



Digital Assessment in Arabic Language and Literature Learning: A Systematic Scoping Review

Chambali¹, Muhsin Riyadi², Alifiyah Hamidah³

UIN Siber Syekh Nurjati Cirebon, Indonesia

chambali@uinssc.ac.id; oberth1979@gmail.com; alifiyahmdh03@gmail.com

Abstract

Digital assessment is increasingly reported in Arabic language education, yet how it has been studied, especially for Arabic literature learning, remains unclear. This scoping review used the Population-Concept-Context framework and PRISMA-ScR reporting to map digital assessment tools described in Arabic language and literature learning research. Records were retrieved through Consensus, Scholar Gateway, and targeted searches of DOAJ, Garuda, and Moraref, covering 2016-2026 publications in Indonesian, English, and Arabic. Of 33 records identified, 14 met the eligibility criteria. Most studies examined generic quiz or gamification platforms, such as Kahoot, Quizizz, Wordwall, and Socrative, mainly at primary and secondary levels, while only a few targeted specific language skills. No study addressed digital assessment for Arabic literature learning, and only one reported measurable learning outcomes, which remained below the expected mastery threshold. These findings indicate a fragmented evidence base and a clear gap in literature-focused digital assessment.

Keywords: digital assessment; Arabic language education; Arabic literature.

تحتل صا لختسم

بدلاً ملعت في فصاخبو، متسارد تفيفيك نأ ريغ، تبيرعلا تغللا ملعت في مقرلا مبيقتلا فيظوت ديازتي موهفملاو ناكسلا راطل تيجهنملا تيعلاطتسلا تساردلا هذه تمدختسا. تحضاو ريغ لازت لا، يبرعلا في قدراولا مقرلا مبيقتلا تاولدلاً تطيرخمسرل، PRISMA-ScR غلابلا ريباعم تعبتاو، قايسلاو Scholar و Consensus في تصنم ربع تلاجسلا تعمج. اهبداو تبيرعلا تغللا ملعت شاحباً نيد تاروشنملا، Moraref و Garuda و DOAJ دعاوق في هجوم شحدي لإ تفاضل، Gateway، 14 تفوتسا، ادحم لاجس 33 لصاً ن م. تبيرعلاو تيزيلجنلاو تيسينودنلا 2016 و 2026 في ماء Kahoot لثم، تماع بيعت وأ رابتخا تصنم تاساردلا مظعم تلوانت. تيلهلأ ريباعم تسارد تفدهتسا امنيد، تفسوتملو تينادتبلا نيتلحرملا في ابلاغ، Socrative و Wordwall و Quizizz و غلبت ملو، يبرعلا بدلاً ملعت في مقرلا مبيقتلا تسارد في لوانتت مل. ددحم تيوط قراهم تليلق تاسارد جئاتنلا هذه ريشت. عقتملا ناقتلا يوتسم نود تناكو، سايقلا تلباق ملعت جئاتن ن ع دحاو تسارد يوس. يبرعلا بدلاً طبترملا في مقرلا مبيقتلا في تحضاو عوجفو تئزجتة قلدا دعاوق لإ.

تيساسأ تاملك
يبرعلا بدلاً؛ تبيرعلا تغللا ملعت؛ مقرلا مبيقتلا

Introduction

Arabic language education has undergone a steady shift toward digital tools over the past decade, reshaping how teachers design instruction and how learning is measured (Rahman et al., 2022). Assessment, as the stage that determines whether learning objectives have been reached, is central to this shift, because a well-designed digital assessment can make evaluation more efficient, more engaging, and more responsive to individual student progress (Nisa' & Abu Bakar, 2026). Within Indonesia specifically, Arabic is taught not only as a foreign language but also as the language through which classical Islamic literature and religious texts are studied, which makes the assessment of both language skills and literary understanding equally important for programs preparing future teachers and scholars of Arabic.

A growing number of studies have documented specific digital tools used to evaluate Arabic learning. Game-based platforms such as Kahoot, Quizizz, Wordwall, and Socrative have been reported across primary and secondary levels, generally to support formative evaluation and student motivation (Sofa, 2025; Chalista et al., 2024; Lisaniyah & Salamah, 2021). A smaller set of studies has targeted specific language skills, including a formative digital tool for listening and reading comprehension (Rahmayanti et al., 2024), a full e-course for speaking and listening delivered through a learning management system (Mohammed, 2022), and a single mobile application designed to evaluate all four basic language skills at once (Rizal et al., 2021). At a broader level, Rahman et al. (2022) traced how Arabic assessment in Indonesia has moved from paper-based testing toward web-based and mobile-based models, while noting that this transformation still could not accommodate the assessment of the speaking skill. Two general reviews have also examined technology in Arabic education without focusing specifically on assessment (Amadi et al., 2023; Zamista, 2022).

