Postgraduate Students' Experience with Online and AI-Assisted English Testing at the faculty of Arts and Langugaes: Tripoli University

Maeyouf,Narjis[1] & Dredi, Ibtisam[2]

Narjis.maeyouf@gmail.com ibtisamdredi@gmail.com

Department of English, Faculty of Arts and Languages, University of Tripoli, Libya

ABSTRACT

The shift toward online and AI-assisted English testing, especially during and after the COVID-19 pandemic, has changed the way students experience language assessment. This study explores postgraduate students' experiences with online and AI-based English tests at the Faculty of Arts and Languages, University of Tripoli. A mixed-methods approach was used, including a questionnaire completed by 24 students and eight semi-structured interviews. The survey results showed generally positive opinions toward the convenience of online testing and the important role AI may play in future education. Interview findings supported these results and highlighted issues such as internet instability, fear of test interruption, and the belief that AI cannot fully understand individual student performance in the way teachers can. This study concludes that online and AI-assisted tests are useful tools, but they should be applied carefully to ensure fair, reliable, and student-centered assessment practices. Importantly, the findings contribute local empirical evidence from the Libyan higher education context, where infrastructural constraints play a significant role in shaping assessment experiences.

Keywords: Online assessment, Artificial Intelligence (AI), English language testing, postgraduate students, automated scoring

الملخص

أدّى التوجّه المتزايد نحو اختبارات اللغة الإنجليزية عبر الإنترنت والمدعومة بالذكاء الاصطناعي، ولا سيما خلال جائحة كوفيد-19 وما بعدها، إلى إعادة تشكيل تجارب الطلبة مع التقييم اللغوي. وعلى الرغم من أن هذه الاختبارات توفر قدرًا من المرونة وسهولة الوصول وسرعة الحصول على النتائج، فإنها تثير في الوقت نفسه مخاوف تتعلق بعدالة التقييم، والضغط النفسي المصاحب للاختبار، ومستوى الثقة في أنظمة التصحيح الآلي. تهدف هذه الدراسة إلى استكشاف تجارب طلبة الدراسات العليا مع اختبارات اللغة الإنجليزية عبر الإنترنت والمدعومة بالذكاء الاصطناعي في كلية الآداب واللغات بجامعة طرابلس. اعتمدت الدراسة منهجية بحث مختلطة، شملت استبيانًا شارك فيه 24 طالبًا، إضافة إلى ثماني مقابلات شبه منظَّمة. وأظهرت

نتائج الاستبيان اتجاهات إيجابية عامة نحو سهولة استخدام الاختبارات الإلكترونية، إلى جانب إدراك متزايد لأهمية الدور الذي قد يؤديه الذكاء الاصطناعي في مستقبل التعليم. في المقابل، عبّر الطلبة عن مستويات متباينة من الثقة في التصحيح المعتمد على الذكاء الاصطناعي، وارتفاع في مستويات القلق المرتبطة بضيق الوقت والمشكلات التقنية.

الكلمات المفتاحية: الذكاء الاصطناعي; اختبارات ; طلبة الدراسات العليا

Introduction

The rapid growth of digital technology, especially after the COVID-19 pandemic, has brought major changes to language assessment. Many traditional paper-based English tests such as IELTS and TOEFL were moved online or supplemented with remote and technology-based alternatives. One example is the Duolingo English Test (DET), which gained popularity because it is more affordable, flexible, and provides fast results. At the same time, artificial intelligence (AI) tools were introduced into assessment through automated scoring, speaking evaluation systems, and AI-generated feedback for writing tasks.

These developments have reshaped how postgraduate students experience language testing. Students now prepare for tests that involve online platforms, remote proctoring, and interaction with AI systems rather than human examiners alone. This shift influences their confidence, testing strategies, sense of fairness, and trust in score results. While digital assessments offer convenience and accessibility, they also raise new concerns related to internet access, technical problems, privacy, and the accuracy of machine-generated scoring. For many postgraduate students, whose academic futures depend heavily on test results, these issues are especially important. However, there is still limited research that focuses directly on how these students experience AI-assisted testing in real situations. This study investigates postgraduate students' perceptions, emotional experiences, and evaluations of fairness in online and AI-assisted English language testing, and contributes to institutional discussions on assessment practices at the University of Tripoli by foregrounding postgraduate students' voices. Although existing research has examined the technical reliability and validity of online English tests, there is limited attention to the actual experiences of postgraduate test-takers. Issues such as fairness, digital accessibility, test anxiety,

authenticity of tasks, and trust in automated scoring have not been sufficiently explored from a student perspective. Without a clearer understanding of how postgraduates experience AI-assisted language assessment, universities and test providers may rely on systems that do not fully meet students' academic and emotional needs.

Research Questions

1. How do postgraduate students perceive the fairness, validity, and usefulness of online and AI-assisted English language tests?
2. What are students' experiences when preparing for and taking these assessments?
3. How do online and AI-based test formats affect students' emotions, including anxiety, confidence, and motivation?

