identify students' strengths and weaknesses.

To accurately assess students' proficiency in English, it is crucial for teachers to establish clear learning objectives (Brown, 2004). When diagnostic tests are not aligned with these objectives, there arises a disparity between the taught material and the test content. This misalignment can impede the test's ability to effectively gauge whether students have comprehended the course content or identified areas requiring improvement. Properly aligned diagnostic tests play a pivotal role in enabling teachers to pinpoint learners' weaknesses and implement targeted interventions. Language tests have a longstanding presence in education, serving consistent purposes across various teaching and learning contexts. As Corder (1973) suggests, these tests primarily measure learners' knowledge or competence in the language at specific points in their courses, rather than evaluating teaching materials or instructors directly.

This research also holds significant implications for language teachers and test developers, providing valuable insights into the creation of valid and reliable assessments. It seeks to evaluate the impact of recent curriculum changes in Libyan third-year secondary education on language teaching practice and student outcomes. Understanding these effects is crucial for educators aiming to optimize their teaching methodologies based on empirical evidence derived from assessment results. Moreover, the study contributes practical strategies for interpreting and utilizing test outcomes effectively.

Material and methods

This study was designed to evaluate the English language test developed specifically for Libyan third-year secondary students, ensuring it produces reliable and valid results. A quantitative research design was adopted to gather objective and measurable data aligned with the research questions and objectives. The study took place in two secondary schools located in Albeida city: Al-Motafawiqueen School and Al-Shoala School. These schools were selected purposively to represent typical secondary education settings in the city.

The sample consisted of 40 third-year secondary students, aged between 17 and 18 years, representing both genders and various academic streams, including high-achieving and average-performing students. Participants were selected using convenience sampling based on their availability and willingness to participate. This ensured a diverse and representative sample to reflect the population under study.

Data collection involved three main instruments:

1. English Proficiency Test:

A standardized test with 30 multiple-choice items covering grammar, vocabulary, and reading comprehension, developed following Brown (2004) stages of test development to ensure content validity and alignment with the official Libyan secondary school curriculum.

2. Closed-ended Questionnaire:

Designed to gather students' perceptions of the test's content, format, clarity, and fairness. The questionnaire was constructed based on quantitative principles, facilitating straightforward statistical analysis of responses (Creswell, 2009).

3. Structured Interviews:

Conducted with 10 English language teachers from the selected schools, these interviews aimed to collect qualitative insights into the test's validity, reliability, and curriculum alignment. The interviews were guided by a well-prepared framework to maintain consistency while allowing detailed responses (Kallio et al., 2006).

Data collection procedures were carefully planned: the proficiency test was administered under standardized conditions in regular classrooms. Immediately afterward, students completed the questionnaire to capture their immediate reactions. The structured interviews were held face-to-face in a quiet setting during school hours to encourage open discussion. All data were collected within one month.

Data analysis was performed using SPSS software. Descriptive statistics summarized student performance, questionnaire responses, and interview data, including frequencies, means, percentages, and standard deviations (Field, 2018). Ethical considerations were strictly observed; permissions were obtained from school administrations, and participants were informed about the study's purpose. Participation was voluntary, and confidentiality was guaranteed, with no identifying information collected.

Results and discussion

Proficiency Test Results and Discussion

The proficiency test used in this study showed strong internal consistency, with a Cronbach's Alpha of 0.836 and a validity coefficient of 0.914, indicating that the instrument was both reliable and valid for measuring students' English proficiency. Questions one to ten presented the analysis of students' responses to reading comprehension. The results showed a relatively strong performance on most literal comprehension questions. For instance, 80 – 87% of students correctly answered items like "Tariq

Zaid was" and "*Tariq was rescued*", indicating clear understanding of key factual details. In contrast, more complex or inferential items such as "*Attacked by the shark*" received only 40% correct answers, highlighting areas of weakness in deep comprehension. Figure (1) illustrates the responses of the participants in the study to the question "Tariq Zaid was". Figure (2) demonstrates the responses of the participants in the study to the question " Tariq was rescued ". Figure (3) presents the responses of the participants in the study to the question " when he was attacked by the shark, he ".

