EDUCATIONAL SCIENCES: THEORY & PRACTICE eISSN: 2148-7561, ISSN: 2630-5984

Received: 02 September 2024
Revision received: 02 December 2024
Accepted: 13 December 2024

www.jestp.com
DOI 10.12738/jestp.2024.2.016 ♦ **June** 2024 ♦ 24(2) ♦ 223-236

Article

Technology-Enhanced Language Assessment: Validating Digital Tools for Measuring English Proficiency in Higher Education

Jing Sun
Faculty of Education, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia, 06010.
School of Foreign Lanaguages, Heilong Jinag Institute of Technology, Harbin, China, 150050.
ORCID iD: https://orcid.org/0009-0008-8864-1342
Email: sunjing916213@sina.com

Nur Ehsan Mohd Said*
Faculty of Education, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia, 06010.
ORCID iD: https://orcid.org/0000-0002-2891-327X
Emaiul: nurehsan@ukm.edu.my

Nur Ainil
Faculty of Education, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia, 06010.
ORCID iD: https://orcid.org/0000-0001-6212-7494

Abstract

This study explores students' and instructors' perceptions of the validity, usability, and educational impact of technology-enhanced English language assessments in higher education contexts. Using a qualitative research design, semi-structured interviews were conducted with six participants (three students and three instructors) who had experience with digital English proficiency assessments. Thematic analysis was applied to analyse participant responses and identify patterns related to assessment validity, platform usability, and instructional impact. Participants acknowledged the efficiency and practicality of technology-enhanced assessments in evaluating basic language skills like grammar and vocabulary. However, they questioned their ability to assess higher-order competencies, such as argumentation and interaction. Usability concerns were reported, especially among less digitally literate students, while educators noted the shift in teaching practices toward test-oriented strategies. Mixed perceptions about fairness and accessibility further emphasized the need for more inclusive, pedagogically aligned assessment models. Findings suggest that TEAs should be complemented by human input to ensure comprehensive language evaluation. Institutions must prioritize accessibility and inclusive design, while educators require training to balance assessment preparation with communicative pedagogy. Future research should expand to diverse contexts and integrate performance data for validation.

Keywords

Assessment Criteria, Language Dimensionality, English Language Evaluation Standards, Language Assessment Validity, Technology-Enhanced Assessments.

Correspondence to Nur Ehsan Mohd Said, Faculty of Education, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia, 06010.
E m a i l : nurehsan@ukm.edu.my ORCID iD: https://orcid.org/0000-0002-2891-327x

Citation: Sun, J., Mohd Said, N. E. & Ainil, N. (2024). Technology-Enhanced Language Assessment: Validating Digital Tools for Measuring English Proficiency in Higher Education. *Educational Sciences: Theory and Practice, 24*(2), 223 - 236.
http://dx.doi.org/10.12738/jestp.2024.2.016

Introduction

The fast growth of educational technology is making it more common to apply technology to language assessment. Because English language testing is now mostly online, people are reviewing how well the assessments are meant for learning, access and accuracy. In higher education, since English skills affect students' academic results and ability to move globally, digital assessment tools must be proven valid and useful by collecting feedback from the people who teach and learn.

In language learning and testing, the use of technology-enhanced assessments has increased due to a need for increased accessibility, easy scaling and fast feedback (Chapelle, 2017). TEAs are designed with interactive resources, multiple methods for testing and flexibility to showcase authentic language like no other traditional option (Yang & Qian, 2017). Universities and colleges are using these methods more frequently to assess both when to place students and whether they meet graduation criteria for English (Hu, 2022). Even though technology helps create more personal and lively assessments, some issues about their effectiveness and relevance for all learners are still being raised (Harding, Alderson, & Brunfaut, 2015). It is vital to assess assessment tools with qualitative inquiry to find out if they measure as expected and also learn how users view their use and their effect on teaching and learning. While numbers are strong, qualitative methods can bring out finer ideas and views about how users feel or perceive something (Creswell & Poth, 2016). In contexts where many languages are used in schools, the ways students and instructors view assessment processes are very important for test results and outcomes (Bachman & Palmer, 2022).

Problem Statement

Many schools use digital language assessment tools, but not much research has investigated how students and teachers truly feel about and use them. Commonly, validation studies place more emphasis on psychometric assessment than on factors such as anxiety, digital education and a sense of fairness (Al-Azawei, Serenelli, & Lundqvist, 2016; Emmanouil et al., 2024; Shadiev & Feng, 2024). Such gaps are seen more often outside the West since perceived differences in setup, teaching style and language can change how students and professors approach digital evaluation. Thus, it must be observed what stakeholders think and go through to confirm that these tools work well and are just for all.

Aims and Objectives

This research seeks to understand what students and instructors think about using technology to judge English proficiency. The research investigates whether learners consider these technologies fair, easy to use and useful for education in school. The specific research objectives are:

- To investigate how students and instructors perceive the validity and reliability of technology-enhanced English language assessment tools.
- To explore the usability and accessibility of digital assessment platforms from the perspective of end users.
- To understand the perceived impact of these assessments on teaching practices and students' language learning experiences.

