Ar-Rayyan: Journal of Islamic Education
Volume 3 Nomor 1 Jan-Juni (2026) E-ISSN 3064-3031
DOI: https://doi.org/10.67899/ar.v3i1.226

Submitted: 15-02-2026 Accepted:15-04-2026 Published:15-06-2026

Authentic Assessment in Islamic Religious Education: Measuring Students' Holistic Competencies through Performance-Based Evaluation

Mahmoud Abdullah[1*], Huzaiman Hushin[2], Salma Begum[3], Yousif Mahmadov[4], Abdallah El-Khatib[5]

[1]Universiti Islam Sultan Sharif Ali, Brunei Darussalam
[2]Universiti Kebangsaan Malaysia, Malaysia
[3]International Islamic University Islamabad, Pakistan
[4]Azerbaijan State Pedagogy University, Baku, Azerbijan
[5]Azerbaijan State Pedagogy University, Baku, Azerbijan

*Correspondence email: mahmoudabddullah@unissa.edu.bn

Abstract

Islamic Religious Education (IRE) requires an assessment paradigm capable of measuring not only students' religious knowledge but also their ability to practice Islamic values, demonstrate moral character, communicate responsibly, and perform religious practices appropriately. This study aims to examine how authentic and performance-based assessment can measure students' holistic competencies in IRE. A qualitative systematic literature review was employed by synthesizing journal articles published between 2016 and 2025 concerning authentic assessment, performance-based assessment, competency measurement, and IRE evaluation. The findings indicate that performance-based authentic assessment provides stronger evidence of students' cognitive, affective, spiritual, social, and psychomotor competencies than conventional written tests alone. Effective assessment requires authentic tasks, explicit criteria, analytic rubrics, multiple sources of evidence, continuous feedback, self- and peer-assessment, and contextualized religious practices. The synthesis also identifies major implementation challenges, including teacher workload, assessor subjectivity, limited assessment literacy, large class sizes, and difficulties in assessing internal spiritual dimensions. The novelty of this study lies in proposing an integrated holistic performance-assessment framework for IRE that connects knowledge, religious practice, character, social responsibility, reflection, and real-life application within one evidence-based assessment cycle.
Keywords: Authentic Assessment, Religious Character, Assessment Rubric.

Introduction

Islamic Religious Education (IRE) occupies a distinctive position within educational systems because its learning objectives extend beyond the acquisition of factual religious knowledge. IRE is expected to facilitate students' understanding of Islamic teachings, internalization of religious values, development of moral dispositions, and ability to translate religious principles into responsible action. Consequently, assessment in IRE should not merely determine whether students can remember

DOI: https://doi.org/10.67899/ar.v3i1.226

definitions, verses, concepts, or legal principles. It should also provide evidence concerning how students interpret religious knowledge, perform religious practices, interact with others, and demonstrate ethical behavior in authentic situations. Research on authentic assessment has consistently emphasized that assessment becomes more meaningful when students are required to demonstrate what they know and can do in contexts resembling real-life situations (Villarroel et al., 2018; Wiewiora & Kowalkiewicz, 2019). This perspective is particularly relevant to IRE because religious education inherently involves the relationship between knowledge, values, attitudes, and practice. Thus, the assessment of IRE needs to move from a predominantly test-oriented paradigm toward a more comprehensive evidence-based model.

Conventional assessment practices in education have historically placed considerable emphasis on written examinations because they are relatively easy to administer, score, and report. Such examinations are useful for measuring certain dimensions of conceptual knowledge, but their capacity to capture complex competencies is limited. In IRE, for example, a student may obtain a high score on a written examination about prayer while experiencing difficulties in demonstrating the correct sequence, concentration, discipline, or social meaning of prayer. Similarly, a student may accurately explain the concept of honesty while failing to demonstrate honest behavior when faced with an authentic moral dilemma. Studies of authentic assessment have shown that conventional tests can become disconnected from the complex competencies that educational programs intend to develop (Braun et al., 2020; Jopp, 2020). The problem is therefore not that written tests are inherently inappropriate, but that they should not constitute the sole source of evidence for holistic learning. In IRE, the gap between knowing religious concepts and embodying them is particularly important because the ultimate educational objective includes value internalization. Therefore, authentic performance evidence is necessary to complement cognitive testing.

Authentic assessment offers a conceptual response to this problem by placing students in meaningful tasks through which learning can be observed in action. Authenticity involves more than simply creating an attractive practical activity; it requires alignment between the assessment task, learning outcomes, real-world contexts, and criteria used to judge performance. Villarroel et al. (2018) explain that authentic assessment can be understood through dimensions such as realism, cognitive challenge, evaluative judgment, and relevance to situations beyond the classroom. Wiewiora and Kowalkiewicz (2019) similarly found that authentic assessment can contribute to the development of competencies and identity when assessment tasks require students to engage with realistic problems. In IRE, authenticity can be expressed through activities such as performing ablution, leading prayer, reciting and interpreting Qur'anic passages, conducting charitable activities, resolving ethical cases, producing religious communication materials, or participating in community service. Such activities allow teachers to observe the integration of knowledge, skills, attitudes, and values. Therefore, authenticity provides a bridge between religious learning in the classroom and students' actual lives. The relevance of this approach becomes increasingly significant when educational success is defined in terms of the capacity to apply learning responsibly rather than simply reproduce information.

The need for authentic assessment is also strengthened by the changing orientation of curriculum policy in Indonesia. The transition from the 2013 Curriculum toward the Merdeka Curriculum has encouraged teachers to place greater emphasis on meaningful learning, competency development, feedback, and differentiated learning experiences.

DOI: https://doi.org/10.67899/ar.v3i1.226

Within IRE, this shift requires assessment to provide information that can be used to improve learning rather than merely classify students at the end of instruction. Aisyah et al. (2024) reported that authentic assessment in IRE under the Merdeka Curriculum involves cognitive, affective, and psychomotor domains and can be connected with feedback for learning improvement. This finding indicates that assessment should operate as part of the learning process rather than as an isolated administrative activity. The assessment process can help teachers identify misconceptions, weaknesses in religious practice, and behavioral tendencies that require further guidance. It can also help students understand what constitutes quality performance and how they can improve. Consequently, authentic assessment is consistent with a curriculum orientation that emphasizes student competency and continuous development.

IRE possesses several characteristics that make authentic assessment especially appropriate. Religious learning contains numerous competencies that are inherently practical, including Qur'anic recitation, worship procedures, ethical decision-making, communication, social responsibility, and religious problem-solving. Many of these competencies cannot be adequately inferred from students' answers to multiple-choice questions. For instance, knowing the requirements of prayer is conceptually different from being able to perform prayer correctly and consistently. Similarly, understanding the meaning of charity is different from demonstrating empathy and responsible generosity in a social situation. Tamrin et al. (2021) found that authentic assessment in IRE can involve practical demonstrations, observation, non-test techniques, and performance-based activities, although teachers also encounter challenges in scoring and managing large numbers of students. These findings demonstrate that the practical nature of IRE creates a strong pedagogical justification for performance-based assessment. The assessment method should therefore correspond to the nature of the competency being measured. In this sense, authentic assessment is not simply an alternative assessment technique but a logical consequence of the educational aims of IRE.

Holistic competency in IRE should be understood as a multidimensional construct rather than a simple combination of academic scores. At minimum, it includes cognitive competence, psychomotor competence, affective disposition, spiritual awareness, social competence, and the ability to apply Islamic principles in authentic situations. Cognitive competence concerns understanding and reasoning about Islamic concepts, while psychomotor competence involves performing religious practices accurately. Affective competence concerns attitudes, dispositions, responsibility, empathy, and ethical sensitivity, whereas spiritual competence concerns the internal orientation toward worship, meaning, and relationship with God.(Nabriz et al., 2025) Recent scholarship on IRE assessment increasingly emphasizes the need to integrate cognitive, affective, and psychomotor domains instead of treating them as unrelated assessment categories (Baroroh & Hamani, 2022; Imanudin, 2023). The holistic perspective therefore requires evidence that demonstrates relationships among these domains. In practical terms, the strongest evidence is obtained when students explain, perform, reflect, and apply simultaneously. This approach allows assessment to represent the complexity of religious learning more faithfully.

Performance-based evaluation provides one of the most appropriate mechanisms for obtaining such evidence. Performance assessment asks students to demonstrate a competency through observable actions, products, processes, or performances rather than merely selecting or recalling answers. Braun et al. (2020) emphasize that performance assessment can provide evidence concerning complex learning objectives when tasks and

DOI: https://doi.org/10.67899/ar.v3i1.226

scoring processes are appropriately designed. In IRE, performance tasks can include demonstrating religious rituals, conducting a simulated counseling session, analyzing an ethical dilemma, delivering a khutbah or religious presentation, organizing a social-service project, or producing a reflective portfolio. These tasks can reveal both the final product and the process through which students reach it. Performance assessment can also encourage students to integrate multiple forms of knowledge rather than compartmentalizing learning. However, performance assessment requires clear criteria and systematic scoring to maintain fairness. Thus, the quality of the rubric and assessment design is as important as the performance task itself.

Rubrics constitute an essential component of performance-based assessment because they transform broad educational expectations into observable criteria. Without a clear rubric, teachers may unintentionally score students according to personal impressions, creating inconsistency and reducing assessment credibility. Research on performance assessment indicates that validity and reliability depend substantially on the quality of task design, scoring criteria, rater preparation, and evidence interpretation (Conn et al., 2020; Yudha, 2020). In IRE, rubric criteria might include conceptual accuracy, procedural correctness, communication, ethical reasoning, cooperation, responsibility, reflection, and application of Islamic values. Such criteria should be accompanied by performance descriptors that distinguish different levels of achievement. Rubrics can also communicate expectations to students before they perform the task, thereby making assessment more transparent. Furthermore, students can use rubrics for self-assessment and peer feedback. Therefore, a well-designed rubric functions simultaneously as a measurement instrument, learning guide, and feedback mechanism.

