



TEST RELIABILITY AS A FOUNDATION FOR ACCURATE AND OBJECTIVE ASSESSMENT IN ISLAMIC EDUCATION

Mariana^{1*}, & Dina Hermina²

^{*1&2} Program Studi Magister Pendidikan Agama Islam, UIN Antasari Banjarmasin,

*e-mail: nanamariana438@gmail.com

Submit Tgl: 05-Juni-2026

Diterima Tgl: 06-Juni-2026

Diterbitkan Tgl: 07-Juni-2026

Abstrak: Penelitian ini mengkaji peran reliabilitas tes, khususnya metode tes ulang (*test-retest*), dalam memperkuat kualitas evaluasi pembelajaran Pendidikan Agama Islam. Penelitian ini dilatarbelakangi oleh kebutuhan akan instrumen penilaian yang akurat, konsisten, dan dapat dipertanggungjawabkan guna mendukung proses evaluasi pembelajaran yang berkualitas. Berbagai penelitian sebelumnya lebih banyak menyoroti validitas instrumen dan asesmen autentik, sementara pembahasan mengenai reliabilitas tes sebagai fondasi evaluasi yang objektif dalam Pendidikan Agama Islam masih relatif terbatas. Penelitian ini menggunakan pendekatan kualitatif dengan metode studi pustaka. Data diperoleh dari buku, artikel jurnal ilmiah, dan publikasi akademik yang relevan dalam sepuluh tahun terakhir, kemudian dianalisis menggunakan teknik analisis isi. Hasil penelitian menunjukkan bahwa reliabilitas tes berperan penting dalam menjamin konsistensi, kestabilan, dan keterpercayaan hasil penilaian. Selain itu, metode test-retest memberikan kerangka yang efektif untuk menguji kestabilan skor dari waktu ke waktu serta meminimalkan kesalahan pengukuran. Dalam konteks Pendidikan Agama Islam, instrumen yang reliabel mendukung penilaian yang adil, memperkuat pengambilan keputusan berbasis data, serta sejalan dengan nilai-nilai Islam yang menekankan ketelitian, kejujuran, dan tanggung jawab. Penelitian ini menyimpulkan bahwa reliabilitas tes merupakan komponen esensial dalam pengembangan sistem evaluasi pembelajaran yang berkualitas guna mendukung peningkatan hasil belajar dan mutu Pendidikan Agama Islam.

Kata Kunci: Reliabilitas Tes; Metode Test-Retest; Pendidikan Agama Islam; Evaluasi Pembelajaran; Penilaian Pendidikan

Abstract: This study examines the role of test reliability, particularly the test-retest method, in strengthening the quality of assessment within Islamic Education. The study is motivated by the growing need for accurate, consistent, and accountable assessment instruments to support meaningful learning evaluation. While previous studies have extensively discussed validity and authentic assessment, limited attention has been given to the contribution of test reliability as a foundation for objective assessment practices in Islamic Education. This research employed a qualitative approach using library research methods. Data were collected from books, scientific journal articles, and relevant academic publications published within the last ten years and analyzed through content analysis techniques. The findings reveal that test reliability plays a crucial role in ensuring the consistency, stability, and trustworthiness of assessment results. Furthermore, the test-retest method provides an effective framework for evaluating score stability over time and minimizing measurement errors. In the context of Islamic Education, reliable assessment instruments support fair evaluation, strengthen data-based decision-making, and align with Islamic values emphasizing accuracy, honesty, and accountability. The study concludes that test reliability serves as an essential

component in developing high-quality assessment systems capable of supporting effective learning evaluation and improving educational outcomes in Islamic Education.

