

Implementing Digital Assessment in Arabic Language Learning: A Case Study of Listening, Speaking, Reading, and Writing Skills

A. Samsul Ma'arif¹, Bakri Muhammad Bakhiet²

¹ Universitas Islam Negeri Maulana Malik Ibrahim Malang, Indonesia

² University of The Holy Quran and Islamic Sciences Sudan

¹ syamsulsyafa@uin-malang.ac.id, ² bakrimuhammadbakhiet@gmail.com

Received: Augus 3 2026

Revised: August 15, 2026

Accepted: August 31, 2026

Abstract

The transformation of education through digital technology has reshaped language assessment practices, including the assessment of Arabic language learning. Digital assessment offers new possibilities for conducting language evaluation in ways that are more flexible, interactive, authentic, and responsive to learners' performance. However, its implementation in Arabic language assessment remains complex because the four language skills—*istimā'* (listening), *kalām* (speaking), *qirā'ah* (reading), and *kitābah* (writing)—require different assessment formats, technological affordances, and criteria of performance. This study examines how digital assessment is implemented to assess the four Arabic language skills and identifies the pedagogical and technical challenges that emerge during its implementation. A qualitative case study design was employed to investigate assessment practices, lecturer experiences, student responses, and the use of digital assessment instruments in an Arabic language learning context. Data were collected through semi-structured interviews, classroom observations, and document analysis, and were analyzed thematically to identify patterns across the four language skills. The findings indicate that digital assessment facilitates more flexible task administration, rapid feedback, multimedia-based assessment, and greater opportunities to document students' language performance. Nevertheless, its implementation varies across skills: listening and reading are relatively adaptable to digitally mediated tasks, whereas speaking and writing require more careful consideration of interaction, authenticity, scoring procedures, and technological constraints. The study also identifies challenges related to instrument design, assessment consistency, digital literacy, technological infrastructure, and the alignment between assessment tasks and the specific characteristics of each skill. The study contributes a context-sensitive understanding of digital assessment in Arabic language education by showing that technological adoption alone does not guarantee meaningful assessment; its pedagogical value depends on the alignment among digital tools, language-skill characteristics, assessment principles, and learning objectives.

Keywords: Arabic language skills, *digital assessment*, digital technology, *istimā'*, *kalām*, *kitābah*, *qirā'ah*,



INTRODUCTION / المقدمة

The rapid expansion of digital technology has transformed contemporary education, particularly the ways teachers design, implement, and assess language learning. In language education, assessment is increasingly shifting from conventional paper-based examinations toward digitally mediated practices that enable interactive tasks, multimedia integration, flexible administration, rapid feedback, and systematic documentation of learners' performance (Alavi, Dashtestani, & Mellati, 2022; Fuller et al., 2022; Jurāne-Brēmane, 2023). This shift is particularly significant for language learning because linguistic competence is multidimensional and involves receptive and productive abilities that cannot be adequately represented through a single assessment format. Technology-enhanced environments can broaden the range of evidence available to teachers by incorporating audio and video tasks, interactive texts, digital portfolios, online communication, and other forms of technology-mediated language use (Çetin Köroğlu, 2020; Jurāne-Brēmane, 2023). Nevertheless, the availability of digital tools does not automatically ensure meaningful assessment, since assessment quality depends on the alignment among learning objectives, assessment tasks, technological affordances, learner characteristics, and scoring procedures (Fuller et al., 2022; Jurāne-Brēmane, 2023). This issue becomes particularly important in Arabic language learning, where *istimā'* (listening), *kalām* (speaking), *qirā'ah* (reading), and *kitābah* (writing) involve different cognitive and communicative processes and therefore require differentiated forms of performance evidence.

Nevertheless, the presence of technology does not automatically guarantee better assessment. Research on technology-enhanced assessment has identified validity, reliability, practicality, accessibility, security, pedagogy, and learner experience as interconnected dimensions that need to be considered when digital

assessment is designed and implemented (Alavi, Dashtestani, & Mellati, 2022). Digital assessment should therefore be understood not simply as the transfer of an existing test from paper to screen but as a broader pedagogical process in which technology is selected according to the construct being assessed and the evidence required to support an assessment decision (Fuller et al., 2022; Jurāne-Brēmane, 2023).

The issue becomes more complex in Arabic language education because Arabic proficiency consists of four interconnected but distinctive language skills: *istimā'* (listening), *kalām* (speaking), *qirā'ah* (reading), and *kitābah* (writing) (Ma'arif, 2021). Each skill involves different linguistic processes and consequently requires assessment procedures that correspond to its particular characteristics. Listening requires learners to process spoken input in real time, whereas speaking requires the production of oral language involving pronunciation, fluency, lexical and grammatical control, interaction, and communicative appropriateness. Reading requires learners to construct meaning from written texts, infer information, and interpret linguistic structures, while writing requires the formulation, organization, revision, and communication of ideas through written discourse. Accordingly, the same digital assessment format cannot be assumed to be equally appropriate across all four skills. Technology-enhanced language assessment should instead be designed around the construct of each skill and the type of evidence required to demonstrate competence (Çetin Köroğlu, 2020; Chapelle & Voss, 2016).

The growing interest in technology-enhanced assessment has been accompanied by substantial international research. Digital assessment can facilitate flexible administration, multimedia-based tasks, rapid feedback, and opportunities for formative assessment, but these advantages are accompanied by challenges involving validity, reliability, accessibility, digital literacy, infrastructure, and assessment security (Alavi, Dashtestani, & Mellati, 2022; Fuller et al., 2022; Jurāne-Brēmane, 2023). Jurāne-Brēmane (2023) argues that digital assessment

should be understood as part of the broader transformation of educational practice rather than merely as a technological replacement for traditional testing. Fuller et al. (2022) similarly emphasize the importance of aligning assessment purposes, technology, learners, feedback, and professional judgment. In language assessment, this principle is particularly important because technological affordances may change the nature of the task itself and consequently influence the evidence produced by learners (Chapelle & Voss, 2016).