Despite this expanding body of work, no review has systematically mapped what is actually known about digital assessment across Arabic language and literature learning as a whole. The closest related work is a systematic review by Clarisa et al. (2026), which examined the transformation of digital evaluation specifically for the speaking skill (*mahārah kalām*) and found that authentic, technology-supported speaking assessment is still constrained by limited teacher digital literacy and infrastructure. That review, however, was deliberately narrowed to a single language skill and did not address listening, reading, writing, the linguistic elements of *nahwu* and *sharaf*, or the study of Arabic literature. The present review differs from it by covering all four language skills together with

Arabic literature learning, which allows patterns and gaps to be compared across skills rather than within one skill alone. This broader scope also responds to a methodological gap: a general scoping review of literacy assessment reporting transparency in the Gulf region (Bin Rashed et al., 2025) demonstrates that a PRISMA-based scoping approach is workable for Arabic-related assessment topics, yet no such review has been directed at digital assessment tools themselves within Arabic language and literature education.

This study aims to map the range of digital assessment tools reported in Arabic language and literature learning research, describe the language skills and educational levels these tools address, and identify where the evidence is concentrated or absent. By doing so, this review contributes an evidence map that can guide teachers in selecting digital assessment tools appropriate to a given skill and level, and it identifies specific directions for future research, particularly the near absence of studies on digital assessment for Arabic literature.

The remainder of this article is organized as follows. The Method section describes the review protocol, eligibility criteria, search strategy, and analysis procedure. The Results and Discussion section presents the characteristics of the included studies, the thematic patterns identified, and a discussion relating these findings to existing literature. The Conclusion summarizes the main findings and outlines directions for future research.

Method

This study used a scoping review design guided by the Population-Concept-Context (PCC) framework and reported following the Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR). A scoping review, rather than an effectiveness review, was selected because prior screening of the literature indicated that included studies would vary widely in design, ranging from qualitative case studies to research and development projects, which made a pooled quantitative synthesis inappropriate. The review protocol was not registered in an external registry, which is noted as a limitation in the Conclusion.

Population was defined as learners and educators involved in Arabic language and/or literature education at any level, from primary school to higher education. Concept was defined as the use or evaluation of digital assessment, meaning any tool, platform, or model of assessment that relies on digital technology, applied to the four basic Arabic language skills (listening, speaking, reading, and writing), the linguistic elements of nahwu and sharaf, or Arabic literature (adab, balaghah, and

4 *Chambali et al., Digital Assessment in Arabic Language and Literature Learning*
literary texts). Context was defined as formal or non-formal Arabic education settings, including schools, madrasah, pesantren, and universities, delivered online, offline, or in blended form.

Eligible records were empirical or development studies published between 2016 and 2026 in Indonesian, English, or Arabic that directly examined digital assessment within the scope above. Records were excluded if they were opinion pieces or editorials without data, duplicate publications, discussed Arabic outside an education context (for example, purely computational natural language processing), or mentioned a digital tool only as a suggestion for future research rather than as an object actually studied.

Records were identified through two AI-assisted academic search tools, Consensus and Scholar Gateway, using combinations of terms covering Arabic language and literature education, digital assessment or evaluation, and each of the four language skills. These were supplemented with targeted searches of the Indonesian repositories Garuda and Moraref and of the Directory of Open Access Journals (DOAJ), because Indonesian Arabic-education research is concentrated in these sources. All search queries were logged. Records from all sources were merged and deduplicated using a title-and-author similarity script before screening.

Title and abstract screening was conducted against the eligibility criteria above. Full text was retrieved and reviewed for four included studies; for the remaining ten, eligibility and data extraction relied on the detailed abstracts and metadata available through the search tools, because the original publisher pages could not be reliably located within the scope of this review. This is disclosed as a limitation of the present review: data for these ten studies should be treated as less thoroughly verified than for the four studies retrieved in full text, and readers seeking to build directly on these findings are encouraged to consult the original sources.