Research Problem

Although existing research has examined the technical reliability and validity of online English tests, there is limited attention to the actual experiences of postgraduate test-takers. Issues such as fairness, digital accessibility, test anxiety, authenticity of tasks, and trust in automated scoring have not been sufficiently explored from a student perspective. Without a clearer understanding of how postgraduates experience AI-assisted language assessment, universities and test providers may rely on systems that do not fully meet students' academic and emotional needs.

This study argues that while online and AI-assisted English tests provide important advantages such as flexibility and rapid scoring, they also introduce new challenges that shape postgraduate students' test experiences, perceptions of fairness, and learning outcomes. Understanding these student perspectives is essential for developing ethical, inclusive, and meaningful language assessment practices in the digital age.

Research Aim

This study focuses on postgraduate students who have recently taken online or AI-assisted English language proficiency tests. Using surveys and follow-up interviews, data is collected to explore students' perceptions, emotional responses, and overall experiences with the assessment process.

The importance of this research lies in its emphasis on student voices within discussions of digital assessment. Much of the current literature highlights technical performance indicators, while fewer studies discuss how students feel about or engage with these assessments. By addressing this gap, the study offers insights that may inform educators, policymakers, and testing organizations in developing more ethical, accessible, and student-centered assessment systems.

Literature Review

The widespread adoption of digital testing in higher education accelerated during the COVID-19 pandemic and has since become a permanent feature of language assessment (Khalil et al., 2020; Tejada & Gallardo, 2021). Traditional exams such as IELTS and TOEFL were disrupted due to test-center closures, prompting universities to accept online alternatives (Chapelle, 2020). Among these, the Duolingo English Test (DET) gained recognition because of its lower cost, flexible scheduling, and rapid score reporting (Duolingo, 2023). Unlike center-based exams, the DET allows candidates to take the test remotely and complete all assessment components on a digital platform (Hill et al., 2022). This convenience has made it particularly appealing to postgraduate applicants who often face tight deadlines and logistical constraints.

However, the expanding use of online testing has introduced concerns regarding test security and validity (Elkhatat et al., 2021). Remote proctoring systems using webcams, facial-recognition software, and behavior-tracking technologies have been adopted to protect exam integrity (Khalil et al., 2020). While these systems aim to prevent cheating, they also create new challenges for test-takers, including privacy concerns, fear of surveillance, and anxiety related to technical malfunctions (Russell & Kavanaugh, 2019). Research suggests that such issues may introduce factors unrelated to actual language ability that influence performance (Tejada & Gallardo, 2021).

Students' perceptions of fairness and validity are critical when evaluating digital tests (O'Sullivan, 2012; Weir, 2005). Some studies report that test-takers appreciate the flexibility of online exams but still question whether these assessments accurately represent real academic language use (Tejada & Gallardo, 2021). Research on the DET shows mixed outcomes: while students feel the test provides equal access, many believe it does not fully

capture the communicative skills required for university study (Hill et al., 2022). Furthermore, findings from predictive validity studies indicate only weak associations between DET scores and students' academic success, particularly at undergraduate level (Hill et al., 2022). These results suggest that continued evaluation of online assessments is necessary to ensure accurate score interpretation (Chapelle, 2020).

Test anxiety has also emerged as an important issue in digital assessment environments (Russell & Kavanaugh, 2019). Remote testing can increase stress because students may feel uncomfortable being monitored by invisible systems or fear technical problems beyond their control (Elkhatat et al., 2021). At the same time, some studies demonstrate that well-designed AI-assisted tools may reduce anxiety by offering low-pressure environments and immediate feedback (Fathi & Ebadi, 2020). For example, online writing platforms using automated feedback have been shown to improve motivation and confidence, even when performance outcomes did not differ significantly from traditional testing formats (Gao & Zhang, 2020). These findings suggest that technology, when used thoughtfully, can support emotional well-being as well as academic engagement (Russell & Kavanaugh, 2019).

Artificial intelligence has expanded its role within language assessment beyond scoring to include feedback generation and practice support tools (Stevenson & Phakiti, 2019). Automated Writing Evaluation (AWE) systems allow students to receive immediate feedback on grammar and spelling but tend to overlook higher-level writing features such as coherence, argument development, and creativity (Warschauer & Ware, 2006; Stevenson & Phakiti, 2019). Although newer AI models generate more natural language feedback, challenges remain related to accuracy, depth, and reliability (Gao & Zhang, 2020).

In speaking assessment, AI tools demonstrate promising efficiency but limited ability to evaluate complex communicative tasks (Xi, 2020). Research comparing AI scoring with teacher ratings indicates that automated systems perform best with controlled tasks such as reading aloud, while open-ended speech tasks struggle to be assessed accurately (Knoch & McNamara, 2015; Xi, 2020). Students report increased awareness of pronunciation issues when using AI speaking tools; however, they identify a lack of meaningful feedback as a limitation (Xi, 2020). When human peer feedback was integrated alongside AI assessment,

students achieved stronger speaking improvements, supporting the need for blended assessment approaches (Knoch & McNamara, 2015). Generative AI tools have also entered the assessment preparation space, raising concerns about test security and ethical boundaries (Shin & Gweon, 2021). Scholars argue that institutions must clearly define acceptable AI use during test preparation and high-stakes assessments (Brown & Abeywickrama, 2019). While AI support in low-stakes learning activities may be beneficial, its use during exams risks undermining test validity (Stevenson & Phakiti, 2019).