This distinction between digitalized testing and genuinely technology-enhanced assessment is important. Digitalized testing generally reproduces conventional test structures through an electronic medium, whereas technology-enhanced assessment can exploit affordances that create assessment activities not readily available in paper-based environments (Çetin Köroğlu, 2020; Jurāne-Brēmane, 2023). For example, listening assessment can incorporate authentic audio or audiovisual input, while speaking assessment can use recorded oral performances, video presentations, online interviews, or synchronous interaction. Reading assessment can employ digital texts, hyperlinks, annotation, and multimedia support, whereas writing assessment can document drafting, revision, peer feedback, and portfolio development. The suitability of each technology therefore depends on whether its affordances correspond to the linguistic and cognitive characteristics of the skill being assessed.

The distinction is especially important for listening and speaking. Listening assessment requires controlled access to spoken input while preserving the perceptual and inferential processes that constitute listening comprehension. Digital platforms can provide authentic recordings and multimedia contexts, but technical problems such as inconsistent audio quality, device differences, and network instability can introduce construct-irrelevant variance. Speaking assessment presents a different challenge because oral proficiency is inherently performance-based and includes pronunciation, fluency, interaction, and communicative

meaning. Recording technologies can preserve evidence of oral performance and permit repeated human evaluation, while automated speech technologies may provide additional information about selected dimensions of pronunciation or fluency. However, automated analysis should not automatically be treated as equivalent to communicative assessment because interactional and pragmatic dimensions remain difficult to capture comprehensively through automated procedures.

Reading and writing likewise require different technological affordances. Digital reading environments can provide interactive texts, glossaries, hyperlinks, visual information, and immediate feedback, potentially expanding the range of comprehension processes that can be assessed. However, access to translation tools and online dictionaries may also alter the cognitive demands of the task. Writing assessment can benefit from digital environments because they make drafting, editing, commenting, revision, and portfolio development visible as part of the learning process. At the same time, automated grammar correction, machine translation, and generative artificial intelligence may complicate the interpretation of student-produced texts because the final product may not fully represent the learner's independent writing competence. These differences demonstrate why technology selection must be connected to the construct and purpose of assessment rather than determined solely by technological availability.

In the Indonesian context, the transformation toward digital assessment in Arabic language education has become increasingly visible. Rahman, Masdar, Ilmiani, Habibie, and Samdouni (2022) specifically examined the transformation of Arabic assessment in Indonesia from conventional assessment toward digital assessment and identified the growing use of web-based, computer-based, and mobile-based assessment. Their study provides direct evidence that digital assessment has become part of the changing landscape of Arabic language assessment in Indonesian educational institutions. This development provides an

important national context for investigating how digital assessment is implemented in specific Arabic language learning environments.

The development of digital assessment is also relevant to intensive Arabic language learning environments such as the Program Khusus Perkuliahan Bahasa Arab (PKPBA) at Universitas Islam Negeri Maulana Malik Ibrahim Malang (Ma'arif & Rosikh, 2025). In such an intensive program, assessment needs to capture continuous development across multiple language skills rather than relying exclusively on isolated end-of-course examinations. Digital environments can support this requirement by documenting oral performances, written assignments, interactive activities, and learning portfolios. However, the pedagogical value of such technologies depends on whether the assessment tasks remain aligned with the communicative objectives of intensive Arabic learning.

Recent Arabic-language assessment research also indicates increasing interest in artificial intelligence, automated evaluation, and authentic digital assessment. Studies published in 2025–2026 have explored the use of learning management systems, AI-supported tools, chatbots, and automated technologies for assessing Arabic speaking and writing (Arifuddin & Husna, 2022). These studies suggest opportunities for immediate feedback and expanded assessment capacity, while simultaneously identifying concerns involving validity, reliability, teacher validation, digital literacy, infrastructure, and the limitations of current Arabic-language AI systems (Faizin et al., 2025).

Despite these developments, the existing literature reveals several areas that require further investigation. First, many studies examine digital assessment from a technology-centered perspective, focusing on the adoption of particular platforms or applications rather than examining how technological affordances correspond to the constructs being assessed. Second, Arabic digital assessment research has often examined particular skills or specific technologies, making it difficult to understand how listening, speaking, reading, and writing can be assessed within an integrated

framework. Rahman et al. (2022), for example, provide important evidence concerning the broader transformation of Arabic assessment in Indonesia, but further research is needed to examine how this transformation operates across distinct language skills.

Third, there remains a pedagogical alignment gap between technological affordances and the characteristics of Arabic language skills. The same digital format cannot necessarily be applied equally to listening, speaking, reading, and writing. A multiple-choice platform may efficiently assess selected aspects of reading comprehension or listening discrimination, but it cannot by itself provide sufficient evidence of interactive speaking ability or extended writing competence. Conversely, recorded speaking and portfolio-based writing tasks can generate richer performance evidence but require human judgment, clear rubrics, and greater scoring time. The central issue is therefore not whether a technology is available, but whether its affordances produce evidence that is relevant to the intended language construct.

Fourth, the literature indicates a continuing validity and authenticity challenge. Digital environments make it easier to provide automated scoring, immediate feedback, and standardized procedures, but these features can also encourage the reduction of complex language performance into easily measurable components. This concern is particularly relevant to productive skills, where communicative competence includes dimensions that may not be fully captured by automated scoring. Recent studies of AI-supported Arabic assessment similarly emphasize the continuing importance of teacher validation and human oversight.

Fifth, artificial intelligence has introduced a new dimension into digital language assessment. AI-supported technologies can provide automated feedback, speech recognition, natural language processing, adaptive tasks, and automated analysis of written language. In Arabic language education, recent studies have examined AI-assisted speaking evaluation and AI-supported writing assessment,

indicating growing interest in these applications. However, questions concerning accuracy, bias, transparency, privacy, construct validity, and the appropriate role of teacher judgment remain important. AI should therefore be considered an assessment support mechanism rather than automatically treated as a replacement for professional assessment judgment.

