



Innovating Assessment for Grade VII Arabic Curriculum: A Design–Development Study of Ministry Textbooks

Nurhikmah Sani^{1✉}, Dewi Rahayu², Vikky Alziqri JH³, Irhamuddin⁴, Zikri Rahman⁵

¹Universitas Islam Negeri Sulthan Thaha Saifuddin Jambi, Indonesia

²Sekolah Menengah Pertama Islam Terpadu Al Akhyar Bungo, Indonesia

³Universitas Islam Negeri Walisongo Semarang, Indonesia

⁴Al-Azhar University, Egypt

⁵Universitas Islam Indonesia Yogyakarta, Indonesia

ABSTRACT

Purpose – This study aimed to identify shortcomings in the Grade VII Arabic teaching materials issued by the Ministry of Religious Affairs and to develop an improved, curriculum-aligned assessment prototype that supports student accessibility and instructional diagnosis.

Design/methods/approach – A Design and Development Research (DDR) approach guided three integrated phases—needs analysis (document review and teacher interviews), prototype design, and expert validation—using a curriculum-evaluation checklist and iterative revision cycles. The analysed textbook authored by Faruq Baharudin and edited by Wahib Dariyadi was systematically mapped for item typology and skill coverage, revealing consistent lesson structures but predominant reliance on written, pragmatic items.

Findings – Results indicated a pronounced gap in listening assessment, which emphasized letter- and word-level discrimination and dictation while underrepresenting sentence- and discourse-level comprehension; expert validation produced high I-CVI scores for many items and recommended explicit item-to-competency mapping, graded difficulty markers, and rubric-guided performance tasks. Prototype revisions incorporated sentence-level listening prompts, short-dialogue response items, and rubric-guided speaking and writing assessments to broaden construct coverage and diagnostic capacity.

Research Implications – The study implies that curriculum designers, teacher educators, and policymakers should prioritize item-competency alignment, targeted teacher professional development, and modest ICT provisions to realize performance-based assessment benefits. Limitations include reliance on document analysis, interviews, and expert validation without large-scale classroom trials, and the paper recommends phased pilot testing with mixed-methods evaluation to establish effectiveness and scalability.

Article History

Received 2 February 2026

Revised 28 February 2026

Accepted 3 March 2026

Keyword:

Arabic Assessment;
Curriculum Innovation;
Listening Comprehension.

Introduction

The assessment of Arabic learning curricula occupies a central role in sustaining linguistic heritage, fostering intercultural communication, and enabling equitable access to education across diverse societies (Mohamed, 2021). Grade VII represents a formative stage in which learners consolidate foundational grammar and communicative skills, making assessment practices at this level particularly consequential for long-term proficiency. Despite this importance, many assessment approaches remain traditional and summative, limiting diagnostic insight and adaptive support for heterogeneous learners (Sutarjo, 2025). Concurrent global shifts toward digital pedagogy and blended learning further complicate assessment requirements, demanding instruments that are flexible, valid, and aligned with contemporary



This is an open-access article under the CC BY license

Corresponding Author: Nurhikmah Sani (hikmahsaninur@gmail.com)

How to Cite (APA 7th Edition): Sani, N., Rahayu, D., Alziqri JH, V., Irhamuddin, I., & Rahman, Z. (2026). Innovating Assessment for Grade VII Arabic Curriculum: A Design–Development Study of Ministry Textbooks. *Lingua: Jurnal Keilmuan dan Kependidikan Bahasa Arab*, 12(1), 12–22. <https://doi.org/10.32678/lingua.v12i1.12647>

instructional modalities (Zakharova et al., 2024). Accordingly, innovating assessment for Grade VII Arabic teaching materials is both timely and necessary to address pedagogical, cultural, and policy challenges.

Empirical work demonstrates that psychometrically sound assessment tools can meaningfully relate to academic outcomes and learner strategies, suggesting the value of validated measurement in educational settings. Scores from the Learning and Study Strategies Inventory (LASSI) showed acceptable internal consistencies and significant correlations with students' GPAs, with anxiety and test strategies emerging as significant predictors of academic achievement. These findings imply that assessments capturing cognitive, metacognitive, and affective dimensions can inform instructional interventions. Moreover, the LASSI research underscores the importance of locally validating instruments before applying them to new educational contexts. Such evidence supports the premise that robust assessment design can both diagnose learning needs and predict performance (Alkhateeb & Nasser, 2014).

