



INTERNATIONAL JOURNAL OF TRENDS IN EMERGING RESEARCH AND DEVELOPMENT

INTERNATIONAL JOURNAL OF TRENDS IN EMERGING RESEARCH AND DEVELOPMENT

Volume 4; Issue 3; 2026; Page No. 172-175

Received: 10-02-2026

Accepted: 21-04-2026

Published: 27-05-2026

The Future of English Language Assessment: Adapting Testing Methods for Online Learning Environments

¹Rohit Babulal Masule and ²Dr. Rohit Kumar

¹Research Scholar, Department of English, Monad University, Hapur, Uttar Pradesh, India

²Assistant Professor, Department of English, Monad University, Hapur, Uttar Pradesh, India

DOI: <https://doi.org/10.5281/zenodo.21961741>

Corresponding Author: Rohit Babulal Masule

Abstract

The sudden switch to online learning brought on by the COVID-19 pandemic has completely transformed the way we teach and evaluate English language proficiency. It has put traditional assessment methods, which were designed for controlled, in-person settings, to the test in ways they were never meant to handle. This study dives into the effectiveness, challenges, and equity issues surrounding online English language assessment, using a convergent mixed-methods approach. We combined a quantitative survey involving 250 educators and students with semi-structured qualitative interviews from 15 educators and assessment experts. Surprisingly, participants rated online assessments as somewhat more effective than a neutral baseline ($M = 3.42$, $SD = 1.05$, $t = 6.32$, $p < .001$), although this positive feedback wasn't evenly spread out. Access to technology turned out to be the strongest and most consistent factor influencing both perceived effectiveness and equity concerns ($r = .52$ and $r = -.61$ respectively, both $p < .001$), with significant regional differences showing that North America and Europe had the upper hand compared to South Asia and Sub-Saharan Africa. Opinions on AI-driven grading tools were mixed, landing close to neutral, and perceptions of their accuracy were inversely related to concerns about bias ($r = -.41$, $p < .001$). On the other hand, formative assessment practices were the most positively received innovation, with 75.6% of respondents finding them helpful or very helpful. These results suggest that the future of online English language assessment hinges less on the complexity of any one technology and more on bridging the technological access gap that dictates who gets to benefit from it. The paper wraps up with practical, evidence-based recommendations for educators, institutions, and policymakers.

Keywords: English language assessment, online learning, computer-adaptive testing, automated grading, formative assessment, educational equity, digital divide

1. Introduction

The COVID-19 pandemic brought about a swift and largely unexpected shift from in-person teaching to online learning in educational systems around the globe, fundamentally altering the way language skills are taught and evaluated. Traditional English language tests like IELTS and TOEFL were built around standardized, proctored formats that rely on controlled testing environments, direct supervision, and consistent access to technology (Weir, 2005) ^[12]. However, these assumptions often fall apart in digital learning settings. Alderson (2009) ^[2] points out that standardized tests are inherently too rigid to cater to the diverse proficiency levels

and needs of individual test-takers, a limitation that becomes even more significant online, where learners often struggle with unequal access to devices, internet connectivity, and digital skills (Martin, 2020) ^[8].

At the same time, digital platforms have introduced tools that simply weren't available in paper-based testing: for example, computer-adaptive testing (CAT), which adjusts the difficulty of questions based on real-time performance (Linacre, 2018) ^[7], and AI-driven grading systems that can evaluate written and spoken responses on a large scale (Shermis & Burstein, 2013) ^[10]. While these innovations haven't completely resolved the existing tensions in the

field, they have certainly reframed them. For instance, automated speech recognition continues to spark concerns about its ability to accurately interpret non-standard speech and cultural nuances (Zhang & Lu, 2019)^[14].

1.1 This study is guided by five hypotheses drawn from existing literature

- H1:** Current methods of English language assessment are less effective in online learning environments compared to traditional, in-person assessments.
- H2:** Educators and students encounter significant challenges, including technological barriers and issues of test integrity, when transitioning traditional assessments to online platforms.
- H3:** The use of AI-driven tools greatly enhances the accuracy and efficiency of online English language assessments.
- H4:** Learners from disadvantaged backgrounds face greater obstacles in accessing and excelling in online assessments.