Keywords: *Test Reliability; Test-Retest Method; Islamic Education; Learning Assessment; Educational Evaluation*

Cara mengutip Mariana, & Hermina, D. (2026). Test Reliability as A Foundation for Accurate and Objective Assessment in Islamic Education. *JIS: Journal Islamic Studies*, 4(2), 296–307. <https://doi.org/10.71456/jis.v4i2.1970>

INTRODUCTION

Assessment is an integral component of the educational process because it serves as the primary basis for measuring learning achievement, determining instructional effectiveness, and supporting educational decision-making. The quality of assessment results is strongly influenced by the quality of the instruments employed. An assessment instrument is expected not only to possess validity but also to demonstrate reliability. Reliability refers to the consistency and stability of measurement results when an instrument is administered repeatedly under relatively similar conditions. Without adequate reliability, assessment outcomes may become inconsistent and lead to inaccurate interpretations of students' learning achievement (Nurrisaa et al., 2025, pp. 98–117).

In contemporary educational practice, considerable attention has been devoted to improving assessment quality through authentic assessment, competency-based evaluation, and higher-order thinking skills (HOTS). However, discussions regarding the reliability of assessment instruments often receive less emphasis compared to validity. Many educators focus primarily on developing test items and aligning them with learning objectives while paying limited attention to whether the instruments consistently measure students' abilities (Hidayati et al., 2025, pp. 6074–6086). Consequently, assessment results may contain measurement errors that reduce the accuracy and credibility of educational evaluations.

The issue of reliability is particularly significant in Islamic Education (Pendidikan Agama Islam/PAI), where assessment encompasses not only cognitive achievement but also affective and behavioral dimensions. The complexity of learning outcomes in PAI requires assessment instruments capable of producing stable and trustworthy results. Reliable instruments contribute to fairer evaluations, accurate identification of students' competencies, and more appropriate educational decisions. Therefore, ensuring the reliability of assessment tools becomes an essential prerequisite for improving the quality of learning evaluation in Islamic Education (Nida et al., 2025, pp. 1–17).

From an Islamic perspective, the principle of reliability is closely associated with the values of accuracy, honesty, and accountability. Islamic teachings emphasize the importance of verification and careful examination before making judgments or decisions. These values are reflected in the Qur'anic command to verify information and uphold justice in all forms of assessment and testimony. In educational contexts, such principles encourage the development of evaluation systems that are objective, consistent, and free from unnecessary bias. Thus, the concept of reliability is not merely a statistical requirement but also aligns with the ethical and epistemological foundations of Islamic education (Azwani Masuwai et al., 2022, pp. 1–17).

The growing demand for quality education in the twenty-first century has increased expectations regarding the effectiveness of assessment practices in Islamic Education. Assessment is no longer viewed merely as a tool for assigning grades but as an integral part of the learning process that provides information about students' progress, strengths, and areas requiring improvement. Consequently, educators are expected to employ assessment instruments capable of generating accurate information that can be used to support instructional decisions and improve learning outcomes. The quality of assessment data therefore becomes an important consideration in determining the effectiveness of educational programs and learning activities.

One of the major challenges in Islamic Education assessment lies in the diversity of competencies that must be evaluated. Learning outcomes in Islamic Education involve not only the acquisition of knowledge but also the development of attitudes, values, and religious practices. Such multidimensional competencies require assessment instruments that are carefully designed and systematically implemented. If assessment instruments fail to provide consistent results, educators may encounter difficulties in accurately identifying students' achievements and determining the extent to which learning objectives have been accomplished. Therefore, consistency in measurement becomes a critical requirement in the assessment process.

In addition, the increasing emphasis on evidence-based educational decision-making highlights the importance of trustworthy assessment results. Educational policies, instructional planning, remedial programs, and student development initiatives often rely on information derived from assessment outcomes. Decisions based on inconsistent or unreliable data may reduce the effectiveness of educational interventions and hinder efforts to improve learning quality. Accordingly, strengthening the quality of assessment instruments represents an essential step toward establishing a more effective, transparent, and accountable evaluation system in Islamic Education.