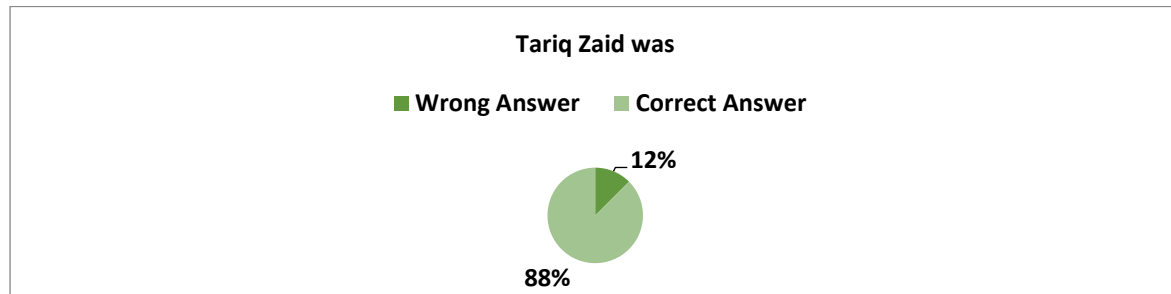


Figure (1): The responses of the participants in the study to the question "Tariq Zaid was".

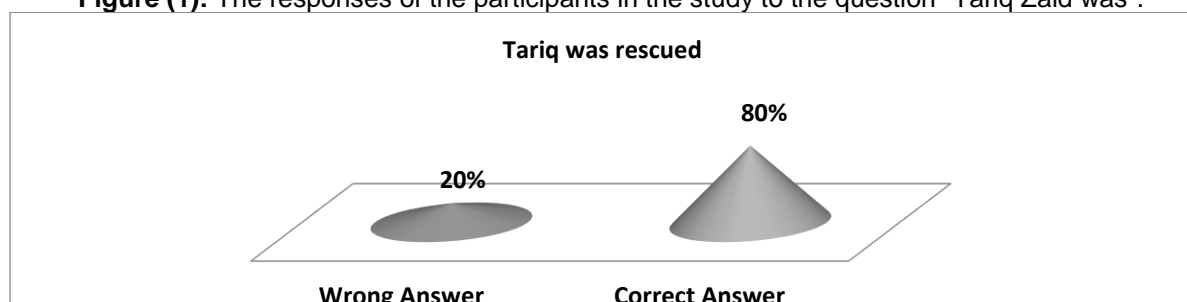


Figure (2): The responses of the participants in the study to the question " Tariq was rescued ".

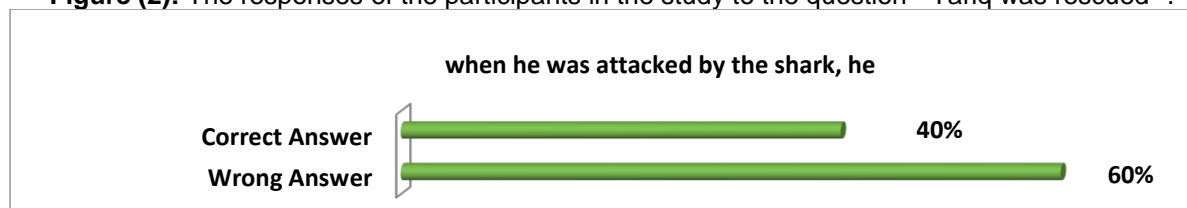


Figure (3): The responses of the participants in the study to the question " when he was attacked by the shark, he ".

The analysis of vocabulary related items (questions 11 to 20) showed varying levels of proficiency. In this section revealed varied levels of lexical and morphologic proficiency among the participants. Correct answers ranged from 50% to 72.5%, with stronger performance on basic grammar related vocabulary. Figure (4) and figure (5) illustrated high scores in recognizing correct prefixes and phrasal verb structure. However, lower scores in context-based items indicated a gap in deeper lexical understanding. These results suggested the need for more focused instruction on vocabulary strategies and word use.

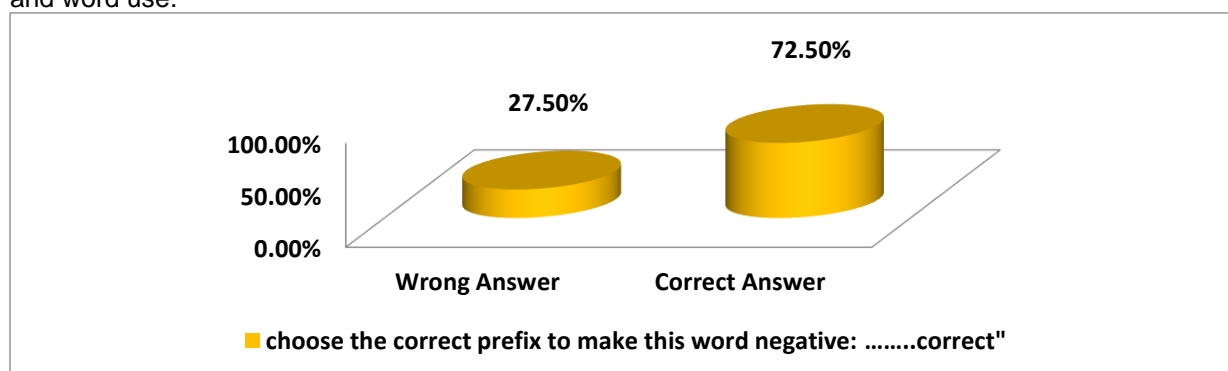


Figure (4): The responses of the participants in the study to the question" choose the correct prefix to make this word negative: Correct".

