

A Comparative Study of Progress Tests and Achievement Tests in Language Assessment: Students' Performance Through the Use of MCQs and SAQs

Naser Abdulkareem Mohammed ^{1*}, Saja Najmaldeen Almoussi ², Shahed Ali Alabed ³,
Maryam Mustafa Alraash ⁴, Shahad Youssef Helmi ⁵
^{1,2,3,4,5} Department of English, Faculty of Education, University of Sabratha, Zolten, Libya
*Email (for reference researcher): naser.alkhafeel@sabu.edu.ly

دراسة مقارنة بين اختبارات التقدم واختبارات التحصيل في تقويم اللغة: أداء الطلاب من خلال استخدام الأسئلة متعددة الاختيارات (MCQs) والأسئلة ذات الإجابات القصيرة (SAQs)

ناصر عبد الكريم محمد^{1*}، سجي نجم الدين الموسي²، شهد علي العابد³، مريم مصطفى الراش⁴، شهد يوسف حلمي⁵
^{1,2,3,4,5} قسم اللغة الإنجليزية، كلية التربية، جامعة صبراتة، زلطن، ليبيا

Received: 12-05-2026; Accepted: 04-08-2026; Published: 16-08-2026

Abstract:

This research examined the effectiveness of progress and achievement tests in assessing English language learning, focusing on the use of Multiple-Choice Questions (MCQs) and Short-Answer Questions (SAQs) in evaluating students' achievement and motivation. The study also explored students' perceptions of assessment practices in EFL contexts and compared formative and summative testing approaches. A descriptive-analytical method was adopted, and data were collected through a questionnaire distributed to 70 randomly selected students from the English Department at the Faculty of Education, Zolten, Sabratha University. The data were analyzed using SPSS Version 27. The findings indicated that students generally viewed the assessment system positively. Progress tests were considered more supportive of learning due to their continuous feedback and role in reducing anxiety, whereas achievement tests were associated with greater stress despite effectively measuring learning outcomes. Results also showed slight statistical differences favoring MCQs, while SAQs were perceived as more effective in assessing critical thinking and deeper understanding.

The research recommends adopting diversified assessment strategies that combine fairness, validity, and pedagogical effectiveness in EFL assessment contexts.

The significance of this study lies in its contribution to improving assessment practices in EFL departments by providing empirical evidence regarding the strengths and limitations of various testing approaches. The findings may help educators, curriculum developers, and policymakers design more balanced, effective, and student-centered assessment systems that align with contemporary educational standards and learners' academic needs.

Keywords: Progress Tests, Achievement Tests, Language Assessment, MCQs, SAQs, EFL Assessment, Formative Assessment, Summative Assessment.

المخلص :

بحثت هذه الدراسة فاعلية اختبارات التقدم واختبارات التحصيل في تقويم تعلم اللغة الإنجليزية، مع التركيز على استخدام أسئلة الاختيار من متعدد (MCQs) والأسئلة قصيرة الإجابة (SAQs) في قياس تحصيل الطلبة ودافعيتهم للتعلم. كما سعت الدراسة إلى استكشاف تصورات الطلبة تجاه ممارسات التقويم في سياقات تعلم اللغة الإنجليزية بوصفها لغة أجنبية (EFL)، وإجراء مقارنة بين أساليب التقويم التكويني والتقويم الختامي.

اعتمدت الدراسة المنهج الوصفي التحليلي، وتم جمع البيانات باستخدام استبانة وزعت على عينة عشوائية مكونة من (70) طالباً وطالبة من قسم اللغة الإنجليزية بكلية التربية زلطن، جامعة صبراتة. وقد جرى تحليل البيانات باستخدام البرنامج الإحصائي SPSS الإصدار 27.

أظهرت النتائج أن الطلبة ينظرون بصورة إيجابية إلى نظام التقويم المعتمد. كما تبين أن اختبارات التقدم تعتبر أكثر دعماً لعملية التعلم بفضل ما توفره من تغذية راجعة مستمرة ودورها في الحد من قلق الاختبارات، في حين ارتبطت الاختبارات التحصيلية بمستويات أعلى من الضغط النفسي، رغم كفاءتها في قياس نواتج التعلم. كذلك كشفت النتائج عن وجود فروق إحصائية طفيفة لصالح أسئلة الاختيار من متعدد (MCQs)، بينما عدت الأسئلة قصيرة الإجابة (SAQs) أكثر فاعلية في تقويم مهارات التفكير الناقد والفهم العميق.

وتوصي الدراسة بتبني استراتيجيات تقويم متنوعة تجمع بين العدالة والصدق والفاعلية التربوية، بما يسهم في تطوير ممارسات التقويم في سياقات تعليم اللغة الإنجليزية بوصفها لغة أجنبية.

وتتبع أهمية هذه الدراسة من إسهامها في تحسين ممارسات التقويم في أقسام اللغة الإنجليزية من خلال تقديم أدلة تجريبية حول مزايا وحدود أنماط الاختبارات المختلفة. ومن المتوقع أن تسهم نتائجها في مساعدة أعضاء هيئة التدريس، ومطوري المناهج، وصناع السياسات التعليمية على تصميم أنظمة تقويم أكثر توازناً وفاعلية، تتمحور حول احتياجات المتعلمين، وتتسجم مع المعايير التربوية المعاصرة والمتطلبات الأكاديمية الحديثة.

الكلمات المفتاحية: اختبارات التقدم، اختبارات التحصيل، تقييم اللغة، أسئلة الاختيار من متعدد، أسئلة الإجابة القصيرة، تقييم اللغة الإنجليزية ك لغة أجنبية، التقييم التكويني (البنائي)، التقييم النهائي (الختامي).

Introduction:

Assessment is an essential element in the process of teaching and learning English as a Foreign Language (EFL), as it provides valuable information about students' progress, achievement, and language proficiency. In recent years, educational institutions have placed greater emphasis on improving assessment practices to ensure effective learning outcomes and maintain academic quality. Assessment not only measures students' academic performance but also influences teaching methods, learning strategies, classroom interaction, and students' motivation.

Among the most common assessment methods in EFL education are progress tests and achievement tests. Progress tests are formative assessments designed to monitor students' continuous learning and provide regular feedback that helps improve performance. In contrast, achievement tests are summative assessments that evaluate students' mastery of course content at the end of a semester or instructional period. Both types of assessment play significant roles in language education; however, their effectiveness in measuring students' actual abilities remains widely discussed.

Question formats also affect students' performance and perceptions. Multiple-Choice Questions (MCQs) are valued for their objectivity and ease of scoring, whereas Short-Answer Questions (SAQs) encourage deeper understanding and critical thinking. This study investigates students' perceptions of these assessment methods and their impact on motivation, fairness, anxiety, and language performance in Libyan EFL university contexts.

Research Objectives:

The main objectives of this research are:

1. To identify the general descriptive perception of English Department students regarding the assessment system implemented in the department.
2. To examine the differences in English Department students' perceptions of the effectiveness of **progress tests** and **achievement tests** in assessing the English language.
3. To investigate the role of **multiple-choice questions (MCQs)** and **short-answer questions (SAQs)** in measuring students' motivation and academic performance in both types of tests.
4. To determine the extent to which students recognize which types of tests provide greater **fairness, validity, and reliability** in evaluating English language learning.
- 5.

Significance of the Research:

This study is significant for educators, learners, and curriculum designers. For educators, it provides insights into the effectiveness of progress and achievement tests and helps in selecting appropriate testing strategies. For learners, it ensures that their performance is evaluated fairly and comprehensively. For curriculum designers, the findings of this research may contribute to the development of more balanced assessment tools that incorporate both MCQs and SAQs effectively.

Statement of the Problem:

Despite the extensive use of progress and achievement tests in education, there remains a lack of clarity regarding their comparative effectiveness, particularly when combined with different test formats such as MCQs and SAQs. Many educators rely heavily on one form of testing without sufficient evidence about which approach better reflects students' learning outcomes. This study seeks to address this gap by systematically comparing the two types of tests and their associated question formats.

Research Questions:

This research is guided by the following research questions:

1. What is the general descriptive perception of English Department students regarding the assessment system adopted in the department?
2. What are the differences in English Department students' perceptions of the effectiveness of **progress tests** and **achievement tests** in assessing English language proficiency?
3. What is the role of **multiple-choice questions (MCQs)** and **short-answer questions (SAQs)** in measuring students' motivation and performance in both types of tests?
4. Which type of test provides greater **fairness, validity, and reliability** in the process of evaluating English language students' learning?

Research Hypotheses:

Based on the problem statement, the research objectives, and the research questions, the present study proposes the following hypotheses:

1. There is a significant difference between progress tests and achievement tests in measuring students' language performance.
2. Multiple-Choice Questions (MCQs) and Short Answer Questions (SAQs) differ significantly in their effectiveness in assessing students' learning outcomes.
3. Progress tests provide more consistent and reliable results in evaluating students' learning than achievement tests.
4. Students perform better on MCQs than on SAQs in both progress and achievement tests.
5. Combining MCQs and SAQs in language assessment leads to a more comprehensive evaluation of students' learning compared to using a single question format.

Theoretical Study:

This research is grounded in several theoretical perspectives relevant to educational assessment. The following sections explore the theories explaining the research topic and define the key subjects related to language testing.

1. Constructivist Learning Theory (Piaget, 1977)

The Constructivist Learning Theory emphasizes that learners actively construct knowledge through interaction with their environment and by integrating new information with prior knowledge. According to Piaget (1977), learning is a cognitive process that involves exploration, reflection, and problem-solving rather than the passive reception of information. Learners are viewed as active participants who develop understanding through meaningful engagement with learning tasks.

In the context of educational assessment, this theory highlights the importance of employing assessment methods that measure deep understanding and higher-order thinking skills. Assessments aligned with constructivist principles focus on students' ability to explain, analyze, and apply knowledge. Therefore, assessment tools such as short-answer questions (SAQs) are considered effective, as they allow learners to express their understanding and demonstrate conceptual comprehension rather than relying solely on memorization.

2. Behaviorist Theory (Skinner, 1953)

The Behaviorist Theory, proposed by Skinner (1953), views learning as a change in observable behavior that occurs through the interaction between stimulus and response. This theory emphasizes the role of repetition, practice, and reinforcement in the learning process. According to the behaviorist perspective, learning is strengthened when correct responses are followed by positive reinforcement, such as rewards or grades.

In educational assessment, behaviorist principles support the use of objective testing methods that focus on measurable outcomes. Tests such as multiple-choice questions (MCQs) are commonly associated with this theory, as they assess learners' ability to select correct responses and provide immediate feedback. Grades and scores function as forms of reinforcement, motivating students to repeat correct behaviors and improve their academic performance.

3. Formative and Summative Assessment Theory (Black & Wiliam, 1998)

The Formative and Summative Assessment Theory (Black & Wiliam, 1998) distinguishes between two essential types of assessments: formative assessments and summative assessments. Formative assessments, such as Progress Tests, are conducted during the learning process to provide regular feedback that helps students identify areas for improvement and adjust their learning strategies. These assessments aim to monitor progress, guide instructional decisions, and encourage continuous improvement. In contrast, summative assessments, like Achievement Tests, are administered at the end of an instructional period to evaluate overall student achievement and determine whether learning objectives have been met. While formative assessments aim to support learning through continuous feedback, summative assessments assess the total knowledge of students they have acquired. Black and Wiliam highlight that formative assessments are important for improving student performance, as the feedback they provide can encourage students and help guide their future learning. This theory is key to the present study, as it sets the basis for comparing Progress Tests and Achievement Tests, while also examining how MCQs and SAQs are used in these assessments to evaluate students' language proficiency at different stages of learning.