For each included study, the following data were extracted: author and year, educational level or setting, research design, the digital tool or platform used, the language skill or content targeted, and the key finding reported. Findings were synthesized narratively and organized into thematic groups based on the type of digital tool and the skill it addressed, consistent with the descriptive, mapping purpose of a scoping review. No formal risk-of-bias or quality appraisal was performed, in line with PRISMA-ScR guidance, which does not require this step for scoping reviews; this is noted as a limitation when interpreting the strength of the reported findings.

Results and Discussion

Search and Selection Results

The search identified 33 unique records: 25 through Consensus, 3 through Scholar Gateway, and 5 through targeted web and repository searches. No duplicates were found across sources. After title and abstract screening against the eligibility criteria, 19 records were excluded and 14 proceeded to data extraction (Figure 1). The most common reason for exclusion was that a study centered on a learning medium, general technology literacy, or teacher technology acceptance rather than on assessment itself (6 records). Other studies were excluded because the assessment instrument described was printed rather than digital (3 records), because the source was itself a general literature review rather than a primary study (3 records), because a digital tool was only recommended for future research rather than actually studied (2 records), or for other reasons detailed in the review log.

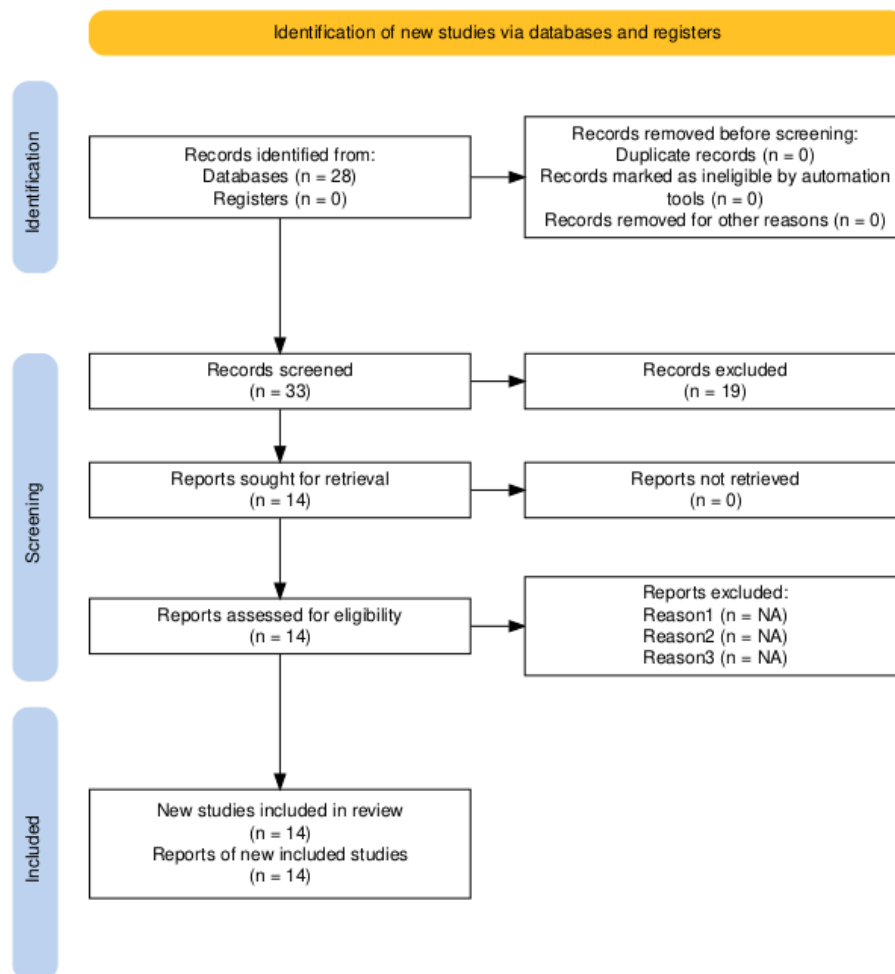


Figure 1. PRISMA 2020 flow diagram of the study selection process.

Characteristics of Included Studies

Table 1 summarizes the 14 included studies. Ten of fourteen studies were conducted at primary or secondary level, three did not specify a level clearly, and one addressed higher education. Eight studies described generic quiz or gamification platforms without targeting a specific skill, three studies targeted one or two specific skills, and three studies addressed system-level or multi-skill digital assessment.