Overall, the literature presents AI-mediated assessment as a process characterized by both opportunity and uncertainty (Chapelle, 2020; Weir, 2005). While digital tools improve accessibility, feedback speed, and scoring efficiency, they also raise persistent questions related to fairness, emotional stress, and score interpretation (O'Sullivan, 2012; Russell & Kavanaugh, 2019). Studies increasingly stress the importance of considering both behavioral engagement how students interact with technology and perceptual engagement how they feel about it (Shin & Gweon, 2021).

Although previous studies have contributed valuable insights into online and AI-assisted language assessment, several limitations remain in current research. Much of the research has focused primarily on technical aspects such as test validity, reliability, and score comparability with traditional exams (Chapelle, 2020; Weir, 2005). While these issues are important, fewer studies have examined how students themselves experience these assessments, particularly in terms of emotional pressure, perceived fairness, and trust in automated scoring systems.

In addition, many studies have concentrated on undergraduate populations or large-scale international testing contexts, leaving postgraduate students relatively underrepresented. This is a significant gap, as postgraduate learners often face higher academic and professional stakes, including admission requirements, scholarship opportunities, and career advancement. As a result, their testing experiences may involve greater levels of anxiety, responsibility, and concern about score accuracy (Hill et al., 2022).

Moreover, much of the existing research has been conducted in contexts with stable technological infrastructure, where access to reliable internet and digital resources is less problematic. Fewer studies have explored how

online and AI-assisted testing is experienced in contexts where technological challenges, such as internet instability and limited access to equipment, are common (Elkhatat et al., 2021). These contextual factors can strongly influence students' perceptions of fairness and their emotional responses during testing.

Finally, while some studies rely on quantitative data to report test outcomes, and others use qualitative methods to explore perceptions, fewer investigations combine both approaches. A mixed-methods perspective allows for a more complete understanding of students' experiences by linking statistical trends with personal accounts and explanations (Shin & Gweon, 2021). Addressing these gaps is necessary to develop assessment practices that are not only efficient and innovative, but also fair, reliable, and sensitive to students' needs.

This study responds to these gaps by centering postgraduate learners' voices. These students often experience particularly high stakes due to program admissions requirements or professional obligations (Hill et al., 2022). By examining their experiences with preparation, performance, and emotional adjustment in AI-assisted testing environments, this research aims to provide insights that move beyond performance numbers and into lived realities. Ultimately, advancing ethical and effective language assessment reform requires collaboration between pedagogy and technology, ensuring that innovation supports, rather than undermines, student confidence and learning (Brown & Abeywickrama, 2019; Chapelle, 2020).

Methodology

This study adopted a mixed-methods approach, combining quantitative survey data with qualitative interview responses. This design was chosen to gain both broad numerical insights into students' perceptions of online and AI-assisted English testing and deeper understanding of their personal experiences. While surveys allowed the researcher to identify common patterns related to fairness, anxiety, and perceived usefulness, interviews provided further explanation of students' emotional responses, concerns, and suggestions for improvement.

Using both methods helped ensure that the study captured a more complete picture of postgraduate students' engagement with digital assessments rather than relying only on statistical results or personal narratives alone.

Two data collection instruments were used in this study:

1. **Online Questionnaire**

A questionnaire was distributed to postgraduate students who had previously taken online or AI-assisted English language proficiency tests. The questionnaire included both closed-ended items using Likert-scale responses and open-ended questions. The close-ended items measured: Students' perceptions of fairness and validity, Levels of anxiety and confidence during testing, Opinions about the usefulness of AI feedback and digital platforms. The open-ended questions allowed participants to describe their test experiences in their own words and provide suggestions for improving the assessment systems.

2. **Semi-Structured Interviews**

Follow-up interviews were conducted with selected participants to gain deeper insight into their testing experiences. Interview questions focused on: Emotional challenges during test preparation and performance, Experiences with remote proctoring and AI scoring, Trust in test results, Views on how online tests compare to traditional exams

Interviews were conducted online and recorded with participants' consent to ensure accuracy during analysis.

The participants were postgraduate students from the Faculty of Arts and Languages at the University of Tripoli who had experience taking AI-assisted or online English language tests for academic purposes. Students from different age groups participated to ensure a range of perspectives.

Participation was voluntary. All students were informed that their responses would be used strictly for research purposes and that no identifying information would be recorded. Pseudonyms were used when reporting interview excerpts.

Given the exploratory aim of the study, convenience sampling was considered appropriate to capture initial insights rather than generalizable outcomes.