Another important issue concerns the relationship between assessment and character formation. IRE is not intended merely to transfer religious information but also to contribute to the formation of morally responsible individuals. If assessment focuses exclusively on memorization, students may learn to optimize examination performance without necessarily internalizing the values represented by the subject matter. Authentic tasks can create opportunities to observe how students respond to ethical situations, cooperate with peers, demonstrate responsibility, and reflect on their behavior. Faizah et al. (2019) demonstrate that authentic assessment can be designed to capture character-related outcomes, indicating that assessment can extend beyond academic achievement. In the context of IRE, this is particularly significant because character development is closely connected with the practical expression of religious values. Nevertheless, teachers must distinguish between observable behavioral evidence and assumptions about students' internal spirituality. Assessment should evaluate demonstrated behaviors, reflections, and contextual evidence rather than claim direct measurement of students' inner faith. This distinction strengthens the ethical and methodological credibility of holistic IRE assessment.

Authentic assessment can also promote student engagement because it gives students opportunities to participate actively in the assessment process. Instead of experiencing assessment as something imposed by the teacher, students can become participants who understand criteria, evaluate their own work, provide peer feedback, and reflect on progress. Jopp (2020) found that technology-supported authentic assessment can enhance engagement, deepen understanding, stimulate creativity, and reduce opportunities for plagiarism. These advantages are relevant to IRE because students can create digital religious projects, reflective portfolios, community-service documentation, podcasts, presentations, or multimedia interpretations of Islamic teachings.(Bakr et al.,

DOI: https://doi.org/10.67899/ar.v3i1.226

2025) The key issue is not technology itself but the pedagogical authenticity of the task. Digital tools should support meaningful evidence of competency rather than merely produce attractive products. Consequently, technology can broaden the range of evidence available to teachers while maintaining the central principles of authentic assessment. This approach is especially valuable for contemporary students who increasingly learn and communicate through digital environments.

Despite its advantages, authentic assessment is not automatically superior simply because it is practical. Poorly designed performance tasks can generate ambiguous evidence, excessive teacher workload, and inconsistent scoring. Tamrin et al. (2021) identified difficulties related to the number of students, personal bias in scoring, and challenges in managing performance scores. Similar concerns are discussed in performance-assessment literature, particularly regarding standardization, validity, and rater judgment (Braun et al., 2019; Conn et al., 2020). These challenges indicate that authentic assessment must be supported by systematic planning. Teachers need training in task design, rubric construction, observation, evidence collection, and feedback. Schools also need mechanisms for moderation so that assessment standards remain reasonably consistent across teachers and classes. Therefore, the implementation of authentic assessment requires institutional support rather than placing the entire responsibility on individual teachers. The challenge should be addressed through better assessment literacy, not by abandoning performance-based evaluation.

Teacher competence is consequently a central determinant of assessment quality. Teachers must understand not only what to assess but also how to construct valid evidence of learning. Baroroh and Hamani (2022) found that PAI teachers develop authentic assessments through stages such as mapping competencies and indicators, selecting appropriate assessment types, preparing instruments, constructing scoring guidance, and implementing the assessment. Their study also indicates that some technically demanding stages, such as extensive validity and reliability testing, may be difficult for teachers to implement in everyday practice. This suggests a need for practical assessment frameworks that are rigorous without becoming unnecessarily burdensome. Teacher professional development should therefore focus on the connection between learning outcomes, authentic tasks, observable indicators, and scoring criteria. Teachers should also learn how to use assessment results to improve instruction. The quality of authentic assessment ultimately depends on teachers' ability to transform educational goals into credible evidence of student performance.

Another concern involves fairness and inclusivity. Performance tasks may advantage students who have greater access to resources, stronger communication skills, or greater confidence in public performance. If authentic assessment is not carefully designed, it can reproduce inequalities rather than provide fair evidence of learning. Assessment criteria should therefore focus on the intended competency rather than irrelevant characteristics. Tasks should be adaptable to different learner needs while preserving the core learning outcome.(Hushin et al., 2025) This principle is consistent with contemporary assessment approaches that emphasize valid interpretation of evidence and appropriate accommodation. In IRE, teachers should be particularly attentive to differences in students' prior religious experiences, linguistic abilities, family contexts, and opportunities to practice religious activities. Authentic assessment should recognize these differences without lowering essential competency standards. Thus, holistic assessment must be accompanied by an equally strong commitment to fairness and educational inclusion.

DOI: https://doi.org/10.67899/ar.v3i1.226

Assessment in IRE should also incorporate feedback because the purpose of assessment extends beyond measurement. Feedback becomes educationally valuable when it helps students understand the gap between their current performance and the desired standard. Watling and Ginsburg (2019) emphasize that assessment and feedback are closely connected to learning processes, while Aisyah et al. (2024) identify feedback as an important dimension of authentic IRE assessment under the Merdeka Curriculum. In performance-based assessment, feedback can address not only whether an answer is correct but also how the student performed, why a particular action was appropriate or inappropriate, and what should be improved. For example, feedback on prayer performance can address procedural accuracy, concentration, and discipline rather than simply assigning a numerical score. Feedback on an ethical project can discuss reasoning, empathy, evidence, and consistency with Islamic principles. Therefore, authentic assessment becomes formative when performance evidence is converted into actionable guidance. This transformation is crucial for making assessment a component of learning rather than a terminal judgment.

The growing literature on IRE assessment nevertheless reveals a research gap. Existing studies have examined authentic assessment implementation, teacher practices, assessment instruments, and the three conventional domains of cognition, affect, and psychomotor performance. However, fewer studies have explicitly integrated spiritual awareness, social responsibility, practical religious competence, reflection, and contextual application into a unified performance-assessment framework. Recent studies have begun to address holistic assessment, but the conceptual connection among different forms of evidence remains insufficiently systematized. This gap is important because IRE competencies are interconnected rather than independent.(Nabriz et al., 2026) A student's ability to perform a religious practice, for example, should ideally be accompanied by understanding of its meaning, ethical implications, and relevance to everyday life. The present study therefore synthesizes contemporary evidence to construct a more integrated framework. Its contribution lies in emphasizing assessment as a multidimensional evidence system rather than a collection of separate scores.

The novelty of this study lies in positioning performance-based authentic assessment as an integrated mechanism for measuring holistic IRE competencies. Rather than treating cognitive, affective, psychomotor, spiritual, and social outcomes as separate assessment components, the proposed approach connects them through authentic tasks and evidence portfolios. This perspective is supported by findings that authentic assessment can promote competency development and meaningful learning when students engage in realistic tasks (Villarroel et al., 2018; Nyinge, 2025). The framework further emphasizes triangulation among teacher observation, student products, performance demonstrations, self-reflection, peer assessment, and contextual evidence. Such triangulation can reduce dependence on a single score and provide a more defensible interpretation of learning. The approach also recognizes that spirituality should be assessed indirectly through observable manifestations, reflection, and consistent practice rather than presumed internal states. Accordingly, the study seeks to contribute both theoretically and practically to contemporary IRE assessment discourse.

Based on the preceding discussion, this study addresses several interrelated questions concerning the conceptual foundations, implementation requirements, and potential outcomes of authentic performance-based assessment in IRE. It examines what constitutes holistic competency in IRE, how authentic performance tasks can generate credible evidence, what assessment instruments and procedures are most appropriate, and

DOI: https://doi.org/10.67899/ar.v3i1.226

what challenges teachers are likely to encounter. The study also considers how authentic assessment can support feedback, character development, religious practice, and meaningful application of Islamic knowledge. A systematic synthesis of recent literature is used to identify patterns across empirical and conceptual studies published during the last decade. The analysis is intended to produce a framework that is sufficiently rigorous for academic discussion while remaining practical for teachers. Ultimately, the study argues that assessment quality should be judged by the quality of evidence it produces about meaningful learning. Authentic performance-based evaluation is therefore proposed as an important pathway toward more comprehensive and educationally useful assessment in IRE.

Theoretical Foundation

Authentic assessment is fundamentally concerned with the quality and relevance of evidence used to make judgments about student learning. It differs from traditional assessment primarily through the nature of the task, the context in which performance occurs, and the complexity of the competencies being demonstrated. Villarroel et al. (2018) conceptualize authentic assessment through a blueprint that emphasizes realism, cognitive challenge, and evaluative judgment. This conceptualization implies that authenticity cannot be reduced to the use of practical activities alone. A task becomes authentic when students are asked to perform meaningful work that resembles how knowledge and skills are used beyond artificial examination settings. Wiewiora and Kowalkiewicz (2019) further demonstrate that authentic assessment can contribute to the development of competencies and professional identity. In IRE, the theory suggests that students should be assessed through situations in which religious knowledge becomes functional and consequential. Thus, authentic assessment provides the conceptual foundation for connecting religious learning with actual practice.

The theory of performance assessment complements authentic assessment by focusing on observable demonstrations of competence. Performance assessment assumes that some learning outcomes are best evidenced by requiring students to perform, create, explain, demonstrate, or solve rather than select answers. Braun et al. (2020) argue that performance assessments are particularly valuable for complex constructs such as critical thinking because they can elicit richer evidence than conventional tests. Braun et al. (2019) similarly explain that well-designed performance tasks can provide evidence concerning what students know and can do with respect to complex learning objectives. For IRE, performance assessment is theoretically appropriate because religious education includes procedural and behavioral competencies that must be demonstrated in action. A student who knows the theory of ablution but cannot perform it correctly has not demonstrated complete mastery of the competency. Performance assessment therefore operationalizes the principle that competency includes the ability to apply knowledge.

Competency-based education provides another theoretical foundation for authentic assessment. Competency is generally understood as the integrated capacity to mobilize knowledge, skills, attitudes, and other resources to respond effectively to situations. From this perspective, learning outcomes cannot be represented adequately through isolated factual recall. Nyinge (2025) found a significant relationship between authentic assessment tools and generic competencies among student teachers, indicating that portfolios, practical work, projects, and teaching practice can contribute evidence of broader competence. This theoretical position is highly compatible with IRE because Islamic education seeks to integrate knowledge and action. Religious competence

DOI: https://doi.org/10.67899/ar.v3i1.226

includes knowing what is right, understanding why it is right, being able to perform it correctly, and demonstrating willingness to enact it responsibly. Authentic performance tasks therefore provide a practical mechanism for observing competency integration. The theoretical implication is that IRE assessment should be organized around meaningful competencies rather than isolated content units.

