Digital and AI-Based Assessment in University ELT Contexts: A Systematic Review

Nebile Uslu KAPLAN[a1], Gülnur Nermin EFEOĞLU [a2]

[a1]*Amasya University, School of Foreign Languages, Amasya, Türkiye*
[a2]*Nevşehir Hacı Bektaş Veli University, Nevşehir, Türkiye*

Abstract
21st century teaching and assessment, focusing on communication and critical thinking, challenges traditional methods which mainly focus on grammar-based and structural exams. At this point, providing essential and continuous feedback, formative assessment has been prefered by language teachers even though they find them more demanding to conduct than the traditional summative assesment. To address these challenges, Digital and AI tools are used for their several advantages. Using a PICO-based approach, this systematic review examines the use, advantages and disadvantages of these tools in university-level EFL assessment practices by searching Scopus, Web of Science and ERIC databases for research articles that are peer-reviewed and empirical. Depending on the database search, out of 488 records have been identified, 53 of which were included in qualitative analysis after the exclusion criteria were applied. The results reveal the overview of the obtained studies as well as the advantages and disadvantages of utilizing these tools for formative assessment practices. The advantages include providing flexibility and personalized feedback, improving language proficiency and learner autonomy, reducing assessment anxiety, while the disadvantages consist of academic integrity, algorithmic limitations and high-demanding nature of these tools in terms of time and effort. As a result, the study provides an essential review of literature and shed light on the effective integration of digital and AI-based assessment into language classes in higher education by revealing that digital tools can be used as adaptive educational resources.
Keywords: Digital tools, AI-based tools, Formative assessment, EFL, Systematic review

Received Date: May 7th, 2026. Acceptance Date: May 28th , 2026.

1. Introduction

In today's technology-based world, a shift in paradigm is occurring in English Language Teaching (ELT), especially at the tertiary level. This shift emerged from the recognition that language is a social, political and cultural phenomenon rather than solely a technical skill. This fact has created new communicative needs under globalization that traditional educational and assessment models fail to meet. While teaching methods in higher education tend to be communicative and critical, the formats of the exams remain traditional and grammar based, which causes an essential structural mismatch (Lihua, 2025; Pascoal & De Almeida Mattos, 2022). Similarly, the pressure of high-stakes testing forces teachers to teach to the tests rather than to implement communicative curricula, and this represents a serious challenge to pedagogical innovation (Ha et al., 2021). Traditional summative assessment fails to provide sufficient support for developmental learning; therefore, the concepts of "assessment for learning" and "sustainable assessment" have emerged (Chung & Choi, 2021; Zhang et al., 2024). The idea is to embed assessment into the education and to equip students with the self-regulation skills needed for lifelong learning (Chung & Choi, 2021).

Formative assessment is widely seen as an essential technique to improve student achievement because it supplies continuous and corrective feedback that guides both teaching and learning by improving test performance and learning gains (Ramdhani et al., 2024; Sortwell et al., 2024), supporting self-regulation, engagement and motivation (Lee et al., 2020; McCallum & Milner, 2020), and guiding teaching and providing mediation (Banihashem et al., 2025; Bulut et al., 2024). However, there also exist some challenges in utilizing the formative assessment practices. Designing

**ADDRESS FOR CORRESPONDENCE: Nebile Uslu KAPLAN, Amasya University, Amasya, Türkiye, E-mail address: kaplanebile@gmail.com ORCID ID:0000-0001-8385-9475.*
**ADDRESS FOR CORRESPONDENCE: Gülnur Nermin EFEOĞLU, Nevşehir Hacı Bektaş Veli University, Nevşehir, Türkiye, E-mail address: gnefeoglu@nevsehir.edu.tr ORCID ID:0009-0009-8859-8230.*

tasks, marking them and giving feedback to each individual according to their needs are perceived as highly time-consuming, especially with large classes and intense curricula (Almahal et al., 2023; Hasan et al., 2025; Lei & Lei, 2025). Therefore, teachers struggle while providing the necessary feedback in time and they tend to use more surface-level feedback (Hasan et al., 2025; McCallum & Milner, 2020). Since the system forces teachers to use summative practices and pressures them to only grade students, it is hard for the instructors to use formative assessment for "assessment for learning" purposes (Almahal et al., 2023; Lei & Lei, 2025). Therefore, these challenges effect the feedback processes negatively, which is highly important for the improvement of the learners in the target language.

At this point, the speed and capacity of digital or AI based tools come into play. Automated scoring systems with the help of AI decrease the time needed for giving feedback and grading by helping teachers to make the process faster and more feasible (Gardner et al., 2021; Vittorini et al., 2020). Teachers and students describe AI tools to be more intelligible and actionable because they provide instant and rich feedback (Hopfenbeck et al., 2023; Saputra et al., 2024). In addition, AI tools also have the potential to solve the limited personalization of feedback by setting the difficulty and sequence of tasks according to the needs and levels of the individual students (Abrar et al., 2025; Hopfenbeck et al., 2023; Saputra et al., 2024). This is also effective in supporting learner autonomy and self-regulated learning (Liao et al., 2024). Likewise, AI tools are good at making formative assessment processes more dynamic and effective, thanks to their features such as providing real-time and personalized feedback and reducing teacher workload.

However, despite the theoretical advantages offered by these technological capabilities, a review of the existing literature reveals that empirical findings remain fragmented and a comprehensive picture of their effects, especially in higher education has not yet been established. Consequently, the systematic, long-term impact of AI tools on formative assessment practices remains unclarified. For this reason, this study aims to fill this gap in the field by systematically reviewing empirical studies in the relevant literature and presenting a systematic approach to assessment practices using digital and AI tools in university-level English language teaching settings. In relation with the purpose of the study, research questions were formulated as follows:

RQ1: How is formative assessment incorporated with digital tools and AI technologies in English language teaching research and what strategies are used to carry it out?

RQ2: How do the AI-driven or digital-tool-based formative assessment methods compare in terms of their pedagogical advantages and disadvantages?

2. Literature

2.1. Technology-Enhanced Language Learning

In the 21st century, the advancements in technology have reshaped the practices in most of the fields, including education, leading to a shift from traditional ones to new teaching methods and techniques which encourage teachers to use various digital tools and platforms for a more effective teaching-learning process (Hyland & Hyland, 2006; Susyla & Jaya, 2023). Changes across all levels of education have also reshaped foreign language instruction as well as ELT (Godwin-Jones, 2019; Jeevarajath & Hema, 2025; Zarate, 2024), which has a unique role as an important skill for communication in the world (Susyla & Jaya, 2023). To support this digital transition, different kinds of digital sources, including "virtual classrooms, online platforms and interactive multimedia content" have been introduced to help language instruction and practices (Rao & Palathil, 2020; Zarate, 2024).

Not only the improvements in technology but also the pandemic and lockdown had an impact upon this transition since the campuses were forced to close across the world which required universities to follow online instruction (Maleki, 2025). In the time of crisis, even social media platforms have been used for academic purposes (Bansal et al., 2023), which accelerated the integration of digital tools as part of the curricula (Celik & Baturay, 2024; Turnbull et al., 2021).

The integration of technology has been achieved with the help of various learner-centered practices such as "game-based learning, flipped classrooms, blended or hybrid learning" (Zarate, 2024). These educational practices are founded on the method of instruction called Computer Assisted Language Learning (CALL) which promotes learner autonomy, engagement, collaboration and personalized learning by using technology in the teaching-learning process from presentation to production and feedback (Hughes et al, 2018, Susyla & Jaya, 2023; Zarate, 2024). Technology-Enhanced Language Learning (TELL) provides contemporary learners with multimodal instructional resources and genuine language exposure, as thoroughly explained by Zarate (2024) in the literature. Susyla and Jaya (2023) also corroborate the fact that such digital tools are successful in enhancing intrinsic motivation and student autonomy in internationalized virtual settings.

But not everyone is happy with this digital shift. Zarate (2024) warns that the presence of digital instruments is not enough to ensure the quality of teaching in online environment, and that language acquisition depends on the face-to-

face relationship, which is often missing in the online environment. Although traditional instructor-led oral explanations still demonstrate superior efficacy in some contexts (Zarate, 2024), prevailing opinion in current ELT research is that the strategic benefits of TELL always outweigh these operational restrictions. Therefore, the focus of the present research has moved from general classroom practices toward the specific role of digital tools in assessment processes.

While there are some concerns, the advantages of TELL are far more numerous than its difficulties. Hence, digital tools have been frequently used in foreign language education in recent years. However, the use of digital tools is not limited to classroom practices or supplementary activities outside the classroom. The use of digital tools in assessment procedures and its impact on language learning have also become a significant research topic in the field of ELT.

2.2. The Evolutionary Shift from Traditional to Digital Assessment

Having a crucial role in the effective learning process, assessment is an integral and indispensable part of education (Amua-Sekyi, 2016; Maleki et al, 2025). With the help of results of assessment practices, the educational process can be shaped by making the necessary adjustments accordingly in the process of curriculum development, adopting teaching methods and setting the policies of education (Abdulrazzaq & Abdellatif, 2023; Zhao, 2024). Although assessment can be carried out in various ways through both formative and summative practices ranging from writing tasks to presentations and projects (Abd Elgalil et al., 2022), the main goal of it does not change. It gives the teachers the opportunity of providing students with feedback, evaluating and personalizing the teaching-learning process and adapting the teaching-learning practices considering the assessment process in order to enhance the quality of education (Kushari & Septiadi, 2022; Maleki et al, 2025; Tosuncuoglu, 2018).

The advancements in technology have inevitably affected the assessment procedures in the field of ELT, and a term called digital assessment has emerged. Digital assessment refers to utilizing digital platforms in the testing procedure from item development and test administration to evaluation and statistical interpretation (Shadiev & Yang, 2020) in order to understand students' linguistic competence and communicative efficacy (Susyla & Jaya, 2023). Instructors have started to make use of online sources to assess their students' performances owing to the technological advancements and the accessibility of modern technology, resulting in digital assessment taking the place of traditional assessment methods (Atmojo & Nugroho, 2020; Eri et al., 2021; Maleki et al., 2025b; Muñoz-Alcón et al., 2023; Nguyen & Nguyen, 2021) since a systematic incorporation of digital assessment tools is essential and should be normalized (Jeevarajath & Hema, 2025). While traditional methods create limitations in time and space, digitalization has made the process more flexible by blurring these boundaries.

Digital assessment offers a more efficient and learner-centered paradigm with a more holistic and adaptive approach (Mansory, 2020; Susyla & Jaya, 2023). Students may be involved in digital assessment procedures in different ways from online questionnaires to e-portfolios (Zarate, 2024) and from interactive assignments to online oral assessments (Susyla & Jaya, 2023). With authentic applications (Jeevarajath & Hema, 2025), which facilitate students to participate in digital scenarios, asynchronous discussions and online interactive projects (Jeevarajath & Hema, 2025) digital assessment paves the way for personalized and adaptive learning (Jeevarajath & Hema, 2025; Wei, 2024) and offers scaffolding, which fosters self-regulation and supports learner engagement (Motteram, 2013; Wei, 2024).

