



<sup>1</sup> Ordu University, Ordu, Türkiye. Lecturer of the Department of Foreign Languages, School of Foreign Languages.

<sup>2</sup> Hacettepe University, Ankara, Türkiye. Associate Professor of the Department of English Language Teaching, Faculty of Education.



## ONLINE TESTS AND TECHNOLOGY IN LANGUAGE ASSESSMENT

TESTES ON-LINE E TECNOLOGIA NA AVALIAÇÃO DE LÍNGUAS

PRUEBAS EN LÍNEA Y TECNOLOGÍA EN LA EVALUACIÓN DEL LENGUAJE

Nurgül BEKDEMİR<sup>1</sup>  
nurgulbekdemir@odu.edu.tr  
İsmail Firat ALTAY<sup>2</sup>  
ifaltay@hacettepe.edu.tr



### How to reference this paper:

Bekdemir, N., & Altay, I. F. (2025). Online tests and technology in language assessment. *Revista on line de Política e Gestão Educacional*, 29(esp.1), e025039. DOI: 10.22633/rpge.v29iesp1.20474

**Submitted:** 19/05/2025

**Revisions required:** 13/06/2025

**Approved:** 30/06/2025

**Published:** 31/07/2025

**ABSTRACT:** Technology has an indispensable part of the language learning and teaching process, including assessment. This study investigates technological tools and platforms in online language testing and assessment within various aspects. It explores both formative and summative assessments at classroom-based and large-scale tests and also online tools having a facilitating effect on alternative assessment. The models, such as Computer-Based Assessment (CBA) and Computer-Adaptive Testing (CAT), are further illustrated. Artificial intelligence, which has a revolutionary effect on language assessment, is scrutinized with the current platforms. Also, critical issues in language assessment, such as reliability and validity issues and basic criticism of online language assessment based on authenticity, interactivity, impact, and practicality, are presented with the advantages and disadvantages. Consequently, the study aims to illustrate how language assessment is transformed through technology and offer significant insights for test designers, educators, and policymakers to enhance testing practices.

**KEYWORDS:** Online Assessment. Computer-based Assessment. Computer Adaptive Testing. Online Alternative Assessment. Artificial Intelligence.

**RESUMO:** A tecnologia é uma parte indispensável do processo de ensino e aprendizagem de línguas, incluindo a avaliação. Este estudo investiga as ferramentas e plataformas tecnológicas na testagem e avaliação online de línguas em diversos aspectos. Explora tanto a avaliação formativa quanto a somativa em testes realizados em sala de aula e em larga escala, bem como ferramentas online que facilitam a avaliação alternativa. Modelos como a Avaliação Baseada em Computador (CBA) e a Avaliação Adaptativa por Computador (CAT) também são ilustrados. A inteligência artificial, que tem um efeito revolucionário na avaliação linguística, é analisada juntamente com as plataformas atuais. Além disso, questões críticas na avaliação linguística, como problemas de confiabilidade e validade e críticas básicas à avaliação online com base em autenticidade, interatividade, impacto e praticidade, são apresentadas com suas vantagens e desvantagens. Consequentemente, o estudo visa ilustrar como a avaliação linguística está sendo transformada pela tecnologia e oferecer insights significativos para designers de testes, educadores e formuladores de políticas para aprimorar as práticas de testagem.

**PALAVRAS-CHAVE:** Avaliação on-line. Avaliação baseada em computador. Avaliação adaptativa por computador. Avaliação alternativa online. Inteligência artificial.

**RESUMEN:** La tecnología es una parte indispensable del proceso de enseñanza y aprendizaje de idiomas, incluida la evaluación. Este estudio analiza herramientas y plataformas tecnológicas en la evaluación y pruebas lingüísticas en línea desde diversas perspectivas. Explora tanto las evaluaciones formativas como sumativas en pruebas aplicadas en el aula y a gran escala, así como herramientas en línea que facilitan la evaluación alternativa. Se ilustran modelos como la Evaluación Basada en Computadora (CBA) y las Pruebas Adaptativas por Computadora (CAT). También se examina la inteligencia artificial, que tiene un efecto revolucionario en la evaluación lingüística, junto con las plataformas actuales. Asimismo, se abordan cuestiones críticas como la fiabilidad, validez y críticas sobre la autenticidad, interactividad, impacto y practicidad de la evaluación en línea, presentando ventajas y desventajas. El estudio busca demostrar cómo la tecnología transforma la evaluación lingüística y ofrecer perspectivas relevantes para diseñadores de pruebas, docentes y responsables de políticas educativas.