No	Author, Year	Level/Setting	Design	Digital Tool	Skill Focus	Key Finding
1	Sofa (2025)	Not specified	Qualitative case study	Kahoot, Quizizz, Google Forms, iSpring	General	Interactive assessment models increased effectiveness of measuring Arabic learning
2	Nisa' & Abu Bakar (2026)	Primary (MI)	Qualitative descriptive	Blended learning platform (unspecified)	General	Blended learning transformed assessment toward cognitive, affective, and psychomotor domains
3	Rahman et al. (2022)	Higher education	Qualitative case study	Web-based and mobile-based models	General	Assessment shifted from paper-based to digital; speaking still not accommodated
4	Aini et al. (2025)	Primary, Grade 1	Quasi-experimental, mixed methods	Wordwall	Vocabulary/general	83% efficiency gain; 94.6% accuracy; 78.6-92.9% positive engagement
5	Chalista et al. (2024)	Secondary	R&D	Competition-based app (Google App-linked)	General	Developed assessment tool judged feasible and practical
6	Afriati et al. (2026)	Not specified	Library-based model development	Kahoot (cloze/completion test)	Vocabulary/structure	Cloze test model built on CEFR framework using Kahoot
7	Samsudin et al. (2025)	Primary (SD)	Qualitative descriptive	Quizizz, YouTube, PowerPoint	General	Digital media used in ongoing formative evaluation practice
8	Rahmayanti et al. (2024)	Secondary (MAN)	R&D, ADDIE model	Formative	Listening, Reading	Formative judged feasible for evaluating istima' and qira'ah
9	Mohammed (2022)	Higher education	Quantitative, course design	LMS-based e-course (NEO LMS)	Speaking, Listening	Full e-course designed; students' perceptions generally positive

No	Author, Year	Level/Setting	Design	Digital Tool	Skill Focus	Key Finding
10	Mursyid et al. (2024)	Not specified	Library-based model development	Quizizz, Mentimeter, Nearpod	General	New 21st-century evaluation model proposed using three platforms
11	Rahman (2022)	Primary (MI), n=30	Classroom action research, 2 cycles	Wondershare Quiz Creator	General	Mastery 63.4%; mean score 65.83 vs KKM 60; implementation not yet optimal
12	Lisaniyah & Salamah (2021)	Primary (SD)	Library-based/conceptual*	Kahoot, Socrative	General	Mapped as digital 4.0 evaluation media; no field-trial data reported
13	Shiddiq et al. (2024)	Secondary (MTs)	R&D, feasibility test	Plotagon, Wordwalls, Genially	General	Web-based gamification media judged feasible by experts
14	Rizal et al. (2021)	Not specified	Library-based/conceptual*	MiSK application	All four skills	MiSK allows evaluation of four basic skills within one application; no field-trial data reported

Table 1. Characteristics of the 14 included studies. * Library-based/conceptual studies described an actual digital tool but did not report field-trial data with learners; see Method for how these were handled.

Thematic Patterns

Four patterns emerged from the included studies. First, the dominant pattern, found in eight of fourteen studies, was the use of generic interactive quiz or gamification platforms, most often Kahoot, Quizizz, Wordwall, and Socrative, applied broadly to Arabic learning without targeting a specific skill (Sofa, 2025; Aini et al., 2025; Chalista et al., 2024; Afriati et al., 2026; Samsudin et al., 2025; Mursyid et al., 2024; Rahman, 2022; Lisaniyah & Salamah, 2021). These tools were consistently reported as motivating and easy to use, most often at primary and secondary level. Second, only three studies targeted specific language skills with a purpose-built digital tool: a formative tool for listening and reading (Rahmayanti et al., 2024), a full e-course for speaking and listening (Mohammed, 2022), and a single application evaluating all four skills together (Rizal et al., 2021). Third, two studies addressed assessment transformation at a system or institutional level rather than a single tool, describing a shift from paper-based to web-based or mobile-based assessment (Rahman et al., 2022) and blended assessment integrating cognitive, affective, and psychomotor domains (Nisa' & Abu Bakar, 2026). Fourth, and most notably, none of the fourteen included studies addressed digital assessment for Arabic literature learning specifically. A related study using augmented reality for Arabic literary

text learning was identified during screening but was excluded because it focused on delivering instruction rather than assessing it, which strengthens rather than weakens the conclusion that literature-focused digital assessment remains unstudied in the literature searched for this review.