Holistic assessment theory emphasizes that learning should be understood through multiple dimensions. Within IRE, the traditional cognitive-affective-psychomotor model provides an important starting point, but contemporary interpretations can expand it to include spiritual, social, ethical, and reflective dimensions. Baroroh and Hamani (2022) emphasize the importance of integrating cognitive, affective, and psychomotor assessment in IRE. Imanudin (2023) similarly argues that authentic assessment can evaluate both learning processes and outcomes and can connect knowledge with everyday application. The holistic perspective therefore rejects the assumption that one score can adequately represent religious learning. Instead, multiple forms of evidence are assembled to create a multidimensional profile. Such a profile can reveal strengths and weaknesses that would remain hidden under a purely cognitive assessment model. Consequently, holistic assessment is not simply an increase in the number of assessment instruments but a change in the conceptualization of what counts as learning.

The cognitive dimension remains essential even within a holistic assessment framework. Authentic assessment should not be interpreted as abandoning tests or conceptual measurement. Students must still understand Islamic concepts, sources, principles, arguments, and contextual interpretations before they can apply them effectively.(Begum et al., 2025) The theory of constructive alignment requires learning outcomes, instructional activities, and assessment evidence to be systematically connected. Aisyah et al. (2024) demonstrate that authentic IRE assessment under the Merdeka Curriculum can combine test and non-test techniques across cognitive, affective, and psychomotor domains. Therefore, cognitive assessment should be retained but positioned within a broader evidence system. Higher-order cognitive processes such as analysis, evaluation, interpretation, and ethical reasoning are especially important for authentic IRE tasks. This means that authentic assessment should measure not only whether students know Islamic teachings but also whether they can reason with and apply them appropriately.

The psychomotor dimension has particular importance because numerous IRE competencies involve correct physical performance. Religious practices such as ablution, prayer, Qur'anic recitation, funeral procedures, and other forms of worship require procedural knowledge and motor execution. Tamrin et al. (2021) found that authentic assessment in IRE includes performance and practical activities, while Baroroh and Hamani (2022) emphasize the integration of skills within authentic assessment. This research supports the proposition that performance should be directly observed rather than inferred from written answers. A performance rubric can specify procedural accuracy, sequence, fluency, safety, discipline, and contextual appropriateness. Repeated performance assessment can also provide evidence of development over time rather than a single snapshot. Thus, psychomotor assessment transforms religious practice from an assumed outcome into observable evidence. The theoretical implication is that practice-based competencies require practice-based assessment.

The affective dimension concerns attitudes, dispositions, values, and behavioral tendencies that influence how students engage with learning and other people. In IRE, affective outcomes may include responsibility, empathy, respect, honesty, discipline,

DOI: https://doi.org/10.67899/ar.v3i1.226

tolerance, cooperation, and willingness to reflect on moral decisions. Faizah et al. (2019) demonstrate that authentic assessment can be designed to evaluate character, suggesting that affective outcomes can be approached through systematic observation and evidence collection.(Begum et al., 2025) However, affective assessment requires careful operationalization because teachers cannot directly observe an internal attitude. Observable indicators and repeated evidence should therefore be used. A single incident should not automatically be interpreted as a stable character trait. In this framework, reflection, peer feedback, teacher observation, and longitudinal evidence can complement one another. Therefore, affective assessment should be interpreted cautiously and developmentally rather than as an absolute judgment of character.

Spiritual competence presents a more complex theoretical issue because spirituality includes internal dimensions that cannot be directly observed. IRE teachers may observe religious practices, ethical decisions, expressions of gratitude, reflection, and participation in worship, but these behaviors should not automatically be treated as definitive evidence of a student's inner faith.(Hushin et al., 2025) The assessment of spiritual learning should therefore focus on demonstrable learning outcomes such as understanding spiritual meanings, articulating personal reflections, recognizing ethical implications, and consistently participating in appropriate practices. This approach preserves the educational importance of spirituality while avoiding unwarranted claims about students' internal states. Contemporary holistic assessment literature supports the need to distinguish between observable evidence and interpretations of complex constructs (Braun et al., 2020; Conn et al., 2020). The theoretical consequence is that spiritual assessment requires epistemic humility. Teachers should report evidence of learning and behavior rather than claim certainty about personal faith. This distinction strengthens both validity and ethical responsibility.

Social competence is another important component of holistic IRE assessment. Islamic teachings are not limited to individual worship but also address relationships, justice, cooperation, compassion, responsibility, and social conduct. Performance-based tasks can make these competencies visible through collaborative projects, community service, conflict-resolution simulations, peer teaching, and ethical problem-solving. Wiewiora and Kowalkiewicz (2019) demonstrate that authentic tasks can support the development of competencies that emerge through interaction with realistic situations. Jopp (2020) likewise shows that authentic projects can increase engagement and provide opportunities for creativity and applied learning. In IRE, social performance therefore provides evidence of whether students can translate religious principles into interpersonal behavior. Group-based assessment must nevertheless distinguish individual contributions from collective products. Consequently, individual reflection and peer assessment should accompany group performance to strengthen validity.

The theory of assessment for learning further establishes that assessment should contribute directly to the learning process. Rather than viewing assessment as a final judgment, assessment for learning emphasizes diagnostic information, feedback, learner participation, and instructional adaptation. Watling and Ginsburg (2019) emphasize the relationship between assessment, feedback, and learning, while Aisyah et al. (2024) identify feedback as an important feature of authentic IRE assessment. In practical terms, a performance task can reveal exactly where students struggle, allowing teachers to provide targeted instruction. Students can also compare their performance against a rubric and identify their own improvement priorities.(Bakr et al., 2025) This creates a feedback loop between performance, reflection, revision, and subsequent performance. In IRE,

DOI: https://doi.org/10.67899/ar.v3i1.226

such a cycle is especially valuable because religious competence often develops through repeated practice. Authentic assessment therefore becomes formative when its evidence is deliberately used to improve future performance.

Rubric theory provides the methodological bridge between complex competencies and reliable judgment. Rubrics clarify the criteria used to evaluate performance and can reduce ambiguity among assessors and students. Yudha (2020), using a Many-Facet Rasch Model approach, demonstrates the importance of examining person, rater, and task-related facets in performance assessment. Djidu et al. (2023) likewise emphasize the importance of construct validity and reliability in performance assessment instruments. These studies show that performance assessment requires attention to measurement quality rather than relying solely on teacher intuition. In IRE, rubrics should include observable indicators and performance levels that correspond directly to learning outcomes. Analytic rubrics may be especially useful when teachers need diagnostic information across several competency dimensions. Thus, rubric design should be regarded as an essential part of assessment pedagogy rather than a bureaucratic scoring procedure.

Validity theory is crucial because authentic assessment involves making interpretations from observed performances. A valid assessment is not simply one that produces consistent scores; it must support defensible interpretations and decisions about student competence. Conn et al. (2020) developed a validity inquiry process to help educators examine the validity and reliability of locally developed performance assessments. Their work highlights the need to consider evidence supporting assessment interpretations rather than treating validity as a one-time statistical property. For IRE, validity requires alignment between the religious competency being claimed and the evidence actually collected. If the objective is ethical reasoning, for example, a memorization test provides weak evidence. If the objective is correct prayer performance, written questions alone provide insufficient evidence. Validity therefore depends on the correspondence among intended competency, task, observation, scoring, and interpretation.

Reliability is also important because authentic performance assessments involve human judgment. Different teachers may interpret the same student performance differently, especially when criteria are vague. Performance assessment literature recommends clear descriptors, assessor training, moderation, multiple observations, and where feasible multiple raters. Yudha (2020) demonstrates the usefulness of many-facet approaches for examining rater agreement and other sources of measurement variation. Conn et al. (2020) likewise emphasize structured processes for improving the credibility of performance assessment interpretations. In IRE, reliability can be strengthened by using detailed rubrics, exemplars, calibration sessions, and repeated observations. However, reliability should not be pursued through excessive standardization that destroys the authenticity of the task. The theoretical challenge is therefore to achieve sufficient consistency while preserving meaningful contextual performance.

Research on Indonesian authentic assessment provides important empirical support for these theoretical principles. Wangid et al. (2017) examined authentic assessment implementation in the 2013 Curriculum and highlighted the complexity of implementing assessment across multiple domains. Pertiwi et al. (2018) found that teachers used observation, self-assessment, peer assessment, written tests, assignments, performance assessment, and portfolios to assess knowledge, attitudes, and skills. These findings show that authentic assessment is best understood as a system of complementary

DOI: https://doi.org/10.67899/ar.v3i1.226

techniques. In IRE, this multimethod principle is especially important because no single technique can capture all dimensions of religious competence. Observation can reveal behavior, performance tasks can reveal practical skills, tests can reveal conceptual understanding, and portfolios can reveal development over time. Therefore, triangulation should be a central theoretical principle of holistic IRE assessment.

The empirical research specifically focused on IRE further demonstrates the feasibility of authentic assessment. Tamrin et al. (2021) found that teachers plan and implement authentic assessments but experience difficulties associated with large numbers of students, scoring management, and personal bias. Baroroh and Hamani (2022) identified systematic steps in authentic assessment development, including mapping competencies, selecting assessment types, preparing instruments, and developing scoring guidance. Imanudin (2023) concluded that authentic assessment in IRE can encompass cognitive, affective, and psychomotor dimensions and can evaluate both learning outcomes and processes. These studies collectively establish that authentic assessment is already practiced in IRE contexts, although implementation remains uneven. The research also suggests that teacher assessment literacy is a critical condition for successful implementation. Therefore, the theoretical framework should incorporate not only student competencies but also teacher and institutional assessment capacity.

