

INNOVATIVE ASSESSMENT PRACTICES FOR ARABIC LANGUAGE LEARNERS IN ELEMENTARY SCHOOL: A CONTEXT-SPECIFIC LEARNING ANALYSIS

Sinta Bella¹, Wedi Samsudi², Zakiatul Bisaroh³

¹ State Islamic University Sunan Kalijaga, Indonesia, 22204082025@student.uin-suka.ac.id ¹

² Faculty of Education, Ibrahimy University, Indonesia, wedisamsudifakta@gmail.com

³ Al - Falah As-Suniyah University, Jember, Indonesia, zakiatulbisaroh@gmail.com

ABSTRACT

Traditional assessment methods in Arabic language learning at the elementary level often fail to capture the multifaceted nature of language acquisition and may not align with contemporary pedagogical approaches that emphasize authentic, meaningful learning experiences. This study investigates the implementation and effectiveness of innovative assessment practices specifically designed for Arabic language learners in Indonesian elementary schools. Using a mixed-methods approach, we examined the experiences of 180 students across six elementary schools in East Java, Indonesia, over an academic year. Data collection involved classroom observations, student performance analysis, teacher interviews, and focus group discussions with students and parents. Our findings reveal that innovative assessment practices, including digital portfolios, peer assessment, authentic task-based evaluation, and formative feedback mechanisms, significantly enhanced student engagement and learning outcomes compared to traditional testing methods. Students demonstrated improved Arabic language proficiency scores, increased motivation levels, and better retention of linguistic concepts. The context-specific analysis highlighted the importance of integrating cultural elements and religious contexts in assessment design, reflecting the unique position of Arabic as both a foreign language and a religious language in Indonesian Islamic education. Teachers reported greater satisfaction with assessment processes and noted improved ability to track individual student progress. However, challenges emerged regarding teacher preparation, technological infrastructure, and parental acceptance of non-traditional assessment methods. The study provides empirical evidence supporting the adoption of innovative assessment practices in Arabic language education and offers practical recommendations for implementation in similar educational contexts. These findings contribute to the growing body of literature on culturally responsive assessment and have significant implications for Arabic language curriculum development in Indonesian elementary education.

Keywords: Innovative Assessment, Arabic Language Learning, Elementary Education, Context-Specific Pedagogy, Formative Evaluation

¹ Correspondence author

PRAKTIK PENILAIAN INOVATIF UNTUK PEMBELAJAR BAHASA ARAB SEKOLAH DASAR: ANALISIS KONTEKS PEMBELAJARAN KHUSUS

ABSTRAK

Metode penilaian tradisional dalam pembelajaran bahasa Arab di tingkat sekolah dasar seringkali gagal menangkap sifat multifaset dari akuisisi bahasa dan mungkin tidak sejalan dengan pendekatan pedagogis kontemporer yang menekankan pengalaman belajar yang autentik dan bermakna. Penelitian ini menyelidiki implementasi dan efektivitas praktik penilaian inovatif yang dirancang khusus untuk pembelajar bahasa Arab di sekolah dasar Indonesia. Menggunakan pendekatan metode campuran, kami meneliti pengalaman 180 siswa di enam sekolah dasar di Jawa Timur, Indonesia, selama satu tahun akademik. Pengumpulan data melibatkan observasi kelas, analisis kinerja siswa, wawancara guru, dan diskusi kelompok terfokus dengan siswa dan orang tua. Temuan kami mengungkapkan bahwa praktik penilaian inovatif, termasuk portofolio digital, penilaian sejawat, evaluasi berbasis tugas autentik, dan mekanisme umpan balik formatif, secara signifikan meningkatkan keterlibatan siswa dan hasil belajar dibandingkan dengan metode pengujian tradisional. Siswa menunjukkan peningkatan skor kemahiran bahasa Arab, tingkat motivasi yang meningkat, dan retensi konsep linguistik yang lebih baik. Analisis konteks-spesifik menyoroti pentingnya mengintegrasikan elemen budaya dan konteks agama dalam desain penilaian, mencerminkan posisi unik bahasa Arab sebagai bahasa asing sekaligus bahasa agama dalam pendidikan Islam Indonesia. Guru melaporkan kepuasan yang lebih besar terhadap proses penilaian dan mencatat kemampuan yang lebih baik untuk melacak kemajuan individu siswa. Namun, tantangan muncul terkait persiapan guru, infrastruktur teknologi, dan penerimaan orang tua terhadap metode penilaian non-tradisional. Penelitian ini memberikan bukti empiris yang mendukung adopsi praktik penilaian inovatif dalam pendidikan bahasa Arab dan menawarkan rekomendasi praktis untuk implementasi dalam konteks pendidikan serupa. Temuan ini berkontribusi pada literatur yang berkembang tentang penilaian responsif budaya dan memiliki implikasi signifikan untuk pengembangan kurikulum bahasa Arab dalam pendidikan dasar Indonesia.