Discussion

These findings extend two existing lines of literature. General reviews of technology in Arabic education (Amadi et al., 2023) and of online assessment across subjects (Zamista, 2022) have described the broad adoption of digital tools, but neither examined assessment within Arabic education specifically enough to reveal the skill-by-skill pattern found here. The most directly comparable prior work is the systematic review of digital evaluation for the speaking skill by Clarisa et al. (2026), which reported that authentic, technology-supported assessment of mahārah kalām remains constrained by teacher digital literacy and infrastructure. The present review both supports and extends that conclusion: consistent with Rahman et al. (2022), who found that Arabic digital assessment in Indonesia could not yet accommodate speaking, the studies identified here show that purpose-built digital tools for listening, reading, and combined four-skill assessment had since emerged (Rahmayanti et al., 2024; Rizal et al., 2021), alongside a dedicated review of speaking assessment specifically (Clarisa et al., 2026). Arabic literature, however, shows no comparable progress; the near-total absence of assessment-focused studies in this area, despite growing use of digital media for literature instruction itself, suggests that digital assessment development in this field has followed language skills before literature, rather than developing in parallel.

The evidence for effectiveness is limited. Of the fourteen included studies, only Rahman (2022) reported a measurable learning outcome from classroom implementation, an average score of 65.83 against a minimum mastery criterion of 60, with the researcher explicitly describing the implementation as not yet optimal. Most other studies reported feasibility, engagement, or motivation rather than measured achievement, which limits the strength of any claim about whether these tools actually improve learning outcomes rather than simply making assessment more convenient to administer. This pattern is consistent with the broader observation that a scoping review, by design, maps what has been studied rather than judging how well it worked; a future systematic review with formal quality appraisal, focused specifically on effectiveness, would be needed to answer that question directly.

This review has several limitations. First, ten of fourteen included studies were assessed at the abstract and metadata level rather than through full text, because

reliable publisher links could not be located within the scope of this review; this is disclosed openly rather than presenting the data as more thoroughly verified than it is. Second, the review relied substantially on two AI-assisted literature search tools alongside targeted repository searches, rather than a fully documented search of named bibliographic databases such as Scopus or Web of Science; this is common practice for a scoping review but means recall cannot be guaranteed to be exhaustive. Third, no formal quality or risk-of-bias appraisal was conducted, consistent with PRISMA-ScR guidance for scoping reviews but limiting judgments about the reliability of individual study findings. Fourth, six reference entries in this review list only the first author, because complete co-author lists could not be verified through the tools available during this review; these should be checked against the original publisher pages before this manuscript is finalized for submission.

Conclusion

This scoping review mapped digital assessment tools reported in Arabic language and literature learning research using a PCC framework and PRISMA-ScR reporting. From 33 records identified, 14 studies met the eligibility criteria. Most examined generic quiz or gamification platforms at primary and secondary level, three studies targeted specific language skills, and two addressed system-level assessment transformation. Only one study reported a measurable learning outcome, and it fell below the expected mastery level. Most importantly, no included study addressed digital assessment for Arabic literature learning, distinguishing this review from a recently published systematic review restricted to the speaking skill and indicating that literature assessment is the area most in need of research attention.

Future research should prioritize the development and evaluation of digital assessment tools specifically for Arabic literature learning, since this review found none. Future studies should also report measurable learning outcomes using research designs capable of testing effectiveness, such as quasi-experimental or experimental designs, since the current evidence base relies mostly on feasibility and perception data. Finally, a systematic effectiveness review with full-text retrieval and formal quality appraisal is recommended as a natural next step once a larger body of primary studies on this topic becomes available.

Acknowledgments

The authors gratefully acknowledge the academic environment of UIN Siber Syekh Nurjati Cirebon, which provided a meaningful foundation for the completion

10 *Chambali et al., Digital Assessment in Arabic Language and Literature Learning*
of this systematic literature review. This study was developed through a careful process of identifying, selecting, evaluating, and synthesizing scholarly publications related to digital assessment in Arabic language and literature learning. The authors also recognize the valuable contribution of previous researchers whose published studies formed the primary evidence base for this review. Their scholarly work enabled the authors to examine current research trends, identify existing gaps, and formulate directions for future studies. All stages of the review and manuscript preparation were conducted independently by the authors.