4. Cognitive Load Theory (Sweller, 1988)

Cognitive Load Theory (Sweller, 1988) suggests that human cognitive capacity is limited, and effective learning occurs when instructional materials and assessments minimize unnecessary cognitive load while maximizing the brain's ability to process and retain information. Sweller identifies three types of cognitive load: intrinsic load (the inherent difficulty of the material), extraneous load (the unnecessary complexity introduced by poor presentation or design), and germane load (the mental effort dedicated to understanding and integrating knowledge). In the context of educational assessments, this theory emphasizes the importance of designing tests that reduce extraneous load and encourage germane load. For instance, overly complex or poorly structured assessments can overwhelm students and hinder their ability to showcase their true understanding. This theory is relevant to this study as it helps explain how MCQs and SAQs can impact students' performance in language assessments. Well-

designed Progress Tests and Achievement Tests should aim to reduce extraneous load by being clear and focused, allowing students to apply their knowledge effectively and demonstrate true proficiency.

Subjects Related to the Topic:

What do we mean by (Testing, Assessment, and Evaluation)?

In education, Testing, Assessment, and Evaluation serve as the primary mechanisms for measuring how effectively students master their assigned materials and reach the required learning goals. While these processes are often associated with assigning grades, they also function as essential tools for students to reflect on and monitor their own academic progress. Although educational professionals distinguish between these three specific terms, they all fundamentally focus on the process of determining a learner's level of knowledge on a given subject. For clarity and simplicity, the term assessment is frequently used as a broad concept to describe the overall measurement of what an individual has learned and understood throughout their studies.

Definition of Testing:

Testing is the use of structured tools to measure learners' knowledge or skills in a specific subject.

Definition of Assessment:

Assessment is the process of collecting information about students' learning to monitor progress and improve instruction.

Definition of Evaluation:

Evaluation is the process of judging learning outcomes based on assessment results.

The Definition of the Test in General

A test can be defined in several ways :

General definition : test is a systematic procedure for measuring a sample of a student's behavior or knowledge.

Cronbach's Definition: A structured way to compare the behavior of two or more people.

Thorndike's Definition: Thorndike is considered among the first to use standardized achievement tests to measure the amount of information a student can memorize and understand.

Test characteristics:

- 1- **Validity:** The ability of the test to measure exactly what it was set up to measure.
- 2- **Objectivity:** Ensuring the results are not affected by the personality or feelings of the teacher/marker.
- 3- **Reliability:** The test should give the same results if repeated under the same conditions.
- 4- **Practicality:** It refers to how easy the test is to design, administer, score, and interpret.

Objectives of the Tests:

Tests aim to achieve the following:

- 1- To determine the level of achievement of students.
- 2- To help students master or acquire information useful in their lives.
- 3- To organize the teacher's work by evaluating their own performance.
- 4- To identify the strengths and weaknesses of students.

Functions of test :

- 1- **Measuring Achievement:** To assess how much of the material students have learned and if they have met the educational goals.
- 2- **Diagnosis:** To identify specific strengths and weaknesses in a student's performance to provide targeted help.
- 3- **Motivation:** To encourage students to study, review regularly, and stay engaged with the subject matter.
- 4- **Placement:** To help place students in the correct level or program that matches their current abilities.
- 5- **Curriculum Evaluation:** To assess the overall effectiveness of the curriculum and the instructional materials used in the teaching process
- 6- **Upgrading:** Deciding if a student is ready to move to the next class or level.
- 7- **Instructional:** Helping students learn from their mistakes through feedback.
- 8- **Self-assessment for Teachers:** Measuring how successful the teacher is in performing their task.
- 9- **Prediction:** Tests serve as a guide to help students identify the specific fields or specialties where they show the most creativity and potential for success.(**Mohammed & Alsayah,2025**)

Progress Tests in Language Assessment:

Progress tests are assessments administered during a course to measure students' development and improvement in language learning over a period of time.

Achievement Tests in Language Assessment:

Achievement tests are designed to measure how successfully learners have achieved the objectives of a specific language course or syllabus.

Multiple-Choice Questions (MCQs):

Multiple-choice questions are test items in which students are required to select the correct answer from a number of given options.

Short-Answer Questions (SAQs):

Short-answer questions are questions that require learners to provide brief responses, usually in the form of a word, phrase, or short sentence.

*** Research Gap and Its Link to Research Questions and Hypotheses**

Despite the growing body of literature on language assessment, most previous studies have examined the effectiveness of question formats, particularly Multiple-Choice Questions (MCQs) and Short-Answer Questions (SAQs), within either achievement tests or progress tests independently. However, limited attention has been given to directly comparing progress tests and achievement tests in terms of their effectiveness in assessing students' language performance when the same question formats are employed.

Moreover, prior research has often focused on the impact of question format alone, without adequately considering how the type of test (progress vs. achievement) interacts with question format to influence students' performance and the overall validity and reliability of language assessment. In addition, empirical studies addressing this issue within higher education EFL contexts, particularly in Arab or similar educational settings, remain scarce.

This gap highlights the need for a systematic investigation into the comparative effectiveness of progress tests and achievement tests in assessing students' language performance through MCQs and SAQs, as well as the potential interaction effect between test type and question format.

Based on the identified research gap, the study proposes the following hypotheses:

There are statistically significant differences in students' language performance due to the type of test administered.

There are statistically significant differences in students' language performance due to question format.

There is a statistically significant interaction effect between test type and question format in assessing students' language performance.

Combining different test types with appropriate question formats results in a more valid and reliable assessment of students' language performance than using either variable independently.

Research Gap

While both progress and achievement tests have been extensively studied, limited research directly compares their effectiveness when combined with different test formats such as MCQs and SAQs. Most existing studies examine either progress or achievement tests in isolation, or focus on test item formats without considering their interaction with assessment types. This study addresses this gap by conducting a comparative analysis, thus contributing to a more comprehensive understanding of language assessment practices.

Literature review:

We present a critical review of previous studies relevant to the present research entitled "**A Comparative Study of Progress Tests and Achievement Tests in Language Assessment: Students' Performance through the Use of MCQs and SAQs.**" The review is thematically organized around the key variables of the study. It first addresses the conceptual and theoretical distinctions between Progress Tests and Achievement Tests in language assessment. It then examines empirical studies that investigated learners' performance in relation to these test types, with particular emphasis on commonly used item formats such as Multiple-Choice Questions (MCQs) and Short-Answer Questions (SAQs). The chapter aims to position the current research within the broader body of assessment literature and to identify research gaps that justify undertaking the present comparative study.

First Section: Foreign Studies on Progress Tests and Achievement Tests

Hughes (2003), "*Testing for Language Teachers*", this study is considered one of the foundational references in the field of language assessment. In his book "*Testing for Language Teachers*", Hughes provides a comprehensive theoretical framework for understanding the nature, purposes, and types of language tests, as well as the essential principles underlying effective test design. The study places particular emphasis on the relationship between test objectives and question types, and how these influence the accuracy of measuring learners' actual language proficiency.

Hughes adopted a theoretical analytical approach, relying on an extensive review of educational literature and practical testing models. Through this approach, he examined different forms of assessment, including Multiple-Choice Questions (MCQs) and Short-Answer Questions (SAQs), highlighting their respective strengths and limitations in terms of reliability, validity, objectivity, and their ability to assess both surface-level knowledge and deeper language competence.

The findings of the study indicate that the quality of a language test is not determined solely by its content, but also by the types of questions it employs and the way they are constructed. Hughes argues that MCQs are highly objective and efficient in terms of scoring and administration; however, they may be limited in their ability to measure higher-order thinking skills and productive language use. In contrast, SAQs are more effective in assessing genuine understanding and learners' ability to produce language, although they may require more time for grading and may introduce a degree of subjectivity in evaluation.

The significance of this study lies in its contribution to establishing clear scientific principles for language test design, as well as clarifying the relationship between test formats and the accuracy of language assessment. It has served as a key reference for many subsequent studies in the field of language testing and evaluation.

Brown (2004), "Language Assessment: Principles and Classroom Practices", this work is widely recognized as a fundamental contribution to the field of language assessment. In his book *"Language Assessment: Principles and Classroom Practices"*, Brown provides a comprehensive and practical framework for understanding how language assessment operates within classroom contexts. His discussion focuses on the principles of effective assessment, the characteristics of well-designed tests, and the role of assessment in enhancing teaching and learning processes.

Brown adopts a descriptive analytical approach, drawing on both theoretical foundations and practical classroom applications. He systematically examines key concepts such as validity, reliability, authenticity, practicality, and washback, emphasizing that effective language assessment must balance these essential principles. The study also explores different types of assessments, including formative (ongoing) and summative (final) assessments, highlighting their respective roles in monitoring student progress and evaluating overall achievement.

One of the central findings of Brown's work is that assessment should not be viewed merely as a tool for measuring learning outcomes, but rather as an integral part of the learning process itself. He argues that well-designed assessments can positively influence students' motivation and learning strategies, while poorly constructed tests may lead to negative washback effects, such as increased anxiety or superficial learning. Additionally, Brown emphasizes the importance of aligning test content and question types with instructional objectives to ensure accurate measurement of students' language abilities.

The study also highlights the importance of using a variety of assessment techniques, including objective formats such as Multiple-Choice Questions (MCQs) and more open-ended formats such as Short-Answer Questions (SAQs) and performance-based tasks. Brown suggests that relying on a single type of question may limit the scope of assessment, whereas combining different formats allows for a more comprehensive evaluation of both receptive and productive language skills.

The significance of Brown's study lies in its practical orientation and its ability to bridge the gap between theory and classroom practice. It provides language teachers with clear guidelines for designing effective assessments that are fair, valid, and supportive of learning. Moreover, it underscores the role of assessment in shaping students' learning experiences and outcomes.

This study is closely related to the current research, as it directly addresses key issues explored in the present study, such as the effectiveness of different test types and the impact of question formats on students' performance and motivation. The current study draws on Brown's theoretical framework to interpret the differences between progress tests and achievement tests, as well as to analyze the roles of MCQs and SAQs in language assessment. Furthermore, Brown's emphasis on validity, fairness, and washback has informed the design of the research instrument and the interpretation of findings related to students' perceptions of assessment practices.

In conclusion, Brown (2004) provides a strong theoretical and practical foundation for the current study, supporting a deeper understanding of how diverse assessment methods can be used to achieve more accurate, fair, and meaningful evaluation of students' language proficiency.

Bachman and Palmer (1996), "Language Testing in Practice", this work is regarded as one of the most influential contributions to the field of language testing. In their book *"Language Testing in Practice"*, the authors present a comprehensive and systematic framework for designing and evaluating language tests, with a strong emphasis on ensuring that assessment practices are both theoretically sound and practically applicable. Their work focuses on the concept of "test usefulness," which integrates several key qualities required for effective language assessment.

Bachman and Palmer adopt a theoretical and analytical approach, drawing on linguistic theory, psychometrics, and applied language testing research. They introduce a multidimensional model of test usefulness that includes essential components such as reliability, validity, authenticity, instructiveness, impact, and practicality. This model highlights that no single aspect alone is sufficient; rather, effective language tests must achieve a balance among all these qualities to ensure accurate and meaningful measurement of learners' language abilities.

