



Challenges Faced by Instructors in Designing Fair and Valid English Language Tests for ELLs: A Case Study at Abu Esa College

Ahlam Mohammed Ali Garaf

Department of English- College of Art - University of Zawia
City - State

EMAIL: a.garaf@zu.edu.ly

تاريخ الاستلام: 2025/12/8 - تاريخ المراجعة: 2025/12/12 - تاريخ القبول: 2025/12/19 - تاريخ النشر: 2026 /1/20

ملخص الدراسة

تهدف هذه الدراسة إلى استكشاف التحديات التي يواجهها مدرّسو اللغة الإنجليزية في تصميم اختبارات عادلة وصحيحة لمتعلمي اللغة الإنجليزية كلغة ثانية في كلية أبو عيسى. اعتمدت الدراسة منهجية البحث الوصفي باستخدام استبانة تتضمن أسئلة مغلقة وأسئلة مفتوحة لجمع البيانات من المدرّسين. أظهرت النتائج أن أبرز التحديات تشمل قلة التدريب في مجال التقييم، ضيق الوقت، كثافة عدد الطلاب، تفاوت مستويات الطلاب، ونقص الدعم المؤسسي. كما أشار المدرّسون إلى أهمية التطوير المهني، والتعاون بين الزملاء، والوصول إلى مصادر تقييم حديثة لتعزيز صحة وعدالة الاختبارات. تؤكد الدراسة تأثير جودة التقييم بشكل كبير على تحفيز الطلاب ونتائج تعلمهم. وقدمت الدراسة توصيات عملية لدعم المدرّسين في تطوير اختبارات أكثر فعالية وعدلاً في سياقات تعلم اللغة الإنجليزية كلغة أجنبية.

الكلمات المفتاحية: تقييم اللغة الإنجليزية، صحة الاختبار، عدالة التقييم، مدرّسو اللغة الإنجليزية كلغة أجنبية، تحديات تصميم الاختبارات

Abstract

This study investigates the challenges English language instructors face in designing fair and valid assessments for English Language Learners (ELLs) at Abu Issa College. Using a descriptive survey approach, data were collected from instructors through a questionnaire consisting of Likert-scale and open-ended items. Findings indicate that major challenges include insufficient training in assessment, time constraints, large class sizes, mixed student proficiency levels, and limited institutional support. Instructors also highlighted the importance of professional development, collaboration, and access to updated assessment resources in enhancing test validity and fairness. The study underscores the significant impact of assessment quality on students' motivation and learning outcomes. Practical recommendations are provided to support instructors in developing more effective and equitable assessments in EFL contexts.

Keywords: English language assessment, test validity, assessment fairness, EFL instructors, challenges in test design.

Introduction

Assessment is a central component of English language education, shaping not only how learners' proficiency is evaluated but also how teaching and learning are guided and improved. In English as a Foreign Language (EFL) contexts, assessment practices carry particular weight, as learners' academic progress, motivation, and confidence are closely tied to how their language abilities are measured. For English Language Learners (ELLs), poorly designed assessments may obscure actual proficiency, generate unnecessary anxiety, and produce results that fail to inform instruction in meaningful ways.

Effective English language tests extend beyond the assignment of grades. When carefully constructed, they provide valid evidence of learners' communicative competence, identify areas of strength and difficulty, and support pedagogical decision-making. Fair and valid assessments ensure that learners are evaluated based on relevant language abilities rather than extraneous factors such as cultural familiarity, test-taking strategies, or ambiguous task design. In this sense, assessment functions not only as a measurement tool but also as a mechanism that can either support or hinder language development.

Despite the importance of sound assessment practices, designing fair and valid language tests remains a complex task for many instructors. Test developers must navigate multiple challenges, including differences in learners' proficiency levels, linguistic and cultural diversity, alignment with learning outcomes, institutional expectations, and limited time for test preparation. These challenges are often intensified in higher education contexts where instructors are responsible for both teaching and assessment, frequently without specialized training in language testing principles.

At Abu Esa College, English language instructors work with a diverse student population whose language abilities, educational backgrounds, and learning needs vary considerably. While instructors are expected to design assessments that accurately reflect course objectives and student proficiency, opportunities for systematic professional development in language assessment are limited. As a result, instructors may rely on personal experience or intuition

when constructing tests, which can affect the fairness, validity, and reliability of assessment outcomes.

Against this backdrop, the present study examines the challenges faced by English language instructors at Abu Esa College in designing fair and valid English language tests for ELLs. By exploring instructors' experiences, perceptions, and practices, the study aims to provide context-sensitive insights into assessment design in EFL higher education settings. Identifying these challenges is a necessary step toward improving assessment quality and supporting more equitable and meaningful evaluation practices for ELLs.

