



Research Article

Reconstructing Islamic Religious Education Test Instruments Through Authentic Digital Assessment

Reza Setya Rachman¹, Dina Hermina², Sheila Rosalia³

- 1 Universitas Islam Negeri Antasari Banjarmasin, Indonesia; rezasetyarachman27@gmail.com
- 2 Universitas Islam Negeri Antasari Banjarmasin, Indonesia; dinahermina@uin-antasari.ac.id
- 3 Universitas Islam Negeri Antasari Banjarmasin, Indonesia; sheilalialiaaa0308@gmail.com

Copyright © 2026 by Authors, Published by **Classroom: Journal of Islamic Education**. This is an open access article under the CC BY License <https://creativecommons.org/licenses/by/4.0/>

Received : March 08, 2026
Accepted : May 14, 2026

Revised : April 10, 2026
Available online : June 14, 2026

How to Cite: Reza Setya Rachman, Dina Hermina, & Sheila Rosalia. (2026). Reconstructing Islamic Religious Education Test Instruments Through Authentic Digital Assessment. *Classroom: Journal of Islamic Education*, 3(1), 115–129. <https://doi.org/10.61166/classroom.v3i1.54>

Reconstructing Islamic Religious Education Test Instruments Through Authentic Digital Assessment

Abstract. The advent of digital transformation and the emergence of generative artificial intelligence appear to have challenged conventional assessment practices within Islamic Religious Education, particularly concerning the treatment of written examinations as the primary evidential measure of students' religious comprehension. This article seeks to effect a reconstruction of assessment instruments in Islamic Religious Education through the integration of written, oral, and performance evaluations within an authentic assessment framework. This study utilizes a descriptive-analytical approach, grounded in a comprehensive examination of extant literature pertaining to educational evaluation, authentic assessment, Islamic Religious Education assessment, digital learning, and artificial intelligence in education. The findings suggest that while written tests may retain utility for assessing conceptual understanding, religious reasoning, and textual literacy, they appear insufficient for encompassing oral competence, worship practices, moral conduct, and the internalization of

Islamic values. Oral tests are needed to verify the depth and authenticity of students' understanding, while performance tests are essential for assessing worship skills, religious behavior, and value-based action. This article proposes the development of an integrative assessment model, which is predicated upon the triangulation of written, oral, and performative evidence, substantially supported by digital technology. The model positions technology as a tool for documentation, feedback, and portfolio management, while teachers remain the central pedagogical interpreters of students' religious learning evidence.

Keywords. Islamic Religious Education, Authentic Assessment, Written Test, Oral Test, Performance Test, Digital Assessment

INTRODUCTION

The digital transformation of education has changed the modalities for the instruction, evaluation, and documentation of Islamic Religious Education. Students today encounter religious knowledge not only through teachers and textbooks, but also through search engines, social media, learning management systems, video platforms, online quizzes, and generative artificial intelligence. This situation creates new opportunities for more flexible and data-supported assessment. At the same time, it challenges the validity and authenticity of conventional assessment practices, especially when written tests are used as the dominant evidence of students' religious understanding.

Islamic Religious Education, known in the Indonesian school context as Pendidikan Agama Islam (PAI), has a distinctive character because it is not limited to the transmission of religious concepts. It also aims to develop faith, moral awareness, worship skills, social responsibility, and the ability to practice Islamic values in daily life. The learning outcomes of Islamic Religious Education and Character Education cover the Qur'an and Hadith, creed, morality, Islamic jurisprudence, and Islamic history, while also emphasizing moderation, verification of information, worship practice, and exemplary conduct (Kementerian Pendidikan Dasar dan Menengah, 2026). This multidimensional orientation means that assessment in Islamic Religious Education cannot be reduced to the measurement of memorized concepts or definitions.

The dominance of written tests creates a serious limitation. Written tests are useful for assessing conceptual understanding, textual literacy, religious reasoning, and the ability to analyze cases. However, they cannot fully capture students' ability to recite the Qur'an, explain religious ideas orally, perform ablution and prayer correctly, demonstrate religious etiquette, or internalize Islamic values through consistent behavior. In assessment theory, the quality of an instrument is determined by its alignment with the learning target and the type of evidence required to support a valid judgment (Brookhart & Nitko, 2019). Therefore, an instrument that is easy to score is not necessarily the most valid instrument for assessing Islamic Religious Education.

The complexity of the challenge escalates in the digital era. Digital platforms facilitate the distribution of assessments, the collection of responses, the storage of

scores, and the provision of feedback with heightened efficiency. Studies on digital assessment in Islamic Religious Education tend to indicate that learning management systems and online quiz platforms can support the management of tests and the documentation of learning outcomes (Magfirah dkk., 2025; Regita dkk., 2024). Nevertheless, digitalization can remain superficial when teachers solely effectuate the transference of paper-based questions to online platforms without redesigning the validity, authenticity, and ethical dimensions of assessment. Technology can accelerate administration, but it cannot automatically ensure that the assessment comprehensively assesses religious understanding, worship performance, moral behavior, and spiritual growth.

