



ISLAMICA : Jurnal Ilmu-Ilmu Agama Islam

Vol. 10 No. 2, Juli - Desember 2026

p-ISSN: 2085-5982, e-ISSN: 2986-4836

Journal Home Page: <https://journal.stai-siliwangi.ac.id/index.php/islamica/index>

Doi: [10.59908/islamica.v10i2.205](https://doi.org/10.59908/islamica.v10i2.205)

Hlm. 115 - 128



This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2026 by the author

Submitted: 01-02-2026 Revised: 30-06-2026 Accepted: 26-09-2026 Published: 02-10-2026

Validitas Instrumen Penilaian Pendidikan