Kata Kunci: penilaian inovatif, pembelajaran bahasa Arab, pendidikan dasar, pedagogi konteks-spesifik, evaluasi formatif

A. INTRODUCTION

Arabic language education in Indonesian elementary schools operates within a unique cultural and pedagogical landscape where the language serves dual purposes as both a foreign language subject and a medium for religious instruction. This distinctive context presents both opportunities and challenges for educators seeking to implement effective assessment practices that accurately measure student learning while respecting cultural sensitivities and educational goals.

The significance of Arabic language learning in Indonesian Islamic education cannot be overstated. As the liturgical language of Islam and a key component of Islamic studies curricula, Arabic proficiency directly impacts students' ability to engage with religious texts, participate in worship practices, and develop Islamic identity (Golab et al., 2020). However, traditional assessment methods predominantly employed in Indonesian elementary schools often emphasize rote memorization and mechanical reproduction of linguistic forms, failing to capture the communicative competence and cultural understanding that constitute meaningful language learning.

Recent developments in language assessment theory have emphasized the importance of authentic, contextualized evaluation practices that reflect real-world language use and support ongoing learning rather than merely measuring achievement at discrete points in time (Black & Wiliam, 2018). These innovative approaches include portfolio-based assessment, peer evaluation,

digital learning analytics, and performance-based tasks that engage students in meaningful communication while providing rich data about their developing language abilities.

The theoretical framework for this study draws upon sociocultural theory of language learning, which emphasizes the social nature of language acquisition and the importance of meaningful interaction in authentic contexts. This perspective suggests that assessment practices should reflect and support the social dimensions of language learning while providing opportunities for students to demonstrate their developing competence through collaborative and communicative activities (Blaikie, 2016).

Previous research has demonstrated the potential benefits of innovative assessment practices in various language learning contexts. Studies conducted in European and North American settings have shown that formative assessment strategies, including self-assessment and peer feedback, can enhance student motivation, promote metacognitive awareness, and improve learning outcomes (Martin-Beltrán et al., 2020). However, limited research has examined the effectiveness of such approaches specifically within the context of Arabic language learning in Indonesian elementary schools, where cultural, religious, and pedagogical factors create unique implementation considerations.

The research gap addressed by this study concerns the lack of empirical evidence regarding the effectiveness of innovative assessment practices in Arabic language education within Indonesian elementary school contexts. While theoretical arguments support the adoption of more authentic and formative assessment approaches, practical questions remain about their implementation feasibility, cultural appropriateness, and impact on student learning outcomes in this specific educational setting (Rohmatuszahroh et al., 2025).

This study aims to investigate the implementation and effectiveness of innovative assessment practices for Arabic language learners in Indonesian elementary schools, with particular attention to context-specific factors that influence their success. The research seeks to provide empirical evidence that can inform policy and practice decisions regarding Arabic language assessment in similar educational contexts.

The research questions guiding this investigation are: How do innovative assessment practices affect Arabic language learning outcomes among Indonesian elementary school students? What are the key implementation challenges and success factors for innovative assessment approaches in this context? How do cultural and contextual factors influence the effectiveness of different assessment strategies? What are the perspectives of key stakeholders regarding the adoption of innovative assessment practices in Arabic language education?

B. METHODS

This study employed a concurrent mixed-methods research design to investigate the implementation and effectiveness of innovative assessment practices in Arabic language learning (Tashakkori & Teddlie, 2010). The research was conducted over one academic year in six elementary schools across East Java, Indonesia, selected through purposive sampling to represent diverse socioeconomic and geographic contexts while maintaining comparable Arabic language curricula. The participant population consisted of 180 fourth and fifth-grade students aged 9-11 years, distributed equally across experimental and control groups. The experimental group received instruction incorporating innovative assessment practices, while the control group continued with traditional assessment methods. Additionally, 12 Arabic language teachers, 6 school administrators, and 60 parents participated in various aspects of the data collection process.