One of the major findings of the study is that language tests should be designed to reflect real-life language use as closely as possible. The authors stress the importance of authenticity in test tasks, arguing that assessment should engage learners in meaningful language use rather than isolated or decontextualized knowledge. In addition, they emphasize instructiveness, which refers to the extent to which test tasks involve the test-taker's language ability, background knowledge, and cognitive processes.

The study also highlights the critical role of test impact (washback), explaining that assessment practices can significantly influence both teaching methods and students' learning behaviors. Positive washback occurs when tests encourage meaningful learning and skill development, whereas negative washback may lead to rote memorization or increased anxiety. Furthermore, the authors underscore the importance of practicality, noting that test design must consider constraints such as time, resources, and scoring procedures.

The significance of this study lies in its comprehensive framework, which has become a cornerstone in modern language testing theory and practice. It provides clear criteria for evaluating the effectiveness of language tests and offers practical guidance for constructing assessments that are not only reliable and valid but also meaningful and contextually appropriate.

This study is highly relevant to the current research, as it directly informs the evaluation of different test types and question formats. The present study draws on Bachman and Palmer's model of test usefulness to analyze the effectiveness of progress tests and achievement tests, as well as the roles of MCQs and SAQs in assessing students' language performance. Their emphasis on authenticity and instructiveness is particularly useful in interpreting the differences between objective and open-ended question formats.

Moreover, the current study benefits from their concept of test impact in explaining students' perceptions of anxiety, motivation, and performance in different testing situations. The principles of reliability and practicality also support the interpretation of findings related to students' preferences for MCQs, while the concept of validity aligns with the study's focus on the ability of SAQs to measure deeper understanding.

In conclusion, Bachman and Palmer (1996) provide a robust theoretical foundation that supports the current study in analyzing language assessment practices from multiple dimensions, contributing to a more comprehensive understanding of how different testing methods and question types affect the accurate evaluation of students' language proficiency.

Green (2014), "Exploring Language Assessment and Testing", he, in his study, investigates the principles, practices, and implications of language assessment in both classroom and institutional contexts. The study aims to provide an in-depth understanding of how different assessment methods such as formative and summative tests, multiple-choice questions, and performance-based tasks affect students' learning outcomes, motivation, and engagement in language learning.

The research adopts a descriptive-analytical methodology, combining literature review with empirical observations of assessment practices in various educational settings. Green examines both theoretical frameworks and practical implementations, analyzing how assessment design, test formats, and scoring procedures influence students' performance and perceptions of fairness, validity, and reliability.

Key findings indicate that the effectiveness of assessment methods depends largely on their alignment with learning objectives and the cognitive processes they target. For instance, tests emphasizing real communicative ability and critical thinking, rather than rote memorization, result in higher student engagement and deeper learning. The study also highlights that multiple-choice tests are efficient for evaluating factual knowledge and language accuracy, whereas performance-based assessments and constructed-response tasks (e.g., short-answer questions) better capture students' reasoning, problem-solving skills, and practical language use.

The significance of Green's study lies in its emphasis on balancing assessment methods to ensure a comprehensive evaluation of language competence. It stresses the importance of reliability, validity, and fairness in testing, while also acknowledging the affective factors such as anxiety, motivation, and test-taking strategies that influence students' performance.

Regarding its relevance to the current study, Green's research provides a strong theoretical and practical foundation for investigating students' perceptions of test types, including multiple-choice and short-answer questions. It supports the current research focus by emphasizing that both test formats serve complementary roles: multiple-choice questions facilitate efficiency and objectivity, while short-answer questions allow for the assessment of deeper understanding and critical thinking. Additionally, Green's insights on fairness, reliability, and validity directly align with the current study's examination of students' perceptions of assessment quality in the Department of English.

Overall, Green (2014) underscores the necessity of a balanced and well-designed assessment system in language teaching, which informs the current research's exploration of student attitudes, motivation, and performance strategies in response to different test types.

Second Section: Arab Studies on Progress and Achievement Tests

Abdel Rahman (2015), "Educational Measurement and Evaluation", the study by Abdel Rahman, presented in the book *Educational Measurement and Evaluation*, is considered an important contribution to the field of educational assessment. It provides a comprehensive overview of the fundamental principles of measurement and evaluation in educational settings, with a particular focus on the design and implementation of effective assessment tools. The study emphasizes the role of testing in accurately measuring students' learning outcomes and improving the quality of the educational process.

Abdel Rahman adopts a descriptive and analytical approach, relying on established theories of educational measurement and practical applications in classroom assessment. The study explores key concepts such as validity,

reliability, objectivity, and fairness, highlighting their importance in constructing sound and effective tests. It also discusses various types of educational assessments, including formative and summative tests, and explains how each type serves different purposes in evaluating student learning.

One of the key findings of the study is that the effectiveness of any test depends largely on the quality of its design and its alignment with instructional objectives. Abdel Rahman stresses that poorly designed tests may fail to measure what they are intended to assess, leading to inaccurate or misleading results. The study also highlights the importance of using diverse question formats, such as objective questions (e.g., MCQs) and subjective questions (e.g., short-answer or essay questions), to ensure a more comprehensive evaluation of students' knowledge and skills.

Furthermore, the study points out that objective questions, such as MCQs, are characterized by ease of scoring and a high degree of reliability and objectivity, whereas subjective questions allow for deeper assessment of understanding, critical thinking, and expressive ability, despite potential challenges related to scoring consistency. Abdel Rahman emphasizes the necessity of balancing these question types to achieve both accuracy and depth in assessment.

The significance of this study lies in its practical contribution to improving assessment practices in educational contexts. It provides educators with clear guidelines for designing valid and reliable tests and underscores the importance of fairness and transparency in the evaluation process. Additionally, it highlights the need to consider students' individual differences and psychological factors when constructing assessment tools.

This study is closely related to the current research, as it directly addresses issues of test design, question types, and the effectiveness of different assessment methods. The present study benefits from Abdel Rahman's theoretical insights in analyzing the differences between progress tests and achievement tests, as well as in evaluating the roles of MCQs and SAQs in measuring students' performance.

Moreover, the study's emphasis on validity, reliability, and fairness supports the interpretation of the current research findings regarding students' perceptions of assessment quality. Its discussion of the strengths and limitations of different question formats aligns closely with the current study's investigation into the comparative effectiveness of MCQs and SAQs.

In conclusion, Abdel Rahman (2015) provides a solid theoretical and practical foundation for the current study, contributing to a deeper understanding of how well-designed assessment tools can enhance the accuracy, fairness, and effectiveness of evaluating students' learning outcomes.

Al-Zahrani (2018), *"The Effectiveness of Formative Tests in Improving English Language Achievement"*, this study investigates the role of formative assessment in enhancing students' performance in learning English. The study focuses on how continuous testing, feedback, and ongoing evaluation can contribute to improving students' academic achievement and overall language proficiency.

Al-Zahrani adopts an experimental research approach, which distinguishes this study from purely theoretical works. The researcher divided participants into experimental and control groups to examine the impact of formative tests on students' achievement. The experimental group was exposed to regular formative assessments accompanied by continuous feedback, while the control group followed traditional assessment methods without systematic formative testing. Data were collected and analyzed using appropriate statistical techniques to measure differences in performance between the two groups.

The findings of the study revealed that formative tests have a significant positive effect on students' English language achievement. Students who were exposed to continuous assessment demonstrated noticeable improvement in their performance compared to those in the control group. The study also found that formative testing helps students identify their strengths and weaknesses, enhances their motivation, and reduces test anxiety by familiarizing them with assessment formats and expectations.

Furthermore, the study emphasizes that formative assessment promotes active learning by encouraging students to engage more consistently with course material. The provision of immediate and constructive feedback was identified as a key factor contributing to improved learning outcomes. In addition, formative tests were found to support better retention of knowledge and more effective development of language skills over time.

The significance of this study lies in its empirical evidence supporting the effectiveness of formative assessment practices in language learning contexts. It highlights the importance of integrating continuous assessment into teaching strategies to enhance both student performance and engagement. The study also contributes to the growing body of research advocating for a shift from traditional, high-stakes testing toward more learner-centered assessment approaches.

This study is highly relevant to the current research, as it directly relates to the concept of progress tests, which represent a form of formative assessment. The current study draws on Al-Zahrani's findings to interpret the effectiveness of progress tests in improving students' learning and reducing anxiety. The positive impact of feedback and continuous evaluation observed in Al-Zahrani's study aligns with the current study's results, which indicate that students perceive progress tests as more supportive of their learning process.

Moreover, the study supports the current research in explaining differences between progress tests and achievement tests, particularly in terms of their impact on motivation and performance. While achievement tests

measure final outcomes, formative tests—such as progress tests—play a crucial role in guiding learning and improving performance over time.

In conclusion, Al-Zahrani (2018) provides strong empirical support for the use of formative assessment in language education, offering valuable insights that reinforce the findings of the current study and contribute to a deeper understanding of effective assessment practices in English language learning.

Hassan (2020), "*Assessment Practices in Teaching English as a Foreign Language*", it investigates the strategies and methods used by English language instructors in assessing students' learning in EFL contexts. The research focuses on identifying the prevailing assessment practices, examining their effectiveness, and exploring how these practices influence students' learning outcomes and motivation.

The study employs a descriptive analytical methodology. Data were collected from a sample of EFL teachers and students using questionnaires and structured interviews to gain insight into the assessment practices commonly applied in classrooms. Quantitative and qualitative analyses were conducted to evaluate trends, effectiveness, and teachers' perceptions of various assessment methods, including progress tests, achievement tests, and alternative assessment tools.

The findings reveal that teachers predominantly rely on traditional assessment methods, such as achievement tests and multiple-choice questions, for grading and monitoring student progress. However, the study also highlights the emerging use of formative assessments, such as ongoing quizzes, self-assessment, and peer-assessment, which provide continuous feedback and support student learning. Hassan's results indicate that formative practices help students better understand their strengths and weaknesses, enhance motivation, and improve language proficiency, while traditional summative tests primarily evaluate final learning outcomes.

The study further emphasizes that the effectiveness of assessment practices is strongly linked to their alignment with learning objectives and the provision of timely feedback. Teachers reported that students respond positively to assessment methods that are perceived as fair, comprehensive, and reflective of real language use, rather than focusing solely on rote memorization or grammar knowledge.

The significance of Hassan's study lies in its practical contribution to improving assessment strategies in EFL classrooms. It underscores the importance of balancing formative and summative assessment methods to achieve more effective learning outcomes and increase student engagement and motivation. The study also provides empirical support for the adoption of diversified assessment practices that cater to students' varied learning needs and abilities.

In relation to the current study, Hassan's research provides a valuable framework for analyzing students' perceptions of assessment methods, particularly in distinguishing between progress tests and achievement tests. The findings support the current research by confirming that formative assessments, like progress tests, not only improve learning outcomes but also reduce anxiety and increase students' motivation. Additionally, the study emphasizes the need for fair, valid, and reliable assessment methods, which directly relates to the current research's focus on evaluating the fairness and effectiveness of multiple-choice and short-answer questions.

Overall, Hassan (2020) reinforces the premise that effective assessment practices in EFL contexts require a combination of continuous evaluation and summative measurement. These insights are critical for interpreting students' perceptions of test types, motivation, and performance strategies, as investigated in the current study.

Mohammed & Alsayah, (2025). "*The role of classroom tests in the assessment of students' learning: Intended objectives and potential defects*", the present study seeks to identify and clarify the role of classroom tests in assessing students' learning, with particular emphasis on their intended purposes and the shortcomings they may involve. Classroom tests are widely used as tools to measure students' comprehension of course content, offer feedback to both students and teachers, and guide instructional planning. Nevertheless, despite their frequent application, such tests may exhibit several limitations, including bias, misalignment with learning objectives, and undesirable washback effects.