**PALABRAS CLAVE:** Evaluación en línea. Evaluación basada en computadora. Evaluación adaptativa por computadora. Evaluación alternativa en línea. Inteligencia artificial.

Article submitted to the similarity system



**Editor:** Prof. Dr. Sebastião de Souza Lemes

**Deputy Executive Editor:** Prof. Dr. José Anderson Santos Cruz

## INTRODUCTION

Technology has had a huge impact on almost every aspect of our lives, including the process of language teaching and learning. Both teachers and students now have easy access to technological tools that support language learning and teaching. Given the well-established importance of assessment in this process, it has become essential to integrate technology into the assessment landscape. Known as information and communication technology-supported (ICT-supported) or online assessment, this approach provides unparalleled opportunities in areas such as online discussions and interaction, automated marking systems, and a wide range of resources, including multimedia (Dunn et al., 2005).

It is also important to remember that learning environments are shaped to address each individual's needs and profile. Online assessment accommodates these needs by offering various assessment types that adjust to pace, timing, the number of test items, multimedia integration, user purpose, and diverse assessment contexts (Chapelle & Douglas, 2006; Dunn et al., 2005). Online assessment is widely used by teachers, students, and institutions. Teachers utilize it for formative and summative assessment purposes, while institutions rely on it for diagnostic tests, placement tests, proficiency tests, midterms, and final examinations, which typically serve as summative evaluations. Moreover, students can track their progress in language learning and prepare for these assessments. One of the key features of online assessment is its user-centric nature, enabling individuals to engage in the learning and assessment process anytime and anywhere, at their own pace.

### *Literature Review*

With technological advancements, language assessment has been transformed. This section explores classroom-based online tests, large-scale assessments, alternative online assessments, computer-based assessments (CBA), and the role of artificial intelligence in language testing.

### *Types of Assessment*

Formative assessment (FA) refers to assessment practices aimed at improving learners' performance (Harmer, 2007) and supporting further learning through teachers' feedback (Fulcher & Davidson, 2007; Stannard & Skip Basiel, 2013). FA relies heavily on feedback that helps learners bridge the gap between their current work and the desired standard (Lin & Lai, 2013; Taras, 2005). Professionals implementing FA need to use measurement tools to provide constructive feedback, expecting learners to revise their work accordingly. This creates a learning cycle that significantly impacts language development (Stannard & Skip Basiel, 2013).

In contrast, summative assessment (SA) takes place at the end of a unit, module, or course and measures what individual learners or groups have achieved and how much progress they have made (Brown, 2000; Harmer, 2007). Some scholars suggest that online SA methods could replace traditional paper-and-pen assessments (Navrat & Tvarozek, 2014). Rather than viewing FA and SA as opposing approaches, it is now widely recommended to combine them for optimal learning outcomes. FA focuses on assessing processes, while SA assesses products (Taras, 2005). By engaging students in feedback and encouraging them to revise their work, FA prepares them for SA, which evaluates how effectively the learning process has supported their progress.

For example, process and product writing represent excellent applications of online FA and SA. In process writing, learners use platforms like Turnitin Draft Coach, Turnitin Feedback Studio, and ETS Criterion to draft, revise, and edit their work based on teacher feedback, aiming to align their work with expected standards. Turnitin Draft Coach, available as an add-on for Microsoft® Word and Google Docs™, provides students with instant feedback, saving educators time and reducing grammar, citation, and academic integrity issues. This creates more formative learning opportunities for students.

Conversely, product writing as a form of SA requires learners to demonstrate their knowledge in a single submission. Tools such as Turnitin Originality Check help ensure the integrity of this work by detecting plagiarism and encouraging original output. Rather than maintaining a strict separation between FA and SA, educators are encouraged to integrate both approaches to maximize students' learning outcomes (Turnitin, n.d.a).

### *Computer-based assessment*

With the rapid integration of computers and the internet into our lives, incorporating them into language assessment has become indispensable. As a significant component of e-learning (Lin & Lai, 2019), computer-based assessment (CBA) offers a wide range of multimedia interfaces and video-based scenarios in contexts where authentic language is used. It also provides learners and test-takers with automatic response analysis and enables them to take tests anytime and anywhere (Burstein et al., 1996; Chapelle & Douglas, 2006).