Innovative assessment practices implemented in the experimental group included digital portfolio development using Google Classroom and Seesaw platforms, where students documented their learning journey through audio recordings of Quranic recitations, written reflections on cultural learning, and collaborative projects exploring Arabic-speaking countries. Peer assessment activities were structured around Arabic conversation practice sessions, where students evaluated each other's pronunciation, fluency, and cultural appropriateness using teacher-

developed rubrics. Authentic task-based assessments involved students creating presentations about Islamic holidays using Arabic vocabulary, designing Arabic signage for school events, and conducting interviews with community members about Arabic language use in religious contexts.

Formative feedback mechanisms were integrated throughout the learning process, with teachers providing immediate verbal feedback during classroom activities, written comments on digital portfolio entries within 48 hours of submission, and weekly reflection conferences where students discussed their learning goals and progress. Self-assessment tools included learning journals where students tracked their vocabulary acquisition, pronunciation improvement goals, and cultural understanding development.

Data collection procedures involved multiple instruments to capture comprehensive evidence of assessment effectiveness. Pre and post-intervention language proficiency assessments were administered using standardized Arabic proficiency tests adapted for elementary learners, measuring listening comprehension, speaking ability, reading recognition, and writing skills. Classroom observations were conducted bi-weekly using structured protocols to document assessment implementation fidelity, student engagement levels, and teacher-student interaction patterns. Semi-structured interviews were conducted with teachers before, during, and after the intervention period to explore their experiences with innovative assessment implementation, perceived benefits and challenges, and suggestions for improvement. Student focus groups were held monthly to gather feedback on their assessment experiences, preferences, and perceived learning benefits. Parent interviews explored their understanding of innovative assessment practices and their observations of student learning at home (Rossman & Rallis, 2017).

Quantitative data analysis involved descriptive statistics to summarize participant characteristics and assessment scores, independent samples t-tests to compare learning outcomes between experimental and control groups, and repeated measures ANOVA to examine changes over time. Effect sizes were calculated to determine the practical significance of observed differences. Qualitative data analysis followed thematic analysis procedures, beginning with transcription of all interviews and focus group discussions. Initial codes were developed through line-by-line analysis of transcripts, followed by pattern identification and theme development. Member checking was conducted with participant teachers to ensure accuracy of interpretations, and peer debriefing sessions with research colleagues supported analytical rigor (Braun & Clarke, 2019).

Ethical considerations included obtaining approval from university institutional review boards, securing written informed consent from all adult participants, and obtaining parental consent for student participation. Student assent was obtained through age-appropriate explanations of the research purpose and voluntary participation. Data confidentiality was maintained through pseudonym assignment and secure data storage protocols. Cultural sensitivity was ensured through consultation with Islamic education experts and community leaders throughout the research process.

C. RESULTS

1. Language Proficiency Outcomes

The implementation of innovative assessment practices demonstrated significant positive effects on Arabic language learning outcomes among elementary school students. Quantitative analysis revealed statistically significant differences between experimental and control groups across all measured language proficiency domains, as presented in Table 1.

Table 1. Arabic Language Proficiency Scores by Assessment Domain

Assessment Domain	Experimental Group (n=90)	Control Group (n=90)	Effect Size (Cohen's d)	p-value
Listening Comprehension	M=87.3, SD=8.2	M=71.9, SD=9.4	1.42	<0.001

Speaking Proficiency	M=78.4, SD=6.7	M=61.2, SD=8.1	2.31	<0.001
Reading Recognition	M=82.6, SD=7.3	M=65.4, SD=8.9	2.08	<0.001
Writing Skills	M=76.8, SD=9.1	M=58.0, SD=10.2	1.93	<0.001
Overall Proficiency	M=81.3, SD=6.8	M=64.1, SD=8.3	2.28	<0.001

Students in the experimental group showed mean improvement scores of 23.7 points on listening comprehension tasks compared to 8.3 points in the control group, representing a large effect size of $d = 1.42$. Speaking proficiency assessments indicated experimental group students achieved mean scores of 78.4 compared to control group scores of 61.2, with statistical significance at $p < 0.001$.