Generative artificial intelligence introduces an additional stratum of pedagogical concern. Students are now capable of generating essays, summaries, reflective texts, or religious arguments through the application of machine assistance. Such texts may appear coherent and academically acceptable, but they do not always represent personal understanding. The Artificial Intelligence Assessment Scale emphasizes that the use of artificial intelligence in assessment may warrant explicit regulation according to learning objectives, task design, and the expected degree of student agency (Perkins dkk., 2024, 2025). For Islamic Religious Education, this issue is not only about plagiarism or academic integrity. It is also about whether assessment can distinguish between borrowed textual production and genuine religious understanding, reflection, and practice.

Prior scholarly investigations have elucidated authentic assessment, digital evaluation, learning management systems, online quizzes, and artificial intelligence in Islamic Religious Education. However, many discussions still treat written, oral, and performance tests as separate instruments rather than as complementary sources of evidence. Studies on authentic assessment in Islamic Religious Education emphasize the need to assess cognitive, affective, and psychomotor domains (Kusnadi, 2018; Marfuah & Febriza, 2019), while studies on digital assessment emphasize the usefulness of digital platforms for efficiency and documentation (Regita, Al-Husein, Alam, & Inayati, 2024; Subehi & Sriyanto, 2021). A discernible lacuna persists, specifically the absence of a comprehensive reconstruction model that integrates written tests, oral tests, and performance tests within a single authentic assessment system that is pertinent to the digital and generative AI era.

The novelty of this article lies in the development of an integrative assessment model for Islamic Religious Education based on the triangulation of written, oral, and performative evidence. Unlike previous studies that tend to discuss authentic assessment, digital evaluation, or test types separately, this article reconstructs written tests, oral tests, and performance tests as a unified assessment system. This model is designed to strengthen validity, authenticity, ethical control, and pedagogical relevance, in assessing Islamic Religious Education in the digital and generative AI era.

This article aims to first, explain the conceptual foundation of test instruments in Islamic Religious Education; second, analyze the roles and limitations of written, oral, and performance tests; third, examine the challenges of digital assessment and

generative AI for assessment validity; and fourth, formulate an integrative model of Islamic Religious Education assessment based on authentic assessment principles. The central argument of this article is that the reconstruction of test instruments in Islamic Religious Education should not be understood as the technical digitalization of assessment tools. It should be understood as a pedagogical effort to ensure that assessment can capture students' conceptual understanding, oral religious competence, worship practice, moral behavior, and internalization of Islamic values.

RESEARCH METHODS

This study uses library research with a descriptive-analytical approach. The object of analysis is the reconstruction of test instruments in Islamic Religious Education through the integration of written, oral, and performance tests within an authentic assessment framework. Library research was selected because the study is conceptual and reconstructive in nature. It focuses on synthesizing theories, previous studies, policy documents, and contemporary debates on digital assessment and generative artificial intelligence (Bowen, 2009; Snyder, 2019; Zed, 2008).

The data sources consist of books on educational evaluation, national education policy documents, curriculum and assessment guidelines, and scholarly articles related to Islamic Religious Education assessment, authentic assessment, digital learning, and artificial intelligence in education. Classical works on evaluation and authentic assessment were used as theoretical foundations, while recent studies were prioritized to explain contemporary challenges related to digital assessment and generative AI. Classical references are limited to foundational concepts, whereas the discussion of current assessment challenges is supported by more recent literature (Miles dkk., 2014).

The literature was selected based on four criteria. First, the relevance of the source was considered across domains such as Islamic Religious Education assessment, educational evaluation, authentic assessment, digital learning, or artificial intelligence in education. Second, a substantive contribution to the conceptual reconstruction of written, oral, or performance-based assessment instruments was requisite. Third, the provision of theoretical, empirical, policy, or methodological relevance constituted another essential criterion. Fourth, the prioritization of contemporary sources was observed when addressing issues pertaining to digital assessment and AI. Sources that did not directly contribute to assessment design, validity, authenticity, or Islamic Religious Education were excluded from the main analysis (Merriam & Tisdell, 2016; Moleong, 2018; Snyder, 2019).

Data collection was conducted through identification, selection, classification, and critical reading of relevant literature. The collected sources were grouped into five themes: educational evaluation, Islamic Religious Education assessment, written, oral, or performance-based assessment instruments; authentic assessment; and digital assessment with AI-related challenges. The analysis was conducted through thematic synthesis. This process involved reducing repetitive concepts, comparing arguments across sources, identifying conceptual gaps, and constructing an

integrative assessment model. The validity of the conceptual analysis was strengthened through source triangulation by comparing theoretical literature, policy documents, and recent empirical studies (Abdurrahman, 2024; Adlini dkk., 2022).