Research on child second language acquisition highlights the developmental specificity and contextual sensitivity of early language learning, indicating that assessment must attend to age-related processes and learning environments. A comprehensive review of fifteen studies illustrated that children's SLA trajectories differ across instructed and naturalistic settings, calling for methodological and pedagogical adaptability in assessment design. This body of work suggests that instruments suitable for adolescents require calibration to developmental stage, classroom practices, and sociolinguistic contexts. Consequently, assessment for Grade VII should balance measurement precision with developmental appropriateness to capture both competency and growth. Integrating such perspectives helps ensure that assessment practices support valid inferences about young learners' Arabic proficiency (Philp et al., 2017).

Evidence from ICT-based initiatives demonstrates that participatory, context-sensitive implementation can strengthen teacher capacity and promote more effective assessment practices. Participatory Action Research and Asset-Based Community Development approaches substantially increased teachers' ICT competence in Arabic assessment contexts, with reported mastery of targeted applications rising from 58% to 90% after intervention (Makruf & Barokah, 2023). Similarly, theory-driven interactive modules designed for primary Arabic instruction have achieved strong content validity and favorable usability outcomes, indicating the potential of multimedia to enhance pedagogical quality (Rajab et al., 2025). A recent systematic synthesis of studies on technology integration further shows gains in engagement and autonomous learning while also highlighting variability in teacher preparedness and infrastructure support (Sapawi & Yusoff, 2025). Taken together, these studies indicate that technology-enhanced assessment can be effective when paired with rigorous design, validation, and capacity-building.

Advances in artificial intelligence and adaptive platforms offer promising avenues for personalized assessment but raise concerns about localization, cultural sensitivity, and governance. Comparative evaluations of modeling approaches in Arabic NLP show that transformer-based systems (e.g., ARABERT) and prompt-based methods achieve markedly higher accuracy in author profiling tasks compared to traditional architectures, demonstrating the maturity of language-specific AI tools (Khouja et al., 2024). Platform-level appraisals indicate that some AI-driven learning systems are recommendable with minor adaptation, yet most require substantial localization for Arabic language support, curriculum alignment, and data governance to be fully effective in regional contexts (Isaifan et al., 2025). Research on internationalization in

higher education similarly reveals that the increasing use of non-native instructional media affects Arabic proficiency and underscores the need for bilingual policy responses and transparent feedback mechanisms (El Nekiti et al., 2026). These findings collectively suggest that leveraging AI and adaptive technologies for Grade VII assessment demands careful consideration of linguistic, cultural, and regulatory factors.

Despite these advances, important limitations and gaps remain that constrain the direct transfer of prior findings to Grade VII teaching materials aligned with Ministry of Religious Affairs books. Much of the existing literature focuses on higher education, primary interventions, or cross-cultural technology reviews rather than on middle-school curricula grounded in national textbook frameworks (Sapawi & Yusoff, 2025). Teacher readiness, infrastructure variability, and curriculum misalignment continue to impede scalable implementation of ICT-based assessment in many contexts, particularly where localized teacher training is limited (Makruf & Barokah, 2023). Additionally, while students and educators express interest in practical, application-oriented modules (e.g., ASP) and multimedia innovations, gaps in communication skills, curriculum design expertise, and market literacy hinder sustainable adoption (Muslimah et al., 2025; Sahrir, 2019). These limitations point to a clear need for context-specific, validated assessment innovations that directly address the characteristics and requirements of Grade VII Arabic materials.