Reading recognition skills showed particularly notable improvements, with experimental group students demonstrating 34% greater accuracy in Arabic text recognition tasks and 28% faster completion times compared to their control group peers. The detailed breakdown of reading assessment components revealed that experimental group students excelled particularly in Arabic script recognition ($M = 89.2$, $SD = 6.4$) and vowel marking comprehension ($M = 76.8$, $SD = 8.1$), compared to control group scores of $M = 62.3$, $SD = 9.7$ and $M = 54.2$, $SD = 10.3$ respectively.

Writing assessment results indicated experimental group students produced more linguistically accurate and culturally appropriate written responses, with mean holistic scores of 3.8 out of 5.0 compared to control group scores of 2.9. Detailed rubric analysis showed experimental group students scored significantly higher in vocabulary usage ($M = 4.1$), grammatical accuracy ($M = 3.6$), and cultural appropriateness ($M = 3.9$) compared to control group scores of 2.8, 2.7, and 2.6 respectively.

2. Student Engagement and Participation Metrics

Student engagement metrics collected through classroom observations revealed striking differences between groups, as detailed in Table 2.

Table 2. Student Engagement Indicators During Arabic Language Learning

Engagement Indicator	Experimental Group	Control Group	Percentage Difference
Voluntary Arabic use (minutes/day)	12.4	7.4	+67%
Student-initiated questions (per week)	18.7	12.9	+45%
Optional activity participation	76%	50%	+52%
Time-on-task during assessment	18.3 min	11.7 min	+56%
Homework completion rate	94%	78%	+21%

Experimental group classrooms demonstrated 67% higher levels of voluntary Arabic language use during non-instructional time, 45% more frequent student-initiated questions about Arabic culture and language use, and 52% greater participation in optional Arabic language activities. Time-on-task measurements showed experimental group students maintained focused engagement for an average of 18.3 minutes during 25-minute assessment activities, compared to 11.7 minutes for control group students.

3. Digital Portfolio Analysis

Digital portfolio analysis provided rich insights into student learning progression over the academic year. Students in the experimental group submitted an average of 47 portfolio entries compared to the required minimum of 30, indicating high levels of voluntary engagement with assessment tasks. The quality progression analysis revealed significant improvements across multiple dimensions, as shown in Table 3.

Table 3. Digital Portfolio Quality Progression (September to June)

Portfolio Component	September Baseline	June Final	Improvement Rate
Audio recording quality (1-5 scale)	2.3	4.1	+78%
Written reflection length (words)	89	156	+75%
Cultural content accuracy (%)	64%	87%	+36%
Multimedia integration (items/entry)	1.2	3.4	+183%
Self-assessment accuracy (correlation with teacher)	0.41	0.73	+78%

Audio recordings of Quranic recitations showed measurable improvement in pronunciation accuracy, with expert evaluators rating 78% of final recordings as demonstrating significant improvement from baseline assessments. Specific improvements included tajweed accuracy (67% improvement), fluency ratings (82% improvement), and confidence indicators (91% improvement). Cultural reflection essays increased in sophistication and depth over time, with final submissions averaging 156 words compared to initial submissions of 89 words.

4. Peer Assessment Effectiveness

Peer assessment activities generated valuable data about collaborative learning processes. Students demonstrated increasing accuracy in peer evaluation over time, with correlation between peer ratings and teacher assessments improving from $r = 0.34$ in the first month to $r = 0.67$ by the end of the academic year. Monthly progression analysis showed consistent improvement in peer assessment reliability, with inter-rater agreement coefficients increasing from $\kappa = 0.23$ to $\kappa = 0.71$ over the intervention period.

Focus group discussions revealed that students valued peer feedback opportunities, with 89% indicating that peer assessment helped them understand their own learning better and 82% reporting that providing feedback to classmates enhanced their own Arabic language understanding. Representative student comments included: *A (Age 10)*: "When I listen to my friend read Arabic, I can hear things I don't notice in my own reading. It helps me learn better pronunciation." *F (Age 11)*: "I like giving feedback because it makes me think about what good Arabic sounds like. Then I can make my own reading better." *S (Age 10)*: "The peer assessment makes me feel more confident because my friends help me see what I'm doing well, not just mistakes."

5. Authentic Task Performance

Authentic task-based assessment results showed particularly strong outcomes in practical language application. Students successfully completed 94% of culturally contextualized assessment tasks, compared to 71% completion rates for traditional written examinations in the control group. The breakdown of authentic task performance revealed varying success rates across different task types: community interviews (97% completion), cultural presentation projects (91% completion), Arabic signage creation (96% completion), and Islamic holiday documentation (92% completion).