Although previous studies have examined educational measurement, test validity, and assessment quality, limited attention has been given to the role of test reliability, particularly the test-retest method, as a framework for strengthening assessment practices in Islamic Education. Therefore, this study aims to analyze the concept of test reliability, explore the implementation of the test-retest method, and examine its contribution to improving the quality of learning assessment in Islamic Education. The findings are expected to provide both theoretical and practical insights for educators, researchers, and assessment developers in designing more reliable and accountable evaluation instruments.

RESEARCH METHOD

Research Design

This study employed a qualitative approach using library research. Library research is a method of inquiry that relies on various forms of literature, including books, scientific journal articles, research reports, conference proceedings, and other relevant academic publications. This approach was selected because the focus of the study is to examine the concept of test reliability, particularly the test-retest method, and its contribution to improving assessment quality in Islamic Education. Through a literature-based approach, the study seeks to develop a comprehensive understanding of theoretical perspectives, empirical findings, and educational implications related to reliability in learning assessment.

Data Sources

The data sources consisted of primary and secondary literature published within the last ten years. Primary sources included scholarly articles discussing educational measurement, test reliability, test–retest procedures, and assessment practices in Islamic Education. Secondary sources comprised textbooks, reference books, policy documents, and supporting academic publications relevant to educational evaluation and Islamic educational studies. Data were collected through documentation techniques by identifying, selecting, reviewing, and organizing literature that addressed the research focus. To ensure relevance and credibility, the selected sources were limited to peer-reviewed publications and authoritative academic references.

Data Analysis Technique

Data analysis was conducted using content analysis. The collected literature was systematically examined to identify key concepts, recurring themes, theoretical frameworks, and research findings related to test reliability and the test–retest method. The analysis process involved data reduction, data categorization, interpretation, and synthesis of findings to generate a coherent understanding of the subject under investigation. Furthermore, the study employed a comparative and analytical approach to relate the concept of reliability to the context of Islamic Education assessment. The results of the analysis were then presented descriptively to provide a comprehensive explanation of how reliable assessment instruments contribute to accurate, objective, and accountable learning evaluation practices.

RESULT AND DISCUSSION

The findings of this study indicate that test reliability constitutes a fundamental component in ensuring the quality of educational assessment. Through the analysis of various scholarly sources, reliability is not merely understood as a statistical coefficient but as an indicator of the consistency, stability, and trustworthiness of assessment results. In the context of Islamic Education, reliable assessment instruments play a strategic role in supporting objective evaluation, minimizing measurement errors, and ensuring fairness in determining students' learning achievement. Furthermore, the findings reveal that the test–retest method provides an effective approach for examining score stability over time, thereby strengthening the credibility of assessment outcomes. Based on the analysis, four major themes emerged, namely the conceptual foundation of test reliability, the implementation of the test–retest method, the contribution of reliability to the quality of Islamic Education assessment, and the relevance of reliability to Islamic educational values.

Test Reliability as the Foundation of Quality Assessment in Islamic Education

Reliability is one of the most important indicators of assessment quality because it reflects the consistency of measurement results obtained from an instrument. A reliable instrument is capable of producing relatively stable scores when administered repeatedly under similar conditions. Within educational measurement theory, reliability is closely related to the extent to which an instrument is free from random measurement errors. The lower the measurement error, the higher the reliability coefficient, indicating greater confidence in the assessment results. Therefore, reliability serves as a prerequisite for generating trustworthy educational data and supporting accurate evaluation practices (Rachmawati & Siswanto, 2025, pp. 26–42).

In Islamic Education (PAI), the importance of reliability becomes increasingly significant because assessment is not limited to measuring cognitive achievement but also encompasses affective, moral, and behavioral dimensions. The multidimensional nature of PAI learning outcomes requires assessment instruments that consistently measure students' competencies without being influenced by irrelevant factors. Instruments with low reliability may lead to inconsistent judgments regarding students' mastery of religious knowledge, attitudes, and practices. Consequently, educational decisions based on such assessments may not accurately represent students' actual competencies (Bahsoan et al., 2025, pp. 121–137).