More recent evidence strengthens the case for holistic performance assessment. Aisyah et al. (2024) reported that authentic assessment in vocational IRE classrooms can integrate cognitive, affective, and psychomotor evidence while providing feedback for learning improvement. Sirajuddin et al. (2025) found that performance-based PAI evaluation in pesantren settings can involve daily worship, discipline, attitudes, and manners and can contribute to moral and spiritual development. Nyinge (2025) found a significant relationship between authentic assessment tools and generic competencies. These studies indicate that authentic assessment can operate across different educational environments and competency structures.(Nabriz et al., 2026) However, they also demonstrate that contextual adaptation is necessary because the meaning of performance varies according to educational setting. A pesantren may emphasize daily worship and manners, while a vocational school may emphasize applied problem-solving and professional conduct. Therefore, the theory of authentic IRE assessment should be flexible enough to accommodate contextual differences while maintaining common principles of validity and authenticity.

The theoretical synthesis suggests that holistic IRE assessment should be organized around evidence rather than isolated techniques. A performance task should begin with a clearly defined competency, establish an authentic context, require observable performance, employ transparent criteria, collect multiple forms of evidence, provide feedback, and include reflection. This sequence creates an assessment cycle rather than a single event. Such an approach is consistent with authentic assessment theory, competency-based education, assessment for learning, and validity theory. It also responds to empirical findings from IRE research concerning the need to integrate cognitive, affective, and psychomotor domains. The central theoretical proposition is that holistic competence emerges from the interaction among knowledge, action, values, reflection, and contextual application. Therefore, assessment should be designed to make these relationships visible. This theoretical foundation becomes the basis for the research method and findings presented in the following sections.

Method

DOI: https://doi.org/10.67899/ar.v3i1.226

This study employed a qualitative systematic literature review to examine contemporary research concerning authentic assessment, performance-based evaluation, and holistic competency measurement in Islamic Religious Education. A literature-based approach was selected because the objective was to synthesize conceptual arguments and empirical findings rather than collect new data from a particular school. The review focused on peer-reviewed journal articles published from 2016 to 2025, thereby representing approximately the most recent decade of scholarship available for the study. The literature included studies directly addressing IRE assessment as well as broader educational studies that provide theoretical or methodological foundations for authentic and performance-based assessment. This combination was necessary because the IRE-specific literature remains smaller than the broader assessment literature. The methodological orientation therefore emphasizes conceptual integration and transferability rather than statistical generalization.

The literature search focused on combinations of keywords including "authentic assessment," "performance-based assessment," "performance assessment," "Islamic Religious Education," "Islamic education assessment," "holistic assessment," "competency assessment," "rubric," and "assessment for learning." Searches were conducted through accessible academic journal platforms and scholarly indexing environments. Priority was given to studies with clear publication information, identifiable authorship, methodological descriptions, and relevance to the research questions. Articles outside the 2016–2025 publication range were excluded unless they were necessary to understand a concept already represented in recent scholarship; however, the substantive reference base was intentionally concentrated on the most recent decade. Studies that only discussed generic examination procedures without relevance to competency-based or authentic assessment were excluded. This procedure helped maintain conceptual consistency across the literature corpus.

The inclusion criteria were established to ensure that the reviewed literature contributed directly to the study's analytical framework. First, studies had to examine authentic assessment, performance assessment, competency assessment, holistic assessment, or assessment in IRE. Second, studies needed to provide conceptual arguments, empirical findings, methodological evidence, or practical implementation information. Third, journal articles had to contain sufficient bibliographic information for accurate APA-style citation. Fourth, preference was given to research that explicitly discussed assessment dimensions, instruments, implementation processes, validity, reliability, feedback, or teacher challenges. This criterion allowed the review to move beyond purely definitional descriptions. The resulting corpus was therefore interpreted as an integrated body of evidence rather than a simple bibliography.

Data extraction was conducted through thematic categorization. Relevant information was organized into categories including definitions of authentic assessment, characteristics of performance tasks, cognitive evidence, psychomotor evidence, affective evidence, spiritual evidence, social competence, assessment instruments, rubrics, feedback, validity, reliability, teacher competence, implementation challenges, and educational outcomes.(Khumaeni et al., 2026) Each article was read for its contribution to one or more of these categories. The extracted evidence was then compared across studies to identify recurring patterns and points of tension. Particular attention was given to findings from IRE studies because they provide contextual evidence concerning religious education. General educational studies were used to strengthen theoretical and methodological interpretation where IRE-specific evidence was limited.

DOI: https://doi.org/10.67899/ar.v3i1.226

The analysis employed thematic synthesis rather than statistical meta-analysis. Thematic synthesis was appropriate because the reviewed studies varied considerably in research design, educational level, participant characteristics, and assessment focus. Some studies used qualitative case studies, some developed assessment instruments, and others investigated relationships between authentic assessment and competency outcomes. Combining their findings statistically would therefore have been methodologically inappropriate. Instead, recurring concepts and findings were identified, compared, interpreted, and integrated into broader themes.(Azhari et al., 2026) The principal themes emerging from the analysis were authenticity and contextual relevance, multidimensional competency, performance evidence, rubric-based scoring, feedback, triangulation, teacher assessment literacy, and implementation constraints. These themes were then used to formulate the proposed holistic performance-assessment framework.

Quality and credibility of the synthesis were addressed through source triangulation. Claims concerning IRE-specific assessment practices were supported primarily by studies conducted directly in Islamic education contexts, including Tamrin et al. (2021), Baroroh and Hamani (2022), Aisyah et al. (2024), and Sirajuddin et al. (2025). Claims concerning performance-assessment validity, reliability, and rubric design were cross-checked against broader assessment research, including Braun et al. (2020), Conn et al. (2020), and Yudha (2020). The use of different types of studies allowed the analysis to distinguish between contextual findings and general assessment principles. Interpretations were also compared across multiple sources rather than relying on a single publication. This process strengthened the plausibility of the synthesized framework. Nevertheless, because the study is literature-based, its conclusions should be interpreted as evidence-informed propositions rather than direct causal findings from a new intervention. The final analytical stage involved constructing a conceptual model of authentic performance-based evaluation for IRE. The model was derived by connecting learning outcomes with authentic tasks, observable indicators, assessment rubrics, multiple evidence sources, feedback, reflection, and follow-up learning. The model recognizes six interconnected competency dimensions: cognitive, psychomotor, affective, spiritual, social, and reflective-application competence. The analysis then examined how each dimension could be represented through authentic performance evidence. The resulting framework was compared with the limitations identified in prior research, especially teacher workload, scoring subjectivity, large classes, and difficulty assessing internal spirituality. The methodological contribution of the study is therefore not the creation of a statistically validated instrument but the development of an evidence-based conceptual framework that can guide future empirical instrument development.

Results and Discussion

The synthesis of the reviewed literature suggests that the educational value of authentic assessment in Islamic Religious Education (IRE) is determined primarily by the extent to which assessment tasks require students to demonstrate meaningful and transferable competencies rather than merely reproduce previously presented information. Authentic assessment is generally characterized by contextual relevance, active participation, intellectual engagement, and the application of knowledge to situations that resemble real-life practices or pedagogically meaningful simulations. Villarroel et al. (2018) argue that authentic assessment should provide opportunities for learners to engage with tasks that reflect the complexity and contextual demands of real-world practice, while Jopp (2020) demonstrates that appropriately designed authentic

DOI: https://doi.org/10.67899/ar.v3i1.226

tasks can promote student engagement, creativity, and deeper learning. These characteristics correspond closely with the nature of IRE, in which learning outcomes encompass not only religious knowledge but also worship practices, ethical reasoning, interpersonal conduct, and social responsibility. Students may consequently demonstrate competence through practical worship, community-based activities, ethical case analysis, religious communication, collaborative projects, and structured reflection. The synthesis therefore indicates that authenticity in IRE should be determined by the meaningfulness and educational realism of the learning context rather than by simply labeling an activity as "practical." This conceptual distinction is essential because an activity may involve physical performance while remaining pedagogically inauthentic if it does not require students to apply knowledge, exercise judgment, or demonstrate meaningful competence.

A central implication emerging from the literature is the multidimensional character of competence within IRE. Religious education cannot be adequately represented through cognitive achievement alone because its educational objectives encompass knowledge, attitudes, practices, values, character, and the capacity to translate religious understanding into responsible action. Baroroh and Hamani (2022) describe authentic assessment in IRE as encompassing cognitive, affective, and psychomotor domains, while Imanudin (2023) emphasizes the importance of evaluating both learning processes and learning outcomes. More recent evidence from the implementation of the Merdeka Curriculum further demonstrates the relevance of integrated assessment approaches that recognize diverse dimensions of student development (Aisyah et al., 2024). These findings support the development of a holistic competency model in which different dimensions are assessed according to their relevance to particular learning outcomes. Holistic assessment, therefore, should not be interpreted as assigning identical numerical weight to every domain. Rather, it requires the systematic collection of appropriate evidence for each competency and the interpretation of that evidence according to clearly established educational purposes. Such an approach shifts the emphasis from producing a single aggregate score toward constructing a more comprehensive profile of student competence.

The literature further indicates that performance-based assessment has particular relevance for competencies involving religious practices and observable behavioral performance. Tamrin et al. (2021) document the use of performance activities and non-test assessment techniques in IRE, including tasks involving practical religious activities. Similarly, Sirajuddin et al. (2025) demonstrate that performance-based evaluation within pesantren settings can provide evidence concerning daily worship, discipline, attitudes, manners, and other forms of religious behavior. These findings establish performance as a legitimate source of assessment evidence when the intended learning outcome involves procedural or behavioral competence. Performance assessment becomes especially appropriate when students are expected not merely to explain a religious practice but to demonstrate their ability to perform it accurately and appropriately. It also provides an opportunity to connect formal classroom learning with the everyday religious practices that characterize students lived educational environments. Consequently, the design of IRE assessment should begin by determining whether the intended competency is most validly demonstrated through written responses, observable performance, or a combination of both. Such alignment between competency and evidence is fundamental to the validity of performance-based assessment.