Quantitative data collected from the questionnaire were analyzed using descriptive statistical techniques. Descriptive statistics were selected because the aim of the study was to identify general trends and patterns in postgraduate students' perceptions rather than to test hypotheses or make predictions. Frequencies, percentages, and mean scores were calculated to summarize participants' responses related to perceived fairness, levels of

anxiety, trust in AI scoring, and overall satisfaction with online and AI-assisted English language tests. This approach allowed for a clear and accessible representation of students' views, which was appropriate given the relatively small sample size.

Likert-scale responses were interpreted by grouping agreement levels to identify overall tendencies. Responses such as "agree" and "strongly agree" were considered indicators of positive perception, while "disagree" and "strongly disagree" reflected negative perception. Neutral responses were analyzed separately, as they were interpreted as indicating uncertainty, limited experience, or mixed feelings rather than clear approval or rejection. This interpretation strategy helped ensure that students' attitudes were represented accurately without oversimplifying their responses.

Qualitative data from open-ended questionnaire items and semi-structured interviews were analyzed using thematic analysis. All responses were read multiple times to ensure familiarity with the data and to identify recurring ideas. Initial codes were generated by grouping similar statements related to fairness, emotional experience, technological challenges, trust in AI scoring, and perceived benefits of online testing. These codes were then compared across different participants and data sources to ensure consistency. Themes were refined by reviewing how frequently they appeared and how clearly they related to the research questions.

The use of both quantitative and qualitative analysis methods allowed the study to address the research questions comprehensively. While descriptive statistics provided an overview of common perceptions and trends, qualitative analysis offered deeper insight into students' experiences and explanations. This integration strengthened the validity of the findings by linking numerical patterns with personal accounts and contextual factors.

Ethical Considerations

Ethical principles were strictly observed throughout the research process. Participants were informed of the purpose of the study and provided informed consent before completing the questionnaire or participating in the interviews. Participation was entirely voluntary, and participants were informed of their right to withdraw from the study at any stage without any penalty.

All collected data were treated with strict confidentiality. No identifying information, including participants' names or personal details, was recorded, and all responses were stored securely and used solely for research purposes. The study focused exclusively on students' perceptions and experiences and did not involve the collection of academic records or official test scores, thereby minimizing potential privacy concerns.

The final sample size was smaller than initially planned due to limited participant availability and voluntary non-participation. Some postgraduate students were not interested or declined to complete the questionnaire, which reduced the total number of responses. All participation decisions were fully respected in accordance with ethical research standards.

Artificial intelligence tools were used in this study solely as supportive research aids. These tools assisted in language editing to enhance clarity and coherence and in organizing the structure of the analysis. AI tools were also used to generate initial drafts of the survey and interview questions, which were carefully reviewed, revised, and validated by the researchers prior to data collection. In addition, AI-assisted transcription tools were used to support the transcription of recorded interview data. All AI-generated transcripts were thoroughly reviewed, corrected, and verified by the researcher against the original audio recordings to ensure accuracy. No AI tools were used for data analysis or interpretation; all research design decisions, data collection, coding, thematic analysis, and interpretation were conducted entirely by the researchers.

Results and Discussion

This study explored postgraduate students' perceptions of online and AI-assisted English language testing through survey responses (N = 24) and eight semi-structured interviews. The results indicate that students hold generally positive attitudes toward the convenience and usefulness of online assessment formats, yet express notable concerns regarding fairness, emotional stress, technological reliability, and the limitations of automated scoring.

The combined findings reveal a complex picture: while participants acknowledged the advantages of flexibility, accessibility, and fast feedback, many remained cautious about the accuracy of AI scoring and

the emotional impact of time pressure and unstable internet connections, particularly within the Libyan context.

Survey Results

1. Trust in Online and AI-Based English Tests

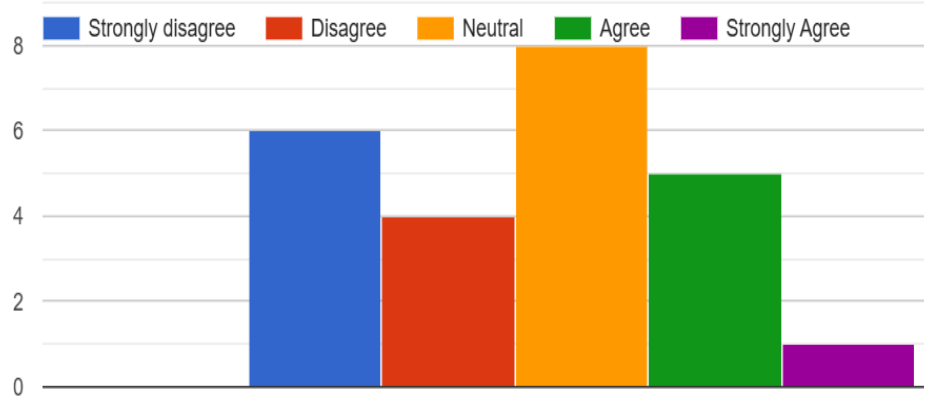

Figure 1. Postgraduate students' trust in online and AI-assisted English language test results (N = 24).