Responding to these gaps, the present study focuses on diagnosing shortcomings in the Ministry of Religious Affairs Grade VII Arabic teaching materials and developing innovative assessment solutions aligned with established evaluative criteria. The primary aim is to identify specific weaknesses in existing textbooks and to design assessment tools that enhance student accessibility, diagnostic clarity, and instructional usefulness as measured against criteria that support ease of learning and pedagogical effectiveness (Objectives of Research). By combining curricular analysis, stakeholder input, and design-based validation, this research intends to produce an assessment prototype that is pedagogically robust, culturally congruent, and practically implementable in Grade VII classrooms. The expected contributions include an empirically grounded model for curriculum-aligned assessment, practical guidelines for teachers and curriculum developers, and evidence to inform policy decisions regarding Arabic language assessment at the middle-school level. Ultimately, the study seeks to bridge empirical insights from recent ICT and SLA research with the specific needs of Grade VII Arabic instruction to promote measurable improvements in student learning outcomes.

Methods

This study employed a Design and Development Research (DDR) approach combined with evaluative curriculum analysis to innovate the assessment component of Grade VII Arabic teaching materials issued by the Ministry of Religious Affairs. The DDR framework was selected because it enables systematic identification of instructional weaknesses and iterative development of improved educational products. The study was conducted in three integrated phases: needs analysis, prototype design and development, and validation with expert and practitioner feedback. This methodological orientation allowed the research to move beyond descriptive critique toward the production of a validated and contextually grounded assessment model. The design aligns with contemporary approaches to curriculum innovation that emphasize evidence-based refinement and stakeholder engagement.

The research procedures were organized sequentially to ensure coherence and methodological rigor. In the first phase, a structured document analysis of the Grade VII Arabic textbook and its existing assessment components was conducted to identify conceptual, linguistic, and pedagogical shortcomings. In the second phase, assessment criteria were formulated based on principles of communicative language teaching, formative assessment, and learner-centered evaluation. A prototype of innovative assessment instruments was then developed, including formative tasks, performance-based assessments, and rubric-guided evaluations aligned with curriculum objectives. In the final phase, the prototype underwent expert review and practitioner consultation to refine clarity, feasibility, and alignment with instructional goals.

Several analytical and developmental tools were utilized throughout the study. A curriculum evaluation checklist was constructed to systematically assess the alignment between learning objectives, instructional content, and assessment tasks. Digital authoring tools were employed to design interactive and structured assessment prototypes suitable for both offline and blended learning environments. Where applicable, learning management system (LMS) features were considered to ensure compatibility with ICT-supported classroom implementation. Semi-structured interview protocols and validation sheets were developed to capture expert and teacher feedback. These instruments collectively supported both qualitative diagnosis and product development.

Data were collected through document analysis, semi-structured interviews, and expert validation questionnaires. Document analysis focused on identifying discrepancies between stated competencies and assessment practices within the existing textbook. Interviews with Arabic language teachers explored practical challenges in implementing current assessment formats and elicited recommendations for improvement. Quantitative data from validation sheets were analyzed using descriptive statistics to determine content validity indices, while qualitative feedback was examined through thematic analysis to identify recurring patterns and suggestions. The integration of qualitative and quantitative findings informed iterative revisions of the assessment prototype.

Several measures were undertaken to ensure the reliability and validity of the study. Content validity was established through expert judgment involving specialists in Arabic language education, curriculum design, and educational assessment, using structured validation instruments. Inter-rater agreement was examined to enhance consistency in evaluation outcomes. Methodological triangulation was applied by integrating document analysis, interview data, and validation results to strengthen interpretive credibility. Additionally, iterative revision cycles were conducted to ensure that the final assessment model demonstrated clarity, curricular alignment, and practical feasibility. These procedures collectively enhanced the trustworthiness and robustness of the research findings.

Results

This section presents the study's principal findings organized by sub-theme, following the sequence of the methods (document analysis → needs analysis interviews → expert validation → prototype revision). Results are reported objectively and without interpretation.

3.1. Descriptive Findings of Textbook Characteristics

The analyzed object is the Grade VII Arabic teaching book authored by Faruq Baharudin, edited by Wahib Dariyadi, and published by the Kementerian Agama

Indonesia. The textbook is composed of six main lessons, each with five consistent subsections: vocabulary (المفردات), texts (النصوص), dialogue (الحوار), structure/syntax (التركيب), and listening (الاستماع). Chapter-end evaluative items are embedded after subsections and are presented as the primary in-book assessment resources. The book's stated preparation criteria reference minimum thresholds concerning requirements, characteristics, and competencies, and assessment is framed around three quality aspects: material, presentation, and language/readability. Descriptive mapping of the textbook content was completed according to the curriculum-evaluation checklist specified in the Methods.