Authentic performance assessment also possesses considerable potential for developing and evaluating higher-order thinking within IRE. Performance tasks should

DOI: https://doi.org/10.67899/ar.v3i1.226

not be restricted to procedural demonstrations because authentic competence may require students to interpret information, analyze complex situations, evaluate alternatives, justify decisions, and construct solutions. Braun et al. (2020) demonstrate that performance assessment can be designed to capture complex cognitive constructs, including higher-order reasoning and critical thinking. Similarly, research on the redesign of authentic examinations emphasizes the importance of realism, cognitive challenge, and evaluative judgment in the development of meaningful assessment tasks. In IRE, these principles can be operationalized through tasks requiring students to analyze contemporary ethical dilemmas, compare alternative interpretations of Islamic principles, justify decisions using Qur'anic or prophetic evidence, or formulate responses to social problems. Such tasks provide teachers with evidence of whether students can transfer religious knowledge beyond the immediate context in which it was learned. Authentic IRE assessment should therefore recognize intellectual performance as an important dimension of religious competence alongside practical religious performance. This orientation is particularly important for preparing students to engage thoughtfully with complex moral and social challenges in contemporary society.

The complexity of performance evidence also underscores the importance of transparent and criterion-referenced judgment. Unlike conventional objective tests, performance tasks frequently produce multidimensional evidence that cannot be evaluated adequately through general impressions or intuitive scoring. Yudha (2020) demonstrates that performance assessment involves multiple sources of variation, including characteristics of the learner, the task, and the assessor, while Conn et al. (2020) emphasize the importance of systematic validity inquiry when interpreting evidence generated through locally developed performance assessments. Within IRE, analytic rubrics can provide greater transparency by specifying the dimensions of competence and describing qualitative differences between performance levels. Depending on the intended learning outcomes, rubric criteria may include conceptual accuracy, procedural competence, ethical reasoning, communication, collaboration, responsibility, and reflective capacity.(Aswa et al., 2026) Providing students with the criteria before they undertake an assessment can further promote transparency by clarifying expectations and supporting self-regulated learning. Rubrics may also facilitate peer assessment and self-assessment because students are given explicit standards against which their performance can be examined. The resulting assessment process moves beyond the conventional model of a teacher simply assigning a score and instead emphasizes the interpretation of observable evidence against clearly articulated competency standards.

The multidimensional nature of IRE competence also necessitates the use of multiple sources of evidence. No individual assessment technique can adequately capture conceptual knowledge, religious practice, moral behavior, social competence, reflective capacity, and developmental progress simultaneously. Written examinations may provide useful evidence of conceptual understanding, whereas performance tasks can demonstrate procedural competence, observations can reveal patterns of behavior, portfolios can document development over time, and reflective activities can provide insight into students' metacognitive and value-oriented thinking. Pertiwi et al. (2018) identify observation, self-assessment, peer assessment, written testing, performance assessment, and portfolio assessment as complementary components of authentic assessment. This supports the principle of assessment triangulation, in which conclusions about student competence are strengthened by converging evidence obtained through different methods. Such triangulation is particularly important in IRE because religious behavior

DOI: https://doi.org/10.67899/ar.v3i1.226

may vary according to social context, emotional condition, and situational demands. Repeated observation combined with performance tasks, reflective evidence, and conventional assessment can therefore provide a more defensible basis for interpretation than reliance on a single assessment event. Accordingly, an evidence portfolio may provide a more meaningful representation of holistic IRE competence than a single numerical examination score.

Within this multidimensional assessment architecture, self-assessment and structured reflection deserve a more prominent role. Authentic assessment is most educationally productive when learners are able to understand the standards by which their performance is judged and actively evaluate their own progress. This principle is particularly relevant to IRE because religious learning involves not only the acquisition of knowledge but also reflection on intention, responsibility, conduct, and personal development. Wiewiora and Kowalkiewicz (2019) demonstrate that authentic assessment can contribute to competency and identity development, while Watling and Ginsburg (2019) emphasize the importance of feedback and learner engagement in assessment processes. Reflective tasks can invite students to examine what they have learned, identify difficulties, explain how their decisions relate to Islamic principles, and formulate specific intentions for improvement.(Azzahro & Darmawan, 2026) Such reflections should not be interpreted as direct or definitive measurements of spirituality because internal religious states cannot be fully captured through written responses. Rather, they constitute evidence of metacognitive awareness, ethical reasoning, and the learner's capacity to connect educational experiences with personal meaning. The incorporation of structured reflection therefore transforms students from passive recipients of assessment judgments into active participants in evaluating and improving their own learning.

The educational effectiveness of authentic assessment is further dependent on the quality and timing of feedback. Assessment has limited formative value when evidence is collected solely for grading purposes and students are given no meaningful opportunity to respond to the results. Aisyah et al. (2024) emphasize the importance of feedback within authentic IRE assessment, while broader assessment-for-learning scholarship highlights feedback as a mechanism through which students can identify gaps between current and desired performance. In performance-based IRE, feedback can address a wide range of dimensions, including procedural accuracy, conceptual misunderstanding, ethical reasoning, communication, collaboration, and behavioral responsibility. Importantly, feedback should be followed by opportunities for revision, repetition, or reassessment so that students can demonstrate improvement. Such an approach transforms assessment from a terminal event into an iterative process of learning and development. The principle is particularly relevant to religious practices because procedural and behavioral competence generally develops through repeated practice rather than through a single demonstration. Authentic IRE assessment should consequently be designed as a cycle of performance, feedback, reflection, revision, and subsequent performance.

Despite its pedagogical potential, the implementation of authentic assessment presents substantial challenges related to teacher workload, scoring consistency, and institutional feasibility. Tamrin et al. (2021) identify large class sizes, potential assessor bias, and difficulties in scoring as significant challenges in the implementation of IRE assessment, while Baroroh and Hamani (2022) highlight the practical complexity of implementing technically demanding assessment procedures within everyday teaching conditions. These findings demonstrate that theoretical sophistication alone does not

DOI: https://doi.org/10.67899/ar.v3i1.226

guarantee successful implementation. An assessment framework must also take account of teachers' available time, class size, institutional resources, and professional competence. Workload can be managed through strategic sampling of competencies, carefully structured group projects accompanied by individual evidence, rotating observation schedules, digital portfolios, and shared assessment rubrics. Collaborative moderation among teachers can further reduce individual scoring burdens while improving consistency in judgment. Consequently, practicality should not be treated as an external consideration but as an integral component of assessment quality alongside validity, reliability, fairness, and educational usefulness.

Another important contribution of authentic assessment lies in its capacity to strengthen the relationship between formal IRE learning and students' lived experiences. Jopp (2020) demonstrates that well-designed authentic tasks can promote engagement, deeper understanding, creativity, and knowledge application, while Nyinge (2025) identifies a significant relationship between authentic assessment tools and the development of broader competencies. In IRE, these educational benefits can be operationalized through projects addressing community service, environmental stewardship, ethical communication, digital religious literacy, social inclusion, charitable initiatives, or collaborative problem-solving. Such activities require students to encounter Islamic principles not merely as theoretical propositions but as frameworks for interpreting and responding to real situations. The resulting assessment evidence can demonstrate whether students are capable of translating religious knowledge into responsible action. This approach also reinforces the formative purpose of Islamic education by connecting classroom learning with character development and social responsibility. Authentic assessment thus becomes more than a mechanism for measuring achievement; it becomes a pedagogical bridge between religious knowledge, personal development, and participation in society. Overall, the findings demonstrate that authentic assessment in IRE should be conceptualized as a multidimensional, evidence-based, and developmentally oriented assessment system. Its effectiveness depends on the alignment of learning outcomes, authentic tasks, observable evidence, transparent criteria, meaningful feedback, and opportunities for subsequent improvement.(Ashfiya' et al., 2026) The literature indicates that the multidimensional nature of Islamic education makes reliance on a single assessment method theoretically and practically inadequate. Cognitive tests remain important for establishing conceptual understanding, but they should be complemented by performance assessment, observation, portfolios, reflective activities, peer assessment, and self-assessment where appropriate. This combination enables teachers to construct a richer and more defensible representation of student competence while reducing the risk of interpreting one isolated performance as a complete indicator of learning. At the same time, implementation must remain sensitive to assessment validity, reliability, fairness, feasibility, and teacher workload. The central contribution of the synthesis is therefore not the replacement of conventional assessment but its repositioning within a broader evidence architecture capable of representing the distinctive educational purposes of IRE.

The broader theoretical implication is that the transformation of IRE assessment requires a shift in the fundamental question underlying educational measurement. Instead of focusing predominantly on **"**What score did the student obtain?", assessment should increasingly ask "What evidence demonstrates what the student understands, can perform, values, and can apply?" This conceptual shift changes assessment from a primarily classificatory mechanism into a process of learning support, self-regulation, and

DOI: https://doi.org/10.67899/ar.v3i1.226

competency development. It simultaneously transforms the teacher's role from a score generator into an interpreter of evidence, provider of formative feedback, and facilitator of improvement. Students become more actively involved in their learning when they understand assessment criteria, monitor their own performance, interpret feedback, and formulate strategies for development.(Raafi et al., 2026) At the institutional level, schools and madrasahs are consequently required to create conditions that make meaningful assessment possible through adequate time, professional development, collaborative moderation, and appropriate technological support. Authentic assessment can therefore contribute not only to improved measurement but also to the development of a more reflective, participatory, and competency-oriented educational culture.