Based on the results in figure (1) above, it illustrates that the students trust the results of online and AI-assisted English tests, responses were mixed. Approximately 45.8% of participants (11 out of 24) expressed either agreement or strong agreement, indicating moderate confidence in digital scoring systems. However, 33.3% of respondents (8 participants) chose neutral options, while 20.8% (5 participants) selected disagreement, suggesting uncertainty rather than full trust.

This divided opinion reflects the tension between students' recognition of technological efficiency and their doubts about the reliability of machine-based evaluation. Several neutral respondents later clarified in open-ended comments that their uncertainty stemmed not from opposition to technology, but from limited personal experience or doubts about technical consistency during assessment.

2. *Preference for Online Tests versus Traditional Paper Exams*

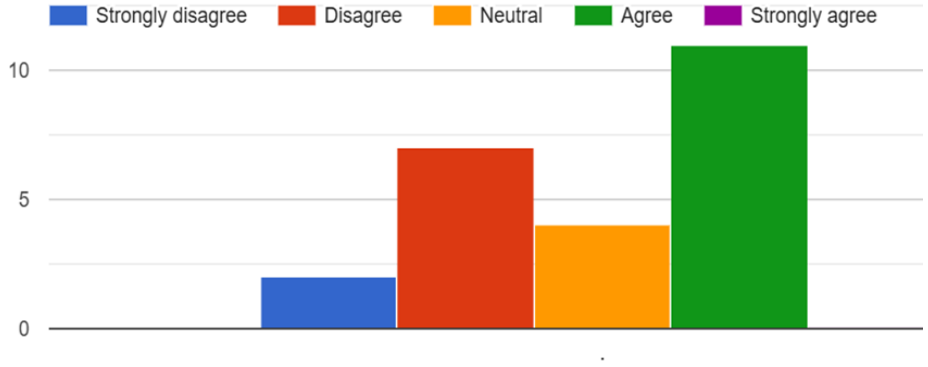

Figure 2. Students' preferences for online versus traditional paper-based English language exams.

Figure (2) shows that participants were not unanimous in preferring online examinations over traditional paper-based tests. A total of 45.8% of respondents (11 participants) indicated a preference for online tests, while 29.2% (7 participants) preferred traditional paper-based exams. Neutral responses accounted for 16.7% (4 participants), and 8.3% (2 participants) expressed a strong preference for online formats.

While some respondents favored online testing due to comfort, convenience, and reduced pressure from physical exam halls, others leaned toward traditional exams due to familiarity, perceived control, and trust in human grading.

Neutral responses were common, suggesting that many students view both formats as serving valid yet different purposes. Rather than perceiving online tests as full replacements for traditional exams, many participants treated them as complementary alternatives. This finding was echoed frequently in interview responses.

2. Future Role of AI-Assisted Testing

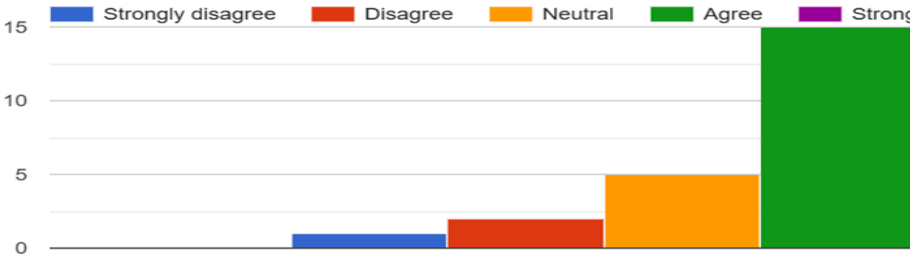

Figure 3. Students' views on the future importance of AI-assisted testing in education.

Figure (3) illustrates that a strong majority of participants agreed that AI-assisted testing will be important in the future of education. Approximately 70.8% of respondents (17 participants) selected agreement or strong agreement, while 20.8% (5 participants) remained neutral. Only 8.4% (2 participants) disagreed with this statement. This indicates widespread acknowledgment that technology will play an increasing role in language assessment.

However, agreement on the importance of AI did not imply unconditional approval. Instead, participants believed that while AI assessment tools are inevitable and valuable, they require careful regulation, ethical frameworks, and clearer guidelines to ensure fairness and accuracy.

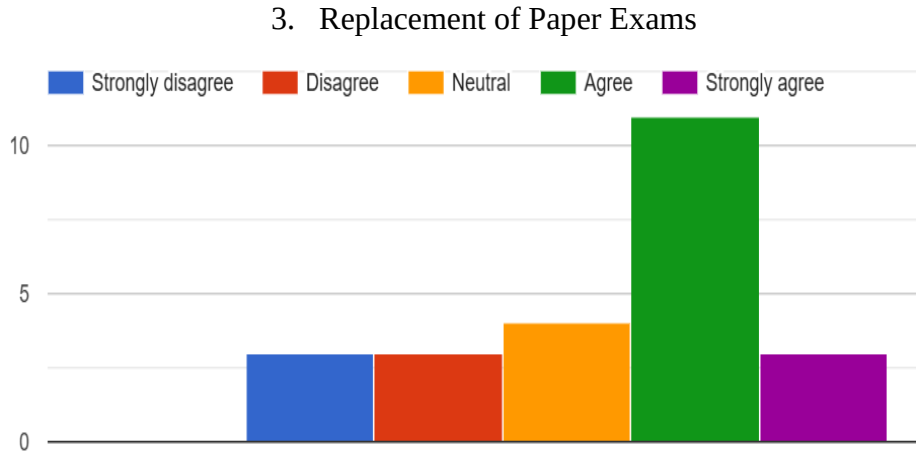

Figure 4. Student perceptions of whether online testing can replace traditional paper-based exams.