RESULTS AND DISCUSSION

A. Test instruments in Islamic Religious Education

Educational evaluation is a systematic process of collecting, interpreting, and using information to make decisions about learning. Within the Indonesian regulatory framework, assessment is delineated as the process of information collection and processing designed to ascertain students' learning requisites and outcomes (Republik Indonesia, 2022). Furthermore, the national assessment guideline positions assessment as an integral component of the learning cycle, rather than merely an activity relegated to the conclusion of instructional periods (Badan Standar, 2025). This orientation, consequently, appears consistent with extant assessment theory, which posits that assessment may furnish meaningful information for educators, learners, and educational institutions alike (Brookhart & Nitko, 2019; Mardapi, 2012).

In Islamic Religious Education, assessment has a broader function because the learning goals include knowledge, speech, practice, morality, and value internalization. The capacity of a student to articulate the definition of ablution may not necessarily ensure the accurate execution of the ritual ablution itself. (Andriani, Ayyubi, Apriyanti, Nurhikmah, & Rahmawati, 2025) A student who can explain prayer may not necessarily understand its ethical and spiritual function. A student who can write about honesty may not necessarily demonstrate honesty in learning behavior. Therefore, assessment in Islamic Religious Education should collect multiple forms of evidence. This is why written, oral, and performance tests need to be treated as complementary instruments rather than as isolated tools.

Written tests remain important because they can assess conceptual understanding, textual literacy, and religious reasoning. They may take the form of multiple-choice questions, short answers, essays, case analysis, digital quizzes, or reflective writing. In Islamic jurisprudence, written tests can assess students' understanding of requirements, pillars, recommended acts, and invalidating factors in worship. In morality lessons, case-based written tests can assess students' ability to make ethical judgments according to Islamic values. Digital written tests through learning management systems may also support documentation and efficiency (Regita dkk., 2024). However, written tests are limited when they are used as the only form of assessment. They tend to privilege cognitive evidence and may fail to capture oral competence, worship performance, and affective development.

Oral tests are needed because many competencies in Islamic Religious Education require direct verbal performance. These include Qur'anic recitation, pronunciation, memorization, oral explanation of religious concepts, argumentation, guided dialogue, and short religious speeches (Hidayah dkk., 2025). Oral tests allow teachers to ask follow-up questions, examine reasoning, and verify whether students truly understand what they have written. In the context of generative AI, oral tests also serve as a mechanism for checking the authenticity of students' understanding. When a

written answer appears sophisticated, an oral follow-up can reveal whether the student can explain, defend, and contextualize the response in their own words.

Performance tests are essential because Islamic Religious Education includes skills and behaviors that can only be assessed through action. These include the practice of ablution, prayer, tayammum, Qur'anic recitation, sermon delivery, religious etiquette, social care projects, and other forms of value-based action. Performance tests allow teachers to observe procedure, accuracy, attitude, discipline, and the connection between religious knowledge and practice. According to the logic of authentic assessment, students should demonstrate competence through meaningful tasks that resemble real situations (Wiggins, 1990). For Islamic Religious Education, this means that students should not only know religious teachings, but also be able to perform and reflect upon them.

The integration of written, oral, and performance tests improves assessment validity because each instrument captures a different dimension of learning. Written tests capture conceptual and analytical evidence. Oral tests capture verbal, dialogical, and authenticity-related evidence. Performance tests capture procedural, behavioral, and value-based evidence. When these three forms of evidence are triangulated, teachers can make a more balanced judgment about students' learning. This triangulation is especially important because the aims of Islamic Religious Education are multidimensional and cannot be represented by one type of score alone.

B. Digital and AI challenges in Islamic Religious Education assessment

Digital technology offers clear benefits for assessment. Learning management systems, online forms, digital quizzes, video submissions, e-portfolios, and data-based feedback tools dapat membantu pengajar mengelola asesmen secara lebih efisien. These tools can store students' work, document progress, provide faster feedback, and make assessment evidence more traceable. Studies on online assessment in Islamic Religious Education show that digital platforms can support assessment of learning, assessment for learning, and assessment as learning (Subehi & Sriyanto, 2021). Quiz-based applications may also increase interactivity and provide adaptive forms of evaluation (Magfirah dkk., 2025).