The identified gaps indicate that research on digital assessment in Arabic language education needs to move beyond the question of which technology is used toward the more fundamental question of how technology is pedagogically matched to the characteristics of each language skill. This distinction constitutes the central position of the present study (Sadiyah, 2018). Rather than treating digital assessment as a collection of applications, the study conceptualizes it as an integrated assessment environment involving technological tools, assessment tasks, learner performance, feedback, teacher judgment, and learning evidence. Within this framework, the four Arabic language skills are understood as complementary domains that require differentiated but interconnected assessment strategies.

The present study therefore focuses on the implementation of digital assessment in Arabic language learning across *istimā'*, *kalām*, *qirā'ah*, and *kitābah*. It examines how digital technologies are employed for each skill and identifies the pedagogical benefits and practical challenges associated with their implementation. Particular attention is given to the relationship between technological affordances and skill-specific assessment requirements, including multimedia input for listening, performance documentation and interaction for speaking, interactive texts for reading, and drafting, feedback, and revision for writing.

The contribution of this study lies in connecting the literature on digital assessment with Arabic language pedagogy through a skill-sensitive perspective. Existing research establishes that technology can expand assessment possibilities, while research on Arabic assessment demonstrates the increasing adoption of digital practices in Indonesia and other contexts. However, a more integrated

empirical understanding is needed regarding why particular digital affordances are useful for particular Arabic language skills and what limitations emerge when technologies are transferred across skills without sufficient pedagogical adaptation.

Ultimately, digital assessment should not be understood simply as the electronic replacement of conventional examinations. Its significance lies in its capacity to change how language performance can be elicited, documented, reviewed, and supported through feedback. The educational value of technology is therefore determined not by its sophistication but by its contribution to valid, authentic, and meaningful assessment. In Arabic language education, this principle requires assessment systems that recognize the distinctive characteristics of *istimāʿ*, *kalām*, *qirāʾah*, and *kitābah* while using digital technologies selectively and purposefully.

METHOD / المنهجية

This study employed a qualitative case study design to examine the implementation of digital assessment in Arabic language learning at the Preparatory Program for Arabic Language (PKPBA), Universitas Islam Negeri Maulana Malik Ibrahim Malang. A qualitative case study was selected because the study sought to understand digital assessment as it occurred within a specific educational setting and to explore how lecturers and students experienced, interpreted, and implemented digital assessment across the four Arabic language skills: listening (*istimāʿ*), speaking (*kalām*), reading (*qirāʾah*), and writing (*kitābah*). Rather than measuring the statistical effectiveness of particular digital platforms, the study focused on the pedagogical practices, technological affordances, challenges, and assessment decisions that emerged in the natural context of Arabic language instruction. This approach is consistent with the view that case study research is appropriate when researchers investigate a contemporary phenomenon within its real-life context and when the boundaries between the phenomenon and its context

are not always clearly separated (Ma'arif et al., n.d.). The qualitative orientation also allowed the researchers to capture participants' perspectives and the contextual factors that shaped the use of digital assessment (Creswell & Poth, 2018).

The research was conducted at PKPBA, Universitas Islam Negeri Maulana Malik Ibrahim Malang, an intensive Arabic language program serving first-year university students. The site was selected because digital technologies have increasingly been incorporated into Arabic teaching, learning, and assessment practices within the program. The research participants consisted of Arabic language lecturers, curriculum and assessment personnel, and students who had direct experience with digitally mediated assessment activities. Participants were selected purposively according to their relevance to the research questions. Lecturers were included because they were responsible for designing, administering, and interpreting assessment activities, while curriculum and assessment personnel were involved in the development and coordination of assessment practices. Students were included because they experienced the assessment process directly and could provide information about task accessibility, feedback, interaction, and the use of digital tools. Purposive sampling was considered appropriate because qualitative case studies require information-rich participants who can provide detailed accounts of the phenomenon under investigation rather than statistically representative samples (Creswell & Poth, 2018; Yin, 2018).

Data were collected through observation, semi-structured interviews, and documentation. Observation was used to examine how digital assessment was implemented during Arabic language learning activities, including the types of tasks assigned, digital platforms or media used, lecturer-student interaction, feedback practices, and the ways in which assessment criteria were applied. Particular attention was given to the relationship between technological affordances and the characteristics of each language skill. For listening assessment, for example,

observations considered the use of audio and audiovisual materials, replay functions, and contextual information. For speaking, attention was directed to oral recordings, video performances, presentations, interviews, and interactive activities. Reading observations focused on digital texts, hyperlinks, multimedia resources, annotation, and comprehension activities, whereas writing observations examined digital drafting, revision, feedback, collaborative writing, and portfolio development. This skill-sensitive observation was important because the pedagogical value of digital assessment cannot be determined solely by the availability of technology; it depends on whether the technology supports the cognitive and communicative demands of the language skill being assessed (Alavi et al., 2022; Jurāne-Brēmane, 2023).

Semi-structured interviews were conducted with lecturers, curriculum and assessment personnel, and selected students. The interview protocol addressed several dimensions of digital assessment, including the reasons for adopting digital tools, the perceived suitability of particular technologies for different language skills, assessment criteria, feedback mechanisms, technical difficulties, student participation, lecturer digital competence, and concerns related to authenticity and academic integrity. The semi-structured format allowed the researchers to maintain consistency across interviews while providing sufficient flexibility to explore issues that emerged from participants' experiences. This was particularly important in examining why a particular technology was considered suitable for one language skill but less appropriate for another. Participants were also invited to describe situations in which digital technologies enhanced assessment practices and situations in which technological affordances created new assessment problems.

Documentation was used to complement observational and interview data. Relevant documents included assessment instructions, digital assessment tasks, scoring rubrics, learning management system materials, student assignments, feedback records, digital portfolios, and other assessment-related materials

available within the research setting. These documents were examined to identify the relationship between intended assessment criteria and actual digital assessment practices. Documentation was also used to examine how the four language skills were represented in assessment tasks and whether the technological features of the selected platforms supported the intended learning outcomes. The use of multiple data sources enabled methodological triangulation and reduced reliance on a single participant perspective or data collection technique.