CBA is also referred to as computer-assisted language testing (CALT) (Chapelle & Douglas, 2006; Suvorov & Hegelheimer, 2013) and computer-assisted language assessment (CALA) (Chapelle & Voss, 2008). Since the 1980s, CBA has undergone significant changes and improvements, such as the introduction of computer-adaptive testing, new item types, integrated skills assessments, and automated evaluations (Suvorov & Hegelheimer, 2013, p. 1).

Developing a computer-based test involves several steps: (1) test design, (2) test construction, (3) item tryout, (4) test item delivery, (5) item management, (6) item scoring, (7) item

analysis and interpretation, and (8) score reporting (Burstein et al., 1996, p. 241–242). To explain: Test design defines the framework and determines whether the test covers all four skills to assess overall language competence; Test construction involves preparing the test using technology to display texts and accept and store different types of responses, such as written or spoken; Item tryout refers to piloting the test before full implementation; Test item delivery ensures the verification of test-takers' identities and the secure collection of responses; Item management, a critical step, focuses on securely storing items to protect test integrity; Item scoring is the assessment stage, which can be automated or manually conducted by humans; Item analysis and interpretation allow test-takers to understand their scores and gain insights into their performance; Finally, score reporting uses technology to deliver results quickly, often with test-takers eager to receive their outcomes. In essence, CBA refers to the process of taking tests on computers and can be applied to both formative and summative assessments.

CBA can be structured in three distinct ways: (1) linear, (2) semi-adaptive, and (3) adaptive (Suvorov & Hegelheimer, 2013). Linear tests have two versions: the first is similar to traditional paper-and-pencil tests, where all items are presented in order and test-takers can view the entire page. In the second version, however, test-takers are not allowed to revisit previous questions.

Semi-adaptive and adaptive tests rely on selecting either testlets or individual items, respectively. These tests are considered effective because they adjust the difficulty level based on the test-takers' responses: if a response is correct, the system presents a slightly more challenging item; if incorrect, an easier item is provided. Therefore, computer-adaptive tests (CATs) are recognized for accurately assessing and reporting language proficiency to professionals and institutions.

Additionally, items are retrieved from item banks or databases (Chapelle & Douglas, 2006). It is important to note that item response theory (IRT) plays a critical role in designing CATs. IRT is based on two assumptions: (1) polytomous measurement and (2) dichotomous measurement (Jiao et al., 2012). When items are scored individually as correct or incorrect, dichotomous measurement applies. In contrast, polytomous measurement is used when items contribute to a single construct and are scored on a scale. For instance, Jamieson (2005) provided an example of reading skills assessed dichotomously, with six items scored 0–1 for correctness. Conversely, essay writing is polytomously measured, with scores ranging from 0 to 5 for each category, reflecting the presence of multiple response levels.

### *Large-scales Tests: Internet-based and Computer-based Tests*

The effectiveness of large-scale language assessments depends on maintaining consistency across key areas such as test administration, scoring, reporting, and score interpretation

during placement tests (Kunnan, 2008, p. 135). This section presents examples of language tests conducted online or through computers.

IELTS, offered in two formats—IELTS Academic and IELTS General Training—refers to the International English Language Testing System, jointly administered by Cambridge ESOL, the British Council, and IDP Education Australia (Kunnan, 2008, p. 140). While the two versions share many similarities, IELTS can be taken in either paper-based or computer-based formats, with this section focusing on the latter.

The test is widely accepted by employers, immigration authorities, and professional bodies in numerous countries (IELTS, n.d.a). It evaluates four skills: listening, speaking, reading, and writing. Test-takers complete the listening, reading, and writing sections on a computer, while the speaking section is conducted face-to-face with an examiner. Writing is assessed based on task achievement, coherence and cohesion, lexical resource, and grammatical range and accuracy, while speaking is evaluated for fluency and coherence, lexical resource, grammar, and pronunciation. Results are typically available within seven days (IELTS, n.d.b).

TOEFL Junior® Comprehensive targets students aged 11 and older. This computer-based test assesses reading, listening, speaking, and writing skills. Scores range from 200–300 for each section and 600–900 overall. The e-rater® engine evaluates writing, while the SpeechRaterSM engine assesses speaking performance (Evanini et al., 2015). SpeechRater analyzes fluency, pronunciation, prosody, grammar, accuracy, and content, whereas e-rater examines writing proficiency across ten variables, including grammar, vocabulary, and mechanics (Evanini et al., 2015).