Research Statement

Drawing from my experience as an instructor, this study will examine common difficulties such as aligning test content with learning outcomes, ensuring cultural and linguistic fairness, and developing tasks that accurately reflect students' language proficiency levels. The research will also explore the benefits of well-designed assessments and identify practical strategies to help instructors create tests that support students' learning, motivation, and success. However, this study seeks to provide insights and recommendations that can enhance assessment practices for ELLs in similar educational contexts.

Aims of the Study

The aim of this study is to:

1. Explore the challenges English language instructors at Abu Esa College face in designing fair and valid English language tests for English Language Learners.
2. The perception of the instructors towards the assessment quality on ELL students' learning experience.

Research Questions

1. What challenges do English language instructors at Abu Esa College encounter when designing fair and valid English language tests for ELLs?
2. How do instructors perceive the impact of assessment quality on ELL students' learning experience?

Literature Review

1. The Role of Assessment in English Language Learning

Assessment occupies a foundational role in English language education, functioning as both a measure of learning outcomes and a tool for supporting instructional development. In language teaching contexts, assessment informs decisions about curriculum design, instructional strategies, and learner support. Brown and Abeywickrama (2019) emphasize that effective language assessment must be closely aligned with instructional objectives and designed to generate meaningful information about learners' language abilities.

Beyond measurement, assessment plays a formative role in the learning process. Through systematic feedback, learners gain insight into their progress and areas requiring improvement, which can foster greater engagement and autonomy (Black & Wiliam, 2009). Richards and Schmidt (2010) note that well-designed assessments allow instructors to diagnose learning difficulties and adjust instruction accordingly, making assessment an integral component of pedagogical practice rather than a separate evaluative event.

For ELLs, assessment design requires particular sensitivity. Hughes (2003) argues that language tests should measure language ability directly and avoid interference from non-linguistic factors such as cultural assumptions or unclear task demands. When assessment fails to reflect learners' actual language competence, test results may become misleading and pedagogically unhelpful. Consequently, assessment quality has direct implications for both instructional effectiveness and learner outcomes.

2. Principles of Fairness and Validity in Language Testing

Validity is widely regarded as the most fundamental principle in language assessment. A test is considered valid when it accurately measures the construct it is intended to assess and supports appropriate interpretations of test scores (Fulcher & Davidson, 2007). In the context of English language testing, this means ensuring that assessment tasks reflect authentic language use and correspond closely to course learning outcomes. Tests that emphasize memorization or isolated language elements may fail to capture learners' communicative competence.

Reliability complements validity by ensuring consistency in test results across administrations, tasks, and raters. According to Kunnan (2018), unreliable assessments undermine confidence in test scores and weaken their usefulness for academic decision-making. In classroom contexts, unclear scoring criteria or subjective judgment can compromise reliability, particularly in productive skills such as speaking and writing.

Fairness is closely linked to both validity and reliability and is especially critical for ELL populations. Fair assessments provide all learners with equitable opportunities to demonstrate their language abilities, regardless of linguistic or cultural background. Kunnan (2018) highlights that fairness requires careful attention to test content, task format, and language complexity to prevent bias. Clear instructions, accessible language, and familiar contexts contribute to fairer testing conditions.

Practicality is another important consideration, particularly in institutions with limited resources. Coombe et al. (2020) argue that even well-designed assessments must be feasible in terms of time, cost, and administrative demands. Instructors often need to balance ideal assessment practices with institutional constraints, which can influence test design decisions.

3. Challenges in Designing Language Tests for ELLs

Despite clear theoretical principles, instructors frequently encounter difficulties when translating assessment theory into classroom practice. One major challenge involves aligning test content with instructional objectives and learners' proficiency levels. When assessments do not accurately reflect what has been taught, learners may perceive tests as unfair, and results may fail to represent actual learning outcomes (Coombe et al., 2020).

Cultural and linguistic bias represents another persistent challenge. Kunnan (2018) notes that test items containing culturally specific references, idiomatic expressions, or unfamiliar contexts may disadvantage ELLs and distort assessment results. Such bias can unintentionally shift the focus from language ability to background knowledge, undermining test validity.

The clarity of instructions and task wording also plays a crucial role in assessment quality. Taylor and Weir (2012) argue that ambiguous or overly complex instructions can lead to misunderstanding, causing learners to perform poorly for reasons unrelated to language proficiency. This issue is particularly salient for lower-level learners who may struggle with dense or abstract language.