Significance of Study

This research provides a critical perspective on examining how digital English language assessments are verified in higher education. It adds value by advancing the idea that assessment validity should focus on both statistical reliability and what users think and experience in specific situations. This research uses student and teacher input to represent assessment practices where authenticity, fairness and inclusivity are the main priorities, as noted by Douglas and Hegelheimer (2007). The results obtained will give useful directions for policymakers, test designers and educators working on improving digital assessment tools. Furthermore, the research will increase the small amount of qualitative work in this area, focusing on regions where assessment of local validation is not widely explored.

Literature Review

As digital technology in education grows fast, it has made a big impact on measuring English skills, both helping in new opportunities and creating hurdles. Instead of being peripheral in language testing, technology-enhanced assessments (TEAs) are now playing a major role in many higher education evaluations. The review brings together major research related to how stakeholders assess TEAs, how easy digital platforms are to use and the implications TEAs have for teaching and language learning. In the end, the section presents the theoretical framework and points out the gaps this study aims to fill.

Perceptions of Validity and Reliability of Technology-Enhanced Language Assessments

The study of assessment in research relies on validity and reliability. Validity means an assessment measures what it is meant to measure, while reliability refers to how consistent a test's scores are over different times and situations (Alderson, Clapham, & Wall, 1995; Messick, 1996). The role of technology-enhanced assessments means that these constructs must also include digital delivery, how the language test is delivered, the digital medium used and any adaptive qualities (Hu, 2022). Although controlled studies reveal that various digital language assessments are psychometrically strong (Shadiev & Feng, 2024), this is not always appreciated by users. In particular, Lee (2025) discovered that university students in Korea had reservations about the fairness of automated scoring for both their writing and speaking tests. A similar study has shown that students may be uncertain about the accuracy of TEAs when unusual designs or insufficient feedback create confusion (Al-Issa et al., 2025).

Some teachers are not convinced about the benefit of TEAs for determining how students learn. While some see how TEAs give quick and accurate results (Reinders & White, 2016), others believe they might make language learning too easy and encourage people to work on simple elements, instead of in-depth skills such as conversation or careful reading (Ranalli, 2021). It has further been noted by educators working within multilingual environments that certain test tools developed in English-dominated situations and used abroad show cultural bias (Sutrisno, Abbas, & Annury, 2024). This means that it should be understood the validity and reliability based on what teachers and test-takers think and feel, as well as on statistics. By relying only on numbers, a quantitative validation may miss the small, important differences in users' use and trust of TEA.

Usability and Accessibility of Digital Assessment Platforms

For any educational technology to succeed in high-stakes situations like language testing, it must be usable and accessible. Usability is how easily users can navigate and use the system and accessibility means making the system useful to users with different abilities (Nielsen, 1994). When digital assessments are not simple, they may require extra mental effort that does not test language proficiency and may change how people perform (Kelly, Phipps, & Swift, 2004; Nuraeni et al., 2020). Often, those with little knowledge of computers struggle in writing assignments because of digital issues, rather than language issues (Al-Issa et al., 2025). Digital literacy not being the same for everyone in higher education institutions can cause big equity issues (Hamp-Lyons, 2007; Sa'di, Abdelraziq, & Sharadgah, 2021).

Other studies back up these views from instructors. It seems that many educators believe they need more training to work with digital assessment outcomes effectively (Husain, 2021). How well a TEA fits with the user interface, learning management and the institution's network can decide if it is adopted or left behind (Bayne & Ross, 2014). Another part of accessibility concerns learners with special needs in education. Although some TEAs have useful features for students, others still do not give them what they require because those TEAs do not comply with universal design standards (Bennett, 2015). Qualitative studies show that difficulty or bias with assessments can lessen a learner's motivation to participate (Al-Azawei et al., 2016). As a result, researchers must look into how various users behave in the system, the obstacles they hit and their overall opinion about both the test's fairness and usefulness.

Perceived Impact on Teaching Practices and Learning Experiences

When technology-enhanced assessments are used, the ways teachers deliver instruction and set their priorities also shift. How an assessment is designed may influence the way instructors teach, especially if those assessments guide student performance or the institution's review (Shohamy, 2020). As a result of this

phenomenon, washback, language use can either encourage authentic speaking in class or make teaching about tests, disappointing students who prefer more in-depth learning (Green, 2013). Results from several qualitative research projects have shown that teachers respond to TEAs in mixed ways. Technology tools have allowed teachers in several areas to develop student-centred and more engaging test questions (Yılmaz & Keser, 2017). A good example is that using automated feedback for pronunciation and grammar in courses can support more individualized teaching of speaking (Witt & Young, 2014). Yet, in different situations, teachers may find that strict testing designs and unclear ways of scoring make it hard for them to decide on how to assess in their classroom (Makransky, Terkildsen, & Mayer, 2019).

Learners may become more engaged when digital assessments are designed to be gamified and use tasks from the real world (Li & Hegelheimer, 2013). Even so, students may have technical-related anxiety about connections stopping, issues with the interface and the lack of clear information on how their scores are calculated (Fageeh, 2015). As a result, TEAs influence teaching and learning differently within each context. These aspects are best uncovered through qualitative inquiry which allows us to see how assessments impact learning in ways that go deeper than just the numbers.