This study employs a descriptive approach and relies on a review of relevant literature to explore how classroom tests operate within the broader context of educational assessment. It underscores the necessity of constructing tests that are valid, reliable, and equitable, ensuring that they accurately represent students' actual learning outcomes, while also acknowledging and addressing the inherent weaknesses of conventional testing practices.

Summary of literature review:

We have reviewed a selection of foreign and Arab studies addressing Progress Tests and Achievement Tests in language assessment. The reviewed literature indicates that although Achievement Tests continue to dominate educational practice, Progress Tests play a vital role in enhancing learning and monitoring students' development. Nevertheless, relatively few studies have combined a comparison of these test types with an examination of item formats such as MCQs and SAQs. Accordingly, the present study seeks to bridge this gap by offering a comprehensive comparative analysis of students' performance across Progress and Achievement Tests using different question formats.

Research methodology:

1 Introduction

This chapter presents the methodology employed in conducting the study. It outlines the research design, population and sample, research instruments, procedures for data collection, and methods of data analysis. The chosen methodology ensures that the research objectives are achieved and that the findings are valid and reliable. Since the aim of this study is to measure and evaluate the effectiveness of assessment tools (test questions) and to determine which assessment formats **Multiple Choice Questions (MCQs)** or **Short Answer Questions (SAQs)** more accurately reflect students' actual proficiency from the perspective of English Department students, as well as to identify the extent of variation in students' performance when their language skills are assessed using two different testing formats, the researchers adopted the **descriptive-analytical approach**, as it is considered the most appropriate method for achieving the objectives of the study and addressing its nature.

The study relied on two main types of data:

1. Primary Data

a. Questionnaire:

Primary data were collected from the study population through a questionnaire, which served as the principal instrument for data collection. The collected data were subsequently coded, tabulated, and analyzed using the **Statistical Package for the Social Sciences (SPSS), version 27**, in order to obtain statistical indicators that support the objectives of the study.

b. Personal Interviews:

The researchers also conducted several personal interviews with members of the study sample from the **Department of English, Faculty of Education, Zolten**, with the aim of clarifying the objectives of the research and obtaining additional information that would contribute to enriching the study.

2. Secondary Data

Secondary data consisted of information obtained from books, reports, academic journals, and previous research studies related to the effectiveness of assessment tools and the improvement of test quality. These sources were utilized to provide a solid scientific foundation and to enrich the theoretical background of the research.

3. Research Population

The research population consists of all students enrolled in the **Department of English, Faculty of Education, Zolten, Sabratha University**, during the academic year **2025–2026**.

4. Research Sample

In this study, the researchers adopted a **random sampling technique**. The sample consisted of **70 students** selected from the target population. Questionnaires were distributed to all selected participants, and **70 completed questionnaires** were retrieved and subsequently subjected to **statistical analysis**.

The following tables present the **general characteristics of the research sample**

Table (1): Distribution of the Sample Members by Gender

Gender	Frequency	Percentages %
Male	5	7.1%
Female	65	92.9%
Total	70	100

It is evident from **Table (1)** above that **male participants** represent **7.1%** of the study sample, whereas **female participants** account for **92.9%**. This indicates that the proportion of females is considerably higher than that of males within the total sample.

The following **figure** illustrates the **percentage distribution of the sample according to gender**.

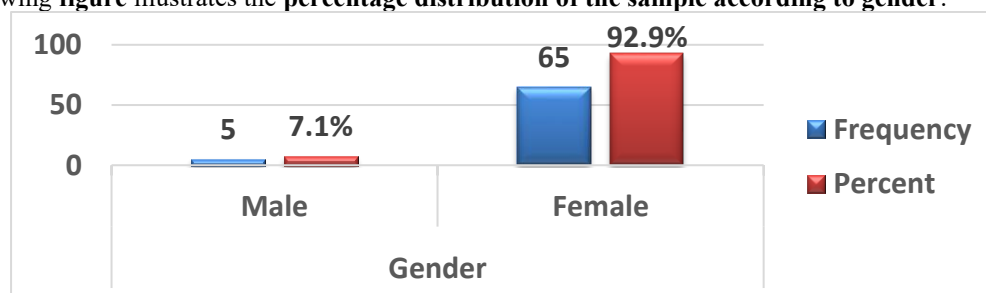


Figure (1): Percentages of Sample Members by Gender

Table (2): Distribution of the Sample Members by semester

Semester	Frequency	Percentages %
3 rd semester	17	24.28%
4 th semester	7	10.00%
5 th semester	23	32.86%
7 th semester	23	32.86%
Total	70	100

Table (2) above exposes that **32.86%** of the sample participants, according to their academic semester, are enrolled in both **the fifth and seventh semesters**, with each representing the highest proportion of the sample. This is followed by students in **the third semester**, who constitute **24.28%** of the sample. Meanwhile, **10%** of the participants are enrolled in **the fourth semester**.

This variation in the academic semesters of the respondents indicates that the study includes students from **different academic semesters**, which enhances the comprehensiveness of the research and strengthens the possibility of accurately assessing students' performance in the targeted language tests.

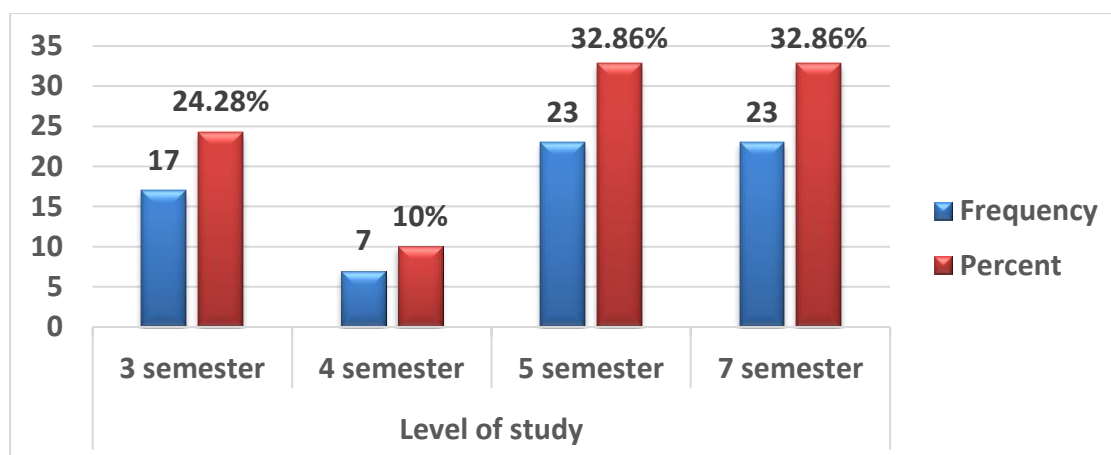


Figure (2): Percentages of Sample Members by Semester

Table (3): Distribution of the Sample Members by (Have you ever taken progress tests in English?)

Have you ever taken progress tests in English?	Frequency	Percentages %
Yes	54	77.1%
No	16	22.9%
Total	70	100

Table (3) above shows that **77.1%** of the sample participants have **previous experience in taking English proficiency (progress) tests**, which represents a relatively high proportion approximately **three quarters of the sample**. In contrast, **22.9%** of the students reported that they **had not previously taken such tests**.

This finding suggests that the majority of students in the department possess **a good level of familiarity with the nature and variety of English progress tests**, reflecting the practice of **continuous assessment** and the accurate monitoring of **ongoing language development** among students in the department.

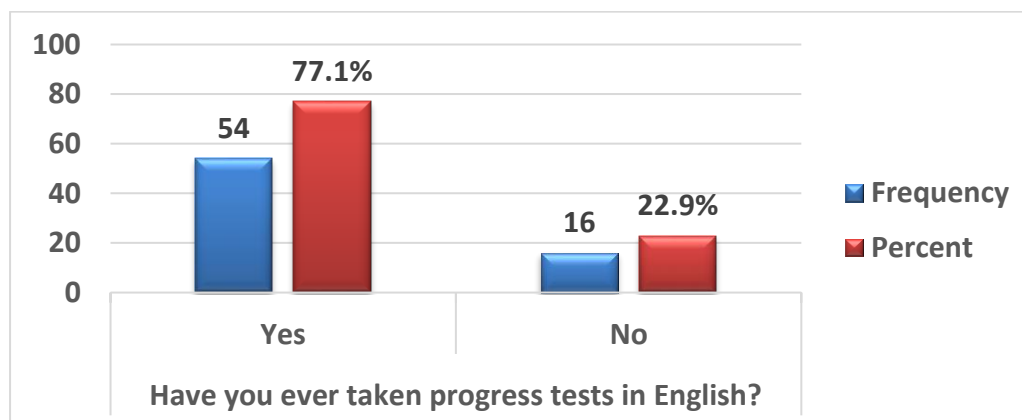


Figure (3): Percentages of Sample Members by Question

Table (4): Distribution of the Sample Members by (Have you ever taken achievement tests in English?)

Have you ever taken achievement tests in English?	Frequency	Percentage %
Yes	45	64.3%
No	25	35.7%
Total	70	100

Table (4) above shows that **64.3%** of the sample participants have **previous experience in taking English achievement tests**, which represents a relatively high proportion. This indicates that **more than half of the respondents are familiar with the nature and variety of achievement tests**. In contrast, **35.7%** of the students reported that they **had not previously taken achievement tests**.

This finding may reflect the influence of **assessment practices and measurement accuracy** on students' performance in achievement tests.

The following **figure** illustrates the **percentage distribution of the sample according to their participation in achievement tests**.

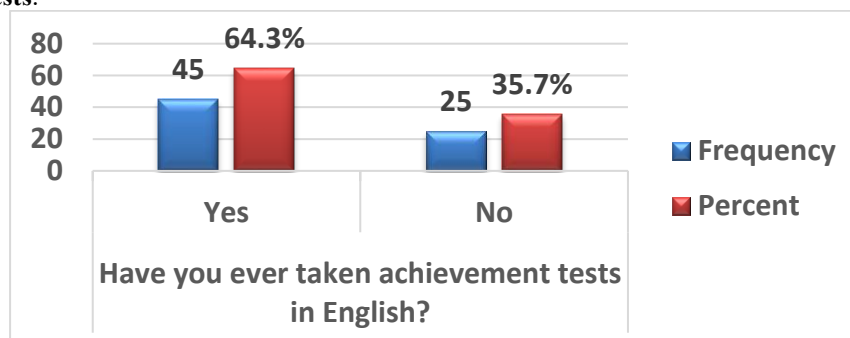


Figure (4): Percentages of Sample Members by Question

5. Research Instrument

The researchers designed a **questionnaire** as the primary data collection instrument, with the aim of obtaining the **evaluations and opinions of students from the Department of English**. The questionnaire was developed to facilitate a comparison between **progress tests and achievement tests in language assessment**, and to measure and analyze students' performance through the use of **Multiple Choice Questions (MCQs)** and **Short Answer Questions (SAQs)** from the perspective of the department's students.

The questionnaire consisted of **two main sections**, as outlined below:

Section One:

This section included **four questions** intended to identify certain **personal characteristics of the research sample**. It covered the following variables: **gender, academic semester, previous experience with progress tests, and previous experience with achievement tests**.

Section Two:

This section consisted of statements designed to measure the **research variables** and to identify the **opinions and evaluations of the respondents regarding progress tests and achievement tests** among students in the Department of English. The questionnaire included **35 statements** aimed at assessing the **effectiveness of progress and achievement tests in language evaluation**.