The research instruments consisted of an observation guide, a semi-structured interview protocol, and a documentation checklist. The observation guide was organized around four analytical dimensions: the nature of the assessment task, the technological affordances employed, the evidence of language competence produced by students, and the challenges encountered during assessment. The interview protocol similarly addressed pedagogical alignment, technological suitability, feedback, authenticity, accessibility, and assessment challenges. The documentation checklist was used to systematically record information from assessment materials, rubrics, digital portfolios, and relevant learning-management-system content. The instruments were developed in accordance with the research objectives and refined to ensure that the data collected addressed both the general implementation of digital assessment and the distinctive characteristics of listening, speaking, reading, and writing assessment.

Data analysis followed the interactive model of qualitative data analysis proposed by Miles, Huberman, and Saldaña (2014). The process involved data condensation, data display, and conclusion drawing and verification. During data condensation, interview transcripts, observation notes, and documents were coded according to recurring themes related to digital assessment practices, technological affordances, skill-specific assessment requirements, feedback, authenticity, accessibility, and teacher digital competence. The data were then displayed thematically to facilitate comparison across the four language skills. Particular

attention was given to patterns showing how the same technological feature could generate different pedagogical consequences depending on the skill being assessed. For example, replay functionality could provide useful access to authentic listening input but could also alter the cognitive demands of a listening task if repeated playback was unlimited. Similarly, automated language feedback could support writing revision while simultaneously creating concerns about whether the resulting text represented the student's independent linguistic competence.

The final stage of analysis involved drawing and verifying conclusions through continuous comparison among interview data, observations, and documents. Emerging interpretations were repeatedly checked against the available evidence to avoid conclusions based solely on individual participants' perceptions. The analysis therefore moved from descriptive accounts of digital assessment practices toward an interpretive explanation of the relationship between technological affordances and the pedagogical characteristics of the four Arabic language skills. This analytical strategy was particularly important for responding to the central issue of the study: digital assessment should not be considered appropriate simply because it is technologically advanced; its suitability depends on the extent to which the technology produces meaningful evidence of the particular language competence being assessed.

To establish the trustworthiness of the findings, the study employed triangulation across data sources and data collection techniques. Information obtained from lecturers was compared with student accounts, observations, and assessment documents. The researchers also conducted repeated examination of emerging interpretations during the analysis process to ensure consistency between the empirical evidence and the conclusions. These procedures were informed by the principles of credibility, dependability, confirmability, and transferability associated with qualitative inquiry (Lincoln & Guba, 1985). Rather than treating reliability as a statistical property, the study emphasized the transparency and

consistency of the research process and the correspondence between participants' accounts, observed practices, documentary evidence, and the researchers' interpretations.

Ethical considerations were maintained throughout the research process. Participants were informed about the purpose and procedures of the study, and their participation was treated as voluntary. Information obtained from interviews, observations, and assessment documents was used exclusively for research purposes, while participant identities and potentially identifiable information were protected in reporting the findings. Particular attention was given to digital assessment materials because they may contain student work, performance recordings, feedback, and other educational information requiring responsible handling. The researchers therefore treated digital data as research evidence while maintaining appropriate confidentiality and respect for participants' rights.

Through this methodology, the study positioned digital assessment not merely as the transfer of conventional tests into digital platforms but as a pedagogical practice requiring alignment among learning outcomes, assessment tasks, technological affordances, performance evidence, scoring criteria, and feedback. The methodological design consequently provided a basis for analyzing the four Arabic language skills separately while also examining their integration within a broader digital assessment environment.

RESULT AND DISCUSSION / نتائج البحث والمناقشة

1. Digital Assessment of Listening Skill: Authenticity, Control, and the Nature of Real-Time Processing

The findings indicate that digital assessment was particularly useful for listening skill because digital media could provide spoken-language input in forms that were difficult to reproduce consistently through conventional paper-based assessment. Lecturers reported using audio and audiovisual materials to present

dialogues, announcements, conversations, interviews, and other contextualized forms of Arabic communication. The use of digital media enabled lecturers to incorporate authentic or semi-authentic speech, control the duration of listening materials, and provide visual or contextual information when appropriate. These practices demonstrate that the value of technology in listening assessment does not simply lie in replacing an audio recorder or conventional examination with a digital platform; rather, technology expands the range of listening evidence that can be presented to learners.

This finding is consistent with the fundamental nature of listening as a receptive skill involving the perception, processing, interpretation, and construction of meaning from spoken language. Unlike reading, in which learners can generally control the speed of their visual processing, listening often involves temporal constraints because spoken input unfolds sequentially. Digital audio and video therefore provide an affordance that is closely aligned with the temporal nature of listening. Students can hear pronunciation, intonation, pauses, stress, and discourse patterns in ways that cannot be adequately represented through written transcripts alone. Technology-enhanced language assessment has consequently created opportunities for language teachers to use more contextualized and multimedia-based assessment tasks (Alavi, Dashtestani, & Mellati, 2022; Jurāne-Brēmane, 2023).

However, the findings also show that the technological affordance that makes digital listening assessment attractive can create an assessment problem. The replay function, for example, provides flexibility because students can listen to difficult sections more than once. From a learning perspective, repeated listening may facilitate comprehension and reduce unnecessary anxiety. From an assessment perspective, however, unlimited replay can change the cognitive demands of the task. If the intended construct is the learner's ability to comprehend spoken Arabic under relatively natural temporal conditions, unlimited repetition may provide assistance

that is not part of the intended construct. Consequently, the suitability of replay functionality depends on the assessment purpose rather than on the technology itself.

The findings further demonstrate that audiovisual materials must be selected carefully. Video can provide contextual information through facial expressions, gestures, setting, and visual cues, which may support comprehension. At the same time, excessive visual information may enable students to infer answers without fully processing the spoken Arabic. This creates an important distinction between assessing listening comprehension and assessing general multimedia comprehension (Arifuddin et al., 2025). Therefore, the use of video is pedagogically appropriate when visual information forms part of the communicative context being assessed, but it requires careful task design when the primary objective is to measure auditory comprehension.