The advantages of digital assessment are numerous. In terms of pedagogy, digital assessment uses a variety of testing methods to suit different learning needs and thus tailors the educational journey through a personalized approach (Jeevarajath & Hema, 2025; Susyla & Jaya, 2023; Zarate, 2024). With instant feedback meeting their own needs, students get the opportunity to analyze their strengths and weaknesses for the specific subject (Deng et al., 2024; Jeevarajath & Hema, 2025; Susyla & Jaya, 2023; Zarate, 2024). This promotes self-directed learning (Jurane-Bremane, 2023) and encourages students to take charge of their own learning process, which is an essential skill for lifelong education (Jeevarajath & Hema, 2025). Students can learn from their mistakes, skip the classes they have mastered before, focus on the subjects they need to study more and compare their results, which results in continuous improvement (Jeevarajath & Hema, 2025; McGuire, 2021). In this sense, it can be said that students become more autonomous and self-regulated learners with the help of digital assessment practices since they can observe their own progress by tracking their speed and time required for each target subject (Jeevarajath & Hema, 2025; Theobald M, Bellhäuser H.). From a psychological perspective, immediate feedback grants learners get a sense of achievement after each practice (Zarate, 2024) and with gamified components and interactive elements, they enjoy and engage in the process with an enhanced enthusiasm (Jeevarajath & Hema, 2025; Zarate, 2024). They have an experience that is more comfortable, relaxing and easy to concentrate, which makes them involved in the process more actively (Muñoz-Alcón et al., 2023). Beyond pedagogical and psychological advantages, these tools are highly accessible and convenient (Deng et al., 2024; Muñoz-Alcón et al., 2023) for both teachers and students with reasonable costs and little physical effort (Maleki, 2025; McGuire, 2021; Muñoz-Alcón et al., 2023), which eliminates physical distance and dissolves borders in a digital space, resulting in bridging cultures through shared goals, fostering intercultural

proficiency (Jeevarajath & Hema, 2025), and thus more authentic evaluation of language use (Susyla & Jaya, 2023). With increased flexibility (Maleki et al, 2025; Maleki, 2025), students take tests according to their own schedule (Akimov & Malin, 2020; Alruwais et al., 2018; Butler-Henderson & Crawford, 2020; Muñoz-Alcón et al., 2023), which saves time for both students and teachers (Khan and Khan, 2018; Muñoz-Alcón et al., 2023). Ultimately, digital assessment fosters inclusivity (Susyla & Jaya, 2023) by appealing to different learning profiles, including students with special needs (Magic Box, 2023). They also achieve equity with an indifferent and uniform grading, which provides a more accurate reflection of students' competencies (Jeevarajath & Hema, 2025; Susyla & Jaya, 2023).

However, technological tools come with their challenges as well. From an infrastructural perspective, since digital assessment is based heavily on internet data, technical problems may occur and there may be limitations in connection via internet, which directly affects the practicality and reliability of them (Susyla & Jaya, 2023; Yulianto & Majid Mujtahid, 2021). This situation may cause anxiety for both students and teachers. Students may also feel anxious because of their economical, technical conditions and their technological abilities together with time pressure (Deng et al., 2024). These systemic vulnerabilities tend to take the form of psychological obstacles, like in the case of technology intimidating learners by depriving conversations in speech evaluations of their naturalness and warmth (Muñoz-Alcón et al., 2023). At the instructional level, teachers also experience some difficulties in terms of psychology and practicality if they lack teacher training related to digital assessment, resulting in heightened stress level and discomfort (Joshi et al., 2020; Kumar and Sajja, 2020; St-Onge et al., 2021). Since they cannot capture the spontaneous interaction with their students, they may not explain the misunderstandings in the learning process. This absence of feedback limits personalization and may harm the relationship between teachers and students, which is crucial for student involvement and motivation in the process (Hickey & Harris, 2021; Winstone & Boud, 2020).

Institutional integrity is another essential issue in addition to psychological and pedagogical issues, as academic dishonesty may be one of the most critical issues related to digital assessment procedures (Adzima, 2021; Mellar et al., 2018; Perwitasari et al., 2021). In the classroom, there is no physical supervision and the chance of cheating increases which is an unsatisfactory circumstance for the efficacy of the teaching-learning process (Maleki, 2025).

To mitigate these multidimensional challenges, the literature emphasizes targeted intervention strategies. There should be strategies to overcome the challenge of time pressure, to improve the technicality and to decrease the stress level in the assessment procedures (Deng et al., 2024). Both teachers and students should be involved in training programs regarding technology enhanced language learning and digital assessment to improve their technological skills and enhance their familiarity with digital platforms (Deng et al., 2024). Also, the research strongly emphasizes that for more efficient and optimized assessment practices, there is a need for strategic integration of digital tools as a complementary source without underestimating the human element and the essential role of teacher in both assessment and instruction procedures (Bahrani, 2011; Bahari 2021; Elmahdi & Al-Hattami, 2018, Potok & Spjut, 2023).

2.3. The Transition to Intelligence: AI in Assessment Practices

As a technological paradigm deeply intertwined with human cognitive processes, AI has become increasingly integrated into our lives in recent years, and it has been creating machines imitating the processes of human intelligence (Dogan et al., 2023). The progress in AI technology created a new era for the educational landscape with various services (Shaalan & Ahmad, 2025). In the field of ELT, students can make use of different generative AI tools from ChatGPT to Gemini in order to manage assignments, address difficult questions and produce original content with pinpoint accuracy (Smolansky et al., 2023). From the teachers' point of view, AI can be used to evaluate assignments by automatically grading, produce feedback with reasonable explanations and examples in real-time, and provide students with various suggestions (Asadi et al., 2025). However, it can be said that AI has reshaped education with both strategic advantages and critical risks (Cotton et al., 2024; Nguyen Thanh et al., 2023).

AI tools are now known for their several facilities in the literature of foreign language education. In terms of affective factors, AI tools provide learners with interactive and engaging activities which makes the learning process more interesting (Rusmiyanto et al., 2023; Zarate, 2024). Since learners are involved in an interactional learning experience which appeals to their own preferences and interests, AI enhances their engagement, motivation and self-efficacy, which are crucial in the learning process of a foreign language (Asadi et al., 2025; Firat, 2023). At a cognitive level, AI has been found beneficial in different higher mental skills like analytical reasoning and creativity rather than replacing human effort (Shaalan & Ahmad, 2025). From a pedagogical perspective, the provision of immediate and individualized feedback, which is often lacking in with the traditional methods in the classroom (Lee et al., 2022), students are assisted in the language learning process according to their own needs (Asadi et al., 2025; Rusmiyanto et al., 2023; Zarate, 2024). Therefore, it can be said that AI is a mediator since it provides learners with guided support which creates a scaffolded learning environment within the Zone of Proximal Development (ZPD) (Asadi et al., 2025; Shabani, 2016) in a short period of time and a high rate of accuracy (Asadi et al., 2025).

However, the integration of AI within educational frameworks has sparked intense discourse concerning its pedagogical advantages versus its potential drawbacks (Smolansky et al., 2023; Zohny et al., 2023). One of the major negative outcomes of the use of AI in education is the over-dependence on it, as it leads to a decline in the creativity of learners and other psychological issues like feeling worried and anxious all the time, as they are not sure about their answers submitted to their teachers (Shaalan & Ahmad, 2025). From a pedagogical perspective, AI tools are good at surface level questions, but they typically do not provide (Grassini, 2023; Wang & Demszky, 2023).

Beyond these cognitive and pedagogical limitations, institutional integrity remains a critical concern, as the use of AI has given rise to another growing concern which is directly linked to academic integrity, plagiarism and ethics (Asadi et al., 2025; Le, 2024; Shaalan & Ahmad, 2025). Therefore, striking a balance in the use of AI tools has become an essential requirement for such practices (Habib et al., 2024).

To achieve this balanced framework, a synergistic combination of AI tools and teacher assessment practices might be considered in the foreign language education context since immediate and tailored feedback from AI tools and teachers' expertise within the context may constitute a significantly greater improvement in learners' learning process by providing an optimized, adaptive and learner-centric evaluative structure (Asadi et al., 2025). This integrated feedback strategy offers a more holistic solution for skill improvement rather than relying on any single instructional method (Asadi et al., 2025).

The theoretical and practical insights discussed in this literature review part highlight the transformative potential of digital tools and AI. However, a rigorous synthesis of recent empirical evidence is necessary to evaluate the actual outcomes of these tools in formative language assessment. To synthesize these scattered findings and provide a cohesive framework for researchers and practitioners, the following section details the systematic methodology employed to identify, select and analyze the most relevant empirical studies in this domain.

3. Method

3.1. General Background

This study systematically reviews the literature to examine the use of AI-enhanced and digital-tool-based formative assessment practices in English language teaching contexts, paying special emphasis to the pedagogical uses and advantages as well as the disadvantages of this approach. High-impact academic databases, including Scopus, Web of Science and ERIC were searched for pertinent empirical research. These databases are well known for being reliable resources for research on education and ELT and assessment. Together, they index the most prestigious publications and give users access to a vast collection of empirical, peer-reviewed research that is pertinent to language evaluation, use of AI and digital tools and formative assessment. To ensure a thorough and focused selection of studies aligned with the study questions, the search strategy was guided by the PICO framework, which focuses on the Population, Intervention, Comparison and Outcomes. It was then improved using Boolean operators (AND, OR, NOT).

3.2. Search Strategy

Using the PICO framework, the related scientific publications were obtained from top academic databases, namely Scopus, Web of Science and ERIC, because of their wide cover of research. The search approach focused on research on formative assessment procedures in ELT, integrating the use of digital and AI-based tools. Terms like "formative assessment," "artificial intelligence," "digital tools" and "English language assessment" were among the keyword combinations that were employed. The keywords were taken from abstracts, publication titles and indexed terms. To narrow down the search, truncation symbols, parentheses and boolean operators (AND, OR, NOT) were used. Furthermore, as indicated in Table I, synonyms and keyword variations were added to guarantee a thorough investigation of related studies.

Table 1. Database Search Strings Used for the Study

Databases	Strings	Filters
Scopus	(assessment OR "language assessment" OR "assessment practice*" OR "classroom assessment" OR "formative assessment") AND ("digital tool*" OR "educational technolog*" OR "technology-enhanced" OR "computer-assisted" OR "learning management system*" OR "artificial intelligence" OR AI OR "AI-based" OR "automated assessment" OR "intelligent tutoring system*") AND	PUBYEAR > 2014 AND PUBYEAR < 2026

	(EFL OR ELT OR "English as a foreign language") AND ("higher education" OR universit* OR undergraduate*)	
Web of Science (WoS)	(assessment OR "language assessment" OR "formative assessment") AND (digital OR technolog* OR online OR "technology-enhanced" OR "learning management system*" OR "educational technolog*" OR "artificial intelligence" OR AI OR "AI-based assessment" OR "automated assessment") AND (EFL OR ELT OR "English as a foreign language") AND ("higher education" OR universit* OR undergraduate*)	**Timespan:** 2014–2026 **Document type:** Article **Research areas:** Education Educational Research, Linguistics, Language Linguistics
ERIC	(assessment OR "language assessment" OR "formative assessment") AND ("digital tools" OR "educational technology" OR online OR "learning management systems" OR "artificial intelligence" OR AI OR "automated assessment" AND (EFL OR ELT OR "English as a foreign language") AND ("higher education" OR "university students")	**Between:** 2014-2026

Data was collected in January 2026

As can be seen in Table I, the strings used for the search in all databases included “formative assessment”, “artificial intelligence”, “digital tools”, “English language learning” and other related terms. These terms created the basis for the inclusion criteria as well. In order to identify contemporary empirical research that demonstrates the current benefits and instructional functions of digital and AI-supported assessment tools in university EFL contexts, the study examined articles published between 2014 and 2026. Prior to 2014, the use of digital tools and AI for ELT assessment was relatively new, with minimal testing and little involvement in developing practices. Additionally, recent thorough re-analyses in computer-assisted language learning and educational technology began their research in the mid-2010s. This shift underlines that era as the time when digital test research made the most significant development. In order to ensure consistency in both the period and the concepts, the 2014–2026 window was selected.