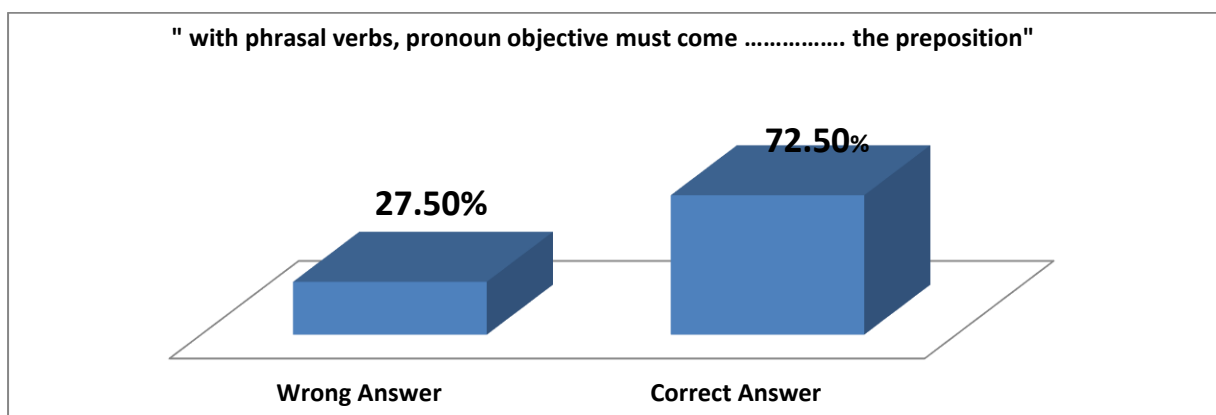


Figure (5): The responses of the participants in the study to the question "with phrasal verbs, pronoun objective must come "The preposition".

The analysis of grammar items (21-30) showed that participants demonstrated varying levels of mastery across past tenses, quantifiers, and sentence structure. Most respondents performed well in recognizing quantifiers like many and much, with over 77% answering correctly in several items. However, questions involving verb tense items usage in specific contexts, such as emergencies or complex conjunctions, showed lower accuracy. As seen in figure (6). The highest success rate reached 85%, while the lowest was 52.5%. the findings suggested that learners needed further reinforcement in applying grammar rules accurately in real-life contexts.

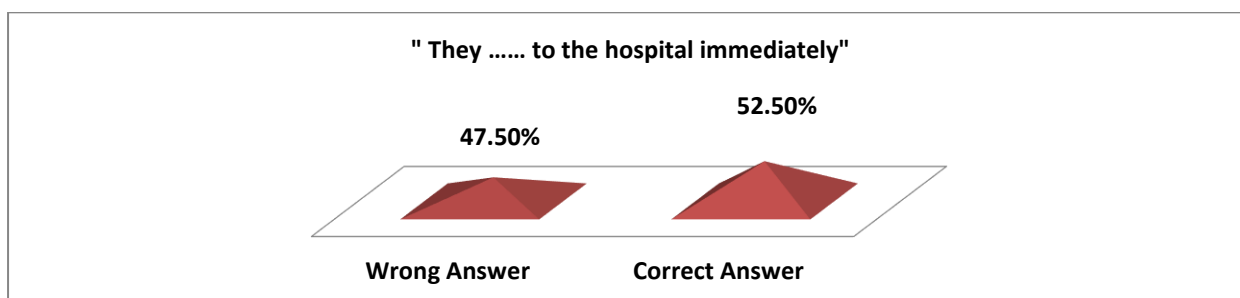


Figure (6): The responses of the participants in the study to the question "They ...to the hospital immediately".