OTE (Oxford Test of English) is a general English proficiency test that uses computer-adaptive testing for the reading and listening sections and task randomization for the writing and speaking sections. The test covers all four skills. In the listening and reading sections, question types include multiple choice, fill-in-the-gaps, and matching. In the writing section, test-takers are asked to compose an email and an essay or review. The speaking section requires them to discuss everyday topics and give short speeches on assigned subjects (OTE, n.d.a). Scores range from 51 to 140, and results—accepted by universities and educational institutions—are released within 14 days (OTE, n.d.b).

TOEIC stands for Test of English for International Communication and is designed for employees to assess their general English proficiency level. It is also administered by ETS. There are three different types of TOEIC: Listening & Reading, Speaking & Writing, and TOEIC Bridge (ETS, n.d.a). The first two are computer-based tests, which are the focus here.

The TOEIC Listening and Reading Test, lasting two and a half hours, includes seven parts: short conversations about work life in the listening section and reading passages related to business contexts (ETS, n.d.b). In the speaking section, test-takers respond to prompts and share their opinions on given topics, while the writing section comprises three



parts—writing a sentence, an email, and an opinion essay (ETS, n.d.c). The maximum achievable score is 990 points.

PTE Academic, a computer-based test, refers to the Pearson Test of English Academic, taken by students or immigrants at test centers in various countries (Pearson, n.d.a). The test is divided into three sections: Speaking & Writing, Reading, and Listening, all completed within three hours. The first section includes tasks such as personal introduction, reading aloud, retelling a lecture, and writing summaries and essays. The reading section involves multiple-choice questions, reordering paragraphs, and fill-in-the-blank tasks, while the listening section features summarizing spoken text, multiple-choice questions, identifying missing words, and writing from dictation (Pearson, n.d.b).

Assessment is conducted through PTE Academic's automated scoring system, which uses three types of responses: correct, incorrect, and partial scoring. Scores are calculated for each skill and for overall English proficiency, ranging from 10 to 90 (Pearson, n.d.c).

E-YDS, or Electronic Foreign Language Exam, is a proficiency test administered in various languages in Turkey. The exam format is similar to the traditional paper-based YDS, with multiple-choice questions assessing grammar, vocabulary, reading comprehension, and translation skills (OSYM, 2019). The main difference is that e-YDS is computer-based, and results are announced immediately after the exam ends.

TOEFL, also administered by the Educational Testing Service (ETS), stands for Test of English as a Foreign Language. The internet-based version, TOEFL iBT, indicates that it is delivered online. TOEFL scores are accepted for university applications and by institutions in over 150 countries (ETS, n.d.d).

TOEFL iBT evaluates four skills: reading, listening, speaking, and writing. In addition to standard reading and listening tasks, it includes independent and integrated speaking/writing tasks. Independent tasks require test-takers to respond to a given topic by speaking or writing an essay, while integrated tasks assess proficiency in combining skills, such as listening and speaking, reading and listening, and then speaking, or reading, listening, and writing. The total possible score is 120.

For TOEFL iBT scoring, the reading and listening sections are assessed automatically, as they include multiple-choice questions. In contrast, speaking proficiency is evaluated by three to six raters, and writing proficiency is assessed by both human raters and an automated scoring system provided by ETS, ensuring the reliability of the results (ETS, n.d.e).

### *Online Alternative Assessment*

Alternative assessment, which is closely related to authentic language use in real-life contexts, is referred to by different terms in the literature: “authentic assessment”,

“performance assessment”, “direct assessment”, and “descriptive assessment” (Hamayan, 1995, p. 213). It encompasses both the learning process and its outcomes, enabling interactive procedures for assessment. Learners also take greater responsibility for their learning and become more autonomous through alternative assessment (Tzagari, 2004).

For example, self- and peer-assessment, portfolios, and journals are used to evaluate learners’ language development (Brown, 2000; Dikli, 2003). Furthermore, technology has enhanced alternative assessment. Brown (2000) notes that self- and peer-assessment improve learners’ ability to monitor their own and their peers’ performance for revision purposes.