Technical accessibility also emerged as a significant consideration. Problems involving internet connectivity, audio quality, headphones, device compatibility, and differences in students' digital competence can influence performance independently of their actual listening ability. This finding supports broader research showing that technology-enhanced assessment can generate new forms of inequality when access and technological conditions are not adequately controlled (Jurāne-Brēmāne, 2023). Digital listening assessment should therefore distinguish between technological difficulty and linguistic difficulty. A student who cannot clearly hear an assessment recording because of poor equipment should not be interpreted as demonstrating weak listening competence.

The pedagogical implication is that digital technology is particularly suitable for listening assessment when its affordances preserve the essential characteristics of spoken communication. Authentic audio, contextualized audiovisual materials, controlled playback, and appropriate task sequencing can produce richer evidence of listening competence. Nevertheless, technological flexibility must remain subordinate to the construct being assessed. The central principle is therefore not

maximum technological sophistication but alignment between the temporal and communicative characteristics of listening and the technological conditions under which evidence of listening competence is produced.

2. Digital Assessment of Speaking Skill: Performance Evidence, Interaction, and the Limits of Automation

Speaking presented a different pattern from listening because it required students to produce rather than receive language. The findings show that lecturers used audio recordings, video recordings, online interviews, presentations, role plays, and other performance-based activities to assess students' speaking competence. These technologies enabled lecturers to document oral performance and, in some cases, review recordings more than once before assigning a score. Digital recording therefore provided an important advantage: the student's oral performance could be preserved as assessable evidence rather than being evaluated only during a single moment of classroom interaction.

The value of recording becomes particularly apparent when speaking assessment involves several dimensions simultaneously. Speaking competence cannot be reduced to grammatical accuracy or vocabulary knowledge because it also includes pronunciation, fluency, intelligibility, interactional ability, discourse organization, and communicative appropriateness. A recorded performance allows lecturers to examine these dimensions repeatedly and to use a structured rubric when assigning judgments. Digital assessment research in Arabic language learning similarly indicates increasing interest in audio, video, LMS-based assessment, and AI-supported tools for evaluating oral performance (Thayyibah et al., 2025; Mahbubah, 2026).

Nevertheless, the findings indicate that recorded speaking and interactive speaking should not be treated as equivalent forms of evidence. An asynchronous recording allows students to prepare, repeat, edit, and select their best performance. These features may be appropriate when the objective is to assess prepared oral

production, pronunciation, presentation skills, or the organization of spoken discourse. They become less appropriate when the objective is spontaneous interaction, because the ability to respond immediately to another speaker cannot be fully demonstrated through a carefully prepared recording. The distinction is therefore pedagogically important: digital recording can document speaking performance effectively, but the technology does not automatically reproduce the interactional conditions of real communication.

Synchronous digital interviews and online conversations provide a different affordance. They can approximate real-time communication because students must listen to questions, interpret meaning, formulate responses, and produce language within a limited temporal window. This makes synchronous technology more appropriate for assessing interactional competence than asynchronous recording. However, online interaction introduces other variables, including internet stability, microphone quality, latency, unfamiliarity with the platform, and the possibility that technical interruptions may influence the interaction. Assessment designers therefore need to distinguish technological disruption from actual speaking performance.

The role of human judgment is particularly important in speaking assessment. Automated speech-recognition systems can potentially assist with pronunciation analysis, transcription, fluency indicators, and immediate feedback. Recent studies have explored AI-supported assessment of Arabic speaking, demonstrating the growing potential of artificial intelligence for oral-language evaluation (Mahbubah, 2026). Nevertheless, automated analysis may not adequately capture pragmatic appropriateness, communicative intention, interactional sensitivity, or culturally situated meanings. AI-generated scores or feedback should therefore be regarded as supplementary evidence rather than an unquestionable replacement for professional assessment judgment.

This finding also relates to the broader issue of authenticity. A technologically advanced speaking assessment is not necessarily authentic if the task does not require

learners to use language in a manner resembling meaningful communication. A highly polished video recording may demonstrate pronunciation and prepared discourse while providing limited evidence of spontaneous interaction. Conversely, a short online interview may generate richer evidence of interaction but also create greater technical and scoring complexity. The pedagogical challenge is therefore to select the digital format according to the specific speaking construct being assessed.

The findings suggest that digital assessment of speaking is most appropriate when technological affordances are deliberately matched to the intended performance. Asynchronous recording is useful for documenting prepared production and enabling repeated review, while synchronous interaction is more appropriate for spontaneous communication. Rubrics remain essential because they translate complex speaking performance into transparent assessment criteria. AI may support transcription, feedback, and preliminary analysis, but final interpretation should remain connected to human judgment, especially when the assessment concerns communicative competence rather than merely formal linguistic accuracy.

3. Digital Assessment of Reading Skill: Hypertext, Multimedia Support, and the Risk of Construct Modification

The findings show that digital technology substantially changes the conditions under which reading assessment is conducted. Lecturers used digital texts, multimedia materials, hyperlinks, online glossaries, images, interactive questions, and other features to assess students' reading comprehension. Compared with conventional printed texts, digital reading environments can provide immediate access to additional information and can connect learners with multiple layers of textual resources. This creates opportunities to assess not only the ability to identify explicit information but also vocabulary, inference, interpretation, main ideas, and relationships among textual elements.

The compatibility between digital technology and reading assessment arises partly from the nature of reading itself. Reading involves the construction of meaning

through interaction between textual information and the reader's linguistic and cognitive resources. Digital texts can enrich this process by providing contextual information, hyperlinks, visual representations, and additional language support. Research on Arabic reading instruction in the digital era has similarly emphasized the potential of digital strategies and resources to facilitate reading development (Isna(i)ni & Umar, 2023). Digital environments therefore offer a wider range of materials through which reading competence can be assessed.