The findings of this study suggest that reliability contributes directly to the objectivity and fairness of assessment practices. Reliable instruments enable educators to distinguish between actual differences in student achievement and variations caused by measurement error (Prihono et al., 2025, pp. 129–141). This function is particularly important in Islamic Education, where assessment results are frequently used as the basis for determining learning progress, providing feedback, and planning instructional improvements. Through reliable assessment procedures, educators can make more informed and accountable decisions regarding students' educational development.

From an Islamic educational perspective, the principle of reliability aligns with the values of accuracy (*al-diqah*), honesty (*al-shidq*), and responsibility (*al-amanah*). Islamic teachings encourage careful verification before making judgments and emphasize the importance of fairness in evaluating others. These principles support the development of assessment systems that produce consistent and dependable results. Therefore, reliability should not be viewed solely as a technical requirement in educational measurement but also as an ethical commitment to ensuring justice and accountability in the evaluation process. Such a perspective reinforces the relevance of reliability as a foundational element in improving the quality of assessment within Islamic Education.

The Implementation of the Test–Retest Method in Islamic Education Assessment

The test–retest method is one of the most widely used approaches for examining the stability of an assessment instrument over time. This method involves administering the same instrument to the same group of respondents on two separate occasions and subsequently correlating the results. A high correlation coefficient indicates that the instrument consistently measures the intended construct and produces stable outcomes. Consequently, the test–retest method serves as an important procedure for determining whether assessment results can be trusted as a representation of students' actual competencies rather than being influenced by temporary or random factors.

In the context of Islamic Education, the test–retest method can be applied to various forms of assessment, particularly those designed to measure relatively stable aspects of learning. These include students' understanding of Islamic concepts, knowledge of religious teachings, attitudes toward religious values, and learning motivation. Before an instrument is implemented on a larger scale, educators may conduct a pilot test and subsequently administer the same instrument after a predetermined interval. The resulting scores are then analyzed to determine the degree of consistency achieved by the instrument. Such procedures contribute to the development of more dependable assessment tools within Islamic Education (Alfiyah et al., 2025, pp. 2672–2679).

Despite its advantages, the implementation of the test–retest method requires careful

consideration of several methodological factors. One important factor is the time interval between the first and second administration. If the interval is too short, respondents may remember their previous answers, potentially inflating the reliability coefficient. Conversely, if the interval is too long, genuine changes in students' knowledge, attitudes, or abilities may occur, leading to score differences unrelated to instrument quality. Therefore, selecting an appropriate time interval is essential for obtaining an accurate estimate of reliability. Researchers commonly recommend intervals ranging from one to several weeks, depending on the characteristics of the construct being measured.

The findings of this study indicate that the test–retest method offers significant benefits for improving the quality of assessment practices in Islamic Education. Through repeated measurement, educators can identify weaknesses within assessment instruments and make necessary revisions before using them for formal evaluation purposes. Moreover, the method provides empirical evidence regarding the consistency of assessment outcomes, thereby enhancing confidence in educational decisions derived from those results. As a result, the implementation of the test–retest method contributes not only to the technical quality of assessment instruments but also to the overall credibility and accountability of learning evaluation in Islamic Education (Islamaya & Putranta, 2025, pp. 31–60).

Furthermore, the application of the test–retest method in Islamic Education can assist teachers in evaluating the effectiveness of newly developed assessment instruments before they are implemented in actual classroom settings. Instruments designed to measure students' understanding of Islamic teachings, religious attitudes, or worship practices require careful examination to ensure that they consistently capture the intended learning outcomes. By conducting repeated testing procedures, educators can obtain empirical evidence regarding instrument stability and make necessary refinements to improve measurement quality. This process contributes to the development of assessment tools that are both scientifically sound and educationally meaningful.

The use of the test–retest method also supports the professional development of Islamic Education teachers in the field of educational assessment. Through engagement with reliability testing procedures, teachers become more aware of the importance of instrument quality and data accuracy in the evaluation process. Such awareness encourages educators to move beyond conventional assessment practices and adopt a more systematic approach to measuring student achievement. Consequently, assessment activities are transformed from routine administrative tasks into evidence-based practices that support continuous improvement in teaching and learning.