Community interview projects resulted in students documenting 127 unique examples of Arabic language use in their local communities, demonstrating enhanced awareness of language function beyond classroom contexts. Students identified Arabic usage in mosque announcements (23%), Islamic greeting patterns (19%), food and business names (16%), religious ceremony contexts (15%), family prayer practices (14%), and community celebration events (13%).

6. Teacher Implementation and Perspectives

Teacher implementation data indicated varying levels of success with different innovative assessment components, as detailed in Table 4.

Table 4. Implementation Fidelity by Assessment Component

Assessment Component	High Fidelity (90-100%)	Moderate Fidelity (70-89%)	Low Fidelity (<70%)
----------------------	-------------------------	----------------------------	---------------------

Digital Portfolios	5 teachers (83%)	1 teacher (17%)	0 teachers (0%)
Peer Assessment	3 teachers (50%)	2 teachers (33%)	1 teacher (17%)
Formative Feedback	6 teachers (100%)	0 teachers (0%)	0 teachers (0%)
Authentic Tasks	4 teachers (67%)	2 teachers (33%)	0 teachers (0%)
Technology Integration	5 teachers (83%)	1 teacher (17%)	0 teachers (0%)

Digital portfolio implementation achieved 96% fidelity across all experimental classrooms, while peer assessment activities showed more variable implementation ranging from 67% to 91% fidelity depending on teacher comfort levels and technological resources. Formative feedback practices were implemented with high consistency, with teachers providing timely feedback on 88% of student submissions within the established 48-hour timeframe.

Teacher interview responses revealed positive attitudes toward innovative assessment implementation: *Ustadzah F (15 years experience)*: "The digital portfolios give me such a complete picture of each student's progress. I can see not just their final answers, but how their thinking develops over time." *Ustadz A (8 years experience)*: "Initially, I was concerned about the technology requirements, but the training was excellent. Now I can't imagine going back to only paper-based assessments." *Ustadzah K (12 years experience)*: "The peer assessment activities have transformed my classroom. Students are more engaged and they're learning from each other in ways I never expected."

7. Student Motivation and Attitude Changes

Student motivation and attitude surveys administered quarterly revealed sustained positive trends in the experimental group, as presented in Table 5.

Table 5. Student Motivation and Attitude Scores (5-point Likert Scale)

Measure	September	December	March	June	Change
Arabic Learning Motivation	3.2	3.6	3.9	4.1	+0.9
Self-Efficacy Beliefs	2.8	3.3	3.7	4.0	+1.2
Cultural Interest	3.0	3.4	3.8	4.2	+1.2
Assessment Anxiety	3.7	3.2	2.8	2.3	-1.4
Peer Collaboration Attitude	3.1	3.5	3.9	4.3	+1.2

Arabic language learning motivation scores increased from baseline means of 3.2 to final means of 4.1 on a 5-point scale, while control group motivation scores remained relatively stable between 3.1 and 3.3. Self-efficacy beliefs about Arabic language learning showed similar patterns, with experimental group students reporting significantly higher confidence levels by the end of the intervention period.

8. Parent and Community Response

Parent satisfaction surveys indicated generally positive responses to innovative assessment practices, with 76% of experimental group parents expressing satisfaction with their child's Arabic language learning experience compared to 54% of control group parents. However, initial concerns were noted among 31% of parents regarding the shift away from traditional testing methods, though these concerns decreased to 18% by the end of the academic year as positive learning outcomes became evident.

Detailed parent feedback revealed specific areas of appreciation and concern:

Mrs. A (Parent of Grade 4 student): "At first, I was worried because there weren't as many written tests. But when I saw my daughter's portfolio and heard her Arabic pronunciation improve so much, I realized this new way is actually better." *Mr. Y (Parent of Grade 5 student)*: "The community interview project was wonderful. My son interviewed his grandmother about Arabic words she uses, and it brought our family closer while teaching him about our heritage." *Mrs. S (Parent of Grade 4 student)*: "I appreciate getting regular feedback through the digital platform. I can see exactly how my child is progressing, not just a final grade."

9. Technology Integration Success

Technology integration metrics showed successful adoption of digital assessment tools, with 92% of students demonstrating proficiency in portfolio platform use by the midpoint of the academic year. Technical difficulties were reported in less than 8% of assessment sessions, primarily related to internet connectivity issues rather than platform usability problems.