Nevertheless, the implementation of digital tools does not inherently guarantee the production of high-quality assessment. A test that only moves from paper to screen may still measure low-level memorization (O'Leary dkk., 2018). A digital quiz that only rewards speed may not capture religious reasoning. A video submission may document a performance, but it still requires a clear rubric and teacher interpretation. Therefore, the key issue is not whether a teacher uses technology, but whether technology is used to support valid, authentic, and ethical assessment. Digital tools should expand the forms of learning evidence, not reduce religious education to administratively convenient numbers (S dkk., 2025). Furthermore, the integration of artificial intelligence introduces critical hurdles, such as the potential for algorithmic bias and the fundamental limitation of technology in capturing the nuanced, spiritual dimensions of religious internalization.

Generative AI intensifies the problem of authenticity. Students can use AI to

generate essays, reflections, summaries, or answers to religious questions. The result may be grammatically polished and logically organized, but it may not show the student's own understanding. This does not mean that AI must always be banned. The AI Assessment Scale suggests that AI use should be regulated according to learning objectives and task design (Perkins dkk., 2025). Some tasks may allow AI as a brainstorming tool, while others must require independent reasoning, oral verification, or direct performance. In Islamic Religious Education, the decision must be guided by the purpose of the assessment. If the goal is to assess memorization, worship practice, moral reflection, or personal understanding, unrestricted AI use would weaken validity.

AI also raises ethical questions. These include academic integrity, bias, data privacy, unequal access, teacher readiness, and the risk of replacing human pedagogical judgment with automated scoring. UNESCO emphasizes that AI in education should be guided by a human-centered perspective, ethical awareness, pedagogical responsibility, and protection of learners (Miao & Cukurova, 2024; Miao & Holmes, 2023). These principles are particularly relevant to Islamic Religious Education because the subject deals not only with information, but also with religious values, moral formation, and spiritual sensitivity. Automated tools may help teachers manage cognitive data, but they cannot fully interpret sincerity, *adab*, moral growth, or the meaning of worship in students' lives.

Another challenge is the digital divide. Not all students have equal access to devices, stable internet connections, quiet learning spaces, or sufficient digital literacy. If digital assessment is designed without considering these differences, it may create unfair outcomes. Students with better access may appear more competent simply because they can submit polished digital products, while others may be disadvantaged by technical limitations (Koh dkk., 2023). Therefore, digital assessment in Islamic Religious Education should be designed with fairness, accessibility, and flexibility. Teachers may need to provide alternative forms of submission, combine online and offline evidence, and avoid over-reliance on digital tools that may exclude certain students.

These challenges show that digital assessment and AI do not remove the need for teacher judgment. Instead, they increase the need for more careful assessment design. Written tests must be redesigned to emphasize reasoning, contextual analysis, and reflection. Oral tests must be used to verify authenticity and depth. Performance tests must be strengthened to assess worship skills, moral action, and value-based behavior. Digital tools should support documentation and feedback, while teachers remain responsible for interpreting learning evidence in relation to the aims of Islamic Religious Education. Furthermore, practitioners must prioritize algorithmic transparency and mitigate potential biases to ensure that automated systems do not inadvertently marginalize specific cultural or theological perspectives (Rahmawati dkk., 2025).

C. Authentic assessment as an integrative framework

Authentic assessment provides an appropriate framework for reconstructing test

instruments in Islamic Religious Education because it emphasizes meaningful tasks, direct performance, clear criteria, and contextual evidence. Wiggins (1990) argues that authentic assessment differs from traditional testing because students are required to demonstrate competence through valuable tasks rather than merely select or recall answers (Wiggins, 1990). Gulikers et al. (2004) further explain that authenticity in assessment can be understood through five dimensions: the task, the physical context, the social context, the assessment result, and the criteria. These dimensions are highly relevant to Islamic Religious Education because religious learning is not only conceptual but also performative, social, ethical, and reflective (Gulikers dkk., 2004).

Authentic assessment does not mean abandoning written tests. Instead, it places written tests within a broader evidentiary system. A written answer may show that a student understands the meaning of prayer. An oral explanation may show that the student can articulate the meaning in their own language. A performance task may show that the student can perform prayer correctly and respectfully (Mukmin & Nuraini, 2024). A reflective journal may show how the student relates prayer to discipline, gratitude, and moral behavior. Each form of evidence contributes to a more complete picture of learning. This may be considered the foundational rationale for integration (Astuti & Ismail, 2025).

Rubrics are necessary to make authentic assessment transparent and accountable. Without rubrics, oral and performance tests can become subjective and inconsistent (Brookhart, 2013; Brufau Alvira dkk., 2025; Jonsson & Svingby, 2007). A rubric for Qur'anic recitation may include pronunciation, tajwid, fluency, rhythm, and adab. A rubric for prayer practice may include the sequence of movements, accuracy of recitation, calmness, discipline, and understanding of meaning. A rubric for moral reflection may include depth of reflection, relevance to Islamic values, honesty, and connection with daily behavior. Clear rubrics allow students to understand expectations and enable teachers to justify assessment decisions.