For peer assessment, platforms such as Turnitin Feedback Studio and ETS Criterion offer valuable opportunities for language learners to comment on their peers’ work using guided online tools (Turnitin, n.d.b). For self-assessment, Turnitin Draft Coach and ETS Criterion are effective platforms where learners can refine their academic writing skills. These platforms allow learners to identify areas requiring correction and revise their work until they achieve an optimal version. The system provides instant feedback, enabling learners to focus on improving their skills without delays.

Instead of traditional portfolios, learners are encouraged to use e-portfolios like the European Language Portfolio (ELP), LinguaFolio Online, and ePEARL. These tools help students record assignments, track their progress, and manage their learning process. E-portfolios also promote self-assessment, which is essential in the learning journey.

Journals, often referred to as reflective practices, have a significant impact on both learners and teachers. Technology can support journal keeping through blogs and social media platforms, where videos and audio recordings can be shared. Tools like Blogger, WordPress, and Edmodo Blog allow learners to enhance both writing and speaking skills.

Overall, compared to traditional testing methods, online alternative assessment enriches the learning process through multimedia integration and promotes academic growth.

### *Artificial Intelligence (AI) in Language Assessment*

Artificial intelligence (AI) is transforming language assessment. With the integration of natural language processing (NLP) and machine learning (ML) algorithms, language skills can now be assessed faster and in a more personalized way. Over the past decade, AI-based scoring systems—also called automated scoring systems or automarkers—have been widely adopted for evaluating speaking and writing skills, as algorithms have been trained to assess essays and spoken responses (Xu et al., 2025).

AI offers many benefits, including fast scoring, on-demand testing without requiring a human examiner, consistent scoring, adaptive testing, instant feedback, and seamless integration of assessment into the learning process (Xi, 2021). However, some challenges remain,



such as concerns about the authenticity and validity of scoring models, the risk of academic dishonesty, and the need for AI literacy.

Authenticity is still a limitation, as AI-based systems require further improvement to handle interaction-based speaking assessments effectively. Validity issues also arise in evaluating complex linguistic features, such as lexico-grammatical quality, discourse organization, pragmatics, and interactional factors like turn-taking. Additionally, misuse of AI tools such as ChatGPT could lead to academic dishonesty by producing human-like content (Karani & Mwancha, 2025). Developing AI literacy is essential for test users to avoid misuse and ensure appropriate application in language assessment (Xu et al., 2025).

Several platforms have adopted AI-based language assessment, particularly for writing and speaking skills. The Duolingo English Test (DET), for example, is an adaptive AI-based exam that assesses language proficiency in a short time. Productive skills like writing and speaking are evaluated through automated scoring systems.

ETS has developed SpeechRater® and e-rater®, which are used in large-scale exams such as TOEFL to assess speaking and writing reliably. These tools are widely adopted by learners worldwide. Write & Improve by Cambridge helps learners develop their writing skills by providing instant AI-generated feedback. For speaking, Cambridge is also piloting Speak & Improve, a research project aligned with CEFR levels, designed for speaking practice. Tools like TalkPal AI and SpeakNow AI further enhance language learning with speaking and pronunciation practice, with SpeakNow AI also offering assessments based on CEFR levels.

These examples demonstrate how AI-based tools can make language assessment more accessible, flexible, and personalized, supporting learners in achieving greater proficiency.

### *Basic Criticism of Technology Usage in Language Assessment*

This section focuses on evaluating CBA, including online assessment and CAT. Five criteria are discussed: (1) reliability and validity, (2) authenticity, (3) interactivity, (4) impact, and (5) practicality.

Starting with reliability and validity, several studies offer different perspectives on whether CBA is valid and reliable or not (Dunkel, 1991; Huff & Sireci, 2001; Öz, 2018; Parhizgar, 2012; Roever, 2001; Shraim, 2018). Test reliability depends on various factors such as clear instructions, familiarity with the CBA format, environmental conditions during test administration, and individual factors (Dunkel, 1991). Scoring also plays a critical role in test reliability. As the human factor can be eliminated in CBA, grading bias is reduced (Dunkel, 1991; Öz, 2018; Shraim, 2018), meaning rater reliability is less of a concern in computer-based delivery. However, the validity of tools such as e-rater and SpeechRater has been questioned (Parhizgar, 2012), and for now, they are generally not used as the sole means of assessing writing and speaking skills.

Another important point is that computers support “partial-credit scoring, which enables more precise measurement and a wider range of scores” (Chapelle & Douglas, 2006). Additionally, Öz (2018) observed that test-takers can listen to audio passages individually using headphones, which improves sound quality and reduces external distractions. Altay and Altay (2017) also found no significant impact on learners’ scores when reading texts online, supporting the reliability of technology use in assessment.