Building upon these findings, the present study conceptualizes the Holistic Performance-Based Islamic Religious Education Assessment Framework as an integrated model comprising six interconnected stages: competency identification, authentic task design, performance evidence collection, rubric-based judgment, reflective feedback, and follow-up improvement.(Halil & Aini, 2026) The framework integrates cognitive understanding, religious practice, character formation, spiritual reflection, social responsibility, and the practical application of Islamic principles within a unified assessment architecture. Its principal novelty lies in conceptualizing these dimensions not as isolated assessment domains but as mutually reinforcing forms of evidence that collectively represent the complexity of IRE competence. The framework also recognizes the epistemological limitations of assessing internal spirituality by restricting formal assessment claims to evidence that students can demonstrate, articulate, reflect upon, and apply. This approach provides a balance between the distinctive spiritual purposes of Islamic education and the methodological requirements of defensible educational assessment. Future empirical research should therefore examine the framework through instrument-development studies, validity and reliability testing, classroom-based interventions, inter-rater agreement analysis, and longitudinal studies of competency development. Such investigations would provide stronger empirical evidence concerning the effectiveness, feasibility, and contextual adaptability of holistic performance-based authentic assessment across diverse Islamic educational settings.

Integrated Findings Matrix

Holistic Competency	**Authentic Performance Task**	**Evidence Collected**	**Recommended Assessment Technique**	**Main Indicator**
Cognitive competence	Ethical case analysis and religious problem-solving	Written reasoning, oral explanation, argumentation	Analytic rubric + teacher questioning	Accuracy, analysis, justification
Psychomotor competence	Ablution, prayer, Qur'anic recitation, funeral practice	Direct performance and video/documentation	Performance rubric	Procedural accuracy, fluency, sequence
Affective competence	Cooperative religious project	Observation, peer	Observation rubric	Responsibility, empathy, discipline

DOI: https://doi.org/10.67899/ar.v3i1.226

Holistic Competency	Authentic Performance Task	Evidence Collected	Recommended Assessment Technique	Main Indicator
		assessment, reflection		
Spiritual-learning competence	Worship reflection and meaning-making activity	Reflection journal, discussion, contextual evidence	Reflective rubric	Understanding, sincerity-related reflection, consistency
Social competence	Community service or social-action project	Product, observation, peer feedback	Project rubric	Cooperation, communication, social responsibility
Ethical competence	Moral dilemma simulation	Decision, explanation, reflection	Scenario-performance rubric	Ethical reasoning and value application
Communication competence	Khutbah, presentation, podcast, religious campaign	Presentation/product	Presentation rubric	Clarity, evidence, relevance, respectful communication
Reflective competence	Learning portfolio	Portfolio and self-evaluation	Portfolio rubric	Growth, self-evaluation, improvement planning
Integrative competence	Community-based Islamic project	Multiple evidence sources	Holistic rubric	Integration of knowledge, practice, values, and application

The synthesis of the literature suggests that holistic assessment in Islamic Religious Education (IRE) should be conceptualized as an integrated system of evidence rather than as a single assessment instrument or a narrowly defined "holistic test." Such an approach recognizes that religious competence encompasses multiple dimensions, including conceptual understanding, religious practice, moral character, spiritual reflection, social responsibility, and the ability to apply Islamic principles in authentic situations. Because these dimensions cannot be adequately represented through one assessment method, a combination of complementary evidence sources is required to develop a comprehensive profile of student competence. Performance tasks, written assessments, observation, portfolios, reflective activities, and peer or self-assessment can therefore serve different but interconnected purposes within the same assessment system.

DOI: https://doi.org/10.67899/ar.v3i1.226

Braun et al. (2020) emphasize that complex competencies require assessment approaches capable of capturing multiple forms of performance evidence, while Conn et al. (2020) stress the importance of ensuring that interpretations derived from such evidence remain valid and defensible. Accordingly, the proposed framework operationalizes holistic assessment as a structured evidence architecture in which each assessment activity is explicitly connected to predetermined competencies and meaningful learning outcomes.

A major implication of this framework is the need to reconceptualize cognitive assessment without abandoning conventional written examinations altogether. Written tests remain valuable for determining whether students have acquired essential theological concepts, Qur'anic knowledge, prophetic traditions, and principles of Islamic jurisprudence.(Alamsyah et al., 2026) Nevertheless, their educational contribution becomes limited when they primarily measure factual recall without requiring interpretation, reasoning, evaluation, or application. Contemporary assessment theory increasingly emphasizes the importance of evaluating complex cognitive processes through contextualized and performance-oriented tasks (Braun et al., 2020). Within IRE, students should therefore be challenged to interpret Qur'anic principles in relation to contemporary social issues, evaluate competing ethical positions, and formulate reasoned responses to authentic problems. Aisyah et al. (2024) similarly demonstrate the continued relevance of cognitive assessment within broader approaches to IRE evaluation, particularly when cognitive outcomes are connected with other dimensions of learning. The essential transformation, therefore, is not the elimination of cognitive testing but the movement from isolated recall toward evidence of meaningful religious reasoning.

Assessment of religious practice requires an equally significant shift from knowledge-based verification toward direct observation of performance. When an instructional objective concerns a procedural or practical competency, students' ability to explain a religious practice theoretically should not be regarded as sufficient evidence of mastery. For instance, knowledge of the procedural requirements of *ṣalāh* does not necessarily demonstrate the student's ability to perform the practice accurately and consistently. Direct observation supported by explicit performance criteria can provide stronger evidence of procedural competence than written responses alone. Tamrin et al. (2021) and Sirajuddin et al. (2025) provide evidence supporting the relevance of practical and performance-based assessment within Islamic educational contexts. Repeated observation is particularly important because individual performances can be affected by situational variables such as anxiety, unfamiliarity with the assessment setting, or temporary lapses in concentration. Consequently, longitudinal observation can provide a more reliable representation of religious practice while simultaneously creating opportunities for formative feedback and improvement.

The assessment of character and affective development likewise requires attention to patterns of behavior rather than isolated incidents. Islamic education seeks not merely to transmit moral concepts but to cultivate dispositions such as honesty, responsibility, empathy, discipline, cooperation, and respect.(Mukhlis et al., 2026) These dispositions are better represented through repeated behavioral evidence than through students' responses to conventional questionnaires or one-time classroom observations. For example, responsibility may be demonstrated through consistent task completion, responsible use of shared resources, transparency in reporting project outcomes, and accountability for assigned responsibilities. Similarly, empathy may become visible through students' listening behavior, responses to disagreement, and sensitivity toward the needs of others. Faizah et al. (2019) demonstrate the potential of authentic assessment

DOI: https://doi.org/10.67899/ar.v3i1.226

to evaluate character development when behavioral indicators are clearly defined. In the context of IRE, however, such assessment should remain developmental rather than punitive, because the primary purpose of evaluating character is to facilitate moral growth rather than to permanently classify students according to perceived personal deficiencies.

The assessment of spiritual learning presents a more complex methodological challenge because spirituality includes dimensions that are not directly observable or objectively measurable through conventional instruments. A defensible approach therefore requires assessment to focus on manifestations that students can demonstrate, articulate, explain, and apply rather than making claims about the absolute depth or authenticity of their faith. Evidence may include students' ability to explain spiritual meanings, articulate personal reflections, connect worship with ethical behavior, and apply religious principles to authentic moral situations. Such evidence does not constitute a direct measurement of the internal state of faith; rather, it provides an educational basis for evaluating aspects of spiritual learning that can reasonably be observed or communicated. Conn et al. (2020) emphasize that valid assessment requires careful alignment between the evidence collected and the interpretations made from that evidence. Accordingly, assessment of spiritual development in IRE should maintain epistemological humility while nevertheless providing meaningful opportunities for students to demonstrate reflective and value-oriented learning.

Social competence constitutes another essential dimension of holistic IRE because Islamic education is concerned not only with individual religious observance but also with ethical participation in society. Students should be able to translate Islamic principles into responsible interaction with other people and constructive responses to social problems. Authentic projects involving environmental sustainability, social inclusion, charitable activities, digital ethics, community service, or conflict resolution can provide contexts in which such competence becomes observable. Wiewiora and Kowalkiewicz (2019) demonstrate that authentic assessment can support the development of competencies and professional or social identity through realistic activities. Within IRE, students can be assessed on their capacity to connect Islamic principles with practical action while also demonstrating collaboration, communication, responsibility, and ethical decision-making. Because collaborative projects may obscure individual contributions, teacher observation should be complemented by self-assessment and peer assessment. Such an approach broadens the conceptualization of religious competence from individual ritual proficiency toward responsible participation in social life.

Digital technology further expands the possibilities for documenting authentic evidence of learning. Students can compile digital portfolios, record religious practices, produce podcasts, develop educational videos, create public-service campaigns, or submit reflective journals through learning-management platforms.(Wahidah & Cahyadi, 2026) Technology-enhanced authentic assessment can increase opportunities for creativity, documentation, reflection, and feedback when it is appropriately aligned with learning objectives (Jopp, 2020). Digital evidence also allows teachers to revisit student performances and examine developmental changes over time, which may be difficult to achieve through conventional classroom observation alone. Nevertheless, technological sophistication should never be confused with educational quality. A highly polished digital product may demonstrate technical competence without necessarily providing evidence of strong religious understanding, ethical reasoning, or spiritual reflection. Therefore, digital tools should function as mechanisms for collecting and organizing evidence rather than as independent indicators of learning quality.

DOI: https://doi.org/10.67899/ar.v3i1.226

The incorporation of authentic assessment also raises important questions concerning fairness, consistency, and standardization. Authentic tasks naturally allow contextual variation because students may demonstrate competence through different problems, projects, or forms of performance. Such variation should not be regarded as a weakness provided that students are evaluated according to comparable learning outcomes and transparent criteria. Braun et al. (2019) identify the tension between authenticity and standardization in performance assessment, emphasizing the importance of maintaining defensible standards while preserving meaningful contexts. In IRE, for example, students from different institutions may conduct different environmental or social projects while still being assessed according to shared criteria concerning conceptual accuracy, ethical reasoning, practical competence, collaboration, and responsible application. Common rubrics, exemplar performances, teacher moderation, and collaborative assessment discussions can improve consistency without forcing every student to complete an identical task. Fairness, therefore, should be understood not as uniformity of assessment activity but as the consistent application of transparent and defensible standards.