The authentic assessment framework also supports assessment as learning. Students are not only judged; they are invited to reflect on their learning. Reflection is important in Islamic Religious Education because the subject aims to build awareness, not merely performance. After completing a written test, oral explanation, or performance task, students may be asked to write or record a short reflection about what they learned, what difficulties they faced, and how the lesson relates to their daily life. This reflection strengthens the connection between knowledge, practice, and internalization.

Digital technology can strengthen authentic assessment when it is used to document evidence over time. Students may upload a video of worship practice, keep a digital portfolio of reflective journals, record oral explanations, or submit project documentation (Widiyanto & Inayati, 2023). Teachers can use this evidence to track development, provide feedback, and make assessment more transparent. However, the digital portfolio should not become a mere storage folder. It should be organized around learning outcomes, rubrics, teacher feedback, and student reflection. In this way, technology supports the pedagogical goals of assessment rather than replacing them.

Authentic assessment in Islamic Religious Education therefore requires a balanced relationship between instrument design, digital tools, and teacher interpretation. Written tests should measure understanding and reasoning. Oral tests should verify explanation and authenticity. Performance tests should assess religious practice and behavior. Reflection should connect learning with moral and spiritual meaning. Digital tools should document and support feedback. Teachers should interpret the evidence holistically. This arrangement prevents assessment from being reduced to one type of test or one digital score.

D. The reconstruction model of Islamic Religious Education test instruments

The reconstruction model proposed in this article is based on six principles: holistic, authentic, integrative, digital-supportive, ethical, and reflective. The holistic principle means that assessment should cover knowledge, oral competence, performance, attitude, and value internalization. The authentic principle means that assessment tasks should be connected to meaningful religious practices and real-life contexts. The integrative principle means that written, oral, and performance tests should be designed as complementary evidence. The digital-supportive principle means that technology should be used for documentation, feedback, and portfolio management. The ethical principle means that assessment should protect integrity, privacy, fairness, and human dignity. The reflective principle means that assessment should encourage students to connect religious knowledge with personal and social practice.

The model begins with the formulation of learning outcomes. Teachers must first clarify whether the intended competence is conceptual, verbal, performative, affective, or reflective. If the competence is conceptual, written tests may be appropriate. If the competence involves recitation or explanation, oral tests are required. If the competence involves worship or behavior, performance tests are necessary. If the competence involves internalization, reflection and observation should be included. This mapping prevents the common error of using one instrument for all competencies.

The second stage is the design of assessment tasks and rubrics. Written tasks might extend beyond mere recall through the utilization of case analysis, reasoning questions, and contextual reflection. Oral tasks should include follow-up questions that require students to explain, justify, and relate their answers to Islamic values. Performance tasks should assess procedure, accuracy, adab, and meaningful action. Rubrics should be prepared before assessment so that the criteria are clear and consistent. The rubric should also be communicated to students as part of the learning process.

The third stage is digital documentation and feedback. Teachers may use learning management systems, digital forms, video submissions, e-portfolios, or feedback sheets. However, technology should not determine the value of the assessment by itself. The teacher still needs to interpret evidence, compare performance with criteria, provide qualitative feedback, and decide follow-up learning support. This is especially important when assessing affective and spiritual

dimensions, which cannot be reduced to automated scoring.

The fourth stage is validation through triangulation. A student's written answer may be checked through oral questioning. A student's oral explanation may be connected to a performance task. A worship practice may be followed by reflective writing. A digital portfolio may be compared with teacher observation. This triangulation reduces the risk of inaccurate judgment, AI-dependent answers, or superficial performance. It also provides a more comprehensive view of students' development in Islamic Religious Education.

The fifth stage is reflective follow-up. Assessment results should be used to improve learning, not merely to fill grade reports. If students understand the concept of prayer but fail to perform it correctly, the teacher should provide practice-based feedback. If students can perform prayer but cannot explain its meaning, the teacher should strengthen conceptual reflection. If students can write good reflections but cannot explain them orally, the teacher should provide dialogical support. In this way, assessment becomes part of the learning cycle.

Stage	Main focus	Implementation	Expected output
Learning outcome mapping	Identify the assessed competence	Classify competence as conceptual, oral, affective, or reflective	Specific and valid assessment target
Instrument selection	Align test type with competence	Use written tests for concepts, oral tests for explanation, performance tests for worship/action	Relevant and assessment for instrument
Rubric development	Ensure transparent scoring	Develop criteria for accuracy, fluency, procedure, reflection, and consistency	Clear and accountable judgment
Digital support	Document evidence and feedback	Use LMS, digital quizzes, video practice, e-portfolios, and feedback sheets	Traceable learning and evidence
Triangulation and reflection	Verify authenticity and depth	Combine written, oral, performance, observation, and reflection evidence	More valid and meaningful assessment results

Table 1. Reconstruction model of Islamic Religious Education test instruments in the digital era

This model also elucidates the function of generative AI. AI may be used in limited and transparent ways, such as assisting students in the brainstorming of questions, the organization of portfolios, or the reception of formative feedback. However, when the objective involves the assessment of independent comprehension, devotional competencies, memorization, oral explication, or the internalization of values, AI utilization may necessitate restriction or concomitant verification. The teacher should communicate the permitted level of AI use before the task begins. This helps protect assessment validity and encourages ethical digital behavior.