2. Literature Review

2.1 Traditional Assessment and Its Limitations

The modern assessment of the English language has evolved around the key ideas of validity and reliability (Weir, 2005)^[12], leading to the creation of standardized tests like IELTS and TOEFL that follow a consistent, proctored format. However, Alderson (2009)^[2] points out a downside: these fixed-form tests struggle to reflect the diverse proficiency levels, cultural backgrounds, and first-language influences of all test-takers. Additionally, Horwitz (2016)^[6] highlights that the high-stakes nature of traditional testing can lead to significant anxiety that distorts performance. Chapelle and Douglas (2006)^[5] further emphasize that these challenges, along with the resource demands (Shermis & Burstein, 2013)^[10] and the limited ability to provide real-time feedback (Black & Wiliam, 2009)^[3], have created ongoing pressure for technological solutions, even before the pandemic hit.

2.2 The Shift Online and Its Technological Responses

Martin (2020)^[8] describes the shift to online assessment during the pandemic as more of an emergency move than a well-researched transition. This shift highlighted the disparities in access to devices and internet connectivity (Blake, 2013)^[4] and removed the controlled physical environments that traditional testing depended on. In light of this, computer-adaptive testing has been proposed as a solution to the limitations of fixed-form assessments (Linacre, 2018)^[7], while automated grading aims to tackle issues related to scalability (Shermis & Burstein, 2013)^[10]. Additionally, formative assessments that focus on feedback are being considered as a better alternative to high-stakes, one-time testing (Reinders & Balcikanli, 2011)^[9]. However, these solutions come with their own set of challenges: Zhang and Lu (2019)^[14] and Yu (2021)^[13] point out ongoing bias issues in automated speech recognition and AI grading, and Wang and Brown (2021)^[11] warn that the excitement around AI-based assessments has outstripped the evidence needed to support their use in high-stakes situations.

2.3 Equity as a Cross-Cutting Concern

Blake (2013)^[4] and Al-Mahrooqi and Denman (2018)^[1] highlight that the advantages of any digital assessment tool really depend on the infrastructure backing it. This means that unequal access leads to unequal outcomes, no matter how proficient someone is in the language. This collection of research positions technological equity not just as one issue among many, but as a fundamental requirement for the effectiveness of all the other innovations mentioned earlier - a claim that this study aims to test directly.

3. Methodology

3.1 Research Design

This study used a convergent mixed-methods approach, gathering both quantitative and qualitative data at the same time and analyzing them together. This way, we can uncover not just the statistical trends but also the rich, contextual experiences that inform them. We opted for this design because our research questions call for both breadth-spotting patterns across a diverse sample and depth-grasping the real-life experiences that shape those patterns.

3.2 Sample: The quantitative phase involved surveying English language teachers and students engaged in online or hybrid learning environments, using purposive and snowball sampling methods for recruitment. After cleaning the data, we ended up with a solid sample of 250 valid responses, representing various types of institutions from seven different global regions. In the qualitative phase, we employed purposive sampling to bring together 15 educators and assessment experts who have significant practical experience with CAT platforms, automated grading systems, and formative assessment technologies.

The survey focused on five key themes: the background of participants, how effective and valid the assessments were, the technological hurdles and access issues, views on AI-driven tools, and the relationship between formative assessment and equity. We analyzed the data using a mix of descriptive statistics, one-sample and independent-samples t-tests, one-way ANOVA, Pearson correlation, and multiple regression. For the interview data, we took a thematic approach, organizing it around four main areas: experiences with assessment technologies, challenges in implementation, equity and accessibility, and recommendations. We then combined both the quantitative and qualitative findings to get a well-rounded view.

4. Results

Table 1: Distribution of Participants by Role

Role	Frequency	Percentage
Student	149	59.6%
Educator	101	40.4%
Total	250	100%

Table 2: Distribution of Participants by Institution Type

Institution Type	Frequency	Percentage
University	68	27.2%
Public School	64	25.6%
Language Institute	61	24.4%
Private School	57	22.8%
Total	250	100%

Table 3: Distribution of Participants by Region

Region	Frequency	Percentage
South Asia	49	19.6%
Europe	46	18.4%
North America	38	15.2%
Southeast Asia	38	15.2%
Sub-Saharan Africa	28	11.2%
Latin America	26	10.4%
Middle East	25	10.0%
Total	250	100%

4.1 Perceived Effectiveness (H1)

Participants generally found online assessments to be effective, scoring them above the neutral midpoint with an average of 3.42 (SD = 1.05). This difference is statistically significant ($t = 6.32, p < .001$), as 60.4% of respondents rated online assessments as equally or more effective than traditional methods, while only 16.4% disagreed. Interestingly, educators had a more favorable view of effectiveness, averaging 3.58, compared to students who rated it at 3.31 ($t = 2.11, p = .036$).