On the other hand, technical issues such as “slow computers, poor network connectivity, or power outages” can make test-takers perceive CBA as unreliable (Shraim, 2018). Moreover, some studies argue that CBA is limited to certain item types (e.g., multiple-choice, true/false, matching, and short answers) (Cook & Jenkins, 2010; Hodgson & Pang, 2012; Roever, 2001). Others, however, assert that well-constructed CBAs can assess higher-order cognitive skills as well (Huff & Sireci, 2001; Jordan, 2013; Simkin & Kuechler, 2005; Williamson, 2018).

Open-ended items can help mitigate these limitations, although CAT may struggle to assign tasks that effectively measure specific skills (Chapelle & Douglas, 2006). Innovative test items, such as audio/video components and simulated performance tasks, can make CBAs more authentic than traditional paper-based tests, potentially increasing construct validity (Huff & Sireci, 2001; Öz, 2018). Nevertheless, factors like computer literacy, time constraints, test anxiety related to CAT formats, high costs, typing speed, and technical failures can negatively affect test validity (Huff & Sireci, 2001; Roever, 2001).

In language testing, authenticity refers to assessing whether test-takers can use the language effectively in real-life contexts. Since the Internet offers vast amounts of real-life materials, it is relatively easy to integrate authentic tasks into language assessments. Multimedia elements such as audio, images, graphs, and videos make CBTs more realistic and engaging (Chapelle & Douglas, 2006; Shraim, 2018). For instance, test-takers might watch a video of a conversation and answer related questions. The advantage of using video in listening tasks is that it helps visualize the scenario, making the input more authentic and reflective of real-life experiences (Huff & Sireci, 2001).

Interactivity is also crucial in language testing. It allows test-takers to demonstrate their “language and communication strategies, topical knowledge, and affective variables in an integrative way” (Chapelle & Douglas, 2006, p. 95). CBTs enhance interaction through task types such as interpreting graphs and writing essays based on them, which require learners to engage in problem-solving and writing skills. The key factor here is “the interaction between the test-taker and the task” (Bachman & Palmer, 1996, p. 25-26).

Tests also create impacts at both micro and macro levels. At the micro level, tests directly influence individual test-takers. At the macro level, they affect educational systems and society as a whole (Bachman & Palmer, 1996). Chapelle and Douglas (2006) highlight both positive and negative impacts of CBA. On the positive side, increased use of CBA may lead

institutions to provide more computers for language testing. On the negative side, a lack of prior experience with CBA can create anxiety for test-takers, and high costs may prevent some individuals from taking the test even when necessary.

Finally, practicality is an essential factor alongside the others. Practicality refers to considerations such as “financial limitations, time constraints, administration, scoring, and interpretation” (Brown, 2000, p. 386). Developing CBA can be costly due to the need for technical expertise (Chapelle & Douglas, 2006). However, it offers advantages such as automated response analysis, which reduces scoring time, and web-based testing (WBT), which allows learners to take tests anytime and anywhere—enhancing overall practicality.

## FINAL CONSIDERATION

In conclusion, technology has become fully integrated into language assessment. The most commonly used terms in this field—CBA, WBT/ICT-supported testing, online assessment, CAT, and AI-based language assessment—have driven a significant transformation in language testing by providing clear benefits such as faster, more consistent, and adaptive evaluations.

There are both advantages and disadvantages regarding reliability, validity, authenticity, interactivity, impact, and practicality. Reliability can be enhanced through automated scoring, which helps eliminate the influence of human raters and external factors in CBA and online assessment. Although there are arguments about the limited validity of online assessments, validity can be improved by incorporating questions that test higher-order cognitive skills.

The use of multimedia makes tests more authentic than PBT, enabling assessments to measure not only language skills but also topical knowledge, demonstrating whether test-takers can effectively convey meaning in the target language. Like PBT, CBA, and online tests have a significant impact on both test-takers and society.

As for practicality, while the initial development phase may be costly, the test-taking process and scoring are more efficient and practical compared to PBT. Moreover, many of the disadvantages can potentially be mitigated with the rapid pace of technological advancement.