3.3. Selection Strategy

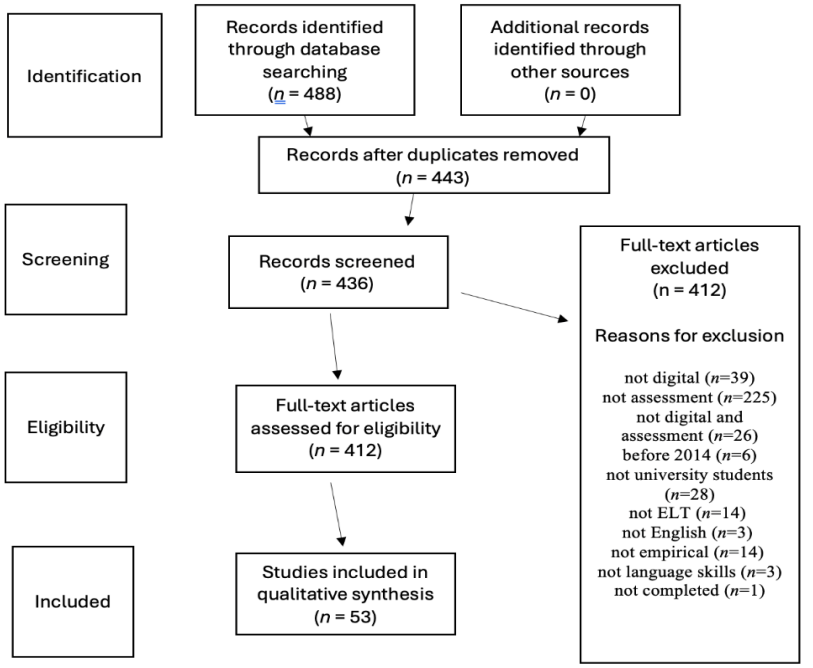
The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) methodology was employed when selecting and screening the studies. To ensure methodological rigor and relevance to the research topics, studies were evaluated according to predetermined inclusion and exclusion criteria, as described in Table II.

Table 2. Inclusion and Exclusion Criteria of the Studies

Inclusion Criteria	Exclusion Criteria
Indexed in Scopus, ERIC or WOS	Not indexed in Scopus, ERIC or WOS
Focused on formative assessment	Not focused on formative assessment
Focused on AI or AI tools	Not focused on AI or AI tools
Related to language learning or assessment	Not related to language learning or assessment
Integrating both AI or digital tools and formative assessment	Not integrating both AI or digital tools and formative assessment
Articles	Not articles (books, conference proceedings, etc.)

To further hone the selection process, the articles found were reassessed by examining titles to determine whether pertinent keywords are included, verifying abstracts for conformity to the methodology, stated outcomes and target population reading the entire manuscript to evaluate methodological soundness and thematic significance. The number of studies obtained, screened and ultimately included in the evaluation is shown in Figure I.

Figure 2. Article Selection Process and Numbers



As Figure 1 illustrates, 53 out of 488 studies were included in the study suitable to the pre-determined criteria. Totally 435 studies were excluded from the research for a variety of reasons including not focusing on formative assessment (n = 225), not focusing on AI or digital tools (n = 39), not related to ELT (n = 14), not integrating both AI or digital tools and formative assessment (n=26), not empirical (n=14) and university level (n = 28).

3.4 Quality Appraisal and Data Synthesis Framework

After the PRISMA screening process, a systematic quality appraisal technique was employed to the final 53 articles by using the Mixed Methods Appraisal Tool to assess methodological quality and bias risk thoroughly (Hong et al., 2018). Two researchers assessed each of the 53 studies separately according to the criteria in alignment with its empirical design. Even though there exist some differences in sample sizes and contextual background, all final publications demonstrated acceptable methodological accuracy to contribute to the synthesis, therefore no studies were excluded based on the quality appraisal scores.

A structured coding scheme that focused on the use, advantages and disadvantages of digital and AI tools was constructed for the data extraction procedure. A random group of roughly 20% of the included articles (*n*=11 studies) was separately coded by both researchers to verify inter-rater coding reliability and extract researcher bias. During this process, a 91% inter-rater agreement rate was obtained and any minimal disagreements about code borders were settled by cooperative discussion and consensus.

3.5 Data Analysis and Thematic Synthesis

The theme analysis framework developed by Braun and Clarke (2006) was used to synthesis the retrieved qualitative and descriptive data under six steps. First, the findings sections of the 53 publications were reviewed repeatedly to become familiar with the data. Second, prior codes were created according to the research questions. Third, these basic codes were grouped under more general topics. Fourth, to justify contextual accuracy, the candidate themes were compared to the original texts. Fifth, the themes were rearranged and designated under the research question-based subtitles. By organizing the developed concepts, the final report was developed.

4. Findings

Modern theoretical frameworks are used in formative assessment to improve the accuracy and responsiveness of evaluating learner development, particularly when paired with AI or digital tools. Suitable to respond to the research questions, this section provides a comprehensive summary of the studies that were included in this review by utilizing

descriptive statistics as in the same direction of the first research question and a detailed examination of the primary theoretical foundations that direct the usage of AI-driven or digital-tool-based formative assessment in English language teaching scenarios as requested to be responded for the second research question.

RQ1: How is formative assessment incorporated with digital tools and AI technologies in English language teaching research and what strategies are used to carry it out?

To address the first research question, the 53 obtained articles focusing on formative assessment utilized by AI or digital tools in ELT were analyzed in terms of four dimensions: assessed area, research design, use of AI or digital tools and participants. An overview of these studies is structured in Table 3.

Table 3. Overview of Studies on AI-Driven or Digital-Tool-Based Formative Assessment in English Language Teaching

Article	Assessed Area	Research design	Use of AI or Digital tools	Participants
Dong-Ok et al. (2025)	essay writing	corpus analysis	Two AI text corpora generated using GPT	English first-year university students from United States and South Korea
Shu-Chiao (2022)	writing	a mixed-methods approach	VocabProfiler 1Checker	33 Chinese-speaking English major students and 31 non-English major EFL students in Southern Taiwan
Alavi et al. (2022)	general language skills	mixed methods study with a quasi-experimental design	Google Forms Oteacher Platform	28 learners of two intact classes of a language school in Kashan, Iran In addition, 345 online questionnaire participants
Öz & Özturan (2018)	content knowledge of ELT teacher candidates	post-test only experimental study	www.classmarker.com	100 ELT student teachers at Hacettepe University
Shadiev et al. (2025)	general language skills	quasi-experiment utilizing a mixed method research approach	TVT or 360VT and a mobile online learning platform	52 postgraduate students from a public normal university and two experienced EFL teachers
Dewi & Slamet (2025)	oral skills	qualitative case study design	Learning Management System (LMS) such as Moodle, Google Classroom, and others	72 first-semester undergraduate students from Indonesia
Li et al. (2024)	writing	quantitative scoring with qualitative feedback	ChatGPT versions 3.5 and 4	30 non-English major undergraduate students and four college English teachers from China
Li et al. (2025)	speaking	univariate and multivariate generalizability theory	ChatGPT4o	40 first-year English major undergraduate students from China
He & Wang (2025)	writing	process-genre action research design	LMS and live-broadcasting platforms	a total of 265 students in three classes of English writing in south China

Borrego (2025)	writing	mixed-method approach with corpus analysis	Trinka. AI	70 English learners from Spain
Nguyen (2024)	students' acceptance of using LMS	descriptive design with a mixed-method approach	LMS	120 students fromVietnam
Saeedi & Soltani (2025)	pragmatics	explanatory mixed method design	Chatlingo: Researcher-made AI chatbot platform	87 undergraduate students
Mahadini et al. (2021)	writing	document analysis	automated Coh-Metrix	81 EFL students' essays
Sun & Shin (2025)	writing	quasi-experimental study	Computer-Assisted and Instructor's Mediational Moves	85 EFL students
Mohamadi (2018).	writing	time series design (with pre and post tests)	electronic writing forum-EWF	130 Iranian EFL students
Silva (2022)	oral skills	an unexplained descriptive pilot-based case study utilizing a task-based performance design.	Screencast-O-Matic	330 Spanish participants
Suhartoyo et al. (2025)	reading	a two-stage approach encompassing research and development (R&D)	The Quizizz	an intermediate reading course students from Indonesia
Alharbi & Meccawy (2020)	general language skills	experiment with a pre- and post-experiment design	Socrative	47 students from a Saudi state university
Fukuda et al. (2024)	general language skills	a cross-sectional, quantitative research design	Web-Enhanced Language Learning (WELL) Questionnaire	173 EFL students in Japan
Wei & Shin (2025)	writing	quasi-experimental study	mediational moves	84 L2 learners from China
Rofiah (2020)	vocabulary	a non-experimental quantitative research design	Socrative	461 first-year students from Thailand
Yavuz et al. (2025)	writing	essay evaluation	ChatGPT (Model GPT4) and Bard (Version: 2023.07.13)	15 experienced EFL instructors
Jin & Liu (2024)	reading	a quasi-experimental design	iSpring QuizMaker	two intact classes of first-year, non-English major undergraduates in a Chinese university
Petchprasert (2021).	writing	pre-test/post-test longitudinal study	Coh-Metrix (a computational text analysis tool)	80 English-major students from Thailand
Zheng et al. (2023)	public speaking anxiety	an explanatory sequential mixed-methods approach with a quasi-experimental designed	Personal Report of Public Speaking Anxiety (PRPSA) EPS performance rubrics	51 English learners
Tran & Ma (2021)	listening skills	qualitative research design	LMS platform (Schoology)	60 first-year English major students from Vietnam