After the preparation of the research instrument (the questionnaire), it was presented to **two expert reviewers from the department**, in addition to the **research supervisor**, for the purpose of evaluating its **content validity**. The reviewers suggested several modifications to some of the questionnaire items, which were subsequently revised by the researchers when preparing the **final version of the questionnaire**.

Once the questionnaire had been verified for **linguistic clarity and accuracy**, the researchers structured the responses according to a **five-point Likert scale**, as follows: (**Strongly Agree – Agree – Neutral – Disagree – Strongly Disagree**).

This scale was selected in accordance with the **nature of the study and the characteristics of the target population**. It should be noted that the **five-point Likert scale is an ordinal scale**, to which the following weights were assigned, as illustrated in the table below.

Table (5): The Five-Point Likert Scale

Level of Availability				
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

Subsequently, the **weighted mean** was calculated in order to determine the **degree of evaluation among the responses of the research sample**, as illustrated in the following table.

Table (6): Weighted Arithmetic Mean

Weighted Mean	Scale	Level of Performance
From 1 to 1.79	Strongly Disagree	Very Low
From 1.80 to 2.59	Disagree	Low
From 2.60 to 3.39	Neutral	Moderate
From 3.40 to 4.19	Agree	High
From 4.20 to 5	Strongly Agree	Very High

6. Validity and Reliability of the Research Instrument (Questionnaire)

Data collection is considered one of the **most important and challenging stages of scientific research**. Selecting an appropriate data collection instrument is therefore a highly precise and critical task. It requires choosing an instrument that demonstrates both **validity and reliability**, since a weak research instrument one lacking validity and reliability may lead to **inaccurate findings**, making the results unreliable for interpretation. Accordingly, researchers must ensure that the research instrument possesses adequate levels of **validity and reliability**. From this perspective, it is essential to examine the validity and reliability of the questionnaire used in this study.

First: Validity

The **validity of a questionnaire** refers to the extent to which its items measure what they are intended to measure, that is, the degree of accuracy with which the instrument captures the purpose for which it was designed. To verify the validity of the questionnaire, the researchers employed two methods:

1. Face Validity (Expert Judgment):

The preliminary version of the questionnaire was presented to a group of **expert reviewers with academic experience and interest in the research topic**. The purpose was to assess the suitability of each item in relation to the dimension to which it belongs within the questionnaire. The reviewers were requested to provide their opinions and comments regarding each item. Based on their feedback, several observations were made concerning some questionnaire items. The researchers carefully considered these comments and incorporated the suggested revisions when preparing the **final version of the questionnaire**. (See Appendix A).

2. Validity Coefficient:

Construct validity refers to the degree of **consistency between the questionnaire items and the dimensions they are intended to measure**, ensuring that each item measures only the concept it was designed to assess. To verify this, the researchers calculated the **validity coefficient for each item and for the overall dimension** by computing the **square root of Cronbach's Alpha coefficient**.

Second: Reliability

The primary objective of assessing the **reliability of the questionnaire** is to ensure that the instrument is suitable for the purposes of the study and possesses stability. Reliability implies that the questionnaire would yield **consistent results if it were administered repeatedly under the same conditions to the same sample**. In other words, reliability reflects the **stability and consistency of results over time**, thereby enabling researchers to obtain **accurate and dependable findings** that can be confidently used in the study.

Cronbach's Alpha Coefficient:

Cronbach's Alpha is one of the most widely used statistical methods for measuring the **internal consistency reliability of a questionnaire**. The value of Cronbach's Alpha ranges between **0 and 1**, and the closer the value is to **1**, the higher the reliability of the instrument. According to **Sekaran (2003)**, a Cronbach's Alpha value of **0.60 or higher** is considered acceptable, and values approaching **1.00** indicate a high level of reliability.

In this study, the researchers used **Cronbach's Alpha** to measure both the **reliability coefficient of the questionnaire and the validity coefficient**. The obtained results are presented as follows.

Table (7): Reliability and Validity Coefficients for Questionnaire

Parts	No. of Items	Cronbach's Alpha (Reliability)	Validity Coefficient
General Perceptions of Language Tests	6	0.980	0.989
Attitudes toward Progress Tests	6	0.986	0.993
Attitudes toward Achievement Tests	6	0.991	0.995
Perceptions of Question Types (MCQs vs SAQs)	6	0.990	0.995
Self-Reflection, Motivation and Performance Strategies	5	0.983	0.991
Perceived Fairness and Validity of Tests	6	0.984	0.992
Questionnaire	35	0.998	0.999

It is evident from **Table (7)** above that all **Cronbach's Alpha coefficients** for the questionnaire dimensions are high, ranging between **0.980 and 0.991**, which indicates a **high and acceptable level of reliability**. Furthermore, the **Cronbach's Alpha coefficient for the entire questionnaire** reached **0.998**, which represents a very high value. In addition, the **validity coefficient** was found to be **0.999**, which is also a high value. These results indicate that the **questionnaire demonstrates a high level of reliability and validity**, and therefore it can be confidently relied upon for the **analysis and interpretation of the research findings**.

7. Statistical Methods Used in the Study

After collecting the questionnaire responses, the researchers coded and processed the data using **statistical techniques appropriate to the nature of the study**, as follows:

1. **Frequencies and percentages** were used to describe the personal characteristics of the research sample.
2. **Arithmetic mean, standard deviation, relative weight, importance, and ranking** were calculated for each questionnaire item as well as for the questionnaire as a whole.
3. **Cronbach's Alpha coefficient** was used to determine the **reliability and validity** of the questionnaire and each of its dimensions.
4. **T-test** was employed to test the statistical significance of differences between variables.

Results and Discussion (General statistic):

1 Statistical Analysis of the Study parts (Questionnaire)

After collecting the study data using the research instrument (the questionnaire), the data were **carefully reviewed** prior to entry into the computer for **statistical analysis**. This analysis aimed to **extract results and achieve the objectives of the study** by analyzing the collected data. The statistical procedures included calculating **arithmetic means, standard deviations, and the percentage distribution of responses** for each questionnaire item.

2 Results Related to the Study Questions

2.1 Results Concerning the First Question:

What is the general descriptive perception of English Department students toward the assessment system applied in the department?

The following tables present the **participants' opinions** regarding the items of the first dimension, which reflect the **students' perceptions of the nature of language tests** from their perspective. The data were classified according to **Part One**.

Part One: General Perceptions of Language Tests

In this section, each item of the first dimension (General Perceptions) was studied individually. **Table (8)** displays the **frequency distribution (frequency and percentage)** of participants' responses to each item within this dimension. **Table (9)** shows the **statistical analysis, including arithmetic mean, standard deviation, relative importance, and predominant trend**, of participants' responses for the same part.

The obtained results are summarized as follows:

Table (8): Frequencies and Percentages of the First part Items

X1	Items		Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1	Language tests accurately measure my English proficiency.	N	13	21	21	9	6
		%	18.6	30	30	12.9	8.6
2	I think tests should include practical language use, not only grammar	N	29	21	8	3	9
		%	41.4	30	11.4	4.3	12.9
3	The level of difficulty in tests is generally appropriate.	N	14	10	28	12	6
		%	20	14.3	40	17.1	8.6
4	I believe both test types (Progress & Achievement) complement each other.	N	12	31	17	5	5
		%	17.1	44.3	24.3	7.1	7.1
5	I get anxious when facing long descriptive questions.	N	32	18	8	5	7
		%	45.7	25.7	11.4	7.1	10
6	My performance improves when I know the test format in advance.	N	42	16	5	2	5
		%	60	22.9	7.1	2.9	7.1

Table (9): Weighted Mean, Standard Deviation, and Level for the Items of the First part

X1	Items	Weighted Mean	Standard Deviation	Level	Relative Importance %.	Rank
1	Language tests accurately measure my English proficiency.	3.37	1.12	Moderate	67.4	5
2	I think tests should include practical language use, not only grammar	3.83	1.14	High	76.6	3
3	The level of difficulty in tests is generally appropriate.	3.20	1.11	Moderate	64	6
4	I believe both test types (Progress & Achievement) complement each other.	3.57	1.04	High	71.4	4
5	I get anxious when facing long descriptive questions.	3.90	1.20	High	78	2
6	My performance improves when I know the test format in advance.	4.26	1.15	Very High	85.2	1
	General Perceptions of Language Tests	3.69	1.03	High	73.8	

From the above tables, the following can be observed:

- Language tests accurately measure my English proficiency:**
According to Table (8), 30% of the participants rated this item as "Agree," while another 30% rated it as "Neutral." Table (9) shows that the **weighted mean** for this item is **3.37**, with a **standard deviation** of **1.12**, indicating that the prevailing response trend is **Neutral**, with a **relative importance** of **67.4%**.
- I believe that tests should include practical use of the language, not only grammar:**
According to Table (8), 41.4% of participants rated this item as "Strongly Agree," and 30% as "Agree." Table (9) shows a **weighted mean** of **3.83**, with a **standard deviation** of **1.14**, indicating a prevailing response trend of **High**, with a **relative importance** of **76.6%**.
- The difficulty level of tests is generally appropriate:**
Table (8) shows that 40% of participants rated this item as "Neutral," and 20% as "Agree." Table (9) indicates a **weighted mean** of **3.20**, with a **standard deviation** of **1.11**, reflecting a prevailing response trend of **Neutral**, with a **relative importance** of **64%**.
- I believe that both types of tests (progress and achievement) complement each other:**
According to Table (8), 44.3% of participants rated this item as "Agree," and 24.3% as "Neutral." Table (9) shows a **weighted mean** of **3.57**, with a **standard deviation** of **1.04**, indicating a prevailing response trend of **High**, with a **relative importance** of **71.4%**.
- I feel anxious when facing long descriptive questions:**
Table (8) shows 45.7% of participants rated this item as "Strongly Agree," and 25.7% as "Agree." Table (9) indicates a **weighted mean** of **3.90**, with a **standard deviation** of **1.20**, showing a prevailing response trend of **High**, with a **relative importance** of **78%**.
- My performance improves when I know the test format in advance:**
Table (8) shows 60% of participants rated this item as "Strongly Agree," and 22.9% as "Agree." Table (9) indicates a **weighted mean** of **4.26**, with a **standard deviation** of **1.15**, reflecting a prevailing response trend of **Very High**, with a **relative importance** of **85.2%**.

Additionally, the overall result for **Part One (General Perceptions of Language Tests)** is "**Agree**", with an **arithmetic mean** of **3.69**, which is above the acceptable mean of 3, a **standard deviation** of **1.03**, and a **relative importance** of **73.8%**.

This indicates that the participants expressed a **high level of evaluation** regarding general perceptions of language tests. A high evaluation reflects **strong satisfaction** among students with the nature of the assessment tests and measurement procedures applied in the department, indicating a **positive general perception of the assessment and evaluation system**. This, in turn, provides **educational legitimacy** for the assessment and evaluation methods used in the Department of English.

2.2 Results Related to the Second Question:

What is the nature of the differences in English Department students' perceptions regarding the effectiveness of progress tests and achievement tests in English language assessment?