Based on figure (4) above shows students' perceptions of whether online testing can replace traditional paper-based exams in the future, responses showed mixed views. About 45.8% of respondents (11 participants) supported the idea of replacement, while 33.3% (8 participants) selected neutral responses. A further 20.9% (5 participants) disagreed, expressing skepticism toward full replacement.

Many respondents explained that replacement would be appropriate only for certain skill areas particularly grammar, vocabulary, and reading comprehension while skills such as speaking and writing would still require greater human involvement. This supports the view that students favor blended assessment models where technology complements rather than replaces teacher evaluation.

Open-Ended Survey Responses

The open-ended survey items provided rich qualitative insights. Many respondents reported positive experiences, describing online tests as : "Comfortable and flexible."

"Convenient because I can take them anytime and anywhere." "Helpful due to quick results and detailed feedback." "Motivating because I can track my progress easily."

Several participants emphasized the effectiveness of AI feedback particularly in grammar, pronunciation, and error correction.
Conversely, negative responses focused primarily on time pressure and ambiguity in automated marking: Some respondents described confusion regarding questions that appeared to have more than one correct answer. Others mentioned feeling anxious due to strict timers or fear of submission errors. Several students stated that they had no direct experience with online testing and therefore felt unable to express full trust.
Cheating concerns emerged from educators among the respondents, with one teacher stating: "Most of the students cheated."
This comment highlights ongoing issues related to academic integrity in unsupervised or weakly proctored digital assessments.

Interview Results
Interview data expanded upon the survey findings and produced five main thematic categories:

1. **Perceptions of Fairness**
Opinions on fairness were mixed. Several participants reported positive experiences, stating that: The content was aligned with course material, Instructions were clear, AI scoring matched their expectations and perceived proficiency levels. "The test was fair. The questions were clear and covered what we studied." (Participant 1)
"The result made sense to me and matched my real level." (Participant 5)
Conversely, other participants expressed concerns:
"The test wasn't fair; the questions didn't match my level." (Participant 8)
"I don't think AI can understand students' real abilities like humans do." (Participant 3)
Participants feared that automated systems could misunderstand answers or fail to distinguish contextual meaning, particularly in writing tasks.
2. Emotional Experience and Test Anxiety
Almost all interviewees described some degree of anxiety before or during online testing. The most frequently mentioned triggers were: Timer pressure, Fear of technical failure, Internet instability, Remote monitoring, Participant comments included:
"Seeing the timer stressed me more than the questions." (Participant 5)
"I was afraid the internet would disconnect." (Participant 1)

"Technology affected my mood because of the timer." (Participant 8)
Group settings or prior familiarity with testing platforms were reported to lessen anxiety: "Working in groups helped me relax." (Participant 6)
"The questions were gradual, so I felt comfortable." (Participant 3)
3. Benefits of Online Testing
Students highlighted several advantages: Accessibility: Ability to test from any location, Efficiency: Immediate scoring and feedback, Practice support: Especially for grammar and pronunciation, Flexibility: Freedom to schedule testing sessions. "They are convenient and easy to access anytime." (Participant 7)
"Quick feedback helps me understand mistakes." (Participant 1)
These characteristics were especially valued by students preparing independently without access to expensive test preparation services.
4. Technological Challenges
Participants repeatedly noted the fragility of technical infrastructure in Libya: Internet instability, Lack of appropriate equipment, Risk of sudden disconnection or system resets. Participant feedback included:
"In Libya most learners suffer from lack of modern equipment." (Participant 2)
"If the internet disconnects, progress resets to zero." (Participant 3)
Student's feared losing completed work or failing unfairly due to problems outside their control.
5. Trust in AI Scoring and Human Judgment
Participants widely expressed respect for teacher grading as more understanding and motivational: "Teachers see the student's struggle and try to motivate." (Participant 3)
While AI scoring was valued for objectivity, students believed it lacked emotional sensitivity and the ability to apply flexible judgment in borderline cases.
These quantitative and qualitative results together reveal several important patterns.

Interpretation of Findings

The findings demonstrate that students are not opposed to AI-assisted assessment itself. Instead, their concerns focus on contextual conditions that influence perceptions of fairness and emotional experience. Technology appears to amplify existing educational inequalities,

particularly in contexts where internet accessibility and technical reliability vary widely between students. As a result, students' acceptance of AI-assisted testing is closely tied to the conditions under which these assessments are implemented rather than to the technology alone.