3.2. Functional roles of the teaching materials

The textbook is documented to serve as a teacher guide that sequences lesson delivery and guides pedagogical choices. It is also described as a student guideline that structures learning content by class level and as an evaluation resource that supplies chapter-level tests and exercises. Observational records from classroom-based checks and document analysis corroborated the presence of routine evaluative sections after each lesson component. No quantitative measures of classroom uptake were produced in this phase; functional roles are reported as categorical findings derived from the document review and practitioner observations. Instruments used to document these roles included structured observation protocols, reading tasks, and speaking prompts.

3.3. Assessment item typology and skill mapping

Assessment items within the textbook target the macro-skills of reading, speaking, writing, and listening and follow a recurring item typology that was catalogued in the document analysis. Vocabulary subsections are typically followed by matching tasks, text subsections by descriptive reading items, and listening subsections by dictation and transcription exercises. The document analysis did not identify an explicit item-to-competency alignment matrix within the textbook materials. Instruments embedded in the book are predominantly written and structured pragmatic tests; performance-based rubrics and interactive oral-assessment formats are largely absent. The mapping of skill domains to assessment techniques, assessed aspects, and typical instruments is summarized in Table 1 below.

Table 1. Mapping of skill domains to assessment techniques, assessed aspects, and typical instruments

Skill domain	Assessment technique (textbook)	Assessed aspect(s)	Typical instrument
Speaking (الكلام)	Conversational-text completion	Cognitive production and oral-response ability	Fill-in/completion items
Vocabulary (المفردات)	Word matching	Vocabulary comprehension/mastery	Matching/pragmatic tests
Reading (القراءة)	Passage reading + descriptive questions	Reading comprehension; written response	Descriptive (essay/short-answer) tests
Structure (التركيب)	Sentence formation tests	Syntactic competence	Sentence-construction items
Listening (الاستماع)	Letter/word identification; dictation	Auditory discrimination and recall	Written transcription/dictation tests
Writing (الكتابة)	Rewriting and composition tasks	Writing accuracy and coherence	Written composition and copying tasks

3.4. Listening assessment: scope and identified gaps

Within the listening subsections, items were concentrated at the letter and single-word recognition levels, requiring identification of hijaiyah letters, discrimination of similarly pronounced letters, and dictation-style transcription. Items designed to assess sentence-level listening comprehension, short-dialogue understanding, or discourse-level inference were limited across the six lessons. The document-based evidence and classroom-observation notes indicate a predominance of productive transcription tasks rather than receptive comprehension or response-eliciting formats in listening subsections. This coverage limitation was consistently recorded during the mapping and needs-analysis phases. No inferential statistical testing was applied in this documentation stage; findings are descriptive and based on content mapping and frequency observations.

3.5. Teacher feedback and expert validation outcomes

Semi-structured interviews with practicing teachers produced recurring remarks that listening tasks emphasize letters and transcription rather than sentence or discourse comprehension; exemplar teacher quotations were recorded verbatim in field notes. Expert validation sheets yielded item-level ratings from a specialist panel; item-level content-validity indices (I-CVI) computed from the validation sheets clustered in the high range for most item types, and the scale-level content-validity average (S-CVI/Ave) indicated overall acceptable content representativeness as reported by experts. Experts recommended inclusion of performance rubrics, explicit mapping of items to competency descriptors, and graded difficulty indicators for listening passages. These validation outputs were summarized using descriptive statistics and thematic categories and served as inputs for prototype revisions. The validation dataset and thematic summaries are archived as part of the project materials.

3.6 Prototype revisions and secondary results

Following the validation phase, the prototype assessment materials were revised to include additional listening item types—explicit sentence-level comprehension prompts and short-dialogue response tasks—and to introduce rubric-guided performance assessments for selected speaking and writing items. Prototype edits also incorporated explicit annotations mapping items to competency descriptors and the addition of graded difficulty markers for listening passages, consistent with expert recommendations. Documentation of prototype revisions records each change as an action item linked to corresponding needs-analysis or expert-validation evidence. Secondary descriptive observations note the original book's limited use of interactive ICT-dependent assessment features and the prototype's initial design considerations for offline and blended delivery modes. No post-revision inferential evaluation is reported in this Results section.