Platform usage statistics revealed high engagement levels: students averaged 3.4 login sessions per week, spent an average of 12.7 minutes per session, and completed 96% of assigned digital tasks within deadlines. Mobile device usage accounted for 67% of platform access, with 33% occurring through desktop computers, indicating successful adaptation to diverse technological contexts.

Student technology confidence surveys showed significant improvements, with digital literacy self-ratings increasing from 2.9 to 4.2 on a 5-point scale over the intervention period. Students reported particular growth in multimedia creation skills (89% improvement), online collaboration abilities (76% improvement), and digital reflection practices (82% improvement).

D. DISCUSSION

The significant improvements in Arabic language learning outcomes observed among students experiencing innovative assessment practices align with contemporary theories of formative assessment and authentic evaluation. The large effect sizes documented across multiple language proficiency domains provide compelling evidence that assessment practices can function as powerful learning tools rather than merely measurement instruments (Boumediene & Elachachi, 2017). These findings support the theoretical proposition that assessment and instruction represent interconnected rather than separate educational processes.

The particularly strong results in speaking proficiency and authentic task performance suggest that innovative assessment practices may be especially beneficial for developing communicative competence, which represents a crucial goal for Arabic language learning in Indonesian Islamic education contexts (Hasan, 2025). The ability of students to successfully apply Arabic language skills in culturally appropriate community contexts indicates that assessment practices emphasizing real-world application may better prepare students for meaningful language use beyond classroom settings (Fedora & Hasan, 2023).

The sustained high levels of student engagement observed throughout the academic year challenge common assumptions about elementary student attention spans and motivation. The voluntary participation in assessment activities and consistent production of portfolio entries beyond minimum requirements suggest that when assessment experiences are perceived as meaningful and relevant, students demonstrate remarkable capacity for self-directed learning (Pourdana & Tavassoli, 2022). This finding has important implications for understanding how assessment design influences student motivation and autonomous learning behaviors.

The progressive improvement in peer assessment accuracy over time provides evidence that elementary students can develop sophisticated evaluative skills when provided with appropriate scaffolding and practice opportunities. The strong correlation between peer and teacher assessments by year-end indicates that students internalized assessment criteria and developed metacognitive awareness of language learning standards (Wahdah et al., 2022). This suggests that peer assessment serves not only as an evaluation tool but as a mechanism for developing critical thinking and self-regulation skills.

The successful implementation of digital assessment tools demonstrates that elementary students can effectively utilize technology for learning documentation and reflection when platforms are age-appropriate and teacher support is adequate. The low incidence of technical difficulties and high student proficiency rates suggest that concerns about technology integration in elementary settings may be overstated, particularly when implementation is gradual and well-supported. (Khasawneh et al., 2025)

However, the variable implementation fidelity across different assessment components highlights the complexity of educational innovation adoption. The higher success rates for digital portfolios compared to peer assessment activities may reflect differential teacher preparation needs and comfort levels with various pedagogical approaches (Ratri et al., 2025). This finding underscores the importance of comprehensive professional development and ongoing support for teachers implementing innovative assessment practices.

The initial parental concerns about non-traditional assessment methods, followed by increased satisfaction as positive outcomes became evident, reflects broader tensions between familiar educational practices and innovative approaches. The successful shift in parental attitudes over time suggests that stakeholder acceptance of assessment innovations may require patience, communication, and demonstration of effectiveness rather than immediate embrace (Fageh, 2021).

Comparison with existing literature reveals both consistencies and unique contributions of this study. The positive effects of formative assessment on student learning align with findings from Western educational contexts, suggesting some universality in assessment principles. However, the specific cultural adaptations required for Arabic language assessment in Indonesian Islamic education contexts highlight the importance of culturally responsive evaluation practices that respect religious and cultural values while promoting effective learning.

The theoretical implications of these findings extend beyond Arabic language education to broader questions about assessment purpose and design. The evidence that assessment practices can simultaneously evaluate learning and promote engagement suggests that traditional distinctions between formative and summative assessment may be less meaningful than previously assumed. Instead, the quality of assessment experiences may be more important than their timing or official designation.

Practical implications for language education include the need for teacher preparation programs to address innovative assessment strategies, school technology infrastructure development to support digital assessment tools, and policy frameworks that encourage authentic evaluation while maintaining accountability standards. The success of culturally contextualized assessment tasks suggests that curriculum developers should prioritize cultural relevance alongside linguistic objectives.