Kamrood et al. (2021)	general language skills	designing and implementing an interventionist online computerized dynamic test	a web-based software (www. dynamic-english-learning.com.)	54 Iranian EFL students
Ghouali & Cecilia (2021)	writing	experimental design with pre- and post-test	Moodle	42 first-year undergraduate EFL students from Algeria
Chang & Lin (2020)	oral ability	quasi-experimental design	Oral Test Rubric and Rubric for Video Projects	60 second-year students from Taiwan
Zou et al. (2025)	speaking	qualitative research design	EAP Talk (https://www.eaptalk.com/)	218 Chinese EFL students
Greenblatt & McDonald (2024)	communicative skills	mixed methods case study	a multimodal video project	an EFL classroom in a Japanese university.
Ebadi & Rahimi (2019)	writing skills	a sequential exploratory mixed-methods approach	Google Docs	6 EFL students
Liu & Aryadoust (2022)	public speaking	quasi-experimental study with a three-group pre-test/post-test design	Public Speaking Competency Instrument (PSCI) and Personal Report of Confidence as a Speaker (PRCS)	74 students from China
Ennadir & Habiballah (2025)	general language skills	comparative-descriptive study	ChatGPT-4	30 participants from Morocco
Elmotri et al. (2025).	writing	mixed-methods approach	surveys, interviews, observations, and written assessments	150 EFL students and 40 instructors from Saudi universities
Topuz et al. (2025)	essay writing	descriptive nonexperimental research	A GenAI tool (ChatGPT 3.5)	210 essays produced by 35 undergraduate students from Türkiye
Ulum (2020)	spoken and written language skills	phenomenography	Versant English Test (VET)	30 students and 10 instructors from Türkiye
Liu et al. (2023).	writing	mixed-method design	a pre- and post-English writing test, questionnaires, interviews	64 students
Chanpet et al. (2020)	media creation	quasi-experimental study	PBL framework instrument	60 pre-service EFL teachers
Al-Abri et al. (2025).	vocabulary	a mixed-methods research design	Mentimeter	275 Level 4 English course students
Ebadi et al. (2023)	listening	a multiple case study design	a web-based CDA software	94 Iranian EFL students
Abdullaeva et al. (2024)	language learning outcome and personal development	a quasi-experimental design	Academic Demotivation Scale (ADS), The Revised Rosenberg Self-esteem Scale, Mindfulness Attention and Awareness Scale (MAAS), A thirty-item objective exam	60 intermediate EFL students

Abdullaeva et al. (2024).	academic emotion regulation (AER), academic resilience (AR), willingness to communicate (WTC), and academic enjoyment (AE), technological literacy and acceptance	an experimental design	A special Telegram group, Academic Emotion Regulation Questionnaire (AERQ), WTC Questionnaire (WTCQ)	67 language institution students in Uzbekistan.
Sayed et al. (2024)	speaking	a concurrent mixed-methods approach	ChatGpt	an intact class with 28 learners at a university in Ethiopia
Sawaki et al. (2024)	writing	an instrumental case study	Two on-demand videos	74 Japanese undergraduate students
Aladini et al. (2024)	academic motivation	cross-sectional quantitative study following a correlational-explanatory design	Core psychological needs satisfaction measure (CPNSM), The Teacher Support Measure (TSM), MacIntyre et al.'s scale, Academic Motivation Scale	185 EFL students from Oman
Khasawneh et al. (2025)	emotion regulation, mindfulness, and language learning attitudes	quasi-experimental study	Mindful Attention Awareness Scale, Academic Emotion Regulation Questionnaire, A researcher-designed 12-item attitude questionnaire focused on AI	69 students from Ethiopia
Gao et al (2025)	writing	an explanatory sequential mixed-method design	DeepSeek-V3 and R1	92 students from China
Al-Obaydi & Pikhart (2025)	writing	a qualitative-dominant mixed method approach	ChatGPT-4	the fourth grade of EFL students
Mirsanjari (2025)	motivation, anxiety, and proficiency	an explanatory sequential mixed-methods design	Intrinsic Motivation Inventory, Language Learning Motivation Questionnaire, Foreign Language Classroom Anxiety Scale, Second Language Writing Anxiety Inventory, Proficiency Tests, Semi-structured interviews	120 EFL learners from Tehran
Lin (2019)	mind-mapping flipped learning	an experimental design	Assessing Writing Performance- Level B2, Cambridge English Language Assessment	57 students
Pourdana (2022)	writing	a repeated-measures design with experimental procedure	Moodle LMS	41 students from Iran

Bhatnagar et al. (2025)	public speaking	a field research approach	Digital Randomizers (DRs)	82 students from diverse disciplines

Table 3 reveals the four basic dimensions of the obtained articles, namely assessed area, research design, use of AI or digital tools and participants. Regarding the assessed area, a primary concentration is identified in productive skills, with essay writing as the most focused area in higher education (Dong-Ok et al., 2025; Gao et al., 2025; He & Wang, 2025; Li et al., 2024; Liu et al., 2023; Mahadini et al., 2021; Mohamadi, 2018; Petchprasert, 2021; Shu-Chiao, 2022; Sun & Shin, 2025; Topuz et al., 2025; Wei & Shin, 2025; Yavuz et al., 2025). This is accompanied by targeted studies of oral skills, public speaking and communicative fluency (Chang & Lin, 2020; Dewi & Slamet, 2025; Greenblatt & McDonald, 2024; Li et al., 2025; Liu & Aryadoust, 2022; Silva, 2022; Ulum, 2020; Zheng et al., 2023; Zou et al., 2025).

Additionally, the scope of investigated areas expands beyond traditional linguistic skills into affective variables, treating the learner's emotional state as an integral part of the feedback including academic emotion regulation, resilience, mindfulness, willingness to communicate, self-esteem and anxiety of the learners (Abdullaeva et al., 2024; Aladini et al., 2024; Khasawneh et al., 2025; Mirsanjari, 2025). Conversely, receptive competencies, specifically reading comprehension and vocabulary acquisition, stayed underdeveloped (Al-Abri et al., 2025; Jin & Liu, 2024; Rofiah, 2020; Suhartoyo et al., 2025).

From a methodological perspective, the field displays a strong preference for mixed-methods structures and quasi-experimental pre-test/post-test designs to triangulate pedagogical efficacy (Alavi et al., 2022; Elmotri et al., 2025; Jin & Liu, 2024; Khasawneh et al., 2025; Liu et al., 2023; Mirsanjari, 2025; Nguyen, 2024; Saeedi & Soltani, 2025; Shu-Chiao, 2022; Sun & Shin, 2025). In the same direction, quantitative data are frequently supported with qualitative data derived from surveys, interviews or longitudinal document analyses. In addition, implementations are also utilized through computer-based Dynamic Assessment (DA) and mind-mapping techniques (Kamrood et al., 2021; Lin, 2019; Sun & Shin, 2025; Wei & Shin, 2025). These designs are integrated with qualitative approaches, including descriptive case and phenomenological studies and corpus analyses (Borrego, 2025; Dewi & Slamet, 2025; Dong-Ok et al., 2025; Ebadi et al., 2023; Sawaki et al., 2024; Ulum, 2020).

Looking at the table, it's clear there is a major shift from basic administrative platforms toward advanced generative systems. LMSs like Moodle, Schoology and Google Classroom are frequently used for digital testing and online writing purposes (Alavi et al., 2022; Dewi & Slamet, 2025; Ghouali & Cecilia, 2021; Mohamadi, 2018; Pourdana, 2022; Tran & Ma, 2021). For the development of vocabulary and content knowledge, gamified quiz applications and interactive tools like Socrative, Quizizz, ClassMarker, Mentimeter, and iSpring QuizMaker are frequently integrated to promote higher engagement of the students (Al-Abri et al., 2025; Alharbi & Meccawy, 2020; Jin & Liu, 2024; Öz & Özturan, 2018; Rofiah, 2020; Suhartoyo et al., 2025). Additionally, for the development of linguistic competences, technological platforms such as VocabProfiler, 1Checker, Coh-Metrix metrics, Screencast-O-Matic, Digital Randomizers, and the Versant English Test are implemented (Borrego, 2025; Bhatnagar et al., 2025; Mahadini et al., 2021; Petchprasert, 2021; Shu-Chiao, 2022; Silva, 2022; Ulum, 2020).

The findings indicate another shift toward Generative AI (GenAI) use for automated scoring and personalized feedback by implementing ChatGPT (GPT-3.5, GPT-4, GPT-4o), DeepSeek (V3 and R1), Bard, and other chatbot platforms like Chatlingo (Al-Obaydi & Pikhart, 2025; Ennadir & Habiballah, 2025; Gao et al., 2025; Li et al., 2024; Li et al., 2025; Saeedi & Soltani, 2025; Sayed et al., 2024; Topuz et al., 2025; Yavuz et al., 2025). Lastly, demographic information gathered through the literature review implies that participants mainly consist of undergraduate students, pre-service teachers and non-English areas (Fukuda et al., 2024; Li et al., 2025; Öz & Özturan, 2018; Rofiah, 2020). Geographically, the research is centered in Asian and Middle Eastern contexts, with prominent studies from China, Iran, Turkey, Saudi Arabia, Taiwan, Vietnam, Thailand and Uzbekistan (Alharbi & Meccawy, 2020; Alavi et al., 2022; Abdullaeva et al., 2024; Elmotri et al., 2025; Petchprasert, 2021; Shu-Chiao, 2022; Topuz et al., 2025; Tran & Ma, 2021). Sample sizes range from micro-level qualitative case studies involving 6 to 30 participants (Ebadi & Rahimi, 2019; Li et al., 2024; Sayed et al., 2024) to mid-scale experimental designs with 40 to 150 learners (Ghouali & Cecilia, 2021; Li et al., 2025; Saeedi & Soltani, 2025) and large-scale surveys involving up to 461 individuals (Rofiah, 2020). This distribution demonstrates that digital and AI-based formative assessment practices are being implemented for a variety of participant profiles and contexts across higher education environments.

RQ2: How do the AI-driven or digital-tool-based formative assessment methods compare in terms of their pedagogical functionality and challenges?

To respond to the second research question, the findings of the 53 obtained articles were analysed through thematic analysis to reveal the advantages and disadvantages of utilizing digital and AI-tool-based formative assessment practices in ELT contexts.

Theme 1. Advantages of Using Digital or AI-based Formative Assessment Tools

The results sections of the 53 included articles were systematically analyzed to detect the pedagogical advantages of technology-enhanced formative assessment practices and the results are presented in Table 4.

Table 4. The Sub-Themes of the First Theme: Advantages of Using Digital or AI-based Formative Assessment Tools

AI or Digital Tools	Their Advantages (Sub-Themes)
AI (GenAI, AWE, Chatbots)	Reliability, Feedback, Anxiety, Efficiency, Objective, Lexical richness, Personalization
Learning Management Systems & Flipped Learning (LMS, Moodle, Portfolio)	Autonomy, Scaffolding, Interaction between teacher and student, Monitoring, Preventing fossilization, Flexibility, Organization
Mobile & Game-Based Assessment (MALL, Quizizz, Mentimeter, Socrative)	Engagement, Mood, Atmosphere, Fluency, Accessibility, Promptness, Usability
Dynamic & Diagnostic Assessment (C-DA, DIA)	Potential growth of ZPD, Diagnosis, Remediation, Cognitive development, Validity, Identification, Interactivity
Peer Assessment & Collaborative Tools (PA-AWE, E-writing)	Reflection, Critical-Thinking, Collaboration, Agency, Mentoring, Transparency

As seen in Table 4, the thematic analysis of the obtained articles determined several benefits of using digital or AI-based formal assessment tools which are given as the sub-themes of the advantages. For the first sub-theme, AI-based tools like Generative AI, Automated Writing Evaluation (AWE), and chatbots are linked to the high level of scoring reliability and objectivity. By providing automated, effective feedback and boosting personalized development opportunities through individualized evaluation benefits, these platforms improve the evaluation, which leads to lowering language learning anxiety among assessed students and lifting their production of vocabulary. The second sub-theme focuses on LMSs and flipped learning platforms like Moodle, e-or portfolios. The analysis showed that these tools rise learner autonomy and help administrative organization. They also offer educational scaffolding, support interaction between teachers and students, facilitate continuous progress monitoring and diminish the fossilization of linguistic errors. The third theme, regarding mobile and game-based assessment tools, comprise MALL, Quizizz, Mentimeter and Socrative and highlights a significant increase in student engagement. These accessible, user-friendly and responsive technologies offer a positive and stress-free classroom atmosphere during evaluation and increase communicative fluency in oral tasks. Fourth, dynamic and diagnostic assessment frameworks like C-DA and DIA show a significant capacity for improving cognitive development and raising the potential growth of the ZPD of the learners. These interactive approaches enhance assessment validity by guiding accurate error detection, diagnosing specific linguistic inadequacies and providing immediate remediation. Finally, peer assessment and collaborative tools such as PA-AWE and electronic writing forums are found to increase critical thinking skills and objective self-reflection practices. By supporting both teacher and peer mentoring, these tools develop learner agency, build meaningful collaboration and increase transparency across assessment practices.