Instructors also face practical constraints, including heavy workloads and limited access to professional development in language assessment. Fulcher and Davidson (2007) emphasize that effective test design requires time for planning, piloting, and revision, processes that are often unrealistic in busy teaching environments. As a result, instructors may rely on intuition or recycled test materials, which can compromise assessment quality.

4. Impact of Assessment Quality on Learner Motivation and Achievement

Assessment quality has a profound influence on learners' motivation, attitudes, and academic performance. When learners perceive assessments as fair, transparent, and aligned with instruction, they are more likely to engage positively with learning tasks and develop confidence in their abilities (Stiggins, 2017). Conversely, assessments perceived as unfair or irrelevant can increase anxiety and reduce motivation, particularly among ELLs (Cheng et al., 2004).

Feedback is a key mechanism through which assessment affects learning. Carless and Boud (2018) argue that meaningful feedback enables learners to interpret assessment results constructively and apply them to future learning. When feedback is clear and actionable, learners are more likely to view assessment as supportive rather than punitive.

Authenticity in assessment tasks further enhances learner engagement. Fulcher and Davidson (2007) suggest that tasks reflecting real-world language use are more motivating and meaningful for learners than decontextualized test formats. For ELLs, authentic assessment can bridge the gap between classroom learning and practical language use, reinforcing the relevance of assessment to real-life communication.

Methodology

Research Design

This study adopted a descriptive survey research design to investigate the challenges English language instructors face when designing fair and valid English language tests for English Language Learners (ELLs) at Abu Issa College. A survey design was considered appropriate because it allows for the systematic collection of instructors' perceptions across multiple dimensions of language assessment, including validity, fairness, reliability, and practicality. Such an approach is commonly used in language assessment research to identify prevailing trends and shared concerns among practitioners (Brown & Abeywickrama, 2019; Kunnan, 2018).

The study employed a mixed-question format, combining closed-ended Likert-scale items with open-ended questions. This design enabled the researcher to capture both quantifiable patterns in instructors' responses and more detailed qualitative insights into their assessment practices and experiences.

Participants

The participants were English language instructors currently teaching ELLs at Abu Issa College. All participants had experience in designing, administering, and grading English language tests across different proficiency levels. The instructors represented a range of teaching experience, which allowed the study to capture diverse perspectives on assessment practices.

Participation was voluntary, and ethical considerations were strictly observed. In line with Creswell (2018), participants were informed about the purpose of the study, assured of anonymity and confidentiality, and informed that their responses would be used solely for academic research purposes.

Research Instrument

Data were collected using a self-administered questionnaire developed specifically for this study and informed by established frameworks in language assessment as this supported by Brown & Abeywickrama, 2019; Coombe et al., 2020. The questionnaire consisted of two main sections.

Section One: Closed-Ended Items

The first section of the questionnaire consisted of 15 Likert-scale items measured on a five-point scale ranging from Strongly Agree to Strongly Disagree. These items targeted key challenges identified in the literature, including instructor training in assessment, workload and time constraints, learner diversity, institutional guidelines, resource availability, collaboration, and grading practices. The items were designed to examine factors that may influence assessment validity, fairness, and reliability, as emphasized by Hughes (2003) and Kunnan (2018).

Section Two: Open-Ended Items

The second section included four open-ended questions that invited instructors to elaborate on their responses, explain the reasons behind assessment challenges, and propose suggestions for improving test design practices. Open-ended questions were included to allow instructors to express context-specific concerns that may not be fully captured through fixed-response items, a practice recommended in educational survey research (Dörnyei & Taguchi, 2010).

This section allowed participants to express their views freely and provide context-specific explanations that could not be captured through fixed-response items.

Data Collection Procedure

The questionnaire was distributed electronically to English language instructors at Abu Issa College. Participants completed the questionnaire at their convenience. Clear instructions were provided at the beginning of the questionnaire to explain its purpose and ensure informed participation. All responses were automatically recorded and stored securely.

Data Analysis

Data from the closed-ended items were analyzed using descriptive statistics, specifically frequencies and percentages, to identify general trends and dominant perceptions among instructors. This approach is appropriate for survey-based research aiming to describe patterns rather than test hypotheses (Brown & Abeywickrama, 2019).

Responses to the open-ended questions were analyzed using thematic analysis, following the procedures outlined by Braun and Clarke (2006). The responses were read repeatedly to achieve familiarity with the data, coded for recurring ideas, and grouped into broader themes related to challenges in test design, institutional constraints, and suggested improvements. The integration of quantitative and qualitative data allowed for a more comprehensive understanding of instructors' experiences with language assessment.