4.2 Technological Challenges (H2)

Participants generally found online assessments to be quite effective, giving them an average score of 3.42 (SD = 1.05), which is above the neutral midpoint. This difference is statistically significant ($t = 6.32, p < .001$), with 60.4% of respondents considering online assessments to be as effective or even more so than traditional methods, while only 16.4% disagreed. Interestingly, educators had a more positive outlook on their effectiveness, scoring them an average of 3.58, whereas students rated them at 3.31 ($t = 2.11, p = .036$).

4.3 AI-Driven Tools (H3)

Participants generally found online assessments to be quite effective, giving them an average score of 3.42 (SD = 1.05), which is above the neutral midpoint. This difference is statistically significant ($t = 6.32, p < .001$), with 60.4% of respondents considering online assessments to be as

effective or even more so than traditional methods, while only 16.4% disagreed. Interestingly, educators had a more positive outlook on their effectiveness, scoring them an average of 3.58, whereas students rated them at 3.31 ($t = 2.11, p = .036$).

4.4 Equity and Access (H4)

The data revealed a significant negative relationship between technology access and equity concerns ($r = -.61, p < .001$), making it the most influential factor in predicting perceived effectiveness in a multiple regression analysis ($\beta = .51, p < .001; R^2 = .28$ for the overall model). We observed notable regional differences in technology access, with equity concerns displaying an opposite trend across different areas (ANOVA $F = 24.61, p < .001$). Additionally, interviews provided real-life examples of issues like device-sharing and the struggle with high connectivity costs.

Table 4: Technology Access and Equity Concern by Region

Region	Mean Tech Access (SD)	Mean Equity Concern (SD)
North America	4.21 (0.61)	2.28 (0.71)
Europe	4.05 (0.68)	2.44 (0.77)
Middle East	3.24 (0.79)	3.02 (0.86)
Southeast Asia	2.98 (0.85)	3.45 (0.83)
Latin America	2.87 (0.81)	3.61 (0.81)
South Asia	2.71 (0.88)	3.88 (0.79)
Sub-Saharan Africa	2.15 (0.92)	4.12 (0.74)

4.5 Formative Assessment (H5)

In the study, formative assessment stood out with impressive ratings, boasting an average score of 3.98 (SD = 0.83). A solid 75.6% of participants found it to be helpful or very helpful. However, there was a slight negative correlation between how helpful participants found formative assessment and their concerns about equity ($r = -.19, p = .003$). This suggests that while formative assessment is beneficial, it doesn't fully address the access-related equity gap highlighted in H4.

Table 5: Summary of Key Correlations

Variable Pair	Pearson r	p-value	Interpretation
Tech access × Perceived effectiveness	.52	< .001	Moderate positive
Tech access × Equity concern	-.61	< .001	Strong negative
AI accuracy perception × AI bias concern	-.41	< .001	Moderate negative
Formative helpfulness × Equity concern	-.19	.003	Weak negative
Digital literacy × Test anxiety (online)	-.47	< .001	Moderate negative

Table 6: Evidence Summary and Hypothesis Outcomes

Hypothesis	Key Evidence	Outcome
H1 - Online assessment less effective	$M = 3.42$ (above neutral), $t = 6.32, p < .001$; educators $M = 3.58$ vs. students $M = 3.31$	Not supported
H2 - Significant adaptation challenges	Student disruption $M = 3.27$ vs. educator $M = 2.86, t = 3.14, p = .002$; institutional readiness gaps	Supported
H3 - AI tools improve accuracy/efficiency	Accuracy $M = 3.12$ (n.s.), $t = 1.84, p = .067$; accuracy × bias $r = -.41, p < .001$	Not supported
H4 - Underprivileged learners disadvantaged	Tech access × equity concern $r = -.61, p < .001$; regression $\beta = .51, p < .001$; ANOVA $F = 24.61, p < .001$	Strongly supported
H5 - Innovation improves fairness/inclusivity	Formative helpfulness $M = 3.98$ (75.6% positive); helpfulness × equity $r = -.19, p = .003$	Partially supported