However, the findings also reveal an important limitation: digital support can alter the construct of reading assessment. Online dictionaries, automatic translation tools, glossaries, search engines, and generative AI can reduce the linguistic processing demands required to understand a text. If a reading assessment intends to determine whether students can infer the meaning of unfamiliar vocabulary from context, immediate access to translation tools can undermine that objective. Likewise, if the objective is to evaluate independent comprehension, excessive scaffolding may transform the task from reading assessment into assisted information retrieval.

This issue demonstrates why digital affordances must be analyzed in relation to assessment purpose. A glossary may be appropriate when the objective is to assess comprehension of a complex text rather than vocabulary knowledge. Hyperlinks may be useful when the objective involves evaluating information across interconnected texts. Multimedia resources may be appropriate when multimodal comprehension forms part of the learning outcome. Conversely, these same technologies can become problematic when the assessment intends to isolate specific reading processes.

The findings also indicate that digital reading environments create opportunities for more authentic assessment. Students can be presented with web pages, digital announcements, online articles, institutional information, or other texts that resemble materials encountered outside the classroom. Such tasks can increase ecological validity because students engage with forms of reading that have practical

communicative purposes. Yet authenticity must not be confused with technological complexity. A digital text is not automatically authentic simply because it appears on a screen; authenticity depends on whether the task reflects meaningful language use and whether the required response corresponds to the intended reading competence.

Another concern relates to navigation. Digital reading frequently requires learners to move among pages, hyperlinks, menus, images, and supplementary resources. These activities can create additional cognitive demands that are unrelated to the reading construct itself. A student may understand the Arabic text but experience difficulty navigating the digital environment. Consequently, digital reading assessments should minimize unnecessary interface complexity when navigation is not itself part of the intended learning outcome.

The pedagogical implication is that digital reading assessment should distinguish between productive technological support and construct contamination. Digital texts, hyperlinks, annotation, multimedia resources, and online glossaries can enrich assessment when they are deliberately incorporated into the construct. However, unrestricted translation and AI tools can provide assistance that changes what the assessment actually measures. The lecturer therefore needs to establish in advance which technological resources are permitted, why they are permitted, and which dimensions of reading competence they are intended to support.

4. Digital Assessment of Writing Skill: Process, Revision, and the Challenge of Authorship

Writing demonstrated perhaps the clearest transformation from product-oriented assessment toward process-oriented assessment. The findings indicate that lecturers used digital documents, online submission systems, collaborative writing environments, portfolios, and electronic feedback to assess students' written Arabic. These technologies made it possible to observe not only the final written product but also stages of drafting, revision, feedback, and improvement. Such affordances are

particularly relevant because writing is a productive skill that develops through recursive processes rather than through a single act of production.

The digital environment therefore provides an important pedagogical advantage for writing assessment. A conventional written examination often provides evidence primarily from the final product. Digital platforms can provide evidence of how students construct that product. Drafts, revisions, lecturer comments, peer feedback, and portfolio development can reveal changes in organization, vocabulary, grammar, coherence, and communicative purpose. This process-oriented approach is consistent with the growing use of LMS platforms, digital portfolios, and authentic assessment practices in Arabic language learning (Thayyibah et al., 2025).

Digital writing assessment also allows feedback to become more continuous. Instead of receiving a score after completing an assignment, students can receive comments during the writing process and use those comments to revise their work. This transforms assessment from a purely summative mechanism into a component of learning. The digital environment is particularly suitable for such a model because feedback can be stored, revisited, and connected to successive versions of the same text.

Nevertheless, writing is currently one of the Arabic language skills most strongly affected by the emergence of generative AI, automatic translation, grammar checkers, and other language technologies. These tools can provide useful linguistic assistance, but they also complicate the interpretation of student writing. A polished Arabic text may reflect the student's own linguistic ability, extensive use of translation technology, AI-generated content, or a combination of these resources. Recent literature on AI and Arabic writing assessment has therefore identified both opportunities for automated support and concerns about validity, authorship, academic integrity, and the need for teacher validation (Noviani et al., 2026; Ma'rifaturohmah & Abu Bakar, 2026).

The central problem is not simply whether students use AI but whether the assessment design makes their use visible and pedagogically meaningful. If students are required to produce a final text without documenting the writing process, it may be difficult to determine the extent of their independent writing competence. A process-oriented digital portfolio can address this problem by requiring students to submit drafts, revisions, reflections, and responses to feedback. Such evidence provides a more comprehensive basis for interpreting the final product because the lecturer can examine how the text developed.

At the same time, AI should not automatically be treated as incompatible with writing assessment. In certain learning contexts, the ability to use AI critically, evaluate generated language, revise inaccurate output, and improve an Arabic text may itself constitute a relevant digital-literacy learning outcome. The assessment problem therefore concerns alignment rather than the mere presence of AI. If the intended outcome is independent Arabic composition, unrestricted generative AI may threaten the construct. If the intended outcome includes critical digital writing practices, controlled AI use may become part of the learning activity.

The findings consequently support a differentiated approach to digital writing assessment. Traditional product assessment remains relevant when independent language production is the primary objective. Digital portfolios and staged assignments are more appropriate when the objective includes writing development, revision, and reflective learning. AI-supported activities may be used when the learning outcome explicitly includes technological literacy, but the assessment criteria must clearly distinguish the student's own linguistic contribution from machine-generated assistance.

5. Cross-Skill Analysis: From Digital Testing to Integrated Digital Assessment

Comparison across the four skills reveals that the suitability of digital assessment cannot be determined at the level of technology alone. The same digital affordance may produce different pedagogical consequences depending on the skill

being assessed. Replay functionality, for example, can support access to listening input but may modify the intended temporal demands of listening comprehension. Recording can strengthen evidence in speaking assessment but may reduce authenticity when spontaneous interaction is the intended construct. Hyperlinks and translation tools can enrich reading assessment while simultaneously reducing the independent linguistic processing required by the task. Digital drafting and AI-supported revision can strengthen writing development while creating uncertainty about authorship and independent competence.