Overall, the proficiency test effectively identified students' strengths and weaknesses. The strongest performance was in grammar, reflecting a solid foundation in this area. Weaknesses appeared clearly in reading comprehension and vocabulary, particularly in lexical variety and inferential understanding, indicating areas for instructional improvement. These findings support the test's validity and reliability as a diagnostic tool, aligning with previous studies such as (Ba-Udan and Mahwar, 2020), which emphasized the importance of balanced assessments that evaluate multiple language domains to guide effective teaching

Results and Discussion

Questionnaire

Responses from 40 participants were analyzed using descriptive statistics (frequencies and percentages) to evaluate the effectiveness of the English language exam from the students' perspective. Regarding students' opinions about the English language exam, the results showed that a large proportion (80%) considered the exam difficult. This suggests a need to review the level of difficulty of the test items. However, 62.5% of the participants stated that they clearly understood the questions, indicating that the wording of the items was relatively clear to most students. In terms of exam length and suitability, 62.5% reported that the exam was appropriate in length to assess their abilities. On the other hand, 77.5% believed that the time allotted was not sufficient, suggesting that time pressure may have affected their performance.

As for the exam content, 70% of students agreed that the test covered what they had learned during the semester. Additionally, 75% felt well-prepared for the test, reflecting a relatively good alignment between the test and the curriculum. On a technical level, 70% of students were satisfied with the clarity of the font and instructions. Nevertheless, 72.5% admitted to feeling anxious during the exam, which

highlights the psychological factors that might influence test performance. Furthermore, 67.5% perceived the test as fair overall.

From an instructional perspective, 75% of students confirmed that teachers had clearly explained the instructions before the test began, showing the teacher's effective role in preparing students. However, only 57.5% felt that the questions matched their level, indicating a possible mismatch between some test items and student abilities. Regarding question types, 72.5% expressed satisfaction with the multiple-choice format, especially in grammar items. Nevertheless, only 55% felt confident answering grammar and vocabulary questions, which reflects a relative weakness in these two skills. Finally, student responses were equally split (50%) on whether the test covered all curriculum units, suggesting differences in learning experiences across schools or teachers.

Taken together, these findings highlight several challenges faced by students in the English exam. The most critical issues include test difficulty, limited time, incomplete syllabus coverage, high levels of test anxiety, and lack of confidence in core language skills. These insights align with the second research question: "What challenges do students face in the English language test for the third year of secondary school?" and emphasize the need to redesign the test to ensure it is fair, well-aligned with the curriculum, and accessible to all students.

Interview

The interview data collected from English language teachers in secondary schools of Albeida City were analyzed using descriptive statistics to explore their perceptions of the English language test. The participants showed a diverse demographic profile, with a balanced distribution across age groups (see Table 1), gender, and teaching experience, predominantly representing experienced teachers with over 10 years of service.

Table (1): The age of participants.

Age	Frequency	Percentage
28-32	3	30%
33-37	2	20%
38-42	2	20%
43-47	1	10%
48-52	2	20%

Most teachers (60%) believed that the English test accurately reflects the curriculum content and students' actual language abilities, though a notable minority expressed concerns about alignment and validity. Opinions on the test difficulty varied, with the majority rating it as moderate or easy as seen in Figure 7, while some found it difficult.

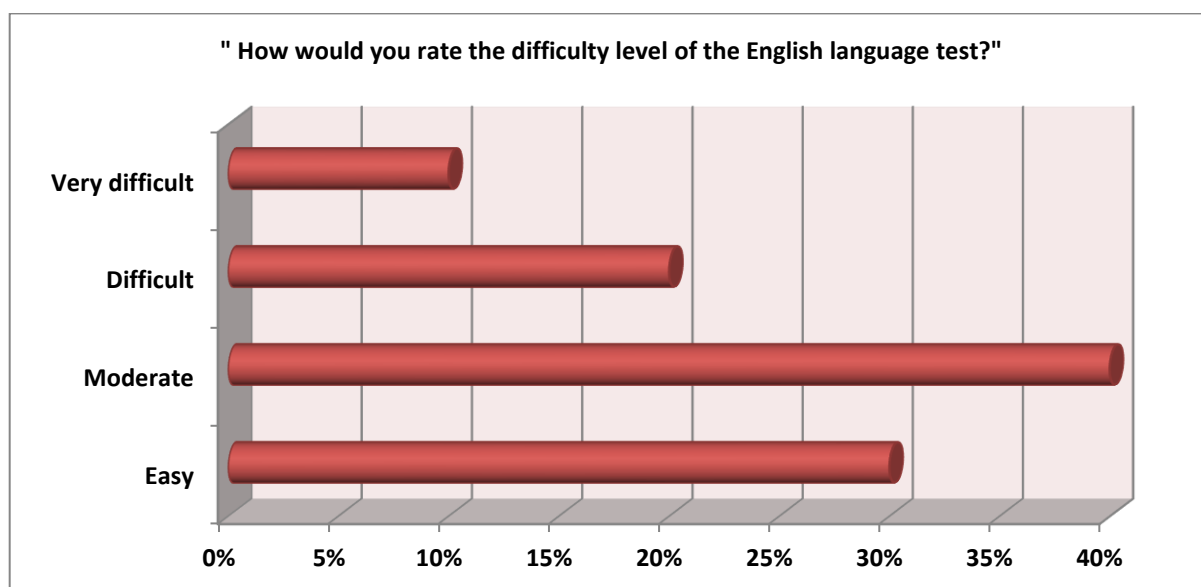


Figure (7): The responses of the participants in the study to the question "How would you rate the difficulty level of the English language test?".