From a broader educational perspective, the implementation of the test–retest method reflects a commitment to quality assurance in Islamic Education assessment. Reliable instruments provide dependable information that can be used to evaluate learning outcomes, monitor student progress, and support educational decision-making. In addition, the availability of consistent assessment data strengthens the credibility of evaluation results among teachers, students, parents, and educational institutions. Therefore, the test–retest method should be viewed not only as a statistical procedure for estimating reliability but also as a practical strategy for enhancing the overall quality and effectiveness of assessment practices in Islamic Education.

The Implementation of the Test–Retest Method in Islamic Education Assessment

Assessment in Islamic Education encompasses broader dimensions than merely measuring cognitive achievement. In addition to evaluating students' understanding of Islamic concepts, educators are also expected to assess religious attitudes, moral behavior, and practical worship competencies. These multidimensional learning outcomes require assessment instruments capable of producing consistent and trustworthy results. Consequently, test reliability becomes a fundamental requirement for ensuring that evaluation outcomes accurately represent students' competencies in Islamic Education rather than being influenced by random measurement errors.

The contribution of reliability is particularly evident in the assessment of cognitive competencies within Islamic Education. Subjects such as Aqidah Akhlak, Fiqh, Al-Qur'an Hadith, and Islamic Cultural History require instruments that consistently measure students' comprehension of concepts, principles, and values. Reliable tests enable educators to determine whether differences in students' scores genuinely reflect variations in learning achievement. Without adequate reliability, assessment results may fluctuate and reduce the accuracy of conclusions regarding students' mastery of Islamic knowledge (Fatmawati, 2025, pp. 109–119).

Beyond the cognitive domain, reliability is equally important in assessing affective aspects of Islamic Education. One of the primary objectives of PAI is the development of religious attitudes, honesty, discipline, responsibility, tolerance, and social awareness. These aspects are commonly measured through observation sheets, attitude scales, self-assessment, and peer assessment instruments (Jannah et al., 2025, pp. 131–143). If such instruments lack reliability, the resulting evaluations may become inconsistent and vulnerable to subjective judgments. Therefore, reliable assessment tools are necessary to ensure that affective evaluations accurately reflect students' character development and internalization of Islamic values.

Reliability also contributes significantly to the assessment of psychomotor competencies, particularly those related to religious practices. In Islamic Education, students are often evaluated on their ability to perform acts of worship such as ablution (wudu), prayer (salat), Qur'anic recitation, and other practical religious activities. These assessments frequently involve performance-based evaluation and teacher observation. Reliable assessment rubrics help maintain consistency in scoring criteria, thereby reducing variations in judgment and ensuring that students are evaluated according to the same standards. As a result, assessment outcomes become more objective and equitable (Fatiha et al., 2026, pp. 195–207).

From an Islamic perspective, reliable assessment instruments support the implementation of justice (al-'adl) in educational evaluation. Islam emphasizes fairness and objectivity in making judgments, including decisions related to learning achievement. Assessment results that are inconsistent or influenced by measurement errors may lead to unfair treatment of students. Conversely, reliable instruments help educators provide evaluations based on actual performance and demonstrated competencies. In this regard, reliability serves not only as a methodological requirement but also as an ethical principle that aligns with the values of justice promoted in Islamic teachings.

Furthermore, the use of reliable assessment instruments reflects the principles of amanah (trustworthiness) and professionalism expected of Islamic Education teachers.

Educators bear the responsibility of ensuring that every evaluation result is accurate, transparent, and accountable. Reliable assessment practices strengthen confidence in educational decisions, support evidence-based learning improvement, and enhance the overall quality of Islamic Education. Therefore, test reliability should be viewed as an essential component of effective assessment systems that promote fairness, accountability, and continuous improvement in Islamic Education learning outcomes.