These differences demonstrate that digital assessment should be understood as a relationship among four components: the intended language construct, the assessment task, the technological affordance, and the evidence produced by the learner. A technology is pedagogically appropriate when its affordances strengthen the relationship between the task and the intended construct. Conversely, technology becomes problematic when it introduces additional performance requirements or external assistance that changes the construct being measured. This perspective is consistent with technology-enhanced assessment research emphasizing the importance of alignment between assessment purposes, technology, and educational practice (Alavi et al., 2022; Jurāne-Brēmane, 2023).

The findings also show that the four skills should not be treated as completely isolated domains. In actual language use, listening, speaking, reading, and writing frequently operate together. A digital project, for example, may require students to listen to an Arabic interview, read supporting information, discuss the content orally, and produce a written response. Such integrated tasks can generate more authentic evidence of communicative competence because they approximate the interconnected nature of language use beyond the classroom. At the same time, integrated assessment requires carefully differentiated criteria so that weaknesses in one skill do not obscure evidence from another.

The transformation toward digital assessment also changes the role of the lecturer. The lecturer is no longer merely an examiner who administers a test and assigns a score. The lecturer becomes a designer of digital assessment environments who must determine the appropriate platform, formulate authentic tasks, establish criteria, monitor technological conditions, interpret digital evidence, provide feedback, and evaluate the role of external technologies such as AI. This expanded role requires digital assessment literacy in addition to conventional assessment knowledge.

The findings further indicate that accessibility and infrastructure remain foundational conditions. Digital assessment cannot produce fair evidence when students experience unequal access to devices, stable internet connections, appropriate audio equipment, or technological support. Consequently, technological sophistication should not become a substitute for pedagogical quality. A simple digital task with clear objectives and reliable access may generate more meaningful assessment evidence than a highly sophisticated platform that creates unnecessary technical barriers.

The issue of authenticity also cuts across all four skills. Authenticity does not mean simply using technology that resembles contemporary digital life. Rather, an assessment becomes more authentic when the task requires students to mobilize language competence in ways that correspond to meaningful communicative practices. Listening to an authentic conversation, participating in an online interview, interpreting a realistic digital text, or producing a purposeful written response can all contribute to authentic assessment when the tasks are appropriately designed. The technology serves authenticity only when it preserves or strengthens the communicative purpose of the language activity.

Finally, the findings indicate that artificial intelligence introduces a new layer of complexity into digital assessment. Recent studies have explored AI tools for assessing multiple Arabic language skills as well as specific skills such as speaking

and writing (Kamal, 2026; Mahbubah, 2026). AI can support automated feedback, speech recognition, text analysis, and assessment administration, but these functions also raise questions about transparency, bias, privacy, accuracy, and the interpretation of student performance. For this reason, AI should be integrated into Arabic assessment through explicit pedagogical purposes and appropriate human oversight rather than being adopted solely because of its technological novelty.

Overall, the findings conceptualize digital assessment in Arabic language learning as an integrated pedagogical system rather than a collection of digital tools. The central issue is the alignment between the characteristics of the four language skills and the affordances of the technologies used to assess them. Listening requires attention to temporal processing and audio quality; speaking requires performance and interaction; reading requires control over digital assistance and navigation; and writing requires attention to process, revision, and authorship. Digital assessment becomes educationally meaningful when these differences are recognized in assessment design, implementation, scoring, and feedback.

CONCLUSION / الخلاصة

This study demonstrates that digital assessment has the potential to transform the evaluation of Arabic language learning from conventional, test-oriented practices into a more flexible, interactive, authentic, and competency-oriented assessment system. The findings across *istimāʿ*, *kalām*, *qirāʾah*, and *kitābah* indicate that digital technology provides different but complementary opportunities for measuring Arabic language competence.

In *istimāʿ*, digital assessment enables the use of authentic audio and multimedia materials that can measure learners' comprehension in more contextualized situations. In *kalām*, audio and video recordings, online interviews, presentations, and digital portfolios provide stronger evidence of students' oral performance while allowing lecturers to review performance repeatedly. In *qirāʾah*,

interactive digital texts facilitate the assessment of comprehension, interpretation, inference, and contextual vocabulary. Meanwhile, digital assessment of *kitābah* supports a process-oriented approach through digital drafting, feedback, revision, and portfolio development. These findings confirm that digital technology can expand the forms of evidence available for evaluating Arabic language proficiency.

However, the study also shows that digital assessment is not automatically superior to conventional assessment. Its effectiveness depends on the quality of assessment design, lecturer digital competence, technological infrastructure, internet accessibility, assessment validity, and the authenticity of student performance. Excessive dependence on automated systems, translation applications, or generative artificial intelligence may reduce the accuracy of assessment if technological assistance replaces rather than supports learners' actual performance.

Therefore, digital assessment should be understood as a pedagogical transformation rather than merely a technological substitution. Effective digital assessment requires alignment among learning outcomes, assessment tasks, technological affordances, scoring criteria, and feedback mechanisms. The central implication of this study is that technology should serve assessment purposes rather than determine them. A balanced digital assessment model for Arabic language learning should combine technological efficiency with authentic performance, formative feedback, transparent rubrics, and professional lecturer judgment. Such an approach can strengthen the validity and meaningfulness of Arabic language assessment while supporting continuous development of learners' communicative competence in contemporary digital learning environments.