Theme 2. Disadvantages of Using Digital or AI-based Formative Assessment Tools

The results sections of the 53 included articles were systematically analyzed to detect the pedagogical disadvantages of technology-enhanced formative assessment practices and the results are presented in Table 5.

Table 5. The Sub-Themes of the Second Theme: Disadvantages of Using Digital or AI-based Formative Assessment Tools

AI or Digital Tools	Their Disadvantages (Sub-Themes)
AI & Automated Scoring (GenAI, DeepSeek, AWE)	Reliability, Bias, Culture, Cognitive limits, Ethics, Complexity for beginners
Learning Management Systems & Flipped Learning (LMS, Moodle, Virtual Learning)	Readiness (Student/Teacher), Infrastructural, Implementation, Interaction deficit, Sustainability (Lack of longitudinal data), Psychological (Technology anxiety)
Mobile & Game-Based Assessment (MALL, Socrative, VET)	Validity, Hardware, Security, Accessibility, Motivation, Connectivity
Collaborative & Dynamic Assessment (Peer-Assessment, C-DA, DIA)	Effectiveness, Social (Peer pressure), Consistency, Time burden for teachers, Design: Practicality issues, Ambiguity
Data Analytics & Computational Tools (Coh-Metrix, DRs)	Interpretive, Generalizability, Isolation, Acceptance, Institutional

As presented in Table 5, utilizing digital and AI-based tools for formative assessment practices contains several disadvantages. The first sub-theme focuses on AI and automated scoring systems including GenAI, DeepSeek, and AWE, and demonstrate difficulties in sustaining grading reliability and eliminating the risks of algorithmic bias. Such tools are often struggling in terms of cultural adaptation, ethical and academic integrity and create immoderate complexity for beginner language learners. Second, the integration of LMSs and flipped applications demand high level of technological readiness for both teachers and students and they require extensive implementation support, often leading to interaction deficits. Furthermore, their pedagogical sustainability is still underdeveloped because of a lack of longitudinal study, while the risk of inducing technology anxiety may result in negative psychological effects. Third, mobile and game-based assessment tools such as MALL, Socrative and the Versant English Test are limited by restrictions regarding hardware uniformity, data security and sudden fall in student motivation. Technical issues like dependency on stable connectivity can create accessibility problems that diminish assessment validity. Fourth, collaborative and dynamic evaluation frameworks such as Peer-Assessment, C-DA, and DIA face distinct practical obstacles. These methods might cause problems for teachers regarding time and administrative constraints derived from the design complexity and practicality issues. For students, their use may create social friction, such as peer pressure. Additionally, variations in scoring consistency can lead to systemic ambiguity regarding the final evaluation results. Lastly, data analytics and computational tools such as Coh-Metrix and Digital Randomizers present explanatory challenges. The focus on automated metrics can provoke student feelings of isolation, hinder institutional acceptance and generate generalizability issues due to isolated, institution-based datasets that prevent the application of findings to broader educational contexts.

5. Discussion

Aiming to systematically review the literature to detect the formative assessment practices through the use of digital and AI-based tools, this study has come up with some significant findings for EFL context. Overall, the study proved the considerable paradigm shift from the traditional teacher-dependent evaluation to technology-based and learner-centered assessment methods in language learning environments. This finding is in alignment with other technology-based research conclusions in the literature in that the technological revolution has a significant role in today's language teaching methods. The reason for this shift from traditional to technology enhanced assessment might be caused by some benefits of utilizing these tools as proved by the findings of this study. First of all, some of the obtained studies showed that the integration of these assessment tools increased learner autonomy (Lin, 2019; Dewi & Slamet, 2025). This is specifically apparent in collaborative assessment implementations by opening a room for self-paced learning and individualized feedback (He & Wang, 2025; Mahadini et al., 2021; Mohamadi, 2018; Petchprasert, 2021; Pourdana, 2022). Likewise, by providing reliable and objective scoring, these technological tools outnumbered the scorings of human raters (Li et al., 2024; Topuz et al., 2025). In addition, the feedback offered by these tools supported the improvement of the learners and contributed to their personalized learning efforts (Gao et al., 2025; Li et al., 2025), which is detected to be a drawback of human evaluation. However, as the findings reveal, human evaluators are more

competent than these tools in terms of cultural issues and emotional gaps (Al-Obaydi & Pikhart, 2025; Ennadir & Habiballah, 2025).

Another positive effect of technology-enhanced assessment is its contribution to psychological well-being by reducing stress caused by the nature of formative evaluation. It is seen that when the learners are employed with these non-judgmental technology-based tools, their affective filter is reduced that leads to the pace for real performance of the learners (Khasawneh et al., 2025; Mirsanjari, 2025). This finding is observable in language abilities like public speaking which is one of the most struggling language skills that cause anxiety among the learners (Alharbi & Meccawy, 2020; Saeedi & Soltani, 2025). This finding seems to be a proof that these technological assessment tools create a safe place for learners in which they can express themselves without the nervousness of being in front of other people. In the same direction, these tools are better for improving learner engagement and motivation in dealing with language assessment methods (He & Wang 2025; Liu & Aryadoust 2022; Shadiev et al., 2025). They increase the collaboration among the peers and result in better critical-thinking skills and constructing knowledge (Liu et al., 2023; Mirsanjai, 2025).

Despite the given benefits, the studies revealed a number of points that need to be considered while using these technological tools. For example, the findings demonstrated that utilizing these tools for language assessment demand both the learners and instructors to have a high level of technological literacy. Without basic knowledge, it would be time-consuming and problematic in terms of implementation processes to handle the necessities caused by technical matters (Abdullaeva et al., 2024; Tran & Ma, 2021). This finding also creates pressure for the learners and instructor regarding practicality. Additionally, these tools are found to be context and institution-dependent (Gao et al., 2025; Mirsanjari, 2025) which diminishes their generalizability and limited capability in detecting learner performance. Moreover, the ethical concerns need to be satisfied before the use of these tools (Elmotri et al., 2025; Khasawneh et al., 2025), which is another obstacle in terms of practicality. As a result of all these findings, the study highlighted the reality that while technology offers many benefits with a variety of advantages in language assessment processes, it should be seen as a pedagogical scaffold for the instructors rather than replacing the traditional practices fully.

5. Conclusion

This systematic review underscores a transformative paradigm evaluation in English foreign language education, characterized by a structural shift from teacher-centered evaluation to technology-mediated and learner-centered formative assessment. The synthesis of the empirical study demonstrates that while technology-driven assessment frameworks are primarily concentrated within productive language competencies in higher education, contemporary evaluative structures are increasingly expanding to capture the holistic development of learners, integrating both linguistic proficiency and affective-psychological dimensions. Individually and collectively, the reviewed literature confirms that digital and AI-based platforms offer a number of pedagogical affordances by optimizing grading consistency, delivering immediate personalized diagnostics and fostering learner autonomy within physically and psychologically secure evaluation spaces.

However, the systemic insights examined in this study show that there are practical and pedagogical challenges associated with incorporating digital technologies. Structural restrictions, such as algorithmic limitations in processing high-level linguistic and cultural nuances, institutional differences in digital readiness, and key risks around academic integrity, fundamentally limit the effectiveness of technology-mediated formative assessment. As a result, the review's conclusions imply that digital and AI-powered platforms should be viewed as adaptable pedagogical scaffolds intended to increase instructional efficiency rather than as independent replacements for human instruction. In the end, the systemic development of institutional protection and ethical standards is essential to the successful sustainability of technology-enhanced evaluation.

5. 1 Limitations and Suggestions

This systematic review has several limitations that should be acknowledged. The potential limitations of the study at hand can be listed as follows:

1) This study is limited to research indexed only in databases named Scopus, Web of Science and ERIC. Therefore, studies in other databases may be excluded, which may decrease the scoop validity.
2) Since the search strategy was based on specific keywords, related studies using different terminologies may have been overlooked during this scanning process.
3) This review is limited to the studies published between 2014 and 2025 in order to reflect the most recent developments in the field. However, the initial studies on the specific subject area might have been overlooked.

4) Only peer-reviewed journal articles were included in the analysis; grey literature such as book chapters, theses and conference proceedings were excluded, which may result in some important findings to have been missed.
5) The inclusion criteria only covered works written in English, which may have led to a potential language bias.
6) Studies for which the full text was unavailable were not included in the analysis process, which may mean a potential data loss in the dataset.

Therefore, future reviews should also include books, theses and conference proceedings to minimize publication bias and map trends in the field more comprehensively. Also, researchers can expand on these findings by reviewing other academic databases not included in this study. Since this study is limited to English-language publications, future studies could examine research published in different languages to make cross-cultural comparisons. Lastly, given the methodological limitations of the studies reviewed, future research may focus on more experimental and longitudinal studies to measure the effectiveness of digital assessment tools and AI-based assessment practices in EFL contexts.