The results further indicate that trust in AI-assisted assessment, test anxiety, and perceptions of fairness are closely interconnected rather than independent factors. Students who expressed lower levels of trust in automated scoring systems often reported higher levels of anxiety during testing. This anxiety was not solely related to language ability, but to concerns about whether AI systems could accurately interpret their responses, particularly in writing and speaking tasks. When students doubted the accuracy or flexibility of automated scoring, their emotional stress increased and their overall perception of test fairness declined.

At the same time, perceptions of fairness were shaped primarily by testing conditions rather than by resistance to AI technologies. Many participants did not reject AI-assisted testing in principle; instead, they emphasized that fairness depended on stable technical systems, clear instructions, and reasonable time limits. When these conditions were met, students were more likely to trust test results and report positive experiences. This suggests that emotional responses such as anxiety can be reduced through improved test design and supportive assessment practices rather than through the removal of AI from assessment.

The local context of Libya plays a significant role in shaping these experiences. Unlike testing environments with reliable digital infrastructure, students in Libya frequently face internet instability, power interruptions, and limited access to appropriate devices. These challenges increase the perceived risk associated with online testing and intensify stress during high-stakes assessments. In this context, technical failures are not viewed as minor inconveniences but as factors that may directly affect academic outcomes, which helps explain why some students questioned the fairness of online testing despite acknowledging its advantages.

Furthermore, postgraduate students face particularly high stakes in English language assessment, as test results may influence admission decisions, degree completion, and professional advancement. This pressure amplifies emotional responses to testing conditions and increases sensitivity to issues of fairness and reliability. The findings therefore highlight the importance of context-aware assessment practices that account for both

technological realities and the academic pressures faced by postgraduate learners.

These findings are consistent with international research reporting mixed student attitudes toward digital testing, where convenience and speed are valued alongside concerns related to anxiety, privacy, and score validity. The preference for blended assessment models observed in this study aligns with existing recommendations that emphasize collaboration between AI systems and human educators rather than full automation. Based on these interpretations, several practical implications emerge:

- Infrastructure investment is essential for equitable assessment.
- Flexible timing systems may reduce test-related anxiety.
- Clear instruction design can minimize confusion and distrust.
- Teacher oversight should remain part of high-stakes evaluation.
- Blended assessment frameworks offer greater reliability and perceived fairness.

Overall, postgraduate students view online and AI-assisted English testing as a useful but imperfect environment. While convenience and rapid feedback are widely appreciated, concerns related to fairness, reliability, and emotional impact remain significant. The effectiveness of AI-supported assessment therefore depends on ethical application, technical stability, and pedagogical integration that prioritizes student-centered evaluation. The findings of this study are largely consistent with previous research on online and AI-assisted language assessment. Similar to earlier studies, participants expressed positive attitudes toward the flexibility, accessibility, and fast feedback provided by online tests (Khalil et al., 2020; Tejada & Gallardo, 2021). As reported by Hill et al. (2022) and Chapelle (2020), students valued the convenience of remote testing but remained cautious about the ability of AI systems to accurately assess complex language skills .

Concerns related to test anxiety, time pressure, and technical problems align with findings from Russell and Kavanaugh (2019) and Elkhatat et al. (2021), who noted that technology-based assessments can increase stress due to monitoring and system instability. At the same time, the preference for blended assessment models observed in this study supports previous research emphasizing the importance of combining AI efficiency with human judgment, particularly in speaking and writing assessment (Knoch & McNamara, 2015; Xi, 2020). Overall, the results support existing

literature while highlighting the impact of local technological conditions on postgraduate students' experiences.

The findings suggest that online and AI-assisted English tests can be effective tools when implemented within supportive and ethical frameworks. Educational institutions may benefit from using digital assessments to improve accessibility and efficiency, especially for low-stakes or placement testing. However, concerns about fairness, anxiety, and technical reliability indicate the need for flexible timing, clear instructions, and adequate technical support.

The results also highlight the importance of maintaining teacher involvement in assessment processes. Blended assessment models that combine AI scoring with human evaluation may enhance trust, fairness, and accuracy, particularly for productive language skills. Finally, the study emphasizes the need for clear ethical guidelines governing AI use in assessment to ensure transparency, data privacy, and student-centered evaluation practices.

Conclusion

This study examined postgraduate students' experiences with online and AI-assisted English language testing in terms of perceived fairness, emotional impact, usefulness, and trust in automated assessment systems. The findings indicate that students generally acknowledge the benefits of digital testing, including convenience, rapid results, flexible scheduling, and opportunities for continuous practice. Many participants appreciated immediate AI feedback and the accessibility of tests such as Duolingo, EF SET, and British Council placement exams.

However, these advantages were consistently balanced by concerns regarding fairness, emotional pressure, and technological reliability. Anxiety caused by strict time limits, fear of internet disruption, difficulties using unfamiliar platforms, and lack of physical test interaction emerged as common issues. Several students expressed uncertainty about fully automated scoring, noting that AI systems cannot account for context, effort, or learning progress in the same way that teachers can. These findings highlight the need for blended assessment approaches that combine the efficiency of AI with the professional judgment of human examiners.