BIBLIOGRAPHY / المراجع

Alavi, Sayyed Mohammad, Reza Dashtestani, and Morteza Mellati. (2022). Crisis and changes in learning behaviours: Technology-enhanced assessment in

- language learning contexts. *Journal of Further and Higher Education*, 46(4), 461–474.
- Arifuddin, N., & Husna, A. (2022). Istiratijiyyat Haqibah al-Injaz fi Ta'lim al-Kitabah lada Thalabah al-Barnamaj al-Mukatstsaf li Ta'lim al-Lughah al-Arabiyah bi Jami'ah Maulana Malik Ibrahim al-Islamiyah al-Hukumiyah Malang. *Lugawiyyat*, 4(2), 118–136.
- Arifuddin, N., Wiranegara, D. A., Ma'arif, A. S., Junaidi, M. R., & Bakhiet, B. M. (2025). Integrating Life Skills in Arabic Language Education: Enhancing Bilingual University Graduates' Employability and Competencies. *Jurnal Al Bayan: Jurnal Jurusan Pendidikan Bahasa Arab*, 17(1), 276–293.
- Creswell, John W., and Cheryl N. Poth. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.
- Çetin Koroğlu, Zeynep. (2020). Assessing language learning in technology-enhanced environments. In Serpil Meri Yılan and Kasim Koruyan (Eds.), *ICT-based assessment, methods, and programs in tertiary education* (pp. 45–64). IGI Global Scientific Publishing. <https://doi.org/10.4018/978-1-7998-3062-7.ch003>
- Faizin, N., Maarif, A. S., & Hanafi, Y. (2025). Considering religious moderation in Islamic law through AI ChatGPT and Bahsul Masail of Nahdlatul Ulama. *Law, Innovation and Technology*, 1–18.
- Fuller, Richard, Viktoria C. T. Goddard, Vishna D. Nadarajah, Tamsin Treasure-Jones, Peter Yeates, Karen Scott, Alexandra Webb, Krisztina Valter, and Eeva Pyörälä. (2022). Technology enhanced assessment: Ottawa consensus statement and recommendations. *Medical Teacher*, 44(8), 836–850. <https://doi.org/10.1080/0142159X.2022.2083489>
- Isnaini, Maulidya, and Mohammad Umar. (2023). استراتيجيات تعليم اللغة العربية لتحسين [Arabic language learning strategies to improve reading skills in the digital era]. *Ma'arif Journal of Education, Madrasah Innovation and Aswaja Studies*, 2(2), 45–54. <https://doi.org/10.69966/mjemias.v2i2.37>
- Jurāne-Brēmane, Anžela. (2023). Digital assessment in technology-enriched education: Thematic review. *Education Sciences*, 13(5), 522. <https://doi.org/10.3390/educsci13050522>
- Kamal, Muhammad Hakim. (2026). Assessment of Arabic language skills using AI tools among gifted talented students. *Malaysian Journal of Social Sciences and Humanities*, 11(1), e003761. <https://doi.org/10.47405/mjssh.v11i1.3761>
- Laudza'i, Ilmia Habibati, M. Yunus Abu Bakar, and Zainur Arifin. (2026). The transformation of Arabic language assessment: A systematic literature review of technology-enhanced assessment (2021–2026). *Aphorisme: Journal of Arabic Language, Literature, and Education*, 7(1), 697–711. <https://doi.org/10.37680/aphorisme.v7i1.9669>

- Lincoln, Yvonna S., and Egon G. Guba. (1985). *Naturalistic inquiry*. SAGE Publications.
- Mahbubah, Latifatul. (2026). Utilization of artificial intelligence (AI) for evaluating students' Arabic speaking skills. *Jurnal Al Bayan: Jurnal Jurusan Pendidikan Bahasa Arab*, 18(1), 134–156. <https://doi.org/10.24042/trqabp68>
- Ma'arif, A. S. (2021). Tathwiir al-kitaab al-ta'limu limahaarotilqiroah alaasasi al-tsaqofah al-Islamiyah fiJamiati Ibnu Sina al-IslamiyahKepanjen Malang. *FASHOHAH*, 1(1), 1–11.
<http://riset.unisma.ac.id/index.php/fashoha/article/view/12604/9981>
- Ma'arif, A. S., & Rosikh, F. (2025). Integrating Arabic Language Learning And Islamic Religious Education For Religious Character Formation. *Ummul Qura Jurnal Institut Pesantren Sunan Drajat (INSUD) Lamongan*, 9(2), 264–270.
- Ma'arif, A. S., Walenta, A. S., Rustiyana, B., Ningsih, Y. R., Rachmawati, D., & Hakim, H. (n.d.). *Metodologi Penelitian Kuantitatif dan Kualitatif*. Agam: PT Srikandi Book Press.
- Ma'rifaturohmah, Khusnul Mukhibbin, and M. Yunus Abu Bakar. (2026). Analisis peluang dan tantangan penilaian otomatis berbasis NLP untuk maharah kitabah (keterampilan menulis) pada siswa dalam pembelajaran bahasa Arab. *Kartika: Jurnal Studi Keislaman*, 6(3), 5075–5087.
<https://doi.org/10.59240/kjsk.v6i3.1177>
- Miles, Matthew B., A. Michael Huberman, and Johnny Saldaña. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). SAGE Publications.
- Noviani, Dwi, Ica Safitri, Erma Sapitri, and Dini Kaspas. (2026). Chatbot AI sebagai instrumen evaluasi autentik maharah kitabah dan kalam dalam pembelajaran bahasa Arab: Sebuah telaah literatur. *Didaktik: Jurnal Ilmiah PGSD STKIP Subang*, 12(3), 427–441. <https://doi.org/10.36989/didaktik.v12i3.16794>
- Rahman, Nur Fuadi, Muhammad Ihsanuddin Masdar, Aulia Mustika Ilmiani, Faizal Habibie, and Soukaina Samdouni. (2022). Transformation of Arabic assessment in Indonesia: Conventional assessment toward digital assessment. *Al-Ta'rib: Jurnal Ilmiah Program Studi Pendidikan Bahasa Arab*, 10(2), 199–212. <https://doi.org/10.23971/altarib.v10i2.4510>
- Sadiyah, H. (2018). Bermain peran (role playing) dalam pembelajaran maharah al-kalam di PKPBA UIN Maliki Malang. *Tarbiyatuna: Jurnal Pendidikan Ilmiah*, 3(2), 1–29.
- Thayyibah, Uyun, Ira Ainul Latifah, Ubaid Ridlo, and Raswan. (2025). Utilization of learning management systems (LMS) and artificial intelligence (AI) in supporting authentic assessment in Arabic language learning: A descriptive study. *Jurnal Pendidikan Bahasa Arab*, 3(2), 108–122.
<https://doi.org/10.59829/3391dv27>
- Yin, Robert K. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE Publications.