References

- Aini, N., et al. (2025). Assessing Arabic learning through a game-based evaluation tool for year one primary school students. *UPGRADE: Jurnal Pendidikan Teknologi Informasi*.
- Afriati, I., et al. (2026). From ancient manuscript to digital platform: Development of completion and cloze test models for Arabic language evaluation using Kahoot media. *Journal of Higher Education and Academic Advancement*.
- Amadi, A. S. M., et al. (2023). Perkembangan pendidikan bahasa Arab di era digital: Sistematis literature review. *Jurnal Motivasi Pendidikan dan Bahasa*.
- Bin Rashed, F. M., et al. (2025). The impact of varying diglossia conceptualizations on transparency in reporting literacy achievement in the GCC region. *Review of Education*. <https://doi.org/10.1002/rev3.70122>
- Chalista, A., Kesuma, G. C., Amrulloh, M. A., Zuliana, E., & Wahidah, Y. L. (2024). Developing Arabic language learning assessment tools using the competition application in secondary school. *ALSUNIYAT: Jurnal Penelitian Bahasa, Sastra, dan Budaya Arab*, 7(2), 382-394. <https://doi.org/10.17509/alsuniyat.v7i2.69424>
- Clarisa, V. D., Nurhalim, Abdussalam, M. M., Zulhannan, & Zuliana, E. (2026). Problematika dan transformasi evaluasi mahārah kalām: Tinjauan sistematis dalam pembelajaran bahasa Arab era digital. *Al Ibrah: Journal of Arabic Language Education*, 9(1), 145-151. <https://doi.org/10.24256/jale.v9i1.10609>
- Lisaniyah, F. H., & Salamah, U. (2021). Media pembelajaran bahasa Arab berbasis digital 4.0 (Kahoot dan Socrative) pada Sekolah Dasar. *Program Studi PGMI, IAINU Tuban*.
- Mohammed, T. (2022). Designing an Arabic speaking and listening skills e-course: Resources, activities and students' perceptions. *Electronic Journal of e-Learning*.
- Mursyid, A. R., et al. (2024). New model of Arabic language evaluation and test system in the development of 21st century learning media. *ELOQUENCE: Journal of Foreign Language*.
- Nisa', K., & Abu Bakar, M. Y. (2026). Utilizing blended learning as an innovation in Arabic language teaching assessment in the society 5.0 era. *Jurnal Pembelajaran Bahasa dan Sastra*.

- Rahman, M. H. (2022). Penggunaan Wondershare Quiz Creator sebagai media evaluasi pembelajaran bahasa Arab. *Al-Waraqah: Jurnal Pendidikan Bahasa Arab*, 3(2), 50-58. <https://doi.org/10.30863/awrq.v3i2.3074>
- Rahman, N. F., Masdar, M. I., Ilmiani, A. M., Habibie, F., & Samdouni, S. (2022). Transformation of Arabic assessment in Indonesia: Conventional assessment toward digital assessment. *Al-Ta'rib: Jurnal Ilmiah Program Studi Pendidikan Bahasa Arab IAIN Palangka Raya*, 10(2), 199-212. <https://doi.org/10.23971/altarib.v10i2.4510>
- Rahmayanti, I., Muhamad, S., La Udin, Y., Nashihah, D., & Qomari, N. (2024). Pengembangan Formative sebagai media evaluasi maharah al-istima' dan al-qira'ah siswa kelas X MAN Kota Batu Malang. *Mantiqu Tayr: Journal of Arabic Language*, 4(1), 257-278. <https://doi.org/10.25217/mantiqu tayr.v4i1.4134>
- Rizal, M., Syihabuddin, & Mad'ali. (2021). Evaluasi pembelajaran bahasa Arab dengan memanfaatkan aplikasi MiSK. *Tsaqofiya: Jurnal Pendidikan Bahasa dan Sastra Arab*, 3(2).
- Samsudin, N. A., et al. (2025). Utilization of Quizizz digital media in Arabic language learning at SD Muhammadiyah Ambarketawang 3. *Eduvest - Journal of Universal Studies*.
- Shiddiq, J., et al. (2024). Feasibility of web-based digital Arabic gamification media for Islamic junior high school students. *Arabiyatuna: Jurnal Bahasa Arab*, 8(1). <https://doi.org/10.29240/jba.v8i1.8946>
- Sofa, A. R. (2025). Application of various interactive assessment models to increase the effectiveness of measuring the Arabic learning process and improve learners' skills. *Ukazh: Journal of Arabic Studies*, 6(1), 20-38.
- Zamista, A. A. (2022). Online assessment as an alternative learning assessment in the digital era: A systematic literature review. *INSANIA*, 27(2).