Theoretical Framework

This research follows Messick's (1996) Unified Theory of Validity, joining content, criterion-related, construct, consequential and face validity into one theory. Says Messick, making a valid judgment requires analysing the test as well as listening to stakeholders and what they think it means for them. By using the framework, the study explores both the construction and the accuracy of digital tools as well as how users experience and use them. The qualitative lens adds to Messick's model by focusing on how assessments influence outcomes, impact stakeholders and the school system and their perceived fairness by learners. Qualitative validation research benefits from using this framework since it fits with the belief that validity varies with social interaction and settings (Roever & McNamara, 2006). It makes it possible to explore both the measurement and experience of TEAs, bringing closer the processes of psychometric review and what users go through.

Literature Gap

While research into digital language assessment is increasing, there are still gaps, mainly in how it is validated qualitatively. In many cases, studies choose to employ numerical validation techniques before seriously listening to what users think (Aryadoust et al., 2020). Moreover, while some of these qualitative studies focus on students or teachers, very few study both groups at once, so the experience of each group is rarely confirmed from a different perspective (Nurlely, 2023). Numerous studies are situated in schools using English or advanced resources, making it unclear how useful they are for other settings in Central Asia, Eastern Europe or the Middle East (Rahimi & Pourshahbaz, 2018). Many of these regions encounter infrastructure and language barriers which play a big role in how TEAs work and are received. Only a few such investigations have used well-known frameworks such as Messick's validity framework, to guide them. Findings are not as coherent or reusable since evaluations examine single issues rather than connecting them to validity. This study fills these gaps by using a thorough, theory-based qualitative method with input from several stakeholders in underrepresented learning environments.

Research Methodology

This chapter lays out the framework used in the present research which investigates stakeholders' perceptions and experiences concerning technology-enhanced English language assessments in higher education. A qualitative, primary research method with semi-structured interviews was used to get detailed information. Results from the data are interpreted using thematic analysis aligned with the aims and theoretical basis of the study.

Research Method and Design

Interpretivism underlies this study, it uses a qualitative method that pays attention to what participants go through and the environment they experience (based on Denzin & Lincoln, 2011). This approach is well-suited for learning about multiple, personal insights into language assessment which are formed by many social,

institutional and technical factors (Tisdell, Merriam, & Stuckey-Peyrot, 2025). Moreover, the Semi-structured interviews formed the core part of the data collected, and a primary research design was used. Thanks to semi-structured interviews, the researcher can ask extra questions because of what the participant says, while the interview is still following a common theme (Kvale & Brinkmann, 2009). This way of research works best when studying digital tools, as focusing on how users experience them helps reveal validity, usability and educational outcomes. Structured surveys and quantitative scales can only get part of our story, but interviews let us share how a person feels, the situation and what a person does (Teo, 2014). Also, qualitative interviews fit with the study's theoretical structure by Messick (1996), as they specifically report stakeholder experiences and what is perceived to follow their use of assessments.

Data Collection Techniques

The six participants in the study included three instructors teaching English and three students "subjects" participating in English-medium instruction at a public university in Eurasia. The study chose this sample to make sure both parts of the assessment, planning/conducting and taking, were present which allows us to compare outcomes and get a full picture of how useable and valid our tools are believed to be (Palinkas et al., 2015). Purposive sampling was used since it lets you pick cases with participants who have personal experience with the phenomenon of interest (Patton, 2002). All participants had completed at least a partial English proficiency assessment that used technology, guaranteeing their answers were meaningful and deep.

All interviews were held for about 40 to 60 minutes at a time and participants spoke in English over a secure video conference system. According to the research goals, an interview protocol was developed about how participants judge the test, use the interface, incorporate it in classrooms and experience the whole process. All audio conversations were first transcribed step by step for closer examination. Participants could ensure their information was correct by seeing a summary of the findings (Birt et al., 2016).

Data Analysis Method

By doing thematic analysis, we studied the data to search, organize and make sense of various themes uncovered (Braun & Clarke, 2006). The study opted for thematic analysis, given that it perfectly fits with its special interest in linking theory with the meanings given by participants (Nowell et al., 2017). To gain a clear perspective, researchers looked at the transcripts of the interviews several times in the beginning. Inductive coding was performed to create the set of codes which were then arranged into meaningful groups that aligned with the research objectives. Davidson and Colley grouped issues such as "interface confusion," "real-world relevance," and "automated scoring bias" with others to make larger groups known as usability concerns and perceived validity. An experienced qualitative researcher worked with me to perfect and finalize the themes to guarantee their trustworthiness. During the entire analysis, a reflexive journal was kept logging the researcher's choices, beliefs and updated understandings. Following this approach enhanced the reliability and examinability of the thematic process (Lincoln & Guba, 1988).

Ethical Consideration

The method was designed following ethical guidelines for the sake of the participants and to ensure good research. Ahead of gathering the data, the study was approved by the university's IRB. Before interviews started, participants were shown the study aims, and their rights to privacy and given an online informed consent form to sign. It was ensured that any names, addresses or school titles present in any conversations were anonymized before reporting (Cohen, Manion, & Morrison, 2002). All data was held in encrypted form and the researcher was the only person with access. Those involved were told they could choose to leave the study at any moment, and no reason had to be given (Abdallah, 2024). So that the researcher's position at the same institution did not create any imbalances, the interviews were conducted only in a fair, non-evaluative manner and location. Participants were told they could say whatever they wanted, without having to worry about academic consequences. Upon request, all information included in publications or presentations was given to participants to promote transparency and encourage everyone to feel they owned the knowledge.