The model can be applied to various Islamic Religious Education topics. For example, in the topic of prayer, written tests can assess students' understanding of requirements, pillars, and the meaning of prayer. Oral tests can assess their ability to explain the purpose of prayer and recite relevant texts. Performance tests can assess

the accuracy and adab of prayer practice. Reflective journals can document how students relate prayer to discipline and moral conduct. The combination of these forms of evidence produces a more valid judgment than written testing alone.

Competence	Written evidence	Oral evidence	Performance evidence	Documentation
Understanding the meaning of prayer	Essay or case analysis about the function of prayer	Explanation of the purpose and value of prayer	Respectful attitude during prayer practice	Written answer, oral record, observation note
Knowledge of the requirements and pillars of prayer	Digital quiz on requirements, pillars, and invalidating acts	Oral explanation of the sequence of prayer	Prayer practice according to Islamic jurisprudence	Quiz result, oral rubric, practice rubric
Internalizing the value of prayer	Reflective writing on prayer and discipline	Dialogue about daily challenges in maintaining prayer	Prayer habit journal or portfolio	Reflection journal, portfolio, teacher feedback

Table 2. Example of integrated assessment design for the topic of prayer

The example shows that assessment in Islamic Religious Education becomes stronger when learning evidence is layered. Written evidence provides conceptual clarity. Oral evidence verifies personal understanding. Performance evidence demonstrates religious practice. Documentation and reflection make development visible over time. This arrangement also provides a response to the problem of generative AI. Even if students use AI to prepare written explanations, they still need to explain, perform, and reflect in ways that reveal their actual competence.

The teacher's role remains central in this reconstruction. Technology can help organize tests, store recordings, display progress, and provide initial feedback. AI can support formative learning and administrative efficiency. However, teachers are still needed to interpret evidence in context, evaluate adab and moral development, identify learning difficulties, and provide human feedback. In Islamic Religious Education, this role cannot be replaced by automated scoring because the subject includes moral, spiritual, and relational dimensions. A valid assessment system must therefore combine technological support with professional pedagogical judgment.

The proposed reconstruction has practical implications. First, teachers should avoid relying only on written tests when assessing Islamic Religious Education. Second, schools should provide rubrics and digital documentation systems that support oral and performance evidence. Third, assessment policy should define the ethical use of AI in student work. Fourth, teacher training should include authentic assessment design, digital portfolio management, and AI literacy. Fifth, assessment results should be used for feedback and learning improvement rather than only for administrative grading. Through these implications, the reconstruction model can help Islamic Religious Education remain pedagogically valid in the digital era.

CONCLUSION

The reconstruction of test instruments in Islamic Religious Education may be necessitated because the subject aims to develop knowledge, oral competence, worship practice, moral behavior, and the internalization of Islamic values. Written tests remain useful for assessing conceptual understanding, religious reasoning, and textual literacy. However, they are insufficient when used as the only evidence of learning, especially in the digital and generative AI era, where written answers can be produced or polished with technological assistance.

This article proposes an integrative model that combines written, oral, and performance tests within an authentic assessment framework. Written tests assess conceptual and analytical evidence. Oral tests verify depth, fluency, argumentation, and authenticity of understanding. Performance tests assess worship skills, religious behavior, and value-based action. Digital technology supports documentation, feedback, and portfolio management, while teachers remain the central interpreters of students' religious learning evidence.

The model emphasizes six principles: holistic, authentic, integrative, digital-supportive, ethical, and reflective. Its main contribution is the triangulation of written, oral, and performative evidence to strengthen validity, authenticity, and pedagogical relevance in Islamic Religious Education assessment. Future empirical studies are needed to test the model in classroom practice, examine its effect on assessment quality, and explore how teachers can implement it across different school levels and digital infrastructure conditions.