References

Abdulrazzaq, D. M. & Abdellatif, M. S. (2023). Online assessment or offline assessment, which one is more aggressive? The impacts on willingness to communicate, test taking anxiety, and language achievement. *Computer-Assisted Language Learning Electronic Journal*, *24*(2), 69–85. https://callej.org/index.php/journal/article/view/24

Abd Elgalil, H. M., Abd El -Hakam, F. E.-Z., Farrag, I. M., Abdelmohsen, S. R., & Elkolaly, H. (2022). Undergraduate students' perceptions of online assessment during COVID -19 pandemic at faculty of medicine for girls, Al-Azhar University, Cairo, Egypt. *Innovations in Education and Teaching International*, *60*(2), 185–195. https://doi.org/10.1080/14703297.2022.2037450

Abdullaeva, B. S., Çakmak, F., & Abdullaev, D. (2024). Paper assessment or online assessment: Exploring the impact of assessment modes on EFL students' language learning outcomes and personal development. *Language Testing in Asia, 14*(1), 35. https://doi.org/10.1186/s40468-024-00309-w

Abrar, M., Aboraya, W., Khaliq, R. A., Subramanian, K. P., Husaini, Y. A., & Husaini, M. A. (2025). AI-Powered Learning Pathways: personalized learning and dynamic assessments. *International Journal of Advanced Computer Science and Applications*, *16*(1), 454–462. https://doi.org/10.14569/ijacsa.2025.0160145

Adzima, K. (2021). Examining online cheating in higher education using traditional classroom cheating as a guide. *The Electronic Journal of e-Learning*, *18*(6), 476–493. https://doi.org/10.34190/jel.18.6.002

Akimov, A., & Malin, M. (2020). When old becomes new: a case study of oral examination as an online assessment tool. *Assessment & Evaluation in Higher Education*, *45*(8), 1205–1221. https://doi.org/10.1080/02602938.2020.1730301

Al-Abri, A., Ranjbaran Madiseh, F., & Morady Moghaddam, M. (2025). Exploring learning-oriented assessment in enhancing students' lexical fluency through MALL. *The Asia-Pacific Education Researcher*, *34*(1), 1–13. https://doi.org/10.1007/s40299-024-00832-7

Aladini, A., Mahmud, R., & Ali, A. A. H. (2024). The importance of needs satisfaction, teacher support, and L2 learning experience in Intelligent Computer-Assisted Language Assessment (ICALA): A probe into the state of willingness to communicate as well as academic motivation in EFL settings. *Language Testing in Asia*, *14*(1), 58. https://doi.org/10.1186/s40468-024-00334-9

Alavi, S. M., Dashtestani, R. & Mellati, M. (2022). Crisis and changes in learning behaviours: Technology-enhanced assessment in language learning contexts. *Journal of Further and Higher Education*, *46*(4), 461–474. https://doi.org/10.1080/0309877X.2021.1985977

Alharbi, A. S., & Meccawy, Z. (2020). Introducing Socrative as a tool for formative assessment in Saudi EFL classrooms. *Arab World English Journal*, *11*(3) 372–384. https://dx.doi.org/10.24093/awej/vol11no3.23

Al-Obaydi, L. H., & Pikhart, M. (2025). AI partner versus human partner: Comparing AI-based peer assessment with human-generated peer assessment in examining writing skills. *Language Testing in Asia*, *15*(1), Article 38. https://doi.org/10.1186/s40468-025-00375-8

Alruwais, N., Wills, G., & Wald, M. (2017). Advantages and challenges of using e-assessment. *International Journal of Information and Education Technology*, *8*(1), 34–37. https://doi.org/10.18178/ijiet.2018.8.1.1008

Amua-Sekyi, E. T. (2016). Assessment, student learning and classroom practice: A review. *Journal of Education and Practice*, *7*(21), 1–6. https://files.eric.ed.gov/fulltext/EJ1109385.pdf

Asadi, M., Ebadi, S., & Mohammadi, L. (2025). The impact of integrating ChatGPT with teachers' feedback on EFL writing skills. *Thinking Skills and Creativity*, *56*, Article 101766. https://doi.org/10.1016/j.tsc.2025.101766

Atmojo, A. E. P., & Nugroho, A. (2020). EFL classes must go online! Teaching activities and challenges during COVID-19 pandemic in Indonesia. *Register Journal*, *13*(1), 49–76. https://doi.org/10.18326/rgt.v13i1.49-76

Australian Tertiary Education Quality and Standards Agency. (2023). *Assessment reform for the age of artificial intelligence*. https://www.teqsa.gov.au/sites/default/files/2023-09/assessment-reform-age-artificial-intelligence-discussion-paper.pdf

Bahari, A. (2020). Computer-assisted language proficiency assessment tools and strategies. *Open Learning: The Journal of Open, Distance and e-Learning*, *36*(1), 61–87. https://doi.org/10.1080/02680513.2020.1726738

Bahrani, T. (2011). Technology as an assessment tool in language learning. *International Journal of English Linguistics*, *1*(2), 295–302. https://doi.org/10.5539/ijel.v1n2p295

Banihashem, S. K., Gašević, D., & Noroozi, O. (2025). A critical review of using learning analytics for formative assessment: Progress, pitfalls and path forward. *Journal of Computer Assisted Learning*, *41*(3), e13150. https://doi.org/10.1111/jcal.13150

Bansal, K., Mathur, T., Mathur, T., Agarwal, S., & Sharma, R. D. (2023). Impact of social media on academics: A fractional order mathematical model. *International Journal of Modelling and Simulation*, *45*(4), 1410–1424. https://doi.org/10.1080/02286203.2023.2286419

Bhatnagar, S., Lata, P., & Sangwan, K. S. (2025). Visual management in ELT classroom assessment: An application of lean philosophy. *Asian-Pacific Journal of Second and Foreign Language Education*, *10*(1), Article 32. https://doi.org/10.1186/s40862-025-00344-w

Borrego, M. M. (2025). Error analysis and artificial intelligence: Exploring error monitoring and error treatment possibilities in EFL. *Revista Brasileira de Linguística Aplicada*, *25*(2), Article e45036. https://doi.org/10.1590/1984-6398202525413

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

Bulut, O., Gorgun, G., & Yildirim-Erbasli, S. N. (2024). The impact of frequency and stakes of formative assessment on student achievement in higher education: A learning analytics study. *Journal of Computer Assisted Learning*, *41*(1), 382–395. https://doi.org/10.1111/jcal.13087

Butler-Henderson, K., & Crawford, J. (2020). A systematic review of online examinations: A pedagogical innovation for scalable authentication and integrity. *Computers & Education*, *159*, Article 104024. https://doi.org/10.1016/j.compedu.2020.104024

Chang, C., & Lin, H. K. (2020). Effects of a mobile-based peer-assessment approach on enhancing language-learners' oral proficiency. *Innovations in Education and Teaching International*, *57*(6), 668–679. https://doi.org/10.1080/14703297.2019.1612264

Chanpet, P., Chomsuwan, K., & Murphy, E. (2020). Online project-based learning and formative assessment. *Technology, Knowledge and Learning*, *25*(3), 685–705. https://doi.org/10.1007/s10758-018-9363-2

Chung, S., & Choi, L. (2021). The development of sustainable assessment during the COVID-19 pandemic: The case of the English language program in South Korea. *Sustainability*, *13*(8), Article 4499. https://doi.org/10.3390/su13084499

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, *61*(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

Çelik, F., & Baturay, M. H. (2024). Technology and innovation in shaping the future of education. *Smart Learning Environments*, *11*(1), Article 39. https://doi.org/10.1186/s40561-024-00339-0

Deng, W., Feng, X., & Feng, H. (2024). Chinese college EFL students' perceptions towards online English listening comprehension assessments in the post-COVID era. *International Journal of English Language and Literature Studies*, *13*(4), 558–574. https://doi.org/10.55493/5019.v13i4.5251

Dewi, O. T. S., & Slamet, J. (2025). Formative ELF-based assessment of spoken communication via learning management system in a culturally diverse classroom. *Discover Education*, *4*(1), Article 516. https://doi.org/10.1007/s44217-025-00431-7

Dogan, M. E., Goru Dogan, T., & Bozkurt, A. (2023). The use of artificial intelligence (AI) in online learning and distance education processes: A systematic review of empirical studies. *Applied Sciences*, *13*(5), Article 3056. https://doi.org/10.3390/app13053056

Ebadi, S., & Rahimi, M. (2019). Mediating EFL learners' academic writing skills in online dynamic assessment using Google Docs. *Computer Assisted Language Learning*, *32*(5-6), 527–555. https://doi.org/10.1080/09588221.2018.152736

Ebadi, S., Karimi, E., & Vakili, S. (2023). An exploration into EFL learners' perspectives on online computerized listening comprehension dynamic assessment. *Language Testing in Asia*, *13*(1), Article 5. https://doi.org/10.1186/s40468-023-00219-3

Elmahdi, I., Al-Hattami, A., & Fawzi, H. (2018). Using technology for formative assessment to improve students' learning. *The Turkish Online Journal of Educational Technology*, *17*(2), 182–188. https://files.eric.ed.gov/fulltext/EJ1176157.pdf

Elmotri, B., Abdelwahab, W., Al-humari, M., Amri, F., Elyasa, Y., Malik, F., Garrouri, S., & Hassanein, O. (2025). Possibilities for improving ESP curriculum design and assessment strategies for Saudi universities with ChatGPT. *Arab World English Journal*, (Special Issue on Artificial Intelligence), 234–250. https://dx.doi.org/10.24093/awej/AI.13

Ennadir, H., & Habiballah, S. (2025). A comparative study of human and AI-assisted assessment using ChatGPT: The case of Moroccan EFL learners. *Arab World English Journal*, *16*(3), 175–187. https://dx.doi.org/10.24093/awej/vol16no3.10

Eri, R., Gudimetla, P., Star, S., Rowlands, J., Girgla, A., To, L., Li, F., Sochea, N., & Bindal, U. (2021). Digital resilience in higher education in response to COVID-19 pandemic: Student perceptions from Asia and Australia. *Journal of University Teaching and Learning Practice*, *18*(5), Article 7. https://doi.org/10.53761/1.18.5.7

Every Learner Everywhere. (2021, May 18). *What is adaptive learning and how does it work to promote equity in higher education?* https://www.everylearnereverywhere.org/blog/what-is-adaptive-learning-and-how-does-it-work-to-promote-equity-in-higher-education/

Firat, M. (2023). How ChatGPT can transform autodidactic experiences and open education. *International Journal of Educational Technology in Higher Education*, *20*(1), Article 62. https://doi.org/10.1186/s41239-023-00431-7

Fukuda, S. T., Lander, B. W., & Pope, C. J. (2024). Exploring student engagement with formative assessment and ICT tools. *LLT Journal: A Journal on Language and Language Teaching*, *27*(2), 911–930. https://doi.org/10.24071/llt.v27i2.8124

Gao, H., Hashim, H., & Yunus, M. (2025). Assessing the reliability and relevance of DeepSeek in EFL writing evaluation: A generalizability theory approach. *Language Testing in Asia*, *15*(1), Article 33. https://doi.org/10.1186/s40468-025-00362-z

Gardner, J., O'Leary, M., & Yuan, L. (2021). Artificial intelligence in educational assessment: 'Breakthrough? Or buncombe and ballyhoo?' *Journal of Computer Assisted Learning*, *37*(5), 1207–1216. https://doi.org/10.1111/jcal.12577

Ghouali, K., & Cecilia, R. R. (2021). Towards a Moodle-based assessment of Algerian EFL students' writing performance. *Porta Linguarum*, (36), 231–248. https://doi.org/10.30827/portalin.v0i36.16103

Godwin-Jones, R. (2019). In a world of SMART technology, why learn another language? *Journal of Educational Technology & Society*, *22*(2), 4–13.