Results

Results of Closed-Ended Questionnaire Items

Table 1 presents a summary of instructors' responses to the Likert-scale items.

Item	Statement	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1	Insufficient assessment training	41%	37%	12%	6%	4%
2	Lack of updated materials	38%	34%	15%	8%	5%
3	Large class sizes	46%	35%	9%	6%	4%
4	Time constraints	49%	36%	8%	4%	3%
5	Mixed proficiency levels	52%	36%	6%	4%	2%
6	Lack of guidelines	40%	34%	14%	8%	4%
7	Limited technology	33%	36%	17%	9%	5%
8	Weak feedback mechanisms	35%	36%	16%	8%	5%
9	Standardized test pressure	29%	34%	19%	11%	7%
10	Limited professional development	45%	37%	10%	5%	3%
11	Heavy teaching workload	50%	36%	7%	4%	3%
12	Alignment with outcomes	42%	37%	11%	6%	4%
13	Limited collaboration	31%	37%	18%	9%	5%
14	Matching real proficiency	47%	36%	9%	5%	3%
15	Grading misalignment	34%	36%	16%	9%	5%

The results indicate strong agreement across most items, particularly those related to workload, mixed proficiency levels, and limited professional development.

Results of Open-Ended Questions

Analysis of the open-ended responses revealed four dominant themes.

a. Misalignment Between Teaching Content and Test Tasks

Several instructors explained that assessment tasks do not always reflect classroom instruction due to time pressure and syllabus coverage demands. Instructor 1 noted that “tests are sometimes designed quickly at the end of the term, which makes it difficult to fully reflect what was practiced in class.” This finding aligns with Fulcher and Davidson’s (2007) observation that limited preparation time often compromises test validity.

Instructor 3 added that institutional expectations to cover a fixed syllabus sometimes result in tests focusing on content coverage rather than communicative competence, a concern also highlighted by Brown and Abeywickrama (2019).

b. Need for Assessment Literacy and Professional Training

A strong theme across responses was the lack of formal training in language assessment. Instructor 2 reported that “most of what I know about test design comes from experience, not from structured training.” This supports Coombe et al.’s (2020) argument that insufficient assessment literacy can lead instructors to rely on intuition rather than systematic design principles.

Participants expressed a clear need for workshops on rubric development, item writing, and validity checking, particularly for productive skills such as speaking and writing.

c. Institutional Constraints and Limited Collaboration

Instructors frequently referred to the absence of clear institutional guidelines and limited collaboration among colleagues. Instructor 4 stated that “each instructor designs tests differently, which affects consistency and fairness.” This reflects Kunnan’s (2018) emphasis on the role of institutional support in ensuring equitable assessment practices.

Several instructors suggested peer review of tests and shared assessment frameworks as possible solutions to improve consistency.

d. Impact on Students’ Motivation and Confidence

Instructors also highlighted how assessment quality affects students’ motivation. Instructor 5 explained that when tests do not reflect students’ real abilities, learners become discouraged

and anxious. This perception is consistent with Cheng et al. (2004), who found that poorly aligned assessments can negatively influence learners' attitudes and performance.

Discussion

The findings of this study reveal that English language instructors at Abu Issa College face multiple, interconnected challenges in designing fair and valid assessments for English Language Learners (ELLs). The analysis of Likert-scale items showed that a high percentage of instructors agreed or strongly agreed that insufficient training in assessment theory, time constraints, heavy teaching workloads, large class sizes, and diverse student proficiency levels significantly hinder their ability to design high-quality tests. Open-ended responses further highlighted the role of limited institutional support, inadequate access to up-to-date materials, and restricted opportunities for collaboration in exacerbating these challenges.

These results align closely with existing research on language assessment in EFL contexts. Consistent with Fulcher and Davidson (2007) and Coombe et al. (2020), instructors' perceptions indicate that formal training in assessment principles is crucial for producing valid and reliable tests. Without systematic knowledge of assessment theory and practical training in rubric design, instructors often rely on intuition or previous test formats, which may not adequately reflect students' language proficiency.

Time and workload pressures emerged as another significant factor. The survey data showed that 85% of instructors reported time constraints as a major challenge, while 86% indicated that heavy teaching workloads reduce the time available for careful test development. These findings are consistent with Kunnan (2018), who emphasizes that practical constraints often limit instructors' ability to align assessment tasks with learning objectives, pilot test items, and ensure grading consistency. Large class sizes and mixed proficiency levels further complicate test design, creating difficulties in producing tasks that are both fair and appropriately challenging for all students (Hughes, 2003).