Reliability and Islamic Educational Values: Accuracy, Justice, and Accountability

The concept of reliability in educational assessment is not merely a technical requirement in measurement theory but also possesses strong relevance to the philosophical foundations of Islamic Education. Reliability emphasizes consistency, stability, and trustworthiness in the assessment process, values that are closely aligned with Islamic teachings concerning truth, justice, and responsibility. In the context of Islamic Education, assessment is not solely intended to measure learning outcomes but also to ensure that educational decisions are made based on accurate and dependable information. Therefore, reliability occupies an important position both from methodological and ethical perspectives.

One of the Islamic values closely related to reliability is accuracy (*al-diqah*). Islamic teachings encourage believers to verify information before accepting or acting upon it. This principle is reflected in the Qur'anic command to investigate the truth of information to avoid making judgments based on uncertain evidence. Within educational assessment, reliability functions as a mechanism for ensuring that the information obtained from students' scores is accurate and consistent. Assessment results that lack reliability may lead educators to draw incorrect conclusions regarding students' competencies and learning progress.

The principle of accuracy is particularly important in Islamic Education because learning outcomes often involve dimensions that are difficult to measure directly. Teachers are expected not only to assess students' mastery of Islamic knowledge but also their understanding, attitudes, and application of religious values in everyday life (Putra et al., 2025, pp. 198–213). Such complex learning outcomes require instruments capable of producing dependable results. Reliable assessment tools help educators obtain more accurate representations of students' competencies and minimize the possibility of misjudgment.

Another fundamental Islamic value reflected in reliability is justice (*al-'adl*). Justice occupies a central position in Islamic teachings and serves as a guiding principle in human interactions, including educational practices. In assessment activities, justice requires that all students be evaluated according to the same standards and criteria. Reliable instruments contribute to the realization of this principle by reducing inconsistencies that may arise from subjective judgments or measurement errors. Consequently, students receive evaluations based on their actual performance rather than external or irrelevant factors (Ulya et al., 2024, pp. 306–312).

In Islamic Education classrooms, fairness in assessment is essential because evaluation outcomes influence students' academic development, motivation, and self-confidence. Inconsistent assessment practices may create perceptions of bias and reduce students' trust in the educational process. Reliability helps prevent such problems by ensuring that assessment results remain stable across different occasions and conditions. As

a result, educational decisions become more equitable and acceptable to all stakeholders involved in the learning process.

The value of justice is also closely related to the responsibility of teachers as evaluators. Teachers are entrusted with the authority to assess students' learning achievements and determine whether educational objectives have been attained. This responsibility requires assessment procedures that are objective and scientifically sound. Reliable instruments provide a stronger foundation for educational judgments and help ensure that assessment outcomes reflect genuine learning achievement. Therefore, reliability supports the implementation of justice not only for students but also for teachers in carrying out their professional duties (Rasyidah & Ramadhan, 2025, pp. 142–162).

In addition to accuracy and justice, reliability reflects the Islamic principle of *amanah* (trustworthiness). *Amanah* refers to the responsibility to perform duties honestly, correctly, and conscientiously. Within educational settings, teachers are entrusted with evaluating students fairly and reporting assessment results accurately. The use of reliable assessment instruments demonstrates a commitment to fulfilling this trust because evaluation outcomes are generated through systematic and dependable procedures rather than arbitrary judgments.

The concept of *amanah* becomes increasingly important in Islamic Education because teachers serve not only as instructors but also as moral exemplars. Their assessment practices communicate important ethical values to students. When teachers employ reliable and transparent evaluation procedures, they model integrity, professionalism, and accountability. Such practices contribute to the formation of a learning culture grounded in honesty and mutual trust, which are essential objectives of Islamic education.

Reliability is also associated with the value of accountability (*mas'uliyah*). Accountability requires that every educational decision be supported by valid and trustworthy evidence. Assessment results often serve as the basis for determining grades, evaluating learning achievement, identifying learning difficulties, and planning instructional improvements. Reliable instruments strengthen the credibility of these decisions by ensuring that the underlying data are consistent and dependable. Consequently, educators can justify their decisions both professionally and ethically.