Grassini, S. (2023). Shaping the future of education: Exploring the potential and consequences of AI and ChatGPT in educational settings. *Education Sciences*, *13*(7), Article 692. https://doi.org/10.3390/educsci13070692

Greenblatt, P., & McDonald, P. (2024). Using multimodal public service announcement video projects as communicative language assessments. *RELC Journal*, *55*(2), 559–566. https://doi.org/10.1177/00336882221111623

Ha, X., Tran, N., & Tran, N. (2021). Teachers' beliefs and practices regarding assessment in English as a Foreign language classrooms in Vietnam. *The Qualitative Report*, *26*(9), 2732–2750. https://doi.org/10.46743/2160-3715/2021.5063

Habib, S., Vogel, T., Anli, X., & Thorne, E. (2024). How does generative artificial intelligence impact student creativity? *Journal of Creativity*, *34*(1), Article 100072. https://doi.org/10.1016/j.yjoc.2023.100072

Hasan, J. R., Habibie, A., & Machmud, K. (2025). Formative assessment and English as a Foreign Language (EFL) teachers' corrective feedback on teaching writing in higher education level: Practice and challenges. *Asia Pacific Journal of Educators and Education*, *40*(2), 69–99. https://doi.org/10.21315/apjee2025.40.2.4

He, M., & Wang, L. (2025). Implementing assessment as learning in online EFL writing classes. *RELC Journal*, *56*(2), 468–482. https://doi.org/10.1177/00336882231174640

Hickey, D., & Harris, T. (2021). Reimagining online grading, assessment, and testing using situated cognition. *Distance Education*, *42*(2), 290–309. https://doi.org/10.1080/01587919.2021.1911627

Hong, Q. N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., ... & Pluye, P. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Education for Information*, *34*(4), 285–291. https://doi.org/10.3233/EFI-180221

Hopfenbeck, T. N., Zhang, Z., Sun, S. Z., Robertson, P., & McGrane, J. A. (2023). Challenges and opportunities for classroom-based formative assessment and AI: A perspective article. *Frontiers in Education*, *8*, Article 1270700. https://doi.org/10.3389/feduc.2023.1270700

Hughes, E., Bradford, J., & Likens, C. (2018). Facilitating collaboration, communication, and critical thinking skills in physical therapy education through technology-enhanced instruction: A case study. *TechTrends*, *62*(3), 296–302. https://doi.org/10.1007/s11528-018-0259-8

Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching*, *39*(2), 83–101. https://doi.org/10.1017/S0261444806003399

Jeevarajath, E. S., & Hema, N. (2025). Integrating digital tools in English language assessments: Insights from tertiary education. *Theory and Practice in Language Studies*, *15*(10), 3307–3318. https://doi.org/10.17507/tpls.1510.17

Jin, C., & Liu, Y. (2024). Diagnosing and promoting learners' L2 inferential reading development through hybrid computerised dynamic assessment in the Chinese EFL classroom. *Computer Assisted Language Learning*. Advance online publication. https://doi.org/10.1080/09588221.2024.2421521

Joshi, A., Vinay, M., & Bhaskar, P. (2020). Impact of coronavirus pandemic on the Indian education sector: Perspectives of teachers on online teaching and assessments. *Interactive Technology and Smart Education*, *18*(2), 205–226. https://doi.org/10.1108/itse-06-2020-0087

Juhana, J. (2025). Shaping the future of English language teaching: Leveraging adaptive pedagogy and technology for revolutionary learning. *Dinasti International Journal of Education Management and Social Science*, *6*(6), 5257–5265. https://doi.org/10.38035/dijemss.v6i6.5498

Jurāne-Brēmane, A. (2023). Digital assessment in technology-enriched education: Thematic review. *Education Sciences*, *13*(5), Article 522. https://doi.org/10.3390/educsci13050522

Kamrood, A. M., Davoudi, M., Ghaniabadi, S., & Amirian, S. Y. R. (2021). Diagnosing L2 learners' development through online computerized dynamic assessment. *Computer Assisted Language Learning*, *34*(7), 868–897. https://doi.org/10.1080/09588221.2019.1645181

Khan, S., & Khan, R. A. (2018). Online assessments: Exploring perspectives of university students. *Education and Information Technologies*, *24*(1), 661–677. https://doi.org/10.1007/s10639-018-9797-0

Khasawneh, M. A. S., Aladini, A., Assi, S. A., & Ajanil, B. (2025). Portfolio assessment in AI enhanced learning environments: A pathway to emotion regulation, mindfulness, and language learning attitudes. *Language Testing in Asia*, *15*(1), Article 5. https://doi.org/10.1186/s40468-025-00331-6

Kruiper, S. M. A., Leenknecht, M. J. M., & Slof, B. (2021). Using scaffolding strategies to improve formative assessment practice in higher education. *Assessment & Evaluation in Higher Education*, *47*(3), 458–476. https://doi.org/10.1080/02602938.2021.1927981

Kumar, D., & SN, R. S. (2020). Qualifying online assessment during COVID-19 pandemic: Reflecting on our experience under the cognitive lens of Miller's pyramid. *Research and Development in Medical Education*, *9*(1), Article 15. https://doi.org/10.34172/rdme.2020.015

Kushari, B., & Septiadi, L. (2022). A learning outcome assessment information system to facilitate Outcome-Based Education (OBE) implementation. *Jurnal Pendidikan Teknologi dan Kejuruan*, *28*(2), 238–250. https://doi.org/10.21831/jptk.v28i2.42339

Le, X. (2024). Assessment reform in higher education: An ethical approach to harness the power of generative artificial intelligence. *Education Research and Perspectives*, *51*, 156–185. https://doi.org/10.70953/erpv51.2412007

Lee, H., Chung, H. Q., Zhang, Y., Abedi, J., & Warschauer, M. (2020). The effectiveness and features of formative assessment in US K-12 education: A systematic review. *Applied Measurement in Education*, *33*(2), 124–140. https://doi.org/10.1080/08957347.2020.1732383

Lee, M., Liang, P., & Yang, Q. (2022). Coauthor: Designing a human-AI collaborative writing dataset for exploring language model capabilities. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (Article 388). Association for Computing Machinery. https://doi.org/10.1145/3491102.3501984

Lei, W., & Lei, Z. (2025). Exploring EFL teachers' beliefs and practices of formative assessment in Chinese context. *PLoS ONE*, *20*(9), Article e0333678. https://doi.org/10.1371/journal.pone.0333678

Li, J., Huang, J., Wu, W., & Whipple, P. B. (2024). Evaluating the role of ChatGPT in enhancing EFL writing assessments in classroom settings: A preliminary investigation. *Humanities and Social Sciences Communications*, *11*(1), Article 612. https://doi.org/10.1057/s41599-024-03108-7

Li, J., Huang, J., & Sheeran, T. (2025). ChatGPT4o as an AI peer assessor in EFL speaking classrooms: Examining scoring reliability and feedback effectiveness. *SAGE Open*, *15*(3). Article 21582440251369938. https://doi.org/10.1177/21582440251369938

Liao, X., Zhang, X., Wang, Z., & Luo, H. (2024). Design and implementation of an AI-enabled visual report tool as formative assessment to promote learning achievement and self-regulated learning: An experimental study. *British Journal of Educational Technology*, *55*(3), 1253–1276. https://doi.org/10.1111/bjet.13424

Lihua, L. (2025). An innovative analysis of English reading instruction based on an integrated approach of teaching, learning, and assessment. *International Educational Research*, *8*(4), 86–95. https://doi.org/10.30560/ier.v8n4p86

Lin, C. J. (2019). An online peer assessment approach to supporting mind-mapping flipped learning activities for college English writing courses. *Journal of Computers in Education*, *6*(3), 385–415. https://doi.org/10.1007/s40692-019-00138-8

Liu, T., & Aryadoust, V. (2022). The effect of in-class and one-on-one video feedback on EFL learners' English public speaking competency and anxiety. *Studies in Language Assessment*, *11*, 25–57.

Liu, C. C., Liu, S. J., Hwang, G. J., Tu, Y. F., Wang, Y., & Wang, N. (2023). Engaging EFL students' critical thinking tendency and in-depth reflection in technology-based writing contexts: A peer assessment-incorporated automatic evaluation approach. *Education and Information Technologies*, *28*(10), 13027–13052. https://doi.org/10.1007/s10639-023-11915-x

Magic Box. (2023, November 9). *Leveraging digital assessment tools for personalized learning*. https://www.getmagicbox.com/blog/leveraging-digital-assessment-tools-for-personalized-learning/

Mahadini, M. K., Setyaningsih, E., & Sarosa, T. (2021). Using conventional rubric and Coh-Metrix to assess EFL students' essays. *International Journal of Language Education*, *5*(4), 260–270. https://doi.org/10.26858/ijole.v5i4.19478

Maleki, A. (2025). Towards a more equitable education: A lens into EFL teachers' pedagogical barriers in online assessments. *International Journal of Lifelong Education*, *44*(2), 186–205. https://doi.org/10.1080/02601370.2024.2438127

Maleki, A., Khodabakhshzadeh, H., Zeraatpishe, M., & Baghaei, P. (2025). Development and validation of the online formative assessment validity scale (OFAVS) in EFL contexts. *Discover Sustainability*, *6*(1), Article 20. https://doi.org/10.1007/s43621-025-01336-0

Mansory, M. (2020). The significance of non-traditional and alternative assessment in English language teaching: Evidence from literature. *International Journal of Linguistics*, *12*(5), 210–224. https://doi.org/10.5296/ijl.v12i5.17782

McCallum, S., & Milner, M. M. (2020). The effectiveness of formative assessment: Student views and staff reflections. *Assessment & Evaluation in Higher Education*, *46*(1), 1–16. https://doi.org/10.1080/02602938.2020.1754761

Mellar, H., Peytcheva-Forsyth, R., Kocdar, S., Karadeniz, A., & Yovkova, B. (2018). Addressing cheating in e-assessment using student authentication and authorship checking systems: Teachers' perspectives. *International Journal for Educational Integrity*, *14*(1), Article 2. https://doi.org/10.1007/s40979-018-0025-x

Mirsanjari, Z. (2025). AI-driven vs. Traditional language assessment: Effects on Iranian EFL learners' motivation, anxiety, and proficiency in a high-stakes exam context. *Language Testing in Asia*, *15*(1), Article 70. https://doi.org/10.1186/s40468-025-00388-3

Mohamadi, Z. (2018). Comparative effect of online summative and formative assessment on EFL student writing ability. *Studies in Educational Evaluation*, *59*, 29–40. https://doi.org/10.1016/j.stueduc.2018.11.001

Motteram, G. (Ed.). (2013). *Innovations in learning technologies for English language teaching*. British Council. https://www.teachingenglish.org.uk/sites/teacheng/files/C607%20Information%20and%20Communication_WEB%20ONLY_FINAL.pdf

Muñoz-Alcón, A. I., Barahona-Esteban, M. N., & Trullén-Galve, F. (2023). Assessing foreign-language students' speaking performance through a digital platform: Challenges and findings. *Education Sciences*, *13*(10), Article 1039. https://doi.org/10.3390/educsci13101039

Nguyen, T. B. (2024). EFL students' acceptance of using learning management system for online testing. In *Proceedings of the International Society for Technology, Education, and Science* (pp. 758–771). ISTES Organization.

Nguyen, U. N., & Nguyen, L. V. (2021). Resilience to withstand COVID-19 crisis: Lessons from a foreign language institution in Vietnam. *Computer Assisted Language Learning Electronic Journal*, *22*(2), 40–55.

Nguyen Thanh, B., Vo, D. T. H., Nhat, M. N., Pham, T. T. T., Trung, H. T., & Xuan, S. H. (2023). Race with the machines: Assessing the capability of generative AI in solving authentic assessments. *Australasian Journal of Educational Technology*, *39*(5), 110–126. https://doi.org/10.14742/ajet.8902

Öz, H., & Özturan, T. (2018). Computer-based and paper-based testing: Does the test administration mode influence the reliability and validity of achievement tests? *Journal of Language and Linguistic Studies*, *14*(1), 67–85.