Regarding the first research question, postgraduate students showed mixed perceptions of fairness and validity. While many believed that question design was appropriate and scoring procedures were objective, fairness was challenged by uneven access to technology, unstable internet connectivity, and rigid testing conditions that may disadvantage some test-takers.

In response to the second question, students' preparation strategies were similar to those used for traditional exams, although increased emphasis was placed on digital practice tools and familiarization with test platforms. Their testing experiences were shaped not only by language ability but also by technical readiness and comfort with online environments.

The third research question revealed a strong emotional dimension to online testing. Time pressure, surveillance technologies, and technical uncertainty increased anxiety levels for many participants, even among those confident in their academic ability. At the same time, some students reported increased motivation and curiosity when receiving immediate feedback and using practice platforms supported by AI.

This study faced several limitations. The sample size was relatively small and focused on postgraduate students from a single university, which limits the generalizability of the findings. Not all participants had direct experience with full AI-assisted testing, with some relying on mock tests or online quizzes instead of formal proficiency exams. Additionally, online data collection depended on self-reported experiences, which may include personal bias or incomplete recall.

Future studies should broaden participant samples across universities and academic disciplines to provide more representative findings. Longitudinal research could explore how continuous exposure to AI-assisted assessment affects language development, emotional resilience, and learning autonomy. Comparative studies examining score reliability between AI-only assessment systems and blended human-AI evaluation models are also recommended. Furthermore, research should investigate the ethical implications of surveillance-based proctoring systems, focusing on data privacy and student well-being. Finally, more studies are needed to examine teachers' perspectives on integrating AI into assessment and how educator training can support ethical and effective digital testing practices.

References

Algwil, K. (2025). The importance of artificial intelligence writing tools in enhancing English writing of EFL students. *Faculty of Languages Journal – Tripoli, Libya*, 1(32).

Brown, H. D., & Abeywickrama, P. (2019). *Language assessment: Principles and classroom practices* (3rd ed.). Pearson Education.

Chapelle, C. A. (2020). *Assessing language through computer technology* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/9781108392047

Duolingo. (2023). *Duolingo English Test: Research and validity overview*. https://englishtest.duolingo.com/research

Elkhatat, A. M., Elsaid, K., & Almeer, S. (2021). Students' attitudes and challenges toward online exams during the COVID-19 pandemic. *Journal of Computer Assisted Learning*, 37(2), 452–463. https://doi.org/10.1111/jcal.12521

Fathi, J., & Ebadi, S. (2020). Exploring EFL learners' experiences of automated written corrective feedback. *Computer Assisted Language Learning*, 33(5–6), 1–26. https://doi.org/10.1080/09588221.2019.1598863

Gao, X., & Zhang, L. J. (2020). Teacher and student beliefs about automated written corrective feedback. *System*, 91, 102244. https://doi.org/10.1016/j.system.2020.102244

Hill, K., O'Sullivan, B., & Rasti, J. (2022). Predictive validity of the Duolingo English Test for academic success in U.S. institutions. *Educational Measurement: Issues and Practice*, 41(1), 20–34. https://doi.org/10.1111/emip.12452

Khalil, R., Mansour, A. E., Fadda, W. A., et al. (2020). The sudden transition to online assessment during the COVID-19 pandemic. *BMC Medical Education*, 20, Article 448. https://doi.org/10.1186/s12909-020-02393-5

Knoch, U., & McNamara, T. (2015). Assessing speaking. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 123–141). De Gruyter Mouton.

O'Sullivan, B. (2012). *Language testing: Theories and practices*. Palgrave Macmillan. https://doi.org/10.1057/9780230353466

Russell, M., & Kavanaugh, R. (2019). Test anxiety in technology-based assessment environments. *Assessment in Education: Principles, Policy & Practice*, 26(5), 482–499. https://doi.org/10.1080/0969594X.2018.1505537

Shin, D., & Gweon, G. (2021). The social implications of AI assessment systems in education. *Computers & Education*, 172, 104263. https://doi.org/10.1016/j.compedu.2021.104263

Stevenson, M., & Phakiti, A. (2019). Automated scoring in second language assessment: A critical review. *Language Testing*, 36(4), 565–591. https://doi.org/10.1177/0265532219885221

Tejada, M. R., & Gallardo, A. V. (2021). Students' perceptions of online testing fairness during the COVID-19 pandemic. *European Journal of Educational Research*, 10(3), 1317–1334. https://doi.org/10.12973/eu-jer.10.3.1317

Warschauer, M., & Ware, P. (2006). Automated writing evaluation: Defining the classroom research agenda. *Language Teaching Research*, 10(2), 157–180. https://doi.org/10.1191/1362168806lr190oa

Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Palgrave Macmillan.

Xi, X. (2020). AI and automated scoring in speaking assessment. *Language Testing*, 37(4), 567–586. https://doi.org/10.1177/0265532220945220