Discussion

The present study asked how the Grade VII Arabic teaching materials published by the Ministry of Religious Affairs conform to contemporary assessment principles and where targeted innovations are required to improve student learning and instructional utility. Previous research underscores the developmental specificity of child second-language acquisition and the need for assessment that is sensitive to instructional context and learner age (Philp et al., 2017). Studies of technology-enhanced Arabic instruction and assessment further indicate that multimedia modules and participatory implementation can strengthen pedagogical outcomes when teacher readiness and

infrastructure are addressed (Makruf & Barokah, 2023; Sapawi & Yusoff, 2025). Taken together, these literatures frame the research question by linking curriculum design, assessment practice, and local implementation conditions.

The study's principal findings indicate that the textbook is systematically organized into six lessons with consistent subsections and embedded chapter-end evaluative items, yet the embedded assessment instruments are predominantly written formats with limited performance-based rubrics. The mapping of item types showed clear correspondence between subsection and item format—vocabulary with matching items, texts with descriptive reading tasks, and listening with dictation and letter/word identification. Listening assessment was concentrated at the letter and single-word recognition level and lacked sufficient sentence- and discourse-level comprehension items. Expert validation returned high item-level content-validity indices for many item types while concurrently recommending explicit item-to-competency mapping and graded difficulty markers for listening passages.

These results align with and extend prior evidence that technology and well-designed instructional materials can improve assessment quality only when paired with intentional design and stakeholder engagement (Makruf & Barokah, 2023; Rajab et al., 2025). The predominance of discrete-letter listening items is consistent with curricular materials that emphasize foundational orthographic and phonemic skills, a focus noted in other descriptive studies of early-stage language textbooks (Philp et al., 2017). However, the documented lack of performance rubrics and interactive oral-assessment formats contrasts with recent modules that demonstrated strong pedagogical validity when multimedia and rubric-guided tasks were implemented (Rajab et al., 2025). The expert panel's recommendations and the prototype revisions therefore both corroborate and refine findings in the literature by indicating specific, implementable design changes.

A plausible explanation for the observed emphasis on letter-level listening and written formats is the combined effect of national textbook production norms, conservative assessment traditions, and resource constraints that limit the incorporation of performance-based and ICT-mediated items. Teacher reports and the needs-analysis data suggest that limited in-service training in formative oral-assessment techniques and variable access to ICT platforms contributed to the maintenance of transcription-style listening tasks (Makruf & Barokah, 2023; Sapawi & Yusoff, 2025). The high I-CVI scores for many item types indicate that content relevance was not the primary deficiency; instead, the gap lay in breadth of construct coverage—particularly receptive oral comprehension and performance assessment. Accordingly, the prototype revisions prioritized adding sentence-level listening prompts and rubric-guided speaking/writing tasks to broaden construct representation and increase diagnostic value.

The modifications introduced in the prototype respond directly to expert recommendations and to empirical patterns found in the document analysis, which may enhance the assessment's capacity to capture both discrete skill and integrative competence. These changes are consistent with evidence that contextually validated, rubric-based assessments and multimedia-supported items foster richer inferences about learner ability when teacher capacity is developed in parallel (Makruf & Barokah, 2023; Rajab et al., 2025). At the same time, the absence of inferential, classroom-level outcome data in the current study necessitates cautious interpretation of the prototype's projected impact on learning outcomes. The present results therefore indicate promise rather than demonstrated effectiveness, pending empirical trials that measure student performance changes attributable to the revised instruments.

Several limitations temper the generalizability of the findings and suggest priorities for subsequent research: the study relied primarily on document analysis, targeted interviews, and expert validation without large-scale classroom trials or experimental comparisons, and no inferential statistical testing of student-level effects was conducted. This methodological scope constrained conclusions about implementation fidelity, scalability, and actual learning gains, tasks that require pilot testing across diverse school contexts and robust pre-post or quasi-experimental designs (Zakharova et al., 2024). Future research should therefore deploy the prototype in representative classroom settings, collect learner-performance metrics, and analyze implementation barriers and facilitators to produce evidence on effectiveness and costed scalability. Such work would also permit measurement of differential effects across learner subgroups and the calibration of difficulty markers and rubrics.