From an institutional perspective, reliable assessment contributes to the overall quality assurance of Islamic Education. Educational institutions increasingly emphasize evidence-based evaluation and continuous improvement. Reliable instruments provide accurate information regarding students' achievement and program effectiveness, enabling schools and educational administrators to make informed decisions. In this way, reliability supports not only individual assessment practices but also broader efforts to enhance educational quality and accountability.

Furthermore, the integration of reliability principles into Islamic Education assessment reflects the harmonious relationship between contemporary educational measurement and Islamic educational philosophy. While reliability originates from modern psychometric theory, its underlying emphasis on consistency, accuracy, and trustworthiness resonates strongly with Islamic values. This compatibility demonstrates that educational measurement can be implemented in ways that uphold both scientific rigor and religious principles. As a result, reliability becomes a bridge connecting methodological excellence with ethical responsibility.

Based on the findings of this study, reliability should be understood as more than a statistical indicator of assessment quality. In Islamic Education, reliability represents a manifestation of accuracy, justice, trustworthiness, and accountability in the evaluation process. The development and implementation of reliable assessment instruments enable educators to make fair and informed decisions while remaining consistent with the moral foundations of Islamic teachings. Therefore, strengthening reliability practices is an essential step toward improving the quality, credibility, and ethical integrity of assessment in Islamic Education

CONCLUSION

This study concludes that test reliability is a fundamental component in ensuring the quality of assessment within Islamic Education. Reliable assessment instruments produce consistent and stable results, thereby enhancing the accuracy, objectivity, and credibility of learning evaluation. The findings indicate that the test–retest method serves as an effective approach for examining score stability over time and identifying the consistency of assessment instruments. In the context of Islamic Education, reliability supports the evaluation of cognitive, affective, and psychomotor competencies by minimizing measurement errors and ensuring that assessment outcomes accurately reflect students' actual learning achievements.

Furthermore, the concept of reliability is closely aligned with the core values of Islamic education, including accuracy (al-diqaq), justice (al-'adl), trustworthiness (amanah), and accountability (mas'uliyah). Reliable assessment practices enable educators to make fair and evidence-based decisions while fulfilling their professional and ethical responsibilities. Therefore, strengthening the reliability of assessment instruments should become a priority in the development of learning evaluation systems within Islamic Education. Such efforts are expected to contribute to the improvement of assessment quality, educational accountability, and the achievement of more meaningful learning outcomes.

REFERENCES

- Alfiyah, N., Zuhriyah, I. A., & Chalishah, N. D. (2025). Validity and Reliability Analysis: Measuring the Accuracy and Consistency of Instruments for the Evaluation of Islamic Religious Education Tests. *JIIP - Jurnal Ilmiah Ilmu Pendidikan*, 8(3), 2672–2679. <https://doi.org/10.54371/jiip.v8i3.7230>
- Azwani Masuwai, Hafizhah Zulkifli, & Ab Halim Tamuri. (2022). Systematic Literature Review on Self-Assessment Inventory for Quality Teaching among Islamic Education Teachers. *Sustainability (Switzerland)*, 12(203), 1–17.
- Bahsoan, A., Tine, N., Santoso, I. R., Neva Lionitha Ibrahim, & Branislav Pupala. (2025). Rasch-Based Cognitive Diagnostic Assessment to Measure Critical Thinking Skills in Islamic Religious Education. *Jurnal Iqra': Kajian Ilmu Pendidikan*, 10(3), 121–137. <https://doi.org/10.25217/ji.v10i3.6648>
- Fatiha, J. I., Rafik, A., & Muhith, A. (2026). Basic Concepts of Learning Evaluation in the Perspective of the Qur'an and Hadith. *Al Qalam: Jurnal Ilmiah Keagamaan Dan Kemasyarakatan*, 20(1), 195–207. <https://doi.org/10.35931/aq.v20i1.5631>