Data Analysis

It provides a summary of thematic findings drawn from interviews with six respondents, who are both students and teachers in higher education. The researchers used Braun and Clarke's (2006) model to search for common themes in the study's data. There are two interview questions for every research objective and the findings are built around these questions. Each main theme is covered in detail with quotes provided by the people interviewed.

Perceptions of Validity and Reliability of Technology-Enhanced Assessments

Respondents gave opinions on how TEAs reflect English proficiency and felt these technological assessment systems to be fair and trustworthy. While many thought the assessments valuable, some pointed out that they were not good at judging the different and detailed ways language is used.

Perceived Alignment with Actual Language Abilities

A major point was that what TEAs indicate is only partly related to actual language ability. It was clear to participants that TEAs can assess basic skills such as grammar and vocabulary, but not as well others such as argumentation, tone or communication skills. Respondent 2 expressed this concern clearly, stating,

"It partially reflects language ability but lacks depth in communicative tasks like conversations or presentations that matter in real classroom situations."

In other words, although basic skills are checked, much of what matters in actual language proficiency is often not included in the results. Similarly, Respondent 5 commented,

"It covered the essential skills, but I felt that open-ended responses were oversimplified, affecting how my true writing ability was judged."

Experts suggest that there is a gap between what is tested and how people use language in real life, so assessments should be wider in their scope for digital tests.

Perceived Fairness and Consistency of Scoring and Feedback

Fairness and uniformity in automated scoring were also a key subject discussed. Some people saw the efficiency of digital scoring, even though others had doubts about its fairness and effectiveness at diagnosis. Respondent 3, an instructor, remarked,

"As a teacher, I noticed consistent scoring, but I couldn't tell if it truly reflected language improvement, especially for complex student writing."

This points to a perceived limitation in the machine's ability to detect nuanced development in student performance. Similarly, Respondent 6 observed,

"It seemed fair for grammar and vocabulary, but I don't think it can evaluate idea development or argument quality in longer essays."

Those surveyed believed that automated systems are not effective at fully examining student responses. It shows that digital assessments should be checked by people or that their guidelines should be made more visible.

Usability and Accessibility of Digital Assessment Platforms

Research in this area investigates the experience of TEA users concerning design, ease of use and inclusion. The subjects covered included how easy it is to use different technologies, digital skills and digital equity.

User Experience with Interface and Navigation

Most participants provided candid feedback on the usability of the platform, with responses reflecting varying degrees of satisfaction. While some found the platform manageable, others encountered technical barriers that hindered their performance. Respondent 1 reported,

"The layout was clean, but loading delays during the speaking section made the experience stressful and affected my concentration during recording."

This illustrates how even well-designed platforms can be undermined by performance issues during real-time assessments. Respondent 5 shared a different but related concern:

"The design was user-friendly, but the speaking section lacked clear instructions and had no retry option, which stressed some of my students."

These responses suggest that even minor issues in interface design or technical guidance can significantly influence user confidence and assessment outcomes. Clearer navigation cues, error recovery options, and trial practice sessions were implied as potential improvements.

Accessibility for Diverse Learner Backgrounds

Participants also discussed the accessibility of TEAs, especially for students with varied digital skills or limited technological access. Several responses indicated that not all learners had an equal footing when engaging with digital platforms. Respondent 3 noted,

"It was manageable for tech-savvy students, but others found it intimidating, especially those with limited experience in online learning environments."

This suggests that digital proficiency is a precondition for success in TEAs, which may disadvantage some learners. Echoing this, Respondent 6 observed,

"Accessibility needs improvement. Students with disabilities or lower computer skills often faced unnecessary challenged that influenced their test performance."

These insights emphasize the digital divide in assessment readiness, reinforcing the need for more inclusive design principles and adaptive supports to ensure fair testing environments for all students, regardless of background.

Perceived Impact on Teaching Practices and Learning Experiences

The final goal of the research is to study what TEAs mean for teaching strategies and how they shape student education. Issues about adjusting instruction, learning with tests in mind and changing classroom activities were discussed.

Influence on Teaching and Learning Strategies

Participants reported adapting their teaching and learning strategies to align with the digital assessment format. Both teachers and students described changes in classroom tasks and preparation practices to better reflect the expectations of the TEAs. Respondent 4, an instructor, shared,

"The assessment influenced me to teach typing and basic tech skills, which aren't language-related but essential for success in the test."

This highlights how technological competencies have become embedded in language instruction due to assessment demands. Respondent 6, a student, added,

"It affected how I prepare, now I use digital platforms for practice, but I feel it takes time away from oral communication activities."

These responses suggest that digital assessment environments may be reshaping pedagogical priorities, sometimes at the expense of balanced language skill development, particularly in speaking and interaction-focused tasks.

Educational Value and Impact on Language Development

Participants shared mixed views on the educational impact of TEAs, with some acknowledging improvements in efficiency and motivation, while others raised concerns about superficial learning and reduced critical engagement. Respondent 1 commented,

"Mixes impact, helpful for vocabulary and grammar, but not effective for speaking and real interaction, which are crucial in language learning."

This view reflects a concern that TEAs prioritize measurable outcomes over communicative competence. Respondent 5 similarly remarked,

"Positive in some ways, but overly standardized. It doesn't consider student' learning styles or creativity in writing and speaking."

They continue to highlight that technology-enhanced approaches are not very flexible in serving all types of students. TEAs might make assessment simpler, but without other activities, they could limit what learners learn about language.