The following tables present the participants' opinions regarding the items of **Part Two** and **Part Three**, which reflect students' perceptions of the differences in language tests from their perspective. The data were classified according to the following dimensions:

Part Two: Perspectives on Progress Tests

In this section, each item of **Dimension Two (Perspectives on Progress Tests)** was analyzed individually. **Table (10)** displays the **frequency distribution (frequency and percentage)** of participants' responses to each item within this dimension. **Table (11)** shows the **statistical analysis**, including **arithmetic mean, standard deviation, relative importance, and prevailing trend**, for the same dimension. The obtained results are summarized as follows:

Table (10): Frequencies and Percentages of the Second part Items

X2	Items		Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1	I find progress tests more helpful than achievement tests.	N	18	21	15	8	8
		%	25.7	30	21.4	11.4	11.4
2	Progress tests help me identify my weaknesses.	N	33	24	5	4	4
		%	47.1	34.3	7.1	5.7	5.7
3	Frequent progress tests reduce my exam anxiety.	N	19	21	22	2	6
		%	27.1	30	31.4	2.9	8.6
4	The feedback I receive after progress tests improves my learning.	N	32	23	5	5	5
		%	45.7	32.9	7.1	7.1	7.1
5	I usually prepare in advance for progress tests.	N	15	23	16	5	11
		%	21.4	32.9	22.9	7.1	15.7
6	I study harder for achievement tests than for progress tests.	N	25	22	10	8	5
		%	35.7	31.4	14.3	11.4	7.1

Table (11): Weighted Mean, Standard Deviation, and Level for the Items of the Second part

X2	Items	Weighted Mean	Standard Deviation	Level	Relative Importance %.	Rank
1	I find progress tests more helpful than achievement tests.	3.47	1.20	High	69.4	5
2	Progress tests help me identify my weaknesses.	4.11	1.01	High	82.2	1
3	Frequent progress tests reduce my exam anxiety.	3.64	1.13	High	72.8	4
4	The feedback I receive after progress tests improves my learning.	4.03	1.11	High	80.6	2
5	I usually prepare in advance for progress tests.	3.37	1.02	Moderate	67.4	6
6	I study harder for achievement tests than for progress tests.	3.77	1.02	High	75.4	3
	Attitudes toward Progress Tests	3.73	1.01	High	74.6	

From the above tables, the following observations can be made:

Part Two: Perspectives on Progress Tests

1. I find progress tests more useful than achievement tests:

According to **Table (10)**, 30% of participants rated this item as "Agree," while 25.7% rated it as "Strongly Agree." **Table (11)** shows a weighted mean of 3.47 with a standard deviation of 1.20, indicating that the prevailing response trend is High, with a relative importance of 69.4%.

2. **Progress tests help me identify my weaknesses:**

Table (10) shows that 47.1% of participants rated this item as “Strongly Agree,” and 34.3% as “Agree.” **Table (11)** shows a weighted mean of 4.11, with a standard deviation of 1.01, reflecting a prevailing response trend of High, with a relative importance of 82.2%.

3. **Frequent testing reduces my anxiety during exams:**

Table (10) indicates that 31.4% of participants rated this item as “Neutral,” and 30% as “Agree.” **Table (11)** shows a weighted mean of 3.64, with a standard deviation of 1.13, reflecting a prevailing response trend of High, with a relative importance of 72.8%.

4. **Feedback after progress tests helps me improve my learning:**

Table (10) shows that 45.7% of participants rated this item as “Strongly Agree,” and 32.9% as “Agree.” **Table (11)** indicates a weighted mean of 4.03, with a standard deviation of 1.11, showing a prevailing response trend of High, with a relative importance of 80.6%.

5. **I usually prepare in advance for progress tests:**

Table (10) shows that 32.9% of participants rated this item as “Agree,” while 22.9% rated it as “Neutral.” **Table (11)** shows a weighted mean of 3.37, with a standard deviation of 1.02, indicating a prevailing response trend of Moderate, with a relative importance of 67.4%.

6. **I study more diligently for achievement tests compared to progress tests:**

Table (10) shows that 35.7% of participants rated this item as “Strongly Agree,” and 31.4% as “Agree.” **Table (11)** shows a weighted mean of 3.77, with a standard deviation of 1.02, indicating a prevailing response trend of High, with a relative importance of 75.4%.

The general result for Part Two (Perspectives on Progress Tests) is “Agree”, with an arithmetic mean of 3.73, which is above the acceptable mean of 3, a standard deviation of 1.01, and a relative importance of 74.6%. This indicates that participants expressed a high level of evaluation regarding progress tests. A high evaluation reflects strong satisfaction among students with the nature of progress tests and the assessment procedures applied in the department, demonstrating a positive general perception of the progress test system used in the Department of English.

Part Three: Perspectives on Achievement Tests

In this section, each item of Part Three (Perspectives on Achievement Tests) was analyzed individually. **Table (12)** presents the frequency distribution (frequency and percentage) of participants’ responses for each item, while **Table (13)** provides the statistical analysis, including arithmetic mean, standard deviation, relative importance, and prevailing trend.

The results obtained are summarized as follows:

Table (12): Frequencies and Percentages of the Third part Items

X3	Items		Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1	Achievement tests reflect my overall learning outcomes.	N	20	19	18	7	6
		%	28.6	27.1	25.7	10	8.6
2	I feel nervous before taking an achievement test.	N	22	26	10	5	7
		%	31.4	37.1	14.3	7.1	10
3	Achievement tests put too much pressure on students.	N	23	15	16	9	7
		%	32.9	21.4	22.9	12.9	10
4	I get anxious when facing long descriptive questions.	N	32	20	7	3	8
		%	45.7	28.6	10	4.3	11.4
5	I find achievement tests more stressful than progress tests.	N	22	18	17	7	6
		%	31.4	25.7	24.3	10	8.6
6	Tests motivate me to improve my language skills.	N	21	23	16	4	6
		%	30	32.9	22.9	5.7	8.6

Table (13): Weighted Mean, Standard Deviation, and Level for the Items of the Third part

X3	Items	Weighted Mean	Standard Deviation	Level	Relative Importance %.	Rank
1	Achievement tests reflect my overall learning outcomes.	3.57	1.12	High	71.4	5
2	I feel nervous before taking an achievement test.	3.73	1.09	High	74.6	2
3	Achievement tests put too much pressure on students.	3.54	1.13	High	70.8	6
4	I get anxious when facing long descriptive questions.	3.93	1.05	High	78.6	1
5	I find achievement tests more stressful than progress tests.	3.61	1.11	High	72.2	4
6	Tests motivate me to improve my language skills.	3.70	1.08	High	74	3
	Attitudes toward Achievement Tests	3.68	1.02	High	73.6	

From the above tables, the following observations can be made:

Part Three: Perspectives on Achievement Tests

1. **Achievement tests reflect my overall learning outcomes:**

According to Table (12), 28.6% of participants rated this item as “Strongly Agree,” while 25.1% rated it as “Agree.” Table (13) shows a weighted mean of 3.57 with a standard deviation of 1.12, indicating that the prevailing response trend is High, with a relative importance of 71.4%.

2. **I feel anxious before taking an achievement test:**

Table (12) shows that 37.1% of participants rated this item as “Agree,” and 31.4% as “Strongly Agree.” Table (13) indicates a weighted mean of 3.73, with a standard deviation of 1.09, reflecting a prevailing response trend of High, with a relative importance of 74.6%.

3. **Achievement tests impose significant pressure on students:**

Table (12) shows that 32.9% of participants rated this item as “Strongly Agree,” while 22.9% rated it as “Neutral.” Table (13) shows a weighted mean of 3.54, with a standard deviation of 1.13, indicating a prevailing response trend of High, with a relative importance of 70.8%.

4. **I feel anxious when facing long descriptive questions:**

Table (12) shows that 45.7% of participants rated this item as “Strongly Agree,” and 28.6% as “Agree.” Table (13) shows a weighted mean of 3.93, with a standard deviation of 1.05, indicating a prevailing response trend of High, with a relative importance of 78.6%.

5. **I find achievement tests more exhausting than progress tests:**

Table (12) shows that 31.4% of participants rated this item as “Strongly Agree,” and 25.7% as “Agree.” Table (13) indicates a weighted mean of 3.61, with a standard deviation of 1.11, showing a prevailing response trend of High, with a relative importance of 72.2%.

6. **Achievement tests motivate me to develop my language skills:**

Table (12) shows that 32.9% of participants rated this item as “Agree,” and 30% as “Strongly Agree.” Table (13) shows a weighted mean of 3.70, with a standard deviation of 1.08, indicating a prevailing response trend of High, with a relative importance of 74%.

The general result for Part Three (Perspectives on Achievement Tests) is “Agree”, with an arithmetic mean of 3.68, which is above the acceptable mean of 3, a standard deviation of 1.02, and a relative importance of 73.6%. This indicates that the participants expressed a high level of evaluation regarding achievement tests. A high evaluation reflects strong satisfaction among students with the nature of achievement tests and the assessment procedures applied in the department, demonstrating a positive general perception of the achievement testing system used in the Department of English.

For the comparative study between the two tests, the (T-test) for differences between means was used to examine the statistical significance at (0.05) between the mean scores of (MCQs) and (SAQs). As shown in the following table:

Table (14): Shows the results of the T-test for significance of differences for averages.

Test	N	Mean	Mean difference	T-value	P-value	Result
MCQs	70	3.73	0.05	2.750	0.008**	significant
SAQs	70	3.68				

** significant at 1%

Table (14) above shows that the **p-value ($P = 0.008$)** is smaller than **0.05**, indicating the presence of a **statistically significant difference** at the 0.05 significance level between the **mean scores of students' opinions on multiple-choice questions (MCQs)** and their **mean scores on short-answer questions (SAQs)**. The difference between the two means was **0.05**, which is considered a significant difference in favor of **MCQs**, with a mean of **3.73** compared to **3.68** for **SAQs**. This indicates that students' performance in both types of tests was slightly better on the **MCQ format**, with a minor but statistically significant difference of **0.05**.

2.3 Results Related to the Third Question:

What is the role of multiple-choice questions (MCQs) and short-answer questions (SAQs) in measuring students' motivation and performance in both types of tests?

The following tables present the participants' opinions regarding the items of **Part Four** and **Part Five**, which reflect the English Department students' perceptions of the role of question types (MCQs and SAQs) in **assessing motivation and performance** in language tests from the students' perspective. The data were classified according to the following dimensions:

Part Four: Perceptions of Question Types (Multiple-Choice vs. Short-Answer)

In this section, each item of **Part Four (Perceptions of Question Types)** was analyzed individually. **Table (15)** displays the **frequency distribution (frequency and percentage)** of participants' responses for each item, while **Table (16)** provides the **statistical analysis**, including **arithmetic mean**, **standard deviation**, **relative importance**, and **prevailing trend** for the same part.