The study's findings have practical implications for curriculum developers, textbook publishers, teacher educators, and policymakers seeking to strengthen Arabic assessment at the middle-school level. At the curriculum-design level, explicit item-to-competency mapping and inclusion of graded listening comprehension tasks are recommended to improve diagnostic precision and instructional alignment. Teacher professional development and modest ICT investments are likely prerequisites to realize the full benefits of performance-based and multimedia-supported assessments (Makruf & Barokah, 2023; Sapawi & Yusoff, 2025). Policy attention to bilingual instructional policy and transparent feedback mechanisms may further support meaningful integration of the revised assessments within schools (El Nekiti et al., 2026). Overall, the study provides actionable design directions and a prioritized agenda for empirical validation to advance Grade VII Arabic assessment practice.

Conclusion

The study aimed to identify shortcomings in the Ministry of Religious Affairs Grade VII Arabic teaching materials and to develop an innovative, curriculum-aligned assessment prototype that improves learner accessibility and instructional usefulness. The analysis found that the textbook is consistently organized into six lessons with embedded chapter-end evaluative items, that assessment items are predominantly written and pragmatic in format, and that listening assessment is concentrated at letter and single-word levels with insufficient sentence and discourse-level comprehension items; expert validation indicated high content relevance for many items while recommending explicit item-to-competency mapping, graded difficulty markers, and performance rubrics. In response to these findings, the prototype revisions introduced sentence-level listening prompts, short-dialogue response tasks, rubric-guided performance assessments for speaking and writing, and explicit item-to-competency annotations to broaden construct coverage and diagnostic value. These design changes have practical implications for curriculum developers and textbook publishers by providing a concrete model for aligning items with competencies and integrating graded listening tasks to improve diagnostic precision, and they underscore the need for teacher professional development and modest ICT support to realize performance-based assessment benefits. The study's principal limitations are its reliance on document analysis, targeted interviews, and expert validation without large-scale classroom trials or inferential testing of student outcomes, which constrains claims about the prototype's effectiveness and scalability. Accordingly, the study recommends phased pilot implementation in representative

Grade VII classrooms with pre–post measurement and mixed-methods evaluation, systematic monitoring of implementation fidelity, targeted teacher training on rubric-based assessment, and exploration of low-bandwidth ICT solutions to support blended delivery. By furnishing an empirically informed assessment prototype and a set of actionable design recommendations, the research contributes a contextually grounded framework that curriculum designers, teacher educators, and policymakers can adapt and test further. Future work should therefore prioritize empirical effectiveness trials, costed scalability studies, and policy-aligned capacity-building to ensure that revised assessment practices translate into measurable improvements in Arabic learning outcomes at the middle-school level.

Declarations

Author contribution statement

Nurhikmah Sani conceptualized the study, designed the research framework, supervised the project implementation, and led the manuscript preparation. Dewi Rahayu contributed to the development of the research instruments, conducted document analysis, and participated in data interpretation. Vikky Alziqri JH was responsible for data collection, including interviews and validation procedures, and assisted in prototype development. Irhamuddin conducted quantitative and qualitative data analysis, including computation of content-validity indices and thematic coding. Zikri Rahman contributed to prototype design refinement, technical editing, and critical revision of the manuscript for intellectual content. All authors reviewed and approved the final version of the manuscript.

Funding statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data availability statement

The datasets generated and analyzed during the current study, including curriculum evaluation checklists, interview transcripts (anonymized), validation sheets, and prototype documentation, are available from the corresponding author upon reasonable request.

Declaration of interests statement

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Additional information

Correspondence and requests for materials should be addressed to hikmahsaninur@gmail.com.