Various topics were brought up by participants in several different interviews. Some attendees said TEAs help assess the foundation, though others noted doubts about tracking more advanced or communicative language. At the same time, because online platforms existed, unskilled users and people less comfortable with technology found these resources hard to utilize. What's more, many schools ended up spending more time teaching tests and nurturing technical learning than on improving students' language and the way they interact. As expected, examples from our study reflect the idea outlined in Messick (1996) that looks at consequential and face validity. Participants in the research believe that digital assessment tools must be seen as fair, work well for users and benefit teaching. By examining these topics with qualitative study, that has been learning how technology-enhanced assessment is implemented in real cases.

Discussion of the Results

The results from the data analysis are discussed in terms of what the related literature already points out. Every part of the chapter is tied to a key objective of the study and includes examples of how participants' opinions back, expand on or challenge known concepts in TEA.

Perceptions of Validity and Reliability

It was seen by participants that TEAs are useful for identifying the basic parts of language such as grammar and vocabulary, but are unsure if they measure more advanced abilities, mainly in speaking and writing. Lee (2025) revealed that in EFL education, students had concerns about how automated scoring can handle challenging and original answers to questions. Sutrisno et al. (2024) recognized that the way people view validity can be collared by culture and context and many respondents mentioned that TEAs did not seem relevant to the language they used for studies. What's more, participants express uncertainty about relying on robots to evaluate their writing skills which agrees with the idea of Aryadoust et al. (2020) that TEA systems are mainly weak at checking rhetorical and abstract parts of writing. While a few respondents recognized how smoothly machines can score, others noted that consistency is not enough to make scoring significant. What these studies tell us is that the benefits of online assessments might mean students are not evaluated as well as they could be in traditional ways (Green, 2013).

Usability and Accessibility

Many of the interviews pointed out that using the assessment platform can be tough for people with limited digital skills and problematic technology access. The concerns are supported by research by Sa'di et al. (2021) who emphasize that learners in higher education do not have equal chances to learn digital literacy. Some individuals who took the assessment found the interface easy to use, though others from rural or low-resource communities faced major difficulties during the evaluation. Based on Al-Azawei et al. (2016), if the person does not include universal design considerations, digital tools present a danger of marginalizing some learners. What's more, the difficulty in using the interface and concerns about performance are similar to what Nuraeni et al. (2020) discovered, struggle with technology can affect how well learners perform in language tasks. Since minor errors on a website can cause stress or result in poor performance, smart design is key when the test means a lot to students.

Impact on Teaching and Learning

Study participants observed how TEAs influenced different teaching methods and also student learning styles. Study participants similarly explained that they added keyboard practice, prepared tasks with multiple possible answers and gave students digital tools to use since the types of assessments often devise the main subject teachers cover. It has long been clear from the research that when assessments drive teaching or washback happens, teachers may emphasize familiar formats instead of improving communication (Shohamy, 2020). Those who studied indicated higher test nerves and less variation in language learning to match the area of the assessment. These findings agree with what Ranalli (2021) reported, that investing more in test strategies tends to come at the expense of developing real language use skills. Green's (2013) theory that the positive or negative results from TEAs can depend on how they are used and integrated, is strengthened by these comments on the mix of educational values.

Conclusion and Recommendation

This study explores whether technology-enhanced tests are accurate, easy to use and valuable in higher education, viewed from the students' and teachers' viewpoints. The results seen through a qualitative approach uncover the way digital tools and the people using them fit together. Whereas many participants appreciated the convenience, quick responses and flexibility in TEAs, several kept voicing doubts about how well they represent a person's actual language skills. The report found that TEAs reliably measure vocabulary and grammar, but do not cover interactive and advanced language abilities. It was recognized by those taking part that automated scoring is not as sophisticated in understanding writing and speaking as human raters. Such worries are most important in the academic world because being able to argue, understand meanings and communicate abstractions is essential. The data also revealed a lot of variation in how easily users could use the websites. Although some users found the process straightforward, others had great difficulties because they lacked the knowledge or the necessary infrastructure. As a result, there is a need to make sure inclusive design is part of assessment technology so that no student falls behind due to digital constraints. Next, the study discovered that TEAs change teachers' and students' actions which can narrow down the overall subject taught to match what's tested in the exam. Even if test performance rises, using shortcuts in language learning risks reducing a student's ability to speak and think critically. According to participants, while TEAs promoted improved efficiency and structure, they decreased the scope for opportunities to speak creatively, act spontaneously and dig deeper into language. The analysis stresses that technology should be combined with proper teaching methods and values. TEAs, being effective in measuring large groups, still require special care to meet user needs, promote equal learning and properly assess people's language abilities. Both educators and policymakers should consider making sure assessment systems support the development of meaningful language, by engaging students and teachers in their development.

Recommendation

It is proposed that educational institutions using technology-based learning continue to regularly validate the use of their assessments both with statistics and with observable examples. User viewpoints should be included in both the design phase and the evaluation of the system and combined scoring systems that include expert input should be used. They also need training to both use TEAs and understand their conclusions so they can adapt their teaching methods better for each learner.

Practical Implications

The findings from this study are useful for colleges, instructors of language courses and those designing assessments. It highlights that institutions should make technological hardware and user support more accessible to everyone. It supports teachers in linking activities for assessment preparation to daily communicative teaching. For those who create apps, the evidence points to the value of easy-to-use interfaces, info that corresponds with local cultures and responsive feedback to improve how easy it is to use and how fair people find the app.