5. Discussion

The evidence didn't back up H1: instead of seeing online assessment as something lesser, participants—especially educators—actually rated it as a bit more effective than a neutral baseline. This aligns with the more hopeful literature on adaptive testing (Alderson, 2009; Linacre, 2018)^[2, 7]. H2 received strong support: technical disruptions and gaps in institutional readiness are distinct yet interconnected barriers that hit students the hardest. H3 didn't hold up—perceptions of AI grading were mixed rather than outright positive, and the negative correlation between how accurate people think it is and their concerns about bias reflects known fairness issues in automated scoring (Zhang & Lu, 2019; Yu, 2021)^[14, 13]. This suggests that these tools need more algorithmic auditing before we can confidently say they're better than human assessment.

H4 stood out as the most solid finding in the study, showing consistency across correlation, regression, ANOVA, and thematic analysis: it's technological access, not any one testing method, that primarily drives perceived effectiveness and equity outcomes. This shifts the perspective on the other results—the positive view of online assessment (H1) and formative assessment (H5) likely comes from the experiences of participants with better resources. Plus, the weak link between formative assessment and reduced equity concerns (H5) reinforces the idea that innovative teaching methods can't replace the need for proper infrastructure investment.

All in all, these findings indicate that enhancing online English language assessment isn't just about fine-tuning CAT algorithms or grading models; it's really about ensuring everyone has the necessary tech access, which is essential for any assessment innovation to work fairly.

6. Conclusion

This study aimed to explore how English language assessment is evolving in the context of online learning environments. It combined survey responses from 250 participants with insights from interviews with 15 educators and assessment experts. The results are anything but straightforward: online assessment isn't simply better or worse than traditional testing; its success hinges on specific factors like proficiency level, the design of the tools used, and the quality of implementation. Throughout the research, one finding emerged with remarkable clarity: technological equity is the bedrock upon which the effectiveness of all other assessment innovations rests. No matter how sophisticated an adaptive test, how meticulously an AI grading system is audited, or how thoughtfully a formative feedback mechanism is crafted, they won't help a learner who doesn't have reliable access to a device or the internet.

As online and hybrid learning continue to grow beyond their pandemic roots, these insights have far-reaching implications that extend beyond language testing. They touch on issues of educational access, economic opportunity, and global mobility for countless learners who depend on English proficiency certification. For future assessment systems to be truly fair, valid, and inclusive, they must combine technological advancements with ongoing, coordinated efforts to bridge the digital divide.

7. References

1. Al-Mahrooqi R, Denman C. Mobile learning for language development: conceptual, theoretical and practical perspectives. Cham: Springer; c2018.
2. Alderson JC. The development of language testing in Europe: a historical and empirical study. Cambridge: Cambridge University Press; c2009.
3. Black P, Wiliam D. Developing the theory of formative assessment. Educational Assessment, Evaluation and Accountability. 2009.
4. Blake RJ. Brave new digital classroom: technology and foreign language learning. Washington, DC: Georgetown University Press; c2013.
5. Chapelle CA, Douglas D. Assessing language through computer technology. Cambridge: Cambridge University Press; c2006.
6. Horwitz EK. Test anxiety and student performance: implications for second language assessment. Language Testing. 2016;33(1):25-42.
7. Linacre JM. Computer-adaptive testing: a methodology whose time has come. London: Routledge; c2018.
8. Martin F. Online assessment in higher education: a review of emerging technologies. Journal of Educational Technology & Society. 2020;23(1):123-134.
9. Reinders H, Balcikanli C. Learning and teaching technology: learner autonomy and new language learning technologies. Language Learning & Technology. 2011;15(3):37-39.
10. Shermis MD, Burstein J. Handbook of automated essay evaluation: current applications and new directions. New York: Routledge; c2013.
11. Wang J, Brown L. AI in English language testing: opportunities and challenges. Language Assessment Quarterly. 2021;18(4):255-274.
12. Weir CJ. Language testing and validation: an evidence-based approach. Basingstoke: Palgrave Macmillan; c2005.
13. Yu S. Challenges and potentials of AI in language testing: a review. Journal of Educational Technology & Society. 2021;24(2):56-65.
14. Zhang X, Lu Z. The limitations of automated speech recognition in high-stakes language assessment. Language Testing. 2019;36(4):503-522.

Creative Commons (CC) License

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY 4.0) license. This license permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.