BIBLIOGRAPHY

- Abdurrahman. (2024). Metode Penelitian Kepustakaan dalam Pendidikan Islam. *Adabuna: Jurnal Pendidikan dan Pemikiran*, 3(2), 102–113. <https://doi.org/10.38073/adabuna.v3i2.1563>
- Adlini, M. N., Dinda, A. H., Yulinda, S., Chotimah, O., & Merliyana, S. J. (2022). Metode Penelitian Kualitatif Studi Pustaka. *Edumaspul: Jurnal Pendidikan*, 6(1), 974–980. <https://doi.org/10.33487/edumaspul.v6i1.3394>
- Astuti, M., & Ismail, F. (2025). Reconstruction of Integrative Methods in Merdeka Curriculum for Islamic Education Subjects. *Al-Hayat: Journal of Islamic Education*, 9(2), 369–390. <https://doi.org/10.35723/ajie.v9i2.67>
- Badan Standar, K., dan Asesmen Pendidikan. (2025). *Panduan Pembelajaran dan Asesmen Pendidikan Anak Usia Dini, Jenjang Pendidikan Dasar, dan Jenjang Pendidikan Menengah* (Edisi revisi 2025). Kementerian Pendidikan Dasar dan Menengah. <https://kurikulum.kemendikdasmen.go.id/service/download.php?id=120&kategori=rujukan>
- Bowen, G. A. (2009). Document Analysis as a Qualitative Research Method. *Qualitative Research Journal*, 9(2), 27–40. <https://doi.org/10.3316/QRJ0902027>

- Brookhart, S. M. (2013). *How to Create and Use Rubrics for Formative Assessment and Grading*. <https://doi.org/10.1002/9781119191681>
- Brookhart, S. M., & Nitko, A. J. (2019). *Educational Assessment of Students* (8th edition). Pearson. <https://www.pearsonhighered.com/assets/preface/0/1/3/4/0134806972.pdf>
- Brufau Alvira, N., Bannister, P., & Santamaría Urbieta, A. (2025). Validating the PANDORA GENAI Susceptibility Rubric for Higher Education Assessment: A Field Test of All Translation and Interpreting BA Assignments. *Higher Education Quarterly*, 79(4), e70056. <https://doi.org/10.1111/hequ.70056>
- Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A Five-Dimensional Framework for Authentic Assessment. *Educational Technology Research and Development*, 52(3), 67–86. <https://doi.org/10.1007/BF02504676>
- Hidayah, S. N., Shofiyuddin, M., & Hasanah, U. (2025). Toward a Religious-Oriented Model of Speaking Assessment: Evidence from English Teachers in Islamic Secondary Schools. *Evaluasi: Jurnal Manajemen Pendidikan Islam*, 9(2), 379–397. <https://doi.org/10.32478/otyg8x34>
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130–144. <https://doi.org/10.1016/j.edurev.2007.05.002>
- Kementerian Pendidikan Dasar dan Menengah. (2026). *Capaian Pembelajaran Pendidikan Agama Islam dan Budi Pekerti Fase D*. Guru Kemendikdasmen. <https://guru.kemendikdasmen.go.id/kurikulum/referensi-penerapan/capaian-pembelajaran/sd-sma/pendidikan-agama-islam-dan-budi-pekerti/fase-d/>
- Koh, K., Lock, J., & DePass, C. (2023). CPI Special Issue: “All That Glitters is Not Gold: Culturally Responsive Online Assessment and Pedagogy in Uncertain Times.” *Cultural and Pedagogical Inquiry*, 15(1), 1–9. <https://doi.org/10.18733/cpi29704>
- Kusnadi, D. (2018). Efektivitas Implementasi Kebijakan Penilaian Autentik pada Pembelajaran Pendidikan Agama Islam. *Ta'dibuna: Jurnal Pendidikan Islam*, 7(1), 20–41. <https://doi.org/10.32832/tadibuna.v7i1.1348>
- Magfirah, I., Afiyati, F., & Bashith, A. (2025). Transformasi Evaluasi Pembelajaran PAI berbasis Digital: Optimalisasi Media Quizizz sebagai Alat Ukur Adaptif. *Al-Qalam: Jurnal Kajian Islam dan Pendidikan*, 17(1), 435–444. <https://doi.org/10.47435/al-qalam.v17i1.3949>
- Mardapi, D. (2012). *Pengukuran, Penilaian, dan Evaluasi Pendidikan*. Nuha Medika. https://perpus-bbgtkjateng.kemendikdasmen.go.id/index.php?id=1507&p=show_detail
- Marfuah, A., & Febriza, F. (2019). Penilaian Autentik pada Pembelajaran Pendidikan Agama Islam (PAI) di Sekolah dan Perguruan Tinggi. *FONDATIA*, 3(2), 35–58. <https://doi.org/10.36088/fondatia.v3i2.301>
- Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative Research: A Guide to Design and Implementation* (4th ed.). Jossey-Bass.
- Miao, F., & Cukurova, M. (2024). *AI competency framework for teachers* (Nos. 978-92-3-100707-1). UNESCO. <https://doi.org/10.54675/ZJTE2084>

- Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research* (Nos. 978-92-3-100612-8). UNESCO. <https://doi.org/10.54675/EWZM9535>
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative Data Analysis: A Methods Sourcebook* (3rd ed.). SAGE Publications.
- Moleong, L. J. (2018). *Metodologi Penelitian Kualitatif* (ed. revisi). PT Remaja Rosdakarya.
- Mukmin, M., & Nuraini, N. (2024). Integrasi Penilaian Tes dan Non-Tes dalam Pendidikan Agama Islam: Menuju Evaluasi Holistik untuk Pembelajaran Berkelanjutan. *Journal of Instructional and Development Researches*, 4(5), 370–379. <https://doi.org/10.53621/jider.v4i5.384>
- Muhammad Yusuf Hilman, & Iskandar Yusuf. (2025). The Influence of Contextual Learning Strategies on Student Learning Outcomes on Moral Creed Material at Mts Al-Muttaqien Balikpapan. *Responsive: Jurnal Manajemen Pendidikan Islam*, 1(1), 25–33. <https://doi.org/10.61166/responsive.viii.5>
- O’Leary, M., Scully, D., Karakolidis, A., & Pitsia, V. (2018). The state-of-the-art in digital technology-based assessment. *European Journal of Education*, 53(2), 160–175. <https://doi.org/10.1111/ejed.12271>
- Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A Framework for Ethical Integration of Generative AI in Educational Assessment. *Journal of University Teaching and Learning Practice*, 21(6). <https://doi.org/10.53761/q3azde36>
- Perkins, M., Roe, J., & Furze, L. (2025). Reimagining the Artificial Intelligence Assessment Scale: A Refined Framework for Educational Assessment. *Journal of University Teaching and Learning Practice*, 22(7). <https://doi.org/10.53761/irm4y757>
- Rahmawati, A. R., Wulandari, R., Sadaruddin, M. A. S., Mustofa, T. A., & Chedimae, H. (2025). Optimization of Artificial Intelligence in Islamic Religious Education: Opportunities and Challenges in Learning Evaluation. *Suhuf: International Journal of Islamic Studies*, 37(2), 274–287. <https://doi.org/10.23917/suhuf.v37i2.11015>
- Regita, R., Al-Husein, M. H., Alam, M. N., & Inayati, N. L. (2024). Penerapan Evaluasi Pembelajaran PAI dan Bahasa Arab dengan Tes Tertulis melalui Media Scola Digital Learning Manajemen System (LMS) di SMA Muhammadiyah Al-Kautsar PK Kartasura. *ALSYS*, 4(1), 88–98. <https://doi.org/10.58578/alsys.v4i1.2471>
- Republik Indonesia. (2022). *Peraturan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Nomor 21 Tahun 2022 tentang Standar Penilaian Pendidikan pada Pendidikan Anak Usia Dini, Jenjang Pendidikan Dasar, dan Jenjang Pendidikan Menengah*. <https://peraturan.bpk.go.id/Details/224425/permendikbudriset-no-21-tahun-2022>
- S, A. M., Asmiatin, A., Khoiri, K., Fattahuddin, M., & Halik, A. (2025). Construction of Theory and Concept of Digital-Based PAI Learning Assessment. *Journal of Pedagogi*, 2(2), 35–43. <https://doi.org/10.62872/aqcdmh98>

- Snyder, H. (2019). Literature Review as a Research Methodology: An Overview and Guidelines. *Journal of Business Research*, 104, 333–339. <https://doi.org/10.1016/j.jbusres.2019.07.039>
- Subehi, R., & Sriyanto, S. (2021). Implementasi Assessment Of, For, dan As Learning dalam Pembelajaran Daring PAI di SMPN 8 Purwokerto. *Alhamra: Jurnal Studi Islam*, 2(2), 111–122. <https://doi.org/10.30595/ajsi.v2i2.10632>
- Uswatun Hasanah, & Nurlaela Rusdin. (2025). The Influence of The Application of The Discussion Method on The Learning Motivation of Intensive PAI Semester IV Students in PAI Learning Planning and Design Courses. *Kasyafa: Jurnal Pendidikan Agama Islam*, 2(2), 149–156. <https://doi.org/10.61166/kasyafa.v2i2.9>
- Widiyanto, A., & Inayati, N. L. (2023). Penerapan Evaluasi Pembelajaran Tes dan Non-Tes Pendidikan Agama Islam di Sekolah Menengah Kejuruan. *Munaddhomah: Jurnal Manajemen Pendidikan Islam*, 4(2), 307–316. <https://doi.org/10.31538/munaddhomah.v4i2.439>
- Wiggins, G. (1990). The Case for Authentic Assessment. *Practical Assessment, Research, and Evaluation*, 2(2). <https://doi.org/10.7275/ffb1-mm19>
- Zed, M. (2008). *Metode Penelitian Kepustakaan*. Yayasan Pustaka Obor Indonesia. <https://books.google.co.id/books?id=zG9sDAAAQBAJ>