Limitations

A main limitation of the study is that the sample is small and related to one specific community, so the findings may not be widely applied. Interviews were done with only six people from a single higher education institution, so it cannot draw broad conclusions across other institutions. Unlike others, this study concentrated only on users' opinions rather than analysing results or direct evidence. Later on, researchers can combine methods to assess learners' performance both in reality and how they feel about their learning.

References

Abdallah, M. M. S. (2024). Revised Ethical Guidelines for Educational Research in Practice: A Qualitative Analysis. *Online Submission.* https://files.eric.ed.gov/fulltext/ED651406.pdf

Al-Azawei, A., Serenelli, F., & Lundqvist, K. (2016). Universal Design for Learning (UDL): A Content Analysis of Peer Reviewed Journals from 2012 to 2015. *Journal of the Scholarship of Teaching and Learning, 16*(3), 39-56. https://doi.org/10.14434/josotl.v16i3.19295

Al-Issa, A. S. M., Baleghizadeh, S., Ghonsooly, B., Haghbin, F., Hashemi, M. R., Holliday, A., et al. (2025). Journal of Language Horizons, Alzahra University. *Journal of Language Horizons, 9*(1), 1-198. https://lghor.alzahra.ac.ir/issue_1431_1432.html

Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge University Press.

Aryadoust, V., Zakaria, A., Lim, M. H., & Chen, C. (2020). An Extensive Knowledge Mapping Review of Measurement and Validity in Language Assessment and SLA Research. *Frontiers in Psychology, 11*, 1941. https://doi.org/10.3389/fpsyg.2020.01941

Bachman, L., & Palmer, A. (2022). *Language Assessment in Practice: Developing Language Assessments and Justifying Their Use in the Real World*. Oxford University Press.

Bayne, S., & Ross, J. (2014). *The Pedagogy of the Massive Open Online Course (MOOC): The UK view*. Higher Education Academy. http://www.heacademy.ac.uk/resources/detail/elt/the_pedagogy_of_the_MOOC_UK_view

Bennett, R. E. (2015). The Changing Nature of Educational Assessment. *Review of Research in Education, 39*(1), 370-407. https://doi.org/10.3102/0091732x14554179

Birt, L., Scott, S., Cavers, D., Campbell, C., & Walter, F. (2016). Member Checking: A Tool to Enhance Trustworthiness or Merely a Nod to Validation? *Qualitative Health Research, 26*(13), 1802-1811. https://doi.org/10.1177/1049732316654870

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Chapelle, C. A. (2017). Evaluation of Technology and Language Learning. In C. A. Chapelle & S. Sauro (Eds.), *The Handbook of Technology and Second Language Teaching and Learning* (pp. 378-392). John Wiley & Sons. https://doi.org/10.1002/9781118914069.ch25

Cohen, L., Manion, L., & Morrison, K. (2002). *Research Methods in Education*. Routledge. https://doi.org/10.4324/9780203224342

Creswell, J. W., & Poth, C. N. (2016). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. Sage Publications. https://edge.sagepub.com/creswellqi4e

Denzin, N. K., & Lincoln, Y. S. (2011). *The Sage Handbook of Qualitative Research*. Sage Publications. https://us.sagepub.com/en-us/nam/the-sage-handbook-of-qualitative-research/book242504

Douglas, D., & Hegelheimer, V. (2007). Assessing Language Using Computer Technology. *Annual Review of Applied Linguistics, 27*, 115-132. https://doi.org/10.1017/S0267190508070062

Emmanouil, K., Zoe, G., Christos, A., & Markos, A. (2024). Linguistic factors affecting L1 language evaluation in argumentative essays of students aged 16 to 18 attending secondary education in Greece. *Assessing Writing, 60*, 100844. https://doi.org/10.1016/j.asw.2024.100844

Fageeh, A. I. (2015). EFL Student and Faculty Perceptions of and Attitudes Towards Online Testing in the Medium of Blackboard: Promises and Challenges. *The JALT CALL Journal, 11*(1), 41-62. https://doi.org/10.29140/jaltcall.v11n1.183

Green, A. (2013). Washback in Language Assessment. *International Journal of English Studies, 13*(2), 39-51. https://doi.org/10.6018/ijes.13.2.185891

Hamp-Lyons, L. (2007). The Impact of Testing Practices on Teaching. In J. Cummins & C. Davison (Eds.), *International Handbook of English Language Teaching* (pp. 487-504). Springer US. https://doi.org/10.1007/978-0-387-46301-8_35

Harding, L., Alderson, J. C., & Brunfaut, T. (2015). Diagnostic assessment of reading and listening in a second or foreign language: Elaborating on diagnostic principles. *Language Testing, 32*(3), 317-336. https://doi.org/10.1177/0265532214564505

Hu, H. (2022). Computer-delivered English Listening and Speaking Test in Zhongkao: Test-Taker Perception, Motivation and Performance. In *Proceedings of SOCIOINT 2022- 9th International Conference on Education & Education of Social Sciences* (pp. 59-75). Ocerints. https://doi.org/10.46529/socioint.202209

Husain, F. N. (2021). Digital Assessment Literacy: The Need of Online Assessment Literacy and Online Assessment Literate Educators. *International Education Studies, 14*(10), 65-76. https://doi.org/10.5539/ies.v14n10p65

Kelly, B., Phipps, L., & Swift, E. (2004). Developing a Holistic Approach for E-Learning Accessibility. *Canadian Journal of Learning and Technology, 30*(3). https://doi.org/10.21432/T2D60S

Kvale, S., & Brinkmann, S. (2009). *Interviews: Learning the Craft of Qualitative Research Interviewing*. Sage Publications.