- Fatmawati, E. (2025). Reconceptualizing Assessment in Islamic Education: A Critical Review of Madrasah Evaluation Practices in the 21st Century. *Journal of Quality Assurance in Islamic Education (JQAIE)*, 5(2), 109–119. <https://doi.org/10.47945/jqaie.v5i2.2376>
- Hidayati, Z., Nurdi, N., Kusaeri, K., Suparto, S., & Malyuna, S. I. (2025). Assessing the Assessor: Investigating Formative Assessment Literacy Among Islamic Education Teachers in Lamongan, Indonesia. *AL-ISHLAH: Jurnal Pendidikan*, 17(4), 6074–6086. <https://doi.org/10.35445/alishlah.v17i4.7933>
- Islamaya, Y., & Putranta, H. (2025). FEASIBILITY TEST ANALYSIS OF MINIMUM COMPETENCY ASSESSMENT (AKM) ISLAMIC RELIGIOUS EDUCATION SUBJECTS ON MARRIAGE. *SUKMA: JURNAL PENDIDIKAN*, 9(1), 31–60.
- Jannah, N., Rosyadi, M., Mislikhah, S., & Hadi, S. (2025). Holistic Evaluation of Islamic Education Programs: Implications of the Kirkpatrick Model in Character-Based Learning. *FALASIFA: Jurnal Studi Keislaman*, 16(02), 131–143. <https://doi.org/10.62097/falasifa.v16i02.2478>
- Nida, S., Rahmawati, N. W., Hidayat, Y., Mandar, Y., Hidayah, A. L., & Damayanti, N. (2025). Innovative Attitude Assessment in Islamic Education: Techniques, Instruments, and Technology Integration from Primary to Higher Education. *HEUTAGOGIA: Journal of Islamic Education*, 5(1), 1–17. <https://doi.org/10.14421/hjie.2025.51-01>
- Nurrisaa, F., Sabdab, S., & Salamahc. (2025). Konsep Kedudukan dan Peran Masyarakat Sebagai Landasan Pendidikan Islam Jurnal Teknologi Pendidikan Dan Pembelajaran (JTPP). *Jurnal Teknologi Pendidikan Dan Pembelajaran (JTPP)*, 02(03), 845–851.
- Prihono, E. W., Latuapo, R., Lapele, F., Dwiningrum, S. I. A., Arlinwibowo, J., & Reyes Jr., M. S. (2025). Assessment and measurement bias in madrasa performance evaluation: Evidence from underdeveloped areas in Indonesia. *REID (Research and Evaluation in Education)*, 11(2), 129–141. <https://doi.org/10.21831/reid.v11i2.71444>
- Putra, A., Efendi, R., Putra, P. H., & Ramadani, L. (2025). Philosophical Dimensions of Artificial Intelligence in Islamic Religious and Legal Education. *Islamika: Jurnal Ilmu-Ilmu Keislaman*, 25(2), 198–213.
- Rachmawati, E., & Siswanto, S. (2025). Developing an instrument to measure students' anti-corruption character using modern test theory. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 29(1), 26–42. <https://doi.org/10.21831/pep.v29i1.84560>
- Rasyidah, N., & Ramadhan, M. F. (2025). Fostering Ethical Values in Islamic Education through Arabic Language Learning: A Conceptual Integration. *Jurnal Cendekia Media Komunikasi Penelitian Dan Pengembangan Pendidikan Islam*, 17(01), 142–162. <https://doi.org/10.37850/cendekia.v17i01.770>
- Ulya, V. F., Afriyanto, F., Suparto, Kusaeri, & Dedduang, V. (2024). MANY-FACET RASCH MODEL (MFRM) FOR DIGITAL CITIZENSHIP MEASUREMENT

ANALYSIS IN ISLAMIC SCHOOL: A BIBLIOMETRIC REVIEW. *MUMTAZ : Jurnal Pendidikan Agama Islam* Volume, 2(3), 306–312.