The results obtained are summarized as follows:

Table (15): Frequencies and Percentages of the Fourth part Items

X4	Items		Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1	MCQ questions test memory rather than understanding.	N	21	15	14	12	8
		%	30	21.4	20	17.1	11.4
2	SAQs encourage me to think critically.	N	24	20	17	3	6
		%	34.3	28.6	24.3	4.3	8.6
3	I prefer SAQs because I can explain my answers.	N	25	22	9	8	6
		%	35.7	31.4	12.9	11.4	8.6
4	I prefer MCQs because they are less time-consuming.	N	21	24	18	0	7
		%	30	34.3	25.7	0	10
5	I perform better in MCQs than in SAQs.	N	27	18	16	3	6
		%	38.6	25.7	22.9	4.3	8.6
6	SAQs require deeper understanding than MCQs.	N	22	18	16	3	11
		%	31.4	25.7	22.9	4.3	15.7

Table (16): Weighted Mean, Standard Deviation, and Level for the Items of the Fourth part

X4	Items	Weighted Mean	Standard Deviation	Level	Relative Importance %.	Rank
1	MCQ questions test memory rather than understanding.	3.41	1.01	High	68.2	6
2	SAQs encourage me to think critically.	3.76	1.02	High	75.2	2
3	I prefer SAQs because I can explain my answers.	3.74	1.05	High	74.8	4
4	I prefer MCQs because they are less time-consuming.	3.75	1.02	High	75	3
5	I perform better in MCQs than in SAQs.	3.81	1.12	High	76.2	1
6	SAQs require deeper understanding than MCQs.	3.53	1.03	High	70.6	5
	Perceptions of Question Types (MCQs vs SAQs)	3.67	1.01	High	73.4	

From the above tables, the following observations can be made:

Part Four: Perceptions of Question Types (Multiple-Choice vs. Short-Answer)

- Multiple-choice questions (MCQs) measure memory more than understanding:**
According to **Table (15)**, 30% of participants rated this item as “Strongly Agree,” while 21.4% rated it as “Agree.”. **Table (16)** shows a weighted mean of 3.41, with a standard deviation of 1.01, indicating that the prevailing response trend is High, with a relative importance of 68.2%.
- Short-answer questions encourage me to think critically:**
Table (15) shows that 34.3% of participants rated this item as “Strongly Agree,” and 28.6% as “Agree.”. **Table (16)** indicates a weighted mean of 3.76, with a standard deviation of 1.02, reflecting a prevailing response trend of High, with a relative importance of 75.2%.
- I prefer short-answer questions because they allow me to explain my answers:**
Table (15) shows that 35.7% of participants rated this item as “Strongly Agree,” and 31.4% as “Agree.”. **Table (16)** shows a weighted mean of 3.74, with a standard deviation of 1.05, indicating a prevailing response trend of High, with a relative importance of 74.8%.
- I prefer multiple-choice questions because they are less time-consuming:**
Table (15) shows that 30% of participants rated this item as “Agree,” and 34.3% as “Strongly Agree.”. **Table (16)** indicates a weighted mean of 3.75, with a standard deviation of 1.02, showing a prevailing response trend of High, with a relative importance of 75%.
- My performance is better on multiple-choice questions compared to short-answer questions:**
Table (15) shows that 38.6% of participants rated this item as “Strongly Agree,” and 25.7% as “Agree.”. **Table (16)** shows a weighted mean of 3.81, with a standard deviation of 1.12, indicating a prevailing response trend of High, with a relative importance of 76.2%.
- Short-answer questions require deeper understanding compared to multiple-choice questions:**
Table (15) shows that 31.4% of participants rated this item as “Strongly Agree,” and 25.7% as “Agree.”. **Table (16)** shows a weighted mean of 3.53, with a standard deviation of 1.03, reflecting a prevailing response trend of High, with a relative importance of 70.6%.

The overall result for Part Four (Perceptions of Question Types – MCQ vs. Short-Answer) is “Agree”, with an arithmetic mean of 3.67, which is above the acceptable mean of 3, a standard deviation of 1.01, and a relative importance of 73.4%.

This indicates that participants expressed a high level of evaluation regarding the variety of question types (MCQs vs. Short-Answer). A high evaluation reflects strong satisfaction among students with the nature and diversity of test questions used in the department, demonstrating a positive general perception of the question type system applied in the Department of English.

Part Five: Personal Impressions, Motivation, and Performance Strategies

In this section, each item of Part Five (Personal Impressions, Motivation, and Performance Strategies) was analyzed individually. **Table (17)** presents the frequency distribution (frequency and percentage) of participants’ responses for each item, while **Table (18)** provides the statistical analysis, including arithmetic mean, standard deviation, relative importance, and prevailing trend for the same part.

The results obtained are summarized as follows:

Table (17): Frequencies and Percentages of the Fifth Domain Items

X5	Items		Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1	I find it difficult to express answers clearly in SAQs.	N	18	19	9	15	9
		%	25.7	27.1	12.9	21.4	12.9
2	Both MCQs and SAQs are necessary to assess different skills.	N	16	31	12	6	5
		%	22.9	44.3	17.1	8.6	7.1
3	I prefer tests that combine both MCQs and SAQs.	N	26	19	14	4	7
		%	37.1	27.1	20	5.7	10
4	I feel confident answering grammar-related MCQs.	N	22	17	13	9	9
		%	31.4	24.3	18.6	12.9	12.9
5	Vocabulary questions are easier to answer in MCQ form.	N	26	28	7	3	6
		%	37.1	40	10	4.3	8.6

Table (18): Weighted Mean, Standard Deviation, and Level for the Items of the Fifth part

X5	Items	Weighted Mean	Standard Deviation	Level	Relative Importance %.	Rank
1	I find it difficult to express answers clearly in SAQs.	3.31	1.03	Moderate	66.2	5
2	Both MCQs and SAQs are necessary to assess different skills.	3.67	1.12	High	73.4	3
3	I prefer tests that combine both MCQs and SAQs.	3.76	1.02	High	75.2	2
4	I feel confident answering grammar-related MCQs.	3.49	1.01	High	69.8	4
5	Vocabulary questions are easier to answer in MCQ form.	3.93	1.04	High	78.6	1
	Self-Reflection, Motivation and Performance Strategies	3.63	1.03	High	72.6	

From the above tables, the following observations can be made:

Part Five: Personal Impressions, Motivation, and Performance Strategies

1. I find it difficult to express myself clearly in short-answer questions:

Table (17) shows that 27.1% of participants rated this item as “Agree,” while 25.7% rated it as “Strongly Agree.”. Table (18) shows a weighted mean of 3.31 with a standard deviation of 1.03, indicating that the prevailing trend of responses is Neutral, with a relative importance of 66.2%.

2. Both multiple-choice and short-answer questions are necessary to assess different skills:

Table (17) shows that 44.3% of participants rated this item as “Agree,” and 22.9% as “Strongly Agree.”. Table (18) indicates a weighted mean of 3.67, with a standard deviation of 1.12, reflecting a prevailing response trend of High, with a relative importance of 73.4%.

3. I prefer tests that combine short-answer and multiple-choice questions:

Table (17) shows that 37.1% of participants rated this item as “Strongly Agree,” and 27.1% as “Agree.”. Table (18) shows a weighted mean of 3.76, with a standard deviation of 1.02, indicating a prevailing trend of High, with a relative importance of 75.2%.

4. I feel confident when answering grammar questions in multiple-choice format:

Table (17) shows that 31.4% of participants rated this item as “Strongly Agree,” and 24.3% as “Agree.”. Table (18) shows a weighted mean of 3.49, with a standard deviation of 1.01, indicating a prevailing trend of High, with a relative importance of 69.8%.

5. Vocabulary questions are easier when presented in multiple-choice format:

Table (17) shows that 40% of participants rated this item as “Agree,” and 37.1% as “Strongly Agree.”. Table (18) shows a weighted mean of 3.93, with a standard deviation of 1.04, indicating a prevailing trend of High, with a relative importance of 78.6%.

The result for Part Five (Personal Impressions, Motivation, and Performance Strategies) is “Agree”, with an arithmetic mean of 3.63, which is above the acceptable mean of 3, a standard deviation of 1.03, and a relative importance of 72.6%.

This indicates that participants expressed a high level of evaluation regarding personal impressions, motivation, and performance strategies. A high evaluation reflects strong satisfaction among students with their personal impressions and performance strategies, demonstrating a positive general perception among students regarding their personal engagement and performance strategies in the Department of English.

Table (19): Frequencies and Percentages of the Sixth part Items

X4	Items		Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1	MCQs are a fair method of assessment.	N	18	29	12	5	6
		%	25.7	41.4	17.1	7.1	8.6
2	SAQs allow teachers to evaluate students' real understanding.	N	23	31	7	5	4
		%	32.9	44.3	10	7.1	5.7
3	The time given for tests is usually sufficient.	N	11	17	12	14	16
		%	15.7	24.3	17.1	20	22.9
4	I feel that my marks in MCQs are more objective than in SAQs.	N	18	23	18	6	5
		%	25.7	32.9	25.7	8.6	7.1
5	I believe MCQs can sometimes be misleading.	N	23	19	19	4	5
		%	32.9	27.1	27.1	5.7	7.1
6	Reading comprehension questions are better tested using SAQs.	N	16	23	13	12	6
		%	22.9	32.9	18.6	17.1	8.6

Results Related to Research Question Four:

Which type of test provides the greatest fairness, validity, and reliability in assessing English language learning?

The following tables present participants' opinions regarding the items of Part Six, reflecting their perceptions of the fairness, validity, and reliability of question types (MCQs vs. Short-Answer) in assessing students' English language learning. The part is classified as follows:

Part Six: Perceptions of Fairness, Validity, and Reliability in Assessment

In this section, each item of Dimension Six (Perceptions of fairness, validity, and reliability of tests) was analyzed individually. **Table (19)** presents the frequency distribution (frequency and percentage) of participants' responses for each item, while **Table (20)** provides the statistical analysis, including arithmetic mean, standard deviation, relative importance, and prevailing trend. The results obtained are summarized as follows:

Table (20): Weighted Mean, Standard Deviation, and Level for the Items of the Sixth part

X4	Items	Weighted Mean	Standard Deviation	Level	Relative Importance %.	Rank
1	MCQs are a fair method of assessment.	3.69	1.12	High	73.8	3
2	SAQs allow teachers to evaluate students' real understanding.	3.91	1.01	High	78.2	1
3	The time given for tests is usually sufficient.	2.90	1.13	Low	58	6
4	I feel that my marks in MCQs are more objective than in SAQs.	3.61	1.11	High	72.2	4
5	I believe MCQs can sometimes be misleading.	3.73	1.12	High	74.6	2
6	Reading comprehension questions are better tested using SAQs.	3.44	1.10	High	68.8	5
	Perceived Fairness and Validity of Tests	3.55	1.01	High	71	

From the above tables, the following observations can be made regarding **Part Six: Perceptions of Fairness, Validity, and Reliability in Assessment:**

- Multiple-choice questions (MCQs) are considered a fair assessment method:**
Table (19) shows that 41.4% of participants rated this item as "Agree," and 25.7% as "Strongly Agree."
Table (20) shows a weighted mean of 3.69 with a standard deviation of 1.12, indicating that the prevailing trend is High, with a relative importance of 73.8%.
- Short-answer questions (SAQs) allow teachers to assess students' true understanding:**
Table (19) shows that 44.3% of participants rated this item as "Agree," and 32.9% as "Strongly Agree."
Table (20) shows a weighted mean of 3.91 with a standard deviation of 1.01, indicating that the prevailing trend is High, with a relative importance of 78.2%.
- The time allocated for tests is usually sufficient:**
Table (19) shows that 24.3% of participants rated this item as "Agree," and 15.7% as "Strongly Agree."
Table (20) shows a weighted mean of 2.90 with a standard deviation of 1.13, indicating that the prevailing trend is Low, with a relative importance of 58%.
- I believe my scores in MCQs are more objective than in SAQs:**
Table (19) shows that 32.9% of participants rated this item as "Agree," and 25.7% as "Strongly Agree."
Table (20) shows a weighted mean of 3.61 with a standard deviation of 1.11, indicating that the prevailing trend is High, with a relative importance of 72.2%.
- MCQs can sometimes be misleading:**
Table (19) shows that 32.9% of participants rated this item as "Strongly Agree," and 27.1% as "Agree."
Table (20) shows a weighted mean of 3.73 with a standard deviation of 1.12, indicating that the prevailing trend is High, with a relative importance of 74.6%.
- Reading comprehension is better assessed through SAQs:**
Table (19) shows that 32.9% of participants rated this item as "Agree," and 22.9% as "Strongly Agree."
Table (20) shows a weighted mean of 3.44 with a standard deviation of 1.10, indicating that the prevailing trend is High, with a relative importance of 68.8%.