Lee, J.-H. (2025). ChatGPT-Assisted Writing: Learners' Perceptions of AI Feedback and Influencing Factors. *Korean Journal of English Language and Linguistics, 25*, 493-515. https://doi.org/10.15738/kjell.25..202504.493

Li, Z., & Hegelheimer, V. (2013). Mobile-Assisted Grammar Exercises: Effects on Self-Editing in L2 Writing. *Language Learning & Technology, 17*(3), 135-156. https://llt.msu.edu/issues/october2013/lihegelheimer.pdf

Lincoln, Y. S., & Guba, E. G. (1988). *Criteria for Assessing Naturalistic Inquiries as Reports*. Sage Publications.

Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction, 60*, 225-236. https://doi.org/10.1016/j.learninstruc.2017.12.007

Messick, S. (1996). Validity of Performance Assessments. In *Technical Issues in Large-Scale Performance Assessment* (pp. 1-18). U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Nielsen, J. (1994). *Usability Engineering*. Morgan Kaufmann. https://www.nngroup.com/books/usability-engineering

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic Analysis:Striving to Meet the Trustworthiness Criteria. *International Journal of Qualitative Methods, 16*(1), 1609406917733847. https://doi.org/10.1177/1609406917733847

Nuraeni, C., Carolina, I., Supriyatna, A., Widiati, W., & Bahri, S. (2020). Mobile-Assisted Language Learning (MALL): Students' Perception and Problems towards Mobile Learning in English Language. *Journal of Physics: Conference Series, 1641*(1), 012027. https://doi.org/10.1088/1742-6596/1641/1/012027

Nurlely, L. (2023). The Role of Technology in Promoting Assessment Literacy in ELT. *Journal of Linguistics, Literacy, and Pedagogy, 2*(1), 52-58. https://doi.org/10.30870/jllp.v2i1.20013

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Administration and Policy in Mental Health and Mental Health Services Research, 42*(5), 533-544. https://doi.org/10.1007/s10488-013-0528-y

Patton, M. Q. (2002). *Qualitative Research & Evaluation Methods: Integrating Theory and Practice*. Sage Publications. https://study.sagepub.com/patton4e/help

Rahimi, M., & Pourshahbaz, S. (2018). *English as a Foreign Language Teachers' TPACK: Emerging Research and Opportunities*. IGI Global Scientific Publishing. https://doi.org/10.4018/978-1-5225-6267-2

Ranalli, J. (2021). L2 student engagement with automated feedback on writing: Potential for learning and issues of trust. *Journal of Second Language Writing, 52*, 100816. https://doi.org/10.1016/j.jslw.2021.100816

Reinders, H., & White, C. (2016). 20 Years of Autonomy and Technology: How Far Have We Come and Where to Next? *Language Learning & Technology, 20*(2), 143-154. https://llt.msu.edu/issues/june2016/reinderswhite.pdf

Roever, C., & McNamara, T. (2006). Language testing: the social dimension. *International Journal of Applied Linguistics, 16*(2), 242-258. https://doi.org/10.1111/j.1473-4192.2006.00117.x

Sa'di, R. A., Abdelraziq, A., & Sharadgah, T. A. (2021). E-Assessment at Jordan's Universities in the Time of the COVID-19 Lockdown: Challenges and Solutions. *Arab World English Journal (AWEJ),* (1), 37-54. https://doi.org/10.24093/awej/covid.3

Shadiev, R., & Feng, Y. (2024). Using automated corrective feedback tools in language learning: a review study. *Interactive Learning Environments, 32*(6), 2538-2566. https://doi.org/10.1080/10494820.2022.2153145

Shohamy, E. (2020). *The Power of Tests: A Critical Perspective on the Uses of Language Tests*. Routledge. https://doi.org/10.4324/9781315837970

Sutrisno, D., Abbas, A., & Annury, M. N. (2024). Enhancing Writing Skills Through Cultural Integration: Exploring the Impact of Culturally-Responsive Writing Instruction in Diverse EFL Classrooms. *Global Synthesis in Education Journal, 2*(3), 1-16. https://doi.org/10.61667/rpm4cp30

Teo, T. (2014). *Encyclopedia of Critical Psychology* (Vol. 1). Springer New York. https://mkontopodis.wordpress.com/wp-content/uploads/2011/03/encyclopediacriticalpsy.pdf

Tisdell, E. J., Merriam, S. B., & Stuckey-Peyrot, H. L. (2025). *Qualitative Research: A Guide to Design and Implementation*. John Wiley & Sons.