The effectiveness of such an assessment system also depends substantially on teachers' assessment literacy. Teachers need to distinguish clearly among learning outcomes, assessment tasks, observable evidence, interpretive judgments, and final grades. Without this distinction, authentic assessment may become vulnerable to subjective judgment, excessive workload, or poorly defined criteria. Baroroh and Hamani (2022) emphasize the importance of systematic assessment development in educational practice, while Conn et al. (2020) highlight the importance of reflective validity inquiry when interpreting performance evidence. Professional development for IRE teachers should consequently address authentic task design, rubric construction, observation strategies, feedback processes, moderation, and ethical assessment practices. Teachers should also be trained to identify high-value evidence rather than attempting to assess every student activity. Institutional support is essential because effective assessment requires adequate time, collaborative planning, and opportunities for teachers to moderate their judgments. Strengthening assessment literacy should therefore be considered an institutional priority rather than an individual responsibility.

A further consideration is the longitudinal nature of competency development. Holistic religious competence does not emerge through a single instructional encounter or assessment event but develops progressively through repeated learning experiences, practice, feedback, reflection, and revision. A portfolio-based approach can document this development by bringing together students' performances, revisions, reflections, and teacher feedback over an extended period. Pertiwi et al. (2018) identify portfolio assessment as an important component of authentic assessment, while Nyinge (2025) highlights the broader contribution of authentic assessment tools to competency development. Within IRE, a longitudinal portfolio could document changes in religious understanding, practical performance, ethical reasoning, communication, collaboration, and reflective capacity. Such documentation allows assessment to become developmental rather than merely classificatory. The portfolio should therefore be understood not simply as a repository of student work but as a structured narrative demonstrating how competence develops over time.

The strongest implementation of performance-based authentic assessment occurs when assessment is integrated into instructional design from the outset rather than introduced as an additional activity after teaching has been completed. This principle

DOI: https://doi.org/10.67899/ar.v3i1.226

reflects the logic of constructive alignment, whereby learning outcomes, instructional activities, and assessment tasks are deliberately coordinated (Villarroel et al., 2018). In IRE, curriculum designers should first identify the knowledge, skills, values, practices, and forms of application that students are expected to develop. Learning experiences can then be designed to provide opportunities for students to construct these competencies, while assessment tasks collect evidence that corresponds directly to the intended outcomes.(Sa'adah et al., 2026) For example, a unit on social justice could culminate in a project requiring students to analyze Islamic sources, investigate a local social problem, formulate an intervention, implement part of the proposed solution, and reflect critically on the experience. Such an assessment simultaneously generates evidence of conceptual knowledge, religious reasoning, collaboration, communication, social responsibility, and reflective capacity. Alignment of this kind transforms assessment from an external measurement mechanism into an integral component of meaningful learning.

Taken together, these findings indicate that authentic assessment in IRE is best conceptualized as an assessment ecosystem rather than as a collection of alternative techniques. The ecosystem begins with clearly defined competencies and proceeds through authentic learning experiences, evidence collection, transparent criteria, performance judgment, formative feedback, reflection, and subsequent improvement. Because IRE encompasses cognitive, practical, affective, spiritual, social, and reflective dimensions, no single assessment method can adequately represent the full range of expected outcomes. Multiple forms of evidence are therefore necessary to strengthen the comprehensiveness and defensibility of assessment judgments. At the same time, the integration of multiple methods requires careful attention to validity, reliability, fairness, feasibility, and teacher workload. Conventional written examinations should consequently not be discarded but repositioned as one component within a broader evidence architecture. The resulting assessment system provides a more nuanced representation of student competence while remaining aligned with the multidimensional educational purposes of Islamic Religious Education.

The broader significance of this framework lies in a fundamental conceptual transformation in how educational achievement is understood. Rather than asking exclusively, "What score did the student obtain?", educators are encouraged to ask, "What evidence demonstrates what the student understands, can perform, values, and can apply?" This shift transforms assessment from a predominantly classificatory mechanism into a process that supports learning, reflection, and personal development. It also redefines the teacher's role from that of a score producer toward that of an evidence interpreter, feedback provider, and facilitator of learning improvement. Students, in turn, become more active participants when they understand assessment criteria, evaluate their own performance, and identify strategies for improvement.(Hasnan et al., 2026) Such a transformation can contribute to a stronger culture of learning because assessment becomes embedded within the processes of inquiry, practice, reflection, and growth. Authentic assessment thus has significance beyond measurement because it can influence how teachers teach and how students understand the purpose of learning itself.

On the basis of the preceding synthesis, this study proposes a Holistic Performance-Based Islamic Religious Education Assessment Framework consisting of six interconnected stages: competency identification, authentic task design, performance evidence collection, rubric-based judgment, reflective feedback, and follow-up improvement. The framework integrates cognitive understanding, religious practice, character development, spiritual learning, social responsibility, and reflective application

DOI: https://doi.org/10.67899/ar.v3i1.226

within a unified assessment architecture. Its principal contribution lies in treating these dimensions as interconnected sources of evidence rather than as separate assessment domains that operate independently. At the same time, the framework acknowledges the methodological limitations associated with assessing internal spirituality by restricting formal assessment claims to evidence that students can demonstrate, explain, reflect upon, and apply. This distinction strengthens the conceptual validity of the framework while preserving the distinctive spiritual objectives of Islamic education. Future empirical research should examine the framework through instrument-development studies, classroom-based interventions, inter-rater reliability analysis, validity studies, and longitudinal investigations of student competency development. Such research would provide stronger empirical foundations for determining how performance-based authentic assessment can be implemented effectively across diverse IRE contexts.

DOI: https://doi.org/10.67899/ar.v3i1.226

Conclusion

Authentic assessment through performance-based evaluation offers a strong pedagogical and methodological foundation for measuring students' holistic competencies in Islamic Religious Education. The literature reviewed in this study demonstrates that IRE assessment should extend beyond conventional cognitive tests to include direct performance, observation, projects, portfolios, reflection, peer assessment, and contextual applications of Islamic knowledge. Such an approach can capture cognitive, psychomotor, affective, spiritual-learning, social, ethical, and reflective competencies more comprehensively. Nevertheless, authentic assessment must be implemented with clear learning outcomes, realistic tasks, transparent rubrics, multiple evidence sources, continuous feedback, and appropriate moderation. The assessment of spirituality should remain focused on observable educational evidence and should avoid unsupported claims about students' internal faith. Therefore, the quality of holistic assessment depends not on the number of instruments used but on the validity and relevance of the evidence they generate. The principal novelty of this study is the formulation of an integrated holistic performance-based assessment framework for IRE that connects knowledge, religious practice, character, social responsibility, reflection, and real-life application. The framework positions assessment as a continuous learning cycle rather than a terminal scoring event. It also provides a practical response to challenges identified in previous research, including teacher workload, scoring subjectivity, large classes, and difficulties in measuring complex competencies. Future research should empirically validate the proposed framework through design-based research, classroom experiments, rubric validation, Many-Facet Rasch analysis, and longitudinal studies. Such research would be particularly valuable for determining the reliability, fairness, and educational impact of holistic performance assessment across elementary schools, secondary schools, madrasahs, vocational institutions, and pesantren. Ultimately, authentic assessment can help IRE fulfill its broader educational mission by demonstrating not merely what students know about Islam, but how they understand, practice, embody, communicate, and apply Islamic values responsibly in real life.

DOI: https://doi.org/10.67899/ar.v3i1.226

Bibliography

Adib, D. A., Rusilowati, A., & Hidayah, T. (2018). Development of authentic appraisal instruments basic skills for playing football of junior high school students. Journal of Research and Educational Research Evaluation, 7(1). https://doi.org/10.15294/jrer.v7i1.21380

Agustiningsih, R., & Rahdiyanta, D. (2019). Implementation of an authentic assessment model to improve the quality of learning assessment. Jurnal Pendidikan Teknologi dan Kejuruan, 25(1). https://doi.org/10.21831/jptk.v25i1.20072

Aisyah, A., Saona, S., & Sadaat, S. H. (2024). Implementation of authentic assessment of Islamic Religious Education in the Merdeka Curriculum: Qualitative studies in the cognitive, affective, and psychomotor domains in vocational high schools. Jurnal Pendidikan Indonesia, 5(4), 189–203. https://doi.org/10.59141/japendi.v5i4.9663

Alamsyah, M. N., Khalilurrahman, & Anshari, M. (2026). Sufism of Ahl al-Sunnah wa al-Jama'ah (ASWAJA) Scholars of the Nusantara: A Comparative Study of the Thought of Shaykh Arsyad al-Banjari, Shaykh Nawawi al-Bantani, and Kiai Haji Hasyim Asy'ari. AL GHAZALI: Jurnal Pendidikan Dan Pemikiran Islam, 6(1), 201–221. https://doi.org/10.69900/ag.v6i1.554

Ashfiya', M. B., Amalia, Z. D., Burda, N., Azzahro, F., Sofanin, A., & Syafi'i, I. (2026). Transforming Educators' Competencies for Human-Centered Learning in the Era of Artificial Intelligence: A Literature Review from the Perspective of Islamic Educational Values. AL GHAZALI: Jurnal Pendidikan Dan Pemikiran Islam, 6(4), 1405–1423. https://doi.org/10.69900/ag.v6i4.625

Aswa, M. R. Al, Maharani, S. N., Ainurrojab, M. I., Azmi, K., Rahman, R. A., Maryam, M., & Mufidah, Q. A. (2026). Implemention of Islamic Religious Education Curriculum Based on Integrated Islamic School Quality Standards at SMAIT Nurul Ilmi Tenggarong. AL GHAZALI: Jurnal Pendidikan Dan Pemikiran Islam, 6(2), 435–450. https://doi.org/10.69900/ag.v6i2.567

Azhari, M., Ma'ruf, H., Yaqin, H., & Hamdan, H. (2026). Transformation of Islamic Religious Education in the Era of the Industrial Revolution 4.0 and Society 5.0. AL GHAZALI: Jurnal Pendidikan Dan Pemikiran Islam, 6(3), 1163–1177. https://doi.org/10.69900/ag.v6i3.583