## REFERENCES

- Altay, İ. F., & Altay, A. (2017). The impact of on-line reading tasks and reading strategies on EFL learners' reading test scores. *Journal of Language and Linguistic Studies*, 13(2), 136-152.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford University Press.
- Brown, H. D. (2000). *Principles of language learning and teaching* (Vol. 4). Longman.
- Burstein, J., Frase, L. T., Ginther, A., & Grant, L. (1996). Technologies for language assessment. *Annual Review of Applied Linguistics*, 16, 240–260. <https://doi.org/10.1017/S0267190500001537>
- Chapelle, C. A., & Douglas, D. (2006). *Assessing language through computer technology*. Ernst Klett Sprachen.
- Chapelle, C. A., & Voss, E. (2008). Utilizing technology in language assessment. In N. H. Hornberger (Ed.), *Encyclopedia of language and education*, 2(7), 123–134.
- Cook, J., & Jenkins, V. (2010). *Getting started with e-assessment*. University of Bath. <http://opus.bath.ac.uk/17712/>
- Council of Europe. (s.d.). *European Language Portfolio (ELP)*. Retrieved March 10, 2020, from <https://www.coe.int/en/web/portfolio>
- Dunn, L., Morgan, C., O'Reilly, M., & Parry, S. (2005). *The student assessment handbook: New directions in traditional and on-line assessment*. Routledge.
- Educational Testing Service. (s.d.-a). *The TOEIC® program*. Retrieved April 15, 2020, from <https://www.ets.org/toeic/>
- Educational Testing Service. (s.d.-b). *Content and format of the TOEIC® Listening and Reading Test*. Retrieved April 15, 2020, from <https://www.ets.org/toeic/test-takers/listening-reading/about/content-format/>
- Educational Testing Service. (s.d.-c). *Content and format of the TOEIC® Speaking and Writing Tests*. Retrieved April 15, 2020, from <https://www.ets.org/toeic/test-takers/speaking-writing/about/content-format/>
- Educational Testing Service. (s.d.-d). *How ETS scores the TOEFL iBT® test*. Retrieved April 15, 2020, from [https://www.ets.org/s/toefl/flash/33910\\_toefl-scoring-video\\_transcript.html](https://www.ets.org/s/toefl/flash/33910_toefl-scoring-video_transcript.html)

- Educational Testing Service. (s.d.-e). *About the TOEFL iBT® test*. Retrieved April 15, 2020, from <https://www.ets.org/toefl/test-takers/ibt/about>
- Evanini, K., Heilman, M., Wang, X., & Blanchard, D. (2015). Automated scoring for the TOEFL Junior® comprehensive writing and speaking test. *ETS Research Report Series*, 2015(1), 1–11. <https://files.eric.ed.gov/fulltext/EJ1109726.pdf>
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment*. Routledge.
- Harmer, J. (2007). *The practice of English language teaching* (4th ed.). Pearson Education Limited.
- Hodgson, P., & Pang, M. Y. C. (2012). Effective formative e-assessment of student learning: A study on a statistics course. *Assessment & Evaluation in Higher Education*, 37(2), 215–225. <https://doi.org/10.1080/02602938.2010.523818>
- International English Language Testing System (IELTS). (s.d.-a). *Computer-delivered IELTS*. Retrieved April 15, 2020, from <https://www.ielts.org/about-the-test/computer-delivered-ielts>
- International English Language Testing System (IELTS). (s.d.-b). *Test format*. Retrieved April 15, 2020, from <https://www.ielts.org/about-the-test/test-format>
- Jamieson, J. (2005). Trends in computer-based second language assessment. *Annual Review of Applied Linguistics*, 25, 228–242. <https://doi.org/10.1017/S0267190505000127>
- Jiao, H., Liu, J., Haynie, K., Woo, A., & Gorham, J. (2012). Comparison between dichotomous and polytomous scoring of innovative items in a large-scale computerized adaptive test. *Educational and Psychological Measurement*, 72(3), 493–509. <https://doi.org/10.1177/0013164411422903>
- Jordan, S. E. (2013). E-assessment: Past, present and future. *New Directions for Adult and Continuing Education*, 2013(140), 87–106. <https://doi.org/10.29311/ndtps.v0i9.504>
- Karani, A. O., & Mwancha, C. N. (2025). The role of ChatGPT in STEM education: Assessing its prospects, challenges, ethical implications, and influence on student creativity and innovation competencies. *International Journal of Education, Technology and Science*, 5(2), 116–138. <https://doi.org/10.5281/zenodo.15489162>
- Lin, J. W., & Lai, Y. C. (2019). User acceptance model of computer-based assessment: Moderating effect and intention-behavior effect. *Australasian Journal of Educational Technology*, 35(1), 1–15. <https://doi.org/10.14742/ajet.4684>
- Navrat, P., & Tvarozek, J. (2014). On-line programming exercises for summative assessment in university courses. In *Proceedings of the 15th International Conference on Computer Systems and Technologies* (pp. 341–348).