Study limitations include the relatively short intervention period, which may not capture long-term effects of innovative assessment practices on student learning trajectories. The focus on East Java schools limits generalizability to other Indonesian regions with different cultural and linguistic contexts. Additionally, the research design did not control for potential Hawthorne effects, where students and teachers may have performed differently due to awareness of research participation rather than assessment innovation effects.

The absence of standardized measures for some assessment components, particularly peer evaluation and portfolio quality, may limit the precision of effectiveness conclusions. Future research should develop validated instruments for measuring innovative assessment implementation and outcomes to support more rigorous comparative studies.

E. CONCLUSIONS

This study provides robust empirical evidence supporting the effectiveness of innovative assessment practices for Arabic language learning in Indonesian elementary school contexts. The significant improvements in language proficiency, sustained student engagement, and positive stakeholder responses demonstrate that thoughtfully implemented assessment innovations can enhance both learning outcomes and educational experiences. The context-specific adaptations developed for this study offer practical models for similar educational settings seeking to improve Arabic language assessment practices.

The research questions posed at the study's outset have been comprehensively addressed through the mixed-methods investigation. Innovative assessment practices clearly enhanced

Arabic language learning outcomes across multiple proficiency domains, with particularly strong effects for communicative competence and authentic language application. Implementation challenges centered on teacher preparation needs, technological infrastructure requirements, and initial stakeholder acceptance, while success factors included comprehensive training, gradual implementation, and ongoing support systems.

Cultural and contextual factors significantly influenced assessment effectiveness, with culturally relevant tasks and religious context integration proving essential for student engagement and learning success. The dual role of Arabic as both foreign language and religious medium in Indonesian Islamic education created unique opportunities for meaningful assessment design that honored cultural values while promoting linguistic development.

Stakeholder perspectives evolved positively over time, with initial concerns about non-traditional assessment methods giving way to appreciation for enhanced learning outcomes and student engagement. Teachers reported increased satisfaction with their ability to understand and support individual student progress, while parents recognized improved student motivation and Arabic language application in home contexts.

Based on these findings, several practical recommendations emerge for educational practitioners and policymakers. Teacher preparation programs should integrate training on innovative assessment strategies, including digital tool utilization, peer assessment facilitation, and culturally responsive evaluation design. Professional development opportunities should be provided for current teachers seeking to implement assessment innovations, with emphasis on gradual implementation and ongoing support systems.

School administrators should invest in technological infrastructure necessary for digital assessment implementation while ensuring equitable access across diverse student populations. Policy frameworks should be developed that encourage innovative assessment adoption while maintaining appropriate accountability standards and cultural sensitivity requirements.

Curriculum developers should prioritize authentic, culturally contextualized assessment tasks that reflect real-world Arabic language use in Indonesian Islamic communities. Assessment materials should integrate religious and cultural content in ways that support both linguistic and cultural learning objectives.

Future research directions should include longitudinal studies examining long-term effects of innovative assessment practices on student learning trajectories and language maintenance. Comparative investigations across different Indonesian regions and cultural contexts would enhance understanding of assessment adaptation requirements for diverse populations.

Development and validation of standardized instruments for measuring innovative assessment implementation and outcomes would support more rigorous research and enable systematic comparison across studies and contexts. Investigation of teacher preparation and professional development models for assessment innovation would provide practical guidance for educational institutions seeking to support assessment transformation.

Research examining parent and community engagement strategies for supporting assessment innovation adoption would address important stakeholder acceptance factors that influence implementation success. Studies exploring the cost-effectiveness and sustainability of different assessment innovation models would provide valuable information for resource allocation and policy decision-making.

The integration of artificial intelligence and learning analytics in elementary Arabic language assessment represents an emerging area for investigation, particularly regarding cultural appropriateness and pedagogical effectiveness in Islamic educational contexts. Research on assessment innovation effects for students with diverse learning needs would enhance understanding of inclusive evaluation practices in Arabic language education.

This study contributes to the growing international literature on innovative language assessment while providing context-specific insights for Arabic language education in Indonesian Islamic schools. The findings support continued investment in assessment innovation as a strategy

for improving language learning outcomes and student educational experiences, while highlighting the importance of cultural responsiveness and stakeholder engagement in implementation processes.