Azzahro, S. N., & Darmawan, D. (2026). The Effect of Teacher Creativity on Students' Learning Interest in Islamic Education at Elementary School. AL GHAZALI: Jurnal Pendidikan Dan Pemikiran Islam, 6(4), 1544–1567. https://doi.org/10.69900/ag.v6i4.624

Bakr, Y. M. A., Raus, N. M., Ibrahim, A. H., & Faizi, A. U. (2025). Islamic Education and Intercultural Competence: Promoting Religious Literacy in Multicultural Societies. Al Husna: Jurnal Ilmiah Pendidikan Agama Islam, 2(1), 86–105. https://doi.org/10.69900/ah.v2i1.23

Baroroh, U., & Hamani, T. (2022). Development of authentic assessment in Islamic Religious Education in elementary school. Nazhruna: Jurnal Pendidikan Islam, 5(3), 940–955. https://doi.org/10.31538/nzh.v5i3.2380

Begum, S., Rashid Saleh Al Tamimi, M., Mohammed Al Dossary, M., El-Khatib, A., & Jaria Maidin, A. (2025). Integrating Maqashid al-Shari'ah into Sustainable

Education: An Islamic Perspective on Global Citizenship Education. Al Husna: Jurnal Ilmiah Pendidikan Agama Islam, 2(1), 65–85. https://doi.org/10.69900/ah.v2i1.22

Braun, H. I., Shavelson, R. J., Zlatkin-Troitschanskaia, O., & Borowiec, K. (2020). Performance assessment of critical thinking: Conceptualization, design, and implementation. Frontiers in Education, 5, 156. https://doi.org/10.3389/feduc.2020.00156

Braun, V., Clarke, V., Hayfield, N., & Terry, G. (2019). Thematic analysis. In P. Liamputtong (Ed.), Handbook of research methods in health social sciences (pp. 843–860). Springer.

Brunaf, C., & colleagues. (2020). Using principles of authentic assessment to redesign written examinations and tests. Innovations in Education and Teaching International, 57(1), 38–49. https://doi.org/10.1080/14703297.2018.1564882

Conn, C. A., Bohan, K. J., Pieper, S. L., & Musumeci, M. (2020). Validity inquiry process: Practical guidance for examining performance assessments and building a validity argument. Studies in Educational Evaluation, 65, 100843. https://doi.org/10.1016/j.stueduc.2020.100843

Djidu, H., Istiyono, E., & Widihastuti. (2023). Quality of performance assessment instruments for educators in higher education: Implementation of factor analysis and generalizability theory. e-Saintika: Jurnal Penelitian dan Pengkajian Ilmu Pendidikan, 7(2). https://doi.org/10.36312/esaintika.v7i2.716

Faizah, U., Zuchdi, D., & Alsamiri, Y. (2019). An authentic assessment model to assess kindergarten students' character. REID: Research and Evaluation in Education, 5(2), 103–119. https://doi.org/10.21831/reid.v5i2.24588

Halil, H., & Aini, A. (2026). Islamic Entrepreneurship among Single-Parent Teachers in Islamic Boarding Schools: A Model for Strengthening Family Economic Resilience Based on Islamic Values at Pesantren Al-Amien Putri I. AL GHAZALI: Jurnal Pendidikan Dan Pemikiran Islam, 6(3), 1110–1127. https://doi.org/10.69900/ag.v6i3.626

Hasnan, U., Yazid, M. E. A. El, Kashada, A., Amer, M. A. B., Ullah, R., & Azmi, A. F. S. (2026). Developing Maqasid al-Shariah-Based Artificial Intelligence Models for Transformative Islamic Education in the Era of Technological Disruption and Knowledge Society. AL GHAZALI: Jurnal Pendidikan Dan Pemikiran Islam, 6(1), 363–378. https://doi.org/10.69900/ag.v6i1.604

Hushin, H., Mahmadov, Y., Bolat, M., & Amer, M. A. B. (2025). Developing Critical Thinking Through Islamic Education: A Systematic Review of Contemporary Approaches. Al Husna: Jurnal Ilmiah Pendidikan Agama Islam, 2(1), 43–64. https://doi.org/10.69900/ah.v2i1.21

Imanudin, B. A. (2023). Penilaian autentik terhadap pembelajaran Pendidikan Agama Islam. Raudhah Proud To Be Professionals: Jurnal Tarbiyah Islamiyah, 8(2), 527–539. https://doi.org/10.48094/raudhah.v8i2.310

Jopp, R. (2020). A case study of a technology enhanced learning initiative that supports authentic assessment. Teaching in Higher Education, 25(8), 942–958. https://doi.org/10.1080/13562517.2019.1613637

Khumaeni, A., Andiyani, F., Ningsih, P. L., Sifa, W., Farhan, M., & Ratoni, R. (2026). Digitalizing Menstrual Fiqh Education: The Use of RIMA Application to Support Haid Cycle Monitoring among Female Islamic Boarding School Administrators. AL GHAZALI: Jurnal Pendidikan Dan Pemikiran Islam, 6(5), 1783–1801. https://doi.org/10.69900/ag.v6i5.686

Mukhlis, M., Sabda, S., Hamdi, F., Salamah, S., & Wahyudin, W. (2026). An Epistemological Analysis of the Thought of K.H. Syukri Unus and Its Contribution to the Development of Islamic Education. AL GHAZALI: Jurnal Pendidikan Dan Pemikiran Islam, 6(4), 1505–1523. https://doi.org/10.69900/ag.v6i4.620

Nabriz, A., Azmi, A. F. S., Kashada, A., Ullah, R., & Chandra, A. (2025). The Future of Islamic Education in Europe: Challenges, Opportunities, and Emerging Paradigms. Al Husna: Jurnal Ilmiah Pendidikan Agama Islam, 2(1), 14–42. https://doi.org/10.69900/ah.v2i1.20

Nabriz, A., Azmi, A. F. S., Kashada, A., Ullah, R., & Chandra, A. (2026). Islamic Religious Education and Digital Citizenship: Preparing Muslim Youth for Ethical Participation in the Global Digital Society. Al Husna: Jurnal Ilmiah Pendidikan Agama Islam, 3(1), 1–25. https://doi.org/10.69900/ah.v3i1.18

Nyinge, B. (2025). Examination of the relationship between authentic assessment and generic competencies among undergraduate science student-teachers in Tanzanian universities. Educational Dimension. https://doi.org/10.55056/ed.707

Pertiwi, P., Marzuki, M., & Sabri, T. (2018). Penilaian autentik dalam pembelajaran tematik di Sekolah Dasar Negeri Kecamatan Sungai Ambawang. Jurnal Pendidikan dan Pembelajaran Khatulistiwa, 7(6). https://doi.org/10.26418/jppk.v7i6.25908

Raafi, H., Syarifah, S., Yalipele, A., & Abdillah, H. (2026). Internalization of Disciplinary Values through Reward and Punishment Approaches in Islamic Religious Education Learning at Junior High Schools: A Multisite Study in Ponorogo. AL GHAZALI: Jurnal Pendidikan Dan Pemikiran Islam, 6(4), 1236–1264. https://doi.org/10.69900/ag.v6i4.636

Sa'adah, U., Zaen, E. S., Firdaus, M., & Yakub, M. (2026). Genealogy of Thought of Islamic Da'wah Organizations in Indonesia: A Historical Analysis of NU, Muhammadiyah, Persis, and Al-Washliyah. AL GHAZALI: Jurnal Pendidikan Dan Pemikiran Islam, 6(3), 1015–1038. https://doi.org/10.69900/ag.v6i3.623

Setiawan, D., Maisarah, Yusnadi, & Damanik, M. R. S. (2019). The development of authentic assessment instruments based on characters in measuring students' competence at primary school. In Proceedings of the International Conference on Social Sciences and Interdisciplinary Studies. https://doi.org/10.2991/icssis-18.2019.61

Sirajuddin, S., Wardani, N., & Rahmah, S. (2025). Evaluasi pembelajaran Pendidikan Agama Islam berbasis kinerja dalam pendidikan pesantren. Ta'dib: Jurnal Pendidikan Islam, 14(2), 211–218. https://doi.org/10.29313/tjpi.v14i2.9381

Tamrin, M., Lubis, R. R., Aufa, A., & Harahap, S. A. (2021). Penilaian autentik pada pembelajaran Pendidikan Agama Islam di SMP Negeri 4 Pematangsiantar. Al-Fikru: Jurnal Ilmiah, 15(2), 127–142. https://doi.org/10.51672/alfikru.v15i2.57

Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2018). Authentic assessment: Creating a blueprint for course design. Assessment & Evaluation in Higher Education, 43(5), 840–854. https://doi.org/10.1080/02602938.2017.1412396

Wahidah, W., & Cahyadi, A. (2026). The Role of Technology in Enhancing the Competence of Islamic Education Teachers in the Digital Era. AL GHAZALI: Jurnal Pendidikan Dan Pemikiran Islam, 6(3), 1178–1196. https://doi.org/10.69900/ag.v6i3.605

Wangid, M. N., Mustadi, A., Senen, A., & Herianingtyas, N. L. R. (2017). The evaluation of authentic assessment implementation of Curriculum 2013 in elementary school. Jurnal Penelitian dan Evaluasi Pendidikan, 21(1), 104–115. https://doi.org/10.21831/pep.v21i1.15779

Watling, C. J., & Ginsburg, S. (2019). Assessment, feedback and the alchemy of learning. Medical Education, 53(1), 76–85. https://doi.org/10.1111/medu.13645

Wiewiora, A., & Kowalkiewicz, A. (2019). The role of authentic assessment in developing authentic leadership identity and competencies. Assessment & Evaluation in Higher Education, 44(3), 415–430. https://doi.org/10.1080/02602938.2018.1516730

Yudha, R. P. (2020). Validity and reliability rubric of performance assessment geometry study in junior high school using the Many-Facet Rasch Model approach. EduMa: Mathematics Education Learning and Teaching, 9(2), 26–36. https://doi.org/10.24235/eduma.v9i2.7100