- Oğrenci Secme ve Yerleştirme Merkezi. (2019). *2019 Yabancı Dil Bilgisi Seviye Tespit Sınavı (YDS) kılavuzu*. Retrieved April 15, 2020, from <https://dokuman.osym.gov.tr/pdfdokuman/2019/YDS-1/kilavuz30012019.pdf>
- Oxford University Press. (s.d.-a). *Oxford Test of English*. Retrieved April 15, 2020, from [https://elt.oup.com/feature/global/oxford\\_test\\_of\\_english/?cc=tr&selLanguage=en](https://elt.oup.com/feature/global/oxford_test_of_english/?cc=tr&selLanguage=en)
- Oxford University Press. (s.d.-b). *Who recognizes the Oxford Test of English?* Retrieved April 15, 2020, from [https://elt.oup.com/feature/global/oxford\\_test\\_of\\_english/recognition?cc=tr&selLanguage=en](https://elt.oup.com/feature/global/oxford_test_of_english/recognition?cc=tr&selLanguage=en)
- Pearson. (s.d.-a). *PTE Academic*. Retrieved April 15, 2020, from <https://pearsonpte.com/the-test/>
- Pearson. (s.d.-b). *Test format*. Retrieved April 15, 2020, from <https://pearsonpte.com/the-test/format/>
- Pearson. (s.d.-c). *How is the test scored?* Retrieved April 15, 2020, from <https://pearsonpte.com/the-test/about-our-scores/how-is-the-test-scored/>
- Stannard, R., & Skip Basiel, A. (2013). A practice-based exploration of technology-enhanced assessment for English language teaching. In Motteram, G. (Ed.), *Innovations in learning technologies for English language teaching* (pp. 145–174). British Council.
- Simkin, M. G., & Kuechler, W. L. (2005). Multiple choice tests and student understanding: What is the connection? *Decision Sciences Journal of Innovative Education*, 3(1), 73–97. <https://doi.org/10.1111/j.1540-4609.2005.00053.x>
- Suvorov, R., & Hegelheimer, V. (2013). Computer-assisted language testing. In A. J. Kunnan (Ed.), *The companion to language assessment* (Vol. 2, pp. 594–613). Wiley-Blackwell.
- Taras, M. (2005). Assessment—summative and formative—some theoretical reflections. *British Journal of Educational Studies*, 53(4), 466–478. <https://doi.org/10.1111/j.1467-8527.2005.00307.x>
- Turnitin Feedback Studio. (s.d.-a). *What can you do with Feedback Studio?* Retrieved March 10, 2020, from <https://www.turnitin.com/products/feedback-studio>
- Turnitin Feedback Studio. (s.d.-b). *How to use PeerMark*. Retrieved March 10, 2020, from <https://www.turnitin.com/videos/how-to-use-peermark>
- Williamson, H. (2018). On-line exams: The need for best practices and overcoming challenges. *The Journal of Public and Professional Sociology*, 10(1), 1–7. <https://doi.org/10.62915/2154-8935.1128>



Xi, X. (2021). Validity and the automated scoring of performance tests. In Fulcher, G. & Harding, L. (Eds.). *The Routledge handbook of language testing* (2nd ed., pp. 513–529). Routledge.

Xu, J., Galaczi, E., Schmidt, E., & Somers, A. (2025). Automarking in language assessment: Key considerations for best practice. *Language Assessment Quarterly*.

*CRediT Author Statement*

---

**Acknowledgements:** No.

**Funding:** This research did not receive any financial support.

**Conflicts of interest:** There is no conflict of interest.

**Ethical approval:** The work respected ethics during the research.

**Data and material availability:** The data and materials used in the work are not publicly available for access.

**Authors' contributions:** 50% each author. Each author contributed equally to the work.

---

**Processing and editing: Editora Ibero-Americana de Educação**  
Proofreading, formatting, normalisation and translation