Witt, S., & Young, S. (2014). Computer-Assisted Pronunciation Teaching Based on Automatic Speech Recognition. In A. van Essen, S. Jager, & J. Nerbonne (Eds.), *Language Teaching and Language Technology* (pp. 25-35). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781315078137-5

Yang, Y., & Qian, D. D. (2017). Assessing English reading comprehension by Chinese EFL learners in computerized dynamic assessment. *Language Testing in Asia, 7*(1), 11. https://doi.org/10.1186/s40468-017-0042-3

Yılmaz, R., & Keser, H. (2017). The Impact of Interactive Environment and Metacognitive Support on Academic Achievement and Transactional Distance in Online Learning. *Journal of Educational Computing Research, 55*(1), 95-122. https://doi.org/10.1177/0735633116656453

Appendices

Interview Responses

Question 1:
From your experience, do you believe that the technology-enhanced English language assessment accurately reflects your (or your students') language abilities? Why or why not?
Respondent 1: I think the assessment reflects basic proficiency, but it doesn't fully capture speaking or critical thinking skills which are essential for academic English use.
Respondent 2: It partially reflects language ability but lacks depth in communicative tasks like conversations or presentations that matter in real classroom situations.
Respondent 3: The tool was fine for grammar and vocabulary, but not for measuring real-life language use like discussion skills or argument construction.
Respondent 4: It was useful for assessing some skills, but students with test anxiety or tech issues may not show their real abilities in this format.
Respondent 5: It covered the essential skills, but I felt that open-ended responses were oversimplified, affecting how my true writing ability was judged.
Respondent 6: I found it fair in general, but limited in assessing complex language use like tone, inference, and rhetorical structure in writing tasks.

Question 2:
How confident are you in the consistency and fairness of the scoring and feedback provided by the digital assessment tool you used? Can you share any examples?
Respondent 1: The scoring seemed fast and consistent, but I'm unsure if it really understood my essay's argument or creativity-feedback felt generic.
Respondent 2: I'm sceptical. I submitted two very different responses and got nearly identical scores, which made me question the tool's sensitivity and fairness.
Respondent 3: As a teacher, I noticed consistent scoring, but I couldn't tell if it truly reflected language improvement, especially for complex student writing.
Respondent 4: I believe the system scores reliably, but it can't recognize context or nuanced meanings that a human rater would easily catch.
Respondent 5: I liked the speed of results, but automated feedback was vague. It didn't help me understand how to improve or what specific mistakes I made.
Respondent 6: It seemed fair for grammar and vocabulary, but I don't think it can evaluate idea development or argument quality in longer essays.

Question 3:
How would you describe your experience navigating and interacting with the digital assessment platform? Were there any specific features that made it easier or harder to use?
Respondent 1: The layout was clean, but loading delays during the speaking section made the experience stressful and affected my concentration during recording.
Respondent 2: It was mostly easy to use, though I had trouble understanding some icons and buttons during the test, especially the audio controls.
Respondent 3: Navigation was straightforward, but it wasn't mobile-friendly. Using a phone made reading passages and typing answers really difficult.
Respondent 4: I appreciated the timer and progress bar, but some students struggled with logging in and technical glitches during writing sections.
Respondent 5: The design was user-friendly, but the speaking section lacked clear instructions and had no retry option, which stressed some of my students.
Respondent 6: I liked the interface overall, but the lack of a pause button during listening made it hard for slower readers to keep up.

Question 4:
Do you think the assessment platform was accessible and user-friendly for all students regardless of their digital skills or background? Why or why not?
Respondent 1: No, I think students with low digital literacy were at a disadvantage, especially those unfamiliar with online typing or navigation tools.
Respondent 2: Some students struggled due to limited internet access and unfamiliarity with tech tools. It wasn't equally accessible for everyone.
Respondent 3: It was manageable for tech-savvy students, but others found it intimidating, especially those with limited experience in online learning environments.
Respondent 4: The system assumes everyone has equal tech skills, but I noticed that rural or older students had more difficulties using it effectively.
Respondent 5: Not really accessible to all. One student couldn't even finish due to unstable internet, which really affected their assessment outcome.
Respondent 6: Accessibility needs improvement. Students with disabilities or lower computer skills often faced unnecessary challenges that influenced their test performance.

Question 5:
How has the use of technology-enhanced assessment influenced your learning (or teaching) strategies in English language courses?
Respondent 1: It made me focus more on short, direct answers and less on developing ideas, because I felt the system preferred concise responses.
Respondent 2: As a teacher, I now integrate more timed writing and practice tests to help students adapt to the digital test environment.
Respondent 3: I started using more digital tools in class to mirror test conditions, especially for listening and reading skills.
Respondent 4: The assessment influenced me to teach typing and basic tech skills, which aren't language-related but essential for success in the test.
Respondent 5: I changed my teaching to include more multiple-choice questions, similar to the test format, even though it limited deeper language exploration.
Respondent 6: It affected how I prepare-now I use digital platforms for practice, but I feel it takes time away from oral communication activities.

Question 6:
Do you think these digital assessments have had a positive, negative, or mixed impact on language development or teaching effectiveness? Please explain.
Respondent 1: Mixed impact-helpful for vocabulary and grammar, but not effective for speaking and real interaction, which are crucial in language learning.
Respondent 2: Mostly positive, as it encouraged more independent practice, but the lack of personal feedback limits deeper learning.
Respondent 3: It improved efficiency but reduced teacher-student interaction. Students rely too much on scores without understanding their language gaps.
Respondent 4: The automated format saves time but removes the human element of feedback that really helps students grow in expressive skills.
Respondent 5: Positive in some ways, but overly standardized. It doesn't consider students' individual learning styles or creativity in writing and speaking.
Respondent 6: I'd say negative overall. It led students to focus on passing tests rather than actually learning how to communicate effectively in English.