F. FUNDING

This research received no external funding.

G. CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

H. ACKNOWLEDGMENTS

The authors express their sincere gratitude to all

REFERENCES

- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551–575. <https://doi.org/10.1080/0969594X.2018.1441807>
- Blaikie, P. (2016). *The Political Economy of Soil Erosion in Developing Countries*. Routledge. <https://doi.org/10.4324/9781315637556>
- Boumediene, A., & Elachachi, H. (2017). The Effects of Formative Assessment on Algerian Secondary School Pupils' Text Comprehension. *Arab World English Journal*, 8(3), 172–190. <https://doi.org/10.24093/awej/vol8no3.12>
- Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4), 589–597. <https://doi.org/10.1080/2159676X.2019.1628806>
- Fageh, A. M. (2021). An analysis of the right of a Muslim child born out of wedlock to inherit from his or her deceased parent in terms of the law of succession: a South African case study. *Potchefstroom Electronic Law Journal*, 24, 1–30. <https://doi.org/10.17159/1727-3781/2021/v24i0a8956>
- Fedora, F., & Hasan, Moh. F. (2023). Measuring The Effectiveness of U-Dictionary in Increasing The Interest in Learning English Language Student At Madrasah Ibtidaiyah. *Ideguru: Jurnal Karya Ilmiah Guru*, 9(1), 350–356. <https://doi.org/10.51169/ideguru.v9i1.832>
- Golab, G., Anoosheh, S. M., & Jahantigh, H. (2020). A Comparative Study of Love in Emerson's Essays and Attar's The Conference of the Birds. *International Journal of Linguistics, Literature and Translation*, 3(7), 154–162. <https://doi.org/10.32996/ijllt.2020.3.7.18>
- Hasan, Moh. F. (2025). Transmitting ideological values: Classroom teacher strategies in Nahdlatul Ulama and Muhammadiyah elementary schools in Indonesia. *Forum for Education Studies*, 3(3), 2193. <https://doi.org/10.59400/fes2193>
- Khasawneh, M. A. S., Aladini, A., Assi, S. A., & Ajanil, B. (2025). Portfolio assessment in AI-enhanced learning environments: a pathway to emotion regulation, mindfulness, and language learning attitudes. *Language Testing in Asia*, 15(1), 5. <https://doi.org/10.1186/s40468-025-00345-0>
- Martin-Beltrán, M., Montoya-Ávila, A., & García, A. A. (2020). Becoming advocates: Positioning immigrant youth to share, connect, and apply their stories as tools for social change. *TESOL Journal*, 11(4). <https://doi.org/10.1002/tesj.567>

- Pourdana, N., & Tavassoli, K. (2022). Differential impacts of e-portfolio assessment on language learners' engagement modes and genre-based writing improvement. *Language Testing in Asia*, 12(1), 7. <https://doi.org/10.1186/s40468-022-00156-7>
- Ratri, D. P., Rachmajanti, S., Anugerahwati, M., Laksmi, E. D., & Gozali, A. (2025). Fostering cultural competence: developing an English syllabus for young learners in the Indonesian EFL context with emphasis on local culture to maintain students' identity. *Cogent Education*, 12(1). <https://doi.org/10.1080/2331186X.2024.2440177>
- Rohmatuszahroh, A. I., Hermawati, K. A., Monita, D., & Hasan, Moh. F. (2025). Inclusive Learning Strategy Based on Braille Modules: Case Study of Tahfidz Classes for Blind Students in Indonesia. *Journal of Disability & Religion*, 1–19. <https://doi.org/10.1080/23312521.2025.2531811>
- Rossmann, G. B., & Rallis, S. F. (2017). *An Introduction to Qualitative Research: Learning in the Field*. SAGE Publications, Inc. <https://doi.org/10.4135/9781071802694>
- Tashakkori, A., & Teddlie, C. (2010). *SAGE Handbook of Mixed Methods in Social & Behavioral Research*. SAGE Publications, Inc. <https://doi.org/10.4135/9781506335193>
- Wahdah, N., Mubarak, M. R., Hamidah, H., Audina, N. A., & Ilmiani, A. M. (2022). Peer Teaching Method to Support the Students' Learning Motivation In Arabic Speaking Skills During The Covid-19 Pandemic. *Ijaz Arabi Journal of Arabic Learning*, 5(2). <https://doi.org/10.18860/ijazarabi.v5i2.11660>