ORCID

Nurhikmah Sani 

Dewi Rahayu 

Vikky Alziqri JH 

Irhamuddin 

Zikri Rahman 

References

- Alkhateeb, Haitham M., & Nasser, Ramzi. (2014). Assessment of Learning and Study Strategies of University Students in Qatar Using an Arabic Translation of the Learning and Study Strategies Inventory. *Psychological Reports*, 114(3), 947–965. <https://doi.org/10.2466/11.03.PR0.114k26w3>
- El Nekiti, A., Mukala, P., Ullah, O., & Yatsyshyn, D. (2026). Advanced Learning Analytics with AI: Analyzing Stakeholder Perceptions of Feedback, Assessment, Engagement, and Faculty Development in MENA Higher Education Using NLP BT - Information System Design: Big Data Analytics and Data Science. In V. Bhateja, M. El Barachi, A. T. Azar, & D. K. Sharma (Eds.), *Proceedings of Eighth International Conference on Information System Design and Intelligent Applications* (pp. 117–129). Springer Nature Singapore. https://doi.org/10.1007/978-981-96-9248-4_9
- Isaifan, R. J., Tawalbeh, A., & Hasna, M. O. (2025). From principles to practice: a novel matrix for evaluating AI-powered learning platforms based on the UNESCO Ethical Impact Assessment tool. *Frontiers in Education*, 10, 1–18. <https://doi.org/10.3389/educ.2025.1640780>
- Khoudja, A. M., Loukam, M., & Belkredim, F. Z. (2024). Assessment of LSTM, ARABERT and Prompt-Based Learning for Gender Author Profiling in Modern Standard Arabic Language. *Ingénierie Des Systèmes d'Information*, 29(6), 2209–2218. <https://doi.org/10.18280/isi.290611>
- Makruf, I., & Barokah, A. (2023). Improving the Quality of ICT-Based Arabic Learning Assessment With Online Applications. *Journal of Higher Education Theory and Practice*, 23(11), 32–40. <https://doi.org/10.33423/jhetp.v23i11.6216>
- Mohamed, S. (2021). The development of an Arabic curriculum framework based on a compilation of salient features from CEFR level descriptors. *Language Learning Journal*, 0(0), 1–15. <https://doi.org/10.1080/09571736.2021.1923781>
- Muslimah, M., Afyuddin, M. S., Cahyono, M. E., & Pradana, S. A. (2025). Bridging Language Learning and Entrepreneurship: Examining Arabic Language Students' Potential in Developing "Arabic for Specific Purposes" Module Projects. *Journal of Arabic Language Learning and Teaching (JALLT)*, 3(1), 13–32. <https://doi.org/10.23971/jallt.v3i1.294>
- Philp, J., Borowczyk, M., & Mackey, A. (2017). Exploring the Uniqueness of Child Second Language Acquisition (SLA): Learning, Teaching, Assessment, and Practice. *Annual Review of Applied Linguistics*, 37, 1–13. <https://doi.org/10.1017/S0267190517000174>
- Rajab, S. S., Mohd-Yusoff, N., & Abdul-Aziz, M. N. (2025). Content Validation and Teacher Perspectives on an Interactive Module for the Arabic Language in Primary Education. *The Eurasia Proceedings of Educational and Social Sciences*, 43, 17–38. <https://doi.org/10.55549/epess.921>
- Sahrir, M. S. (2019). Participation in Research and Innovation Exhibitions: E-Learning Applications in Arabic Language. *Al-Basirah Journal*, 7(1), 39–53. <https://doi.org/10.22452/basirah.vol7no1.4>
- Sapawi, M. S. M., & Yusoff, N. M. R. N. (2025). Integrating technology into the Arabic language curriculum: A systematic review of trends, strategies and cultural dimensions. *Social Sciences & Humanities Open*, 12, 101974. <https://doi.org/10.1016/j.ssaho.2025.101974>
- Sutarjo, J. (2025). Instructional Dynamics of Muḥādaṣah: Evaluating Pedagogical Methods and Institutional Barriers in an Indonesian Pesantren. *Lingua: Jurnal Keilmuan Dan Kependidikan Bahasa Arab*, 11(2), 73–79.

<https://doi.org/10.32678/lingua.v11i2.12501>

Zakharova, N., Frumina, S., Lobuteva, L., & Alwaely, S. (2024). The specifics of integrating distance learning technologies with traditional classroom instruction: How to design educational curricula in modern education? *Heliyon*, *10*(20). <https://doi.org/10.1016/j.heliyon.2024.e38740>