Regarding the test content, half of the teachers felt the test was balanced across grammar, vocabulary, and reading comprehension, though others disagreed or were unsure, indicating room for

improvement. The majority recommended including more real-life tasks and integrating listening and speaking sections to enhance test validity. A strong consensus (80%) supported the need for long-term changes to maintain the test's relevance and reliability, with emphasis on teacher involvement and regular updates. Feedback opportunities for teachers were split equally, pointing to inconsistent participation in test design. Most participants agreed that closer alignment of test tasks with classroom activities would improve student performance (90%). However, only half of the teachers reported receiving training to prepare students fairly for the test, suggesting a need for more consistent professional development. Finally, 70% of teachers believed the test scoring system is fair and objective, though some uncertainty remains, highlighting a need for greater transparency and consistency in assessment practices.

Overall, the findings indicate that the interviews revealed a general consensus on the need to improve the English language test to make it more valid and reliable. This includes redesigning the test to cover all language skills, ensuring content reflects real-life use, simplifying language, increasing teacher involvement in test development, and providing ongoing professional development to enhance reliability in test preparation and scoring.

Conclusion

In conclusion, this study has shed light on the reality of the English language test administered to third-year secondary school students in Libya. It analyzed the results of the proficiency test and surveyed students' opinions and collected teachers' views through interviews on the most prominent challenges associated with the test, as well as its validity and reliability.

The results revealed that the current test suffers from several shortcomings related to the distribution of content, the formulation of questions, and its lack of comprehensiveness of all curriculum units. Furthermore, there are difficulties in translation and students' understanding of the nature of the questions. The study also demonstrated an urgent need to redesign the English language test to ensure its consistency with the objectives of the curriculum and its ability to measure students' actual proficiency in the areas of reading, vocabulary, and grammar.

Findings from questionnaires and interviews highlighted the importance of involving teachers in the test development process and provide them with ongoing training in the field of language assessment. Overall, the study emphasizes the significance of adopting tests that are valid and reliable, and that are consistent with modern educational standards, which will contribute to improving language education outcomes in Libya. Finally, the study presented a set of practical recommendations for developing future language tests and proposed directions for subsequent research that can build on these findings.

References

1. Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford University Press.
2. Ba-Udhan, H. S. A., & Mahwari, W.S.M. (2020). *Designing an English Language Proficiency Test for first- year University students*. International of English Language, literature and Translation Studies, 7 (1), 69-91.
3. Brown, H. D. (2004). *Language Assessment: Principles and Classroom Practices*. San Francisco State University.
4. Corder, S.P. (1973). *Error Analysis and Inter Language*. Oxford University Press.
5. Creswell, J. W. (2009). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE publications
6. Ellis, R. (2003). *Task-Based Language Learning and Teaching*. Oxford University Press.
7. Field, A. (2018). *Discovering Statistic Using IBM SPSS Statistic*. Sage.
8. Fulcher, G. (2010). *Practical Language Testing*. London.
9. Giraldo, F. (2018). *Language Assessment Literacy: Implications for Language Teachers*. Profile: Issues in Teachers' Professional Development, 20(1), 179-195
10. Harmer, J. (2002). *The Practice of English Language Teaching*. Malaysia: Person Educated Limited.
11. Hughes, A. (2003). *Testing for Language Teachers*. Cambridge University Press.
12. Kallio, H., Pietila, A. M., Johnson, M., & Kangasniemi, M. (2006). *Systematic Mythological Review: Developing a Framework for a Qualities Semi-Structured Interview Guide*. Journal of Advanced Nursing. 72 (12), 2954-2965.
13. Smith, M. K. (2018). *Assessment: The Bridge Between Teaching and Learning*. Higher Education, 23(3), 293-307.
14. Sultana, N. (2019). *Language Assessment Literacy: An Uncharted Area for the English language Teachers in Bangladesh*. Language Testing in Asia, 9(1), 2-14.
15. Weir, C. (2005). *Language Testing and Validation*. Palgrave Macmillan.