The overall result for Part Six is "Agree," with an arithmetic mean of 3.55, which is above the acceptable mean of 3, a standard deviation of 1.01, and a relative importance of 71%.

This indicates that participants expressed a high level of evaluation regarding the fairness, validity, and reliability of assessments. A high evaluation reflects strong satisfaction among students with the fairness and reliability of the tests, demonstrating a positive general perception among English department students regarding the fairness, validity, and reliability of the assessment methods used.

3 Levels of Perceptions and Views on English Language Tests:

To determine the levels, ranking, and relative importance of the parts of perceptions and views regarding English language tests (including progress tests and achievement tests), the arithmetic mean, standard deviation, relative importance, and ranking were calculated for all dimensions of perceptions and views, as shown in the following table:

Table (21): Ranking and Levels of Perceptions and Views on English Language Tests

X4	Axis	Weighted Mean	Standard Deviation	Level	Relative Importance %.	Rank
X1	General Perceptions of Language Tests	3.69	1.03	High	73.8	2
X2	Attitudes toward Progress Tests	3.73	1.01	High	74.6	1
X3	Attitudes toward Achievement Tests	3.68	1.02	High	73.6	3
X4	Perceptions of Question Types (MCQs vs SAQs)	3.67	1.01	High	73.4	4
X5	Self-Reflection, Motivation and Performance Strategies	3.63	1.03	High	72.6	5
X6	Perceived Fairness and Validity of Tests	3.55	1.01	High	71	6
	Total	3.66	1.02	High	73.2	

Levels Perceptions of and Views on English Language Tests

The analysis of Table (21) reveals the following findings regarding the levels and ranking of perceptions and views of students on English language tests:

- General perceptions of language tests:**

This dimension ranked second with a mean of 3.69 and a standard deviation of 1.03. The mean exceeds the

hypothetical average (3), indicating that the prevailing level for this variable is high, with a relative importance of 73.8%, reflecting students' positive general perception of the language tests.

2. **Views on progress tests:**

This dimension ranked first with a mean of 3.73 and a standard deviation of 1.01. The mean surpasses the hypothetical average, confirming a high level of perception regarding progress tests, with a relative importance of 74.6%, indicating strong student satisfaction with the progress test system.

3. **Views on achievement tests:**

Ranking third, this dimension had a mean of 3.68 and a standard deviation of 1.02. The mean above the hypothetical average indicates a high level of perception regarding achievement tests, with a relative importance of 73.6%.

4. **Perceptions of question types (MCQs – SAQs):**

This dimension ranked fourth, with a mean of 3.67 and a standard deviation of 1.01. The results reflect a high level of perception concerning different question formats, with a relative importance of 73.4%.

5. **Personal impression, motivation, and performance strategies:**

Ranking fifth, this dimension recorded a mean of 3.63 with a standard deviation of 1.03. The mean above the hypothetical average indicates a high level of perception, with a relative importance of 72.6%, showing positive student attitudes toward motivation and performance strategies.

6. **Perceptions of fairness, validity, and reliability of tests:**

This dimension ranked sixth, with a mean of 3.55 and a standard deviation of 1.01. The mean exceeding the hypothetical average indicates a high level of perception regarding the fairness and reliability of tests, with a relative importance of 71%.

Overall evaluation:

- All parts show positive evaluations, reflecting students' satisfaction with both progress and achievement tests.
- The overall mean of the questionnaire is 3.66, which exceeds the hypothetical mean (3), indicating a prevailing high level of satisfaction, with a standard deviation of 1.02 and a relative importance of 73.2%.
- This demonstrates that participants are generally satisfied with the nature of the tests used to assess their progress in English language learning within the department.

Discussion, Conclusions, and Recommendations:

This section presents a comprehensive discussion of the study findings in relation to the research questions and objectives. It also provides the main conclusions drawn from the results, followed by practical recommendations and suggestions for future research. The chapter aims to interpret the statistical findings in a meaningful way that contributes to improving assessment practices in English language teaching.

Discussion of the Results:

The findings of this study reveal several important insights regarding students' perceptions of language assessment methods, particularly the effectiveness of **Multiple-Choice Questions (MCQs)** and **Short-Answer Questions (SAQs)**.

First, the results indicated that students hold a **generally positive perception** of language tests, with a high overall mean (3.69). This suggests that the current assessment system is acceptable and relatively effective from the students' perspective. However, the moderate responses regarding the accuracy of tests in measuring proficiency indicate that there is still room for improvement in test design.

Second, the findings related to **progress tests** showed a high level of student satisfaction (mean = 3.73). Students reported that progress tests help them identify their weaknesses and improve their learning through continuous feedback. This supports the idea that formative assessment plays a crucial role in enhancing learning outcomes and reducing exam anxiety.

In contrast, **achievement tests** were also positively perceived (mean = 3.68), but they were associated with higher levels of stress and pressure. Students reported feeling anxious before such tests, which suggests that summative assessment, while important, may negatively affect students' emotional state.

Regarding **question types**, the results revealed an interesting balance:

- Students acknowledged that MCQs are easier, faster, and more objective.
 - At the same time, they believed that SAQs promote deeper understanding and critical thinking.
- This dual perception highlights the complementary nature of both question types. While MCQs are efficient for assessing knowledge and recognition, SAQs are more effective in evaluating higher-order thinking skills. Furthermore, the statistical analysis using the **T-test** showed a **significant difference** between MCQs and SAQs ($p = 0.008$), with a slight advantage in favor of MCQs. Although the difference is small (0.05), it indicates that students tend to perform better in MCQs, possibly due to familiarity, guessing strategies, or reduced cognitive load.
- Finally, the findings related to **fairness and validity** showed that students perceive both test types as generally fair. However, they also acknowledged that:
- MCQs can sometimes be misleading.
 - SAQs better reflect real understanding.

This suggests that relying on only one type of question may not provide a complete and accurate evaluation of students' abilities.

Conclusions:

The findings indicate that English Department students generally have positive attitudes toward the assessment methods used in their academic environment. Although students are mostly satisfied with the current testing system, some aspects still need improvement to ensure more accurate evaluation of language proficiency.

The study also shows that progress tests play an important role in supporting learning by identifying weaknesses, providing continuous feedback, and reducing exam anxiety. In contrast, achievement tests, despite their importance in measuring learning outcomes, are often associated with greater stress among students.

Furthermore, the results reveal a significant difference between students' performance in multiple-choice and short-answer questions, with slightly better performance in multiple-choice items. Students consider multiple-choice questions more practical and objective, while short-answer questions require deeper understanding and critical thinking.

The findings further emphasize that using only one question type is insufficient for comprehensive assessment. Combining multiple-choice and short-answer questions creates a more balanced evaluation by measuring a wider range of language skills.

Finally, students consider fairness, validity, and reliability essential in assessment practices. Although the current system is viewed as generally fair, students recognize limitations in each test format, highlighting the need for more balanced and effective assessment methods in English language teaching.

Recommendations:

In light of the study findings, the following recommendations are proposed:

1. Integrate both MCQs and SAQs in test design to ensure a balanced assessment of knowledge and skills.
2. Increase the use of progress tests as a formative assessment tool to support continuous learning.
3. Provide clear instructions and test formats in advance to reduce student anxiety and improve performance.
4. Train instructors in modern assessment techniques to improve test quality and validity.
5. Reduce over-reliance on achievement tests and adopt more diverse assessment strategies.
6. Ensure that test questions measure not only memorization but also critical thinking and language use.
7. Improve time allocation during exams to better match the difficulty level of questions.

Limitations of the Study:

Despite the valuable findings, this study has some limitations:

1. The study was limited to a single department (English Department, Faculty of Education, Zolten), which may limit generalizability.
2. The sample size (70 students) is relatively small.
3. The study relied mainly on self-reported data (questionnaire), which may be subject to bias.
4. Only two types of questions (MCQs and SAQs) were examined, excluding other assessment formats.

Compliance with ethical standards

Disclosure of conflict of interest

The authors declare that they have no conflict of interest.

References:

1. Abdel Rahman, M. (2015). *Educational Measurement and Evaluation*. Cairo: Dar Al-Fikr Al-Arabi, pp. 133–158.
2. Alderson, J. C. (2000). *Assessing Reading*. Cambridge: Cambridge University Press, pp. 201–225.
3. Alkhafel, Naser.A1 and Elkholy, A. R (2022). The Influence of the First Language (Arabic) on Learning English as a Second Language. Teaching and Linguistic issues. American Research Journal of Humanities & Social Science (ARJHSS). Volume-05, Issue-04, pp-33-42
4. Al-Zahrani, A. (2018). The effectiveness of formative tests in improving English language achievement. *Journal of Educational Sciences*, 30(2), 211–236.
5. Andrade, H. L., & Heritage, M. (2018). 'Using Formative Assessment to Enhance Learning, Achievement, and Academic Self-Regulation'.
6. Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice*. Oxford: Oxford University Press, pp. 17–38.
7. Black, P., & Wiliam, D. (1998). "Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*" 5(1), 7–74.
8. Brookhart, S. M.(2001). 'Successful students' formative and summative uses of assessment information'. *Assessment in Education*, 8(2), 153–169.
9. Brown, H. D. (2004). "Language Assessment: Principles and Classroom Practices". Pearson Education.

10. Brown, H. D. (2004). *Language Assessment: Principles and Classroom Practices*. White Plains, NY: Pearson Education, pp. 47–72.
11. Cheng, L., Rogers, T., & Hu, H. (2004). 'ESL/EFL Instructors' Classroom Assessment Practices: Purposes, Methods, and Procedures'. *Language Testing*, 21(3), 360–389.
12. Dayal, H., I. Lingam, G., Sharma, L., Fitoo, B., and Sarai, V. (2018). "Fijian primary school leaders' beliefs and practices about assessment."
13. Dunn, E. K. and W. Mulvenon, S. (2019) "A Critical Review of Research on Formative Assessments: The Limited Scientific Evidence of the Impact of Formative Assessments in Education."
14. Gielen, S., Dochy, F., & Dierick, S. (2003). 'Evaluating the Consequences of Assessment Innovations on Students' Learning: The Case of Self- and Peer Assessment'.
15. Green, A. (2014). *Exploring Language Assessment and Testing*. London: Routledge, pp. 89–112.
16. Guskey, T. R. (2003). How classroom assessments improve learning. "Educational Leadership", 60(5), 6–11.
17. Harlen, W., & Deakin Crick, R. (2003). 'Testing and Motivation for Learning'. *Assessment in Education*, 10(2), 169–207.
18. Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships. "Assessment in Education: Principles, Policy & Practice", 4(3), 365–379.
19. Hassan, S. (2020). Assessment practices in teaching English as a foreign language. *Arab Journal of Educational Studies*, 12(1), pp 55–78.
20. Hughes, A. (2003). *Testing for Language Teachers* (2nd ed.). Cambridge: Cambridge University Press, pp. 5–23.
21. Mohammed, N. A., & Alsayah, F. F. M. (2025). The role of classroom tests in the assessment of students' learning: Intended objectives and potential defects. *Libyan Journal of Contemporary Academic Studies*, 3(2, first part), pp 348-359.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of LJCAS and/or the editor(s). LJCAS and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.