Pascoal, L. A. V., & De Almeida Mattos, A. M. (2022). Critical literacy in ELT classroom testing. *Revista Alicantina de Estudios Ingleses*, *36*, Article 29. https://doi.org/10.14198/raei.2022.36.02

Perwitasari, F., Astuti, N. B., & Atmojo, S. (2021). Online learning and assessment: Challenges and opportunities during pandemic COVID-19. In *Advances in Social Science, Education and Humanities Research* (pp. 520–525). Atlantis Press. https://doi.org/10.2991/assehr.k.210423.077

Petchprasert, A. (2021). Utilizing an automated tool analysis to evaluate EFL students' writing performances. *Asian-Pacific Journal of Second and Foreign Language Education*, *6*(1), Article 1. https://doi.org/10.1186/s40862-020-00105-x

Potok, H., & Spjut, S. (2023). *Formative assessment in a digital environment: Digital tools and feedback as an aid for language development in EFL classrooms* (Independent Project with Specialization in English Studies and Education, Malmö University). DiVA Portal. http://urn.kb.se/resolve?urn=urn:nbn:se:mau:diva-60582

Pourdana, N. (2022). Impacts of computer-assisted diagnostic assessment on sustainability of L2 learners' collaborative writing improvement and their engagement modes. *Asian-Pacific Journal of Second and Foreign Language Education*, *7*(1), Article 11. https://doi.org/10.1186/s40862-022-00138-0

Rao, M., & Palathil, A. (2020). An analysis of customer sentiments towards education technology app: A text mining approach. In *Proceedings of the International Conference on Innovative Computing & Communications (ICICC 2020)*. SSRN. https://doi.org/10.2139/ssrn.3564996

Rofiah, N. L. (2020). Using Socrative for vocabulary tests: Thai EFL learner acceptance and perceived risk of cheating. *The Journal of Asia TEFL*, *17*(3), 966–982. https://doi.org/10.18823/asiatefl.2020.17.3.15.966

Rusmiyanto, R., Huriati, N., Fitriani, N., Tyas, N. K., Rofi'i, A., & Sari, M. N. (2023). The role of artificial intelligence (AI) in developing English language learner's communication skills. *Journal on Education*, *6*(1), 750–757. https://doi.org/10.31004/joe.v6i1.2990

Saeedi, Z., & Soltani, M. (2025). Developing an AI chatbot for language pragmatics instruction: From algorithms to dynamic assessment in an EFL context. *Computer Assisted Language Learning*. Advance online publication. https://doi.org/10.1080/09588221.2025.2532014

Saputra, I., Kurniawan, A., Yanita, M., Putri, E. Y., & Mahniza, M. (2024). The evolution of educational assessment: How artificial intelligence is shaping the trends and future of learning evaluation. *Indonesian Journal of Computer Science*, *13*(6), Article 4465. https://doi.org/10.33022/ijcs.v13i6.4465

Sawaki, Y., Ishii, Y., Yamada, H., & Tokunaga, T. (2024). Developing and validating an online module for formative assessment of summary writing with automated content feedback for EFL academic writing instruction. *Language Testing in Asia*, *14*(1), Article 50. https://doi.org/10.1186/s40468-024-00318-z

Sayed, B. T., Bani Younes, Z. B., Alkhayyat, A., Adhamova, I., & Teferi, H. (2024). To be with artificial intelligence in oral test or not to be: A probe into the traces of success in speaking skill, psychological well-being, autonomy, and academic buoyancy. *Language Testing in Asia*, *14*(1), Article 49. https://doi.org/10.1186/s40468-024-00317-0

Shaalan, I. E. A. W., & Ahmad, A. S. K. (2025). Trustworthiness of EFL assessment of learning in the age of AI: Challenges and solutions. *World Journal of English Language*, *15*(5), Article 1. https://doi.org/10.5430/wjel.v15n5p1

Shabani, K. (2016). Applications of Vygotsky's sociocultural approach for teachers' professional development. *Cogent Education*, *3*(1), Article 1252177. https://doi.org/10.1080/2331186X.2016.1252177

Shadiev, R., & Yang, M. (2020). Review of studies on technology-enhanced language learning and teaching. *Sustainability*, *12*(2), Article 524. https://doi.org/10.3390/su12020524

Shadiev, R., Liu, J., & Yang, L. (2025). The effect of evaluation and feedback in 360-degree video-based VR environments on language learning outcomes. *Computer Assisted Language Learning*, *38*(7), 1561–1591. https://doi.org/10.1080/09588221.2023.2300990

Silva, I. S. (2022). From language teaching to language assessment with the help of technological resources: Higher education students and oral production. *ELIA: Estudios de Lingüística Inglesa Aplicada*, (22), 93–123. https://doi.org/10.12795/elia.2022.i22.04

Smolansky, A., Cram, A., Raduescu, C., Zeivots, S., Huber, E., & Kizilcec, R. F. (2023). Educator and student perspectives on the impact of generative AI on assessments in higher education. In *Proceedings of the Tenth ACM Conference on Learning @ Scale* (pp. 378–382). Association for Computing Machinery. https://doi.org/10.1145/3573051.3596191

Sortwell, A., Trimble, K., Ferraz, R., Geelan, D. R., Hine, G., Ramirez-Campillo, R., Carter-Thuiller, B., Gkintoni, E., & Xuan, Q. (2024). A systematic review of meta-analyses on the impact of formative assessment on K-12 students' learning: Toward sustainable quality education. *Sustainability*, *16*(17), Article 7826. https://doi.org/10.3390/su16177826

St-Onge, C., Ouellet, K., Lakhal, S., Dubé, T., & Marceau, M. (2021). COVID-19 as the tipping point for integrating e-assessment in higher education practices. *British Journal of Educational Technology*, *53*(2), 349–366. https://doi.org/10.1111/bjet.13169

Suhartoyo, E., Afrilyasanti, R., & Mukminatien, N. (2025). Online classroom-based reading assessment: Comprehension and practice development. *Turkish Online Journal of Distance Education*, *26*(1), 177–192. https://doi.org/10.17718/tojde.1465225

Sun, W., & Shin, Y. (2025). Improving organizational skills in Chinese L2 learners through dynamic assessment. *SAGE Open*, *15*(3), Article 21582440251356186. https://doi.org/10.1177/21582440251356186

Theobald, M., & Bellhäuser, H. (2022). How am I going and where to next? Elaborated online feedback improves university students' self-regulated learning and performance. *The Internet and Higher Education*, *55*, Article 100872. https://doi.org/10.1016/j.iheduc.2022.100872

Topuz, A. C., Yıldız, M., Taşlıbeyaz, E., Polat, H., & Kurşun, E. (2025). Is generative AI ready to replace human raters in scoring EFL writing? Comparison of human and automated essay evaluation. *Educational Technology & Society*, *28*(3), 36–50. https://doi.org/10.30191/ETS.202507_28(3).SP04

Tosuncuoglu, I. (2018). Importance of assessment in ELT. *Journal of Education and Training Studies*, *6*(9), 163–167. https://doi.org/10.11114/jets.v6i9.3443

Tran, T. T. T., & Ma, Q. (2021). Using formative assessment in a blended EFL listening course: Student perceptions of effectiveness and challenges. *International Journal of Computer-Assisted Language Learning and Teaching*, *11*(3), 17–38. https://doi.org/10.4018/IJCALLT.2021070102

Tsai, S. (2022). Chinese students' perceptions of using Google Translate as a translingual CALL tool in EFL writing. *Computer Assisted Language Learning*, *35*(5-6), 1250–1272. https://doi.org/10.1080/09588221.2020.1799412

Turnbull, D., Chugh, R., & Luck, J. (2021). Transitioning to e-learning during the COVID-19 pandemic: How have higher education institutions responded to the challenge? *Education and Information Technologies*, *26*(5), 6401–6419. https://doi.org/10.1007/s10639-021-10633-w

Ulum, Ö. G. (2020). A critical deconstruction of computer-based test application in Turkish state university. *Education and Information Technologies*, *25*(6), 4883–4896. https://doi.org/10.1007/s10639-020-10196-8

Vittorini, P., Menini, S., & Tonelli, S. (2020). An AI-based system for formative and summative assessment in data science courses. *International Journal of Artificial Intelligence in Education*, *31*(2), 159–185. https://doi.org/10.1007/s40593-020-00230-2

Wang, R. E., & Demszky, D. (2023). *Is ChatGPT a good teacher coach? Measuring zero-shot performance for scoring and providing actionable insights on classroom instruction* (arXiv:2306.03090). arXiv. https://doi.org/10.48550/arXiv.2306.03090

Wei, S., & Shin, Y. (2025). Analyzing the effects of computer-assisted dynamic assessment on L2 writing tasks for Chinese learners. *Cogent Education*, *12*(1), Article 2510082. https://doi.org/10.1080/2331186X.2025.2510082

Wei, Z. (2024). Navigating digital learning landscapes: Unveiling the interplay between learning behaviors, digital literacy, and educational outcomes. *Journal of the Knowledge Economy*, *15*(3), 10516–10546. https://doi.org/10.1007/s13132-023-01522-3

Winstone, N. E., & Boud, D. (2020). The need to disentangle assessment and feedback in higher education. *Studies in Higher Education*, *47*(3), 656–667. https://doi.org/10.1080/03075079.2020.1779687

Won, D., Shin, Y. K., Kim, H., & Yoo, I. W. (2025). Advancing language assessment with GPT: Is it nonnative-language friendly? *Language Assessment Quarterly*, *22*(1), 1–28. https://doi.org/10.1080/15434303.2024.2444349

Yavuz, F., Çelik, Ö., & Yavaş Çelik, G. (2025). Utilizing large language models for EFL essay grading: An examination of reliability and validity in rubric-based assessments. *British Journal of Educational Technology*, *56*(1), 150–166. https://doi.org/10.1111/bjet.13506

Yulianto, D., & Mujtahid, N. M. (2021). Online assessment during Covid-19 pandemic: EFL teachers' perspectives and their practices. *JET (Journal of English Teaching)*, *7*(2), 229–242. https://doi.org/10.33541/jet.v7i2.2770

Zarate, M. E. C. (2024). Effectiveness of digital assessment tools on the English language learning outcomes of tertiary students. *Frontiers in Health Informatics*, *13*(3), Article 10673. https://doi.org/10.30699/fhi.v13i3.541

Zhang, H., Ge, S., & Saad, M. R. B. M. (2024). Formative assessment in K-12 English as a foreign language education: A systematic review. *Heliyon*, *10*(10), Article e31367. https://doi.org/10.1016/j.heliyon.2024.e31367

Zhao, M. (2024). A study on the measurement and standardized assessment model of student learning outcomes in vocational institutions. *Applied Mathematics and Nonlinear Sciences*, *9*(1), 1–117. https://doi.org/10.2478/amns.2023.2.00400

Zheng, C., Wang, L., & Chai, C. S. (2023). Self-assessment first or peer-assessment first: Effects of video-based formative practice on learners' English public speaking anxiety and performance. *Computer Assisted Language Learning*, *36*(4), 806–839. https://doi.org/10.1080/09588221.2021.1946562

Zohny, H., McMillan, J., & King, M. (2023). Ethics of generative AI. *Journal of Medical Ethics*, *49*(2), 79–80. https://doi.org/10.1136/jme-2023-108909

Zou, B., Lyu, Q., Han, Y., Li, Z., & Zhang, W. (2025). Exploring students' acceptance of an artificial intelligence speech evaluation program for EFL speaking practice: An application of the integrated model of technology acceptance. *Computer Assisted Language Learning*, *38*(5-6), 1366–1391. https://doi.org/10.1080/09588221.2023.2278608