Institutional factors were also critical in shaping assessment practices. A majority of participants noted the lack of clear guidelines and limited professional development opportunities as barriers to consistent, high-quality test design. This supports previous findings that institutional support, including access to resources, clear assessment policies, and opportunities for collaboration, is essential for ensuring fairness and validity (Coombe et al., 2020; Fulcher & Davidson, 2007). Limited access to technological tools was also cited as a

constraint, affecting instructors' ability to develop authentic assessment tasks that reflect real-life language use.

The study further revealed the relationship between assessment quality and student outcomes. Instructors reported that poorly aligned or unfair tests negatively affect student motivation, confidence, and engagement, echoing findings by Stiggins (2017) and Cheng, Rogers, and Hu (2004). Conversely, well-designed assessments were perceived as promoting students' learning and motivation by providing accurate, meaningful feedback and opportunities for self-reflection. This underscores the broader pedagogical impact of assessment design beyond mere grading.

Open-ended responses suggested several avenues for improvement. Instructors emphasized the need for targeted professional development, peer collaboration, and shared assessment frameworks to improve validity and fairness. These recommendations align with the literature on best practices in language assessment, which highlights the importance of collaborative test design, pilot testing, and reflective practice in enhancing assessment quality (Brown & Abeywickrama, 2019; Coombe et al., 2020).

To sum up, the findings suggest that assessment design in EFL contexts is a complex, multifaceted process influenced by instructor knowledge, institutional resources, student diversity, and contextual constraints. Addressing these challenges requires a combination of professional development, clear institutional policies, and collaborative practices to ensure that English language tests accurately reflect learners' abilities while promoting motivation and engagement.

Conclusion

This study examined the challenges English language instructors face in designing fair and valid English language tests for English Language Learners at Abu Issa College. The findings indicate that assessment design is influenced by a complex interaction of pedagogical, institutional, and contextual factors. Instructors identified insufficient training in language assessment, time constraints, heavy workloads, large class sizes, and mixed proficiency levels as major obstacles to developing high-quality assessments.

The results further suggest that limitations in institutional guidelines, professional development opportunities, and assessment resources contribute to inconsistencies in test validity and fairness. Importantly, instructors perceived assessment quality as closely linked to student

motivation and learning outcomes, reinforcing the view that assessment practices play a critical role in shaping learners' academic experiences.

These findings highlight the need for systematic institutional support, including targeted professional development in language assessment, clearer assessment policies, and opportunities for collaboration among instructors. By addressing these areas, higher education institutions can support instructors in designing assessments that more accurately reflect learners' language abilities and promote equitable and meaningful evaluation practices.

While the study is limited to one institutional context, its findings offer insights that may be relevant to similar EFL settings where instructors face comparable challenges. Future research could build on this work by incorporating classroom observations or student perspectives to further enrich understanding of assessment practices in EFL contexts.

References

Brown, H. D., & Abeywickrama, P. (2019). *Language assessment: Principles and classroom practices* (3rd ed.). Pearson Education. Retrieved from <https://discover.library.unt.edu/catalog/b7446609> (Accessed Jan 2026)

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325. <https://doi.org/10.1080/02602938.2018.1463354> (Accessed Jan 2026)

Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: Purposes, methods, and procedures. *Language Testing*, 21(3), 360–389. (SAGE Publications)

Coombe, C., Davidson, P., O'Sullivan, B., & Stoyanoff, S. (2020). *The Cambridge guide to second language assessment* (2nd ed.). Cambridge University Press.

Fulcher, G., & Davidson, F. (2007). *Language testing and assessment*. Routledge.

Harmer, J. (2015). *The practice of English language teaching* (5th ed.). Pearson.

Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge University Press.

Kunnan, A. J. (2018). *Evaluating language assessments*. Routledge.

Richards, J. C., & Schmidt, R. (2010). Longman dictionary of language teaching and applied linguistics (4th ed.). Pearson Education.

Taylor, L., & Weir, C. J. (2012). Examining listening: Research and practice in assessing second language listening (Studies in Language Testing, Vol. 35). Cambridge University Press.

Stiggins, R. (2017). Assessment literacy for educators: Knowledge required and strategies for implementation. (Often cited in assessment literature; locate through institutional library or database if needed.)

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. Educational Assessment, Evaluation and Accountability, 21(1), 5–31.

Carless, D., & Boud, D.'s (2018) article DOI verified via institutional Springer link:
<https://doi.org/10.1080/02602938.2018.1463354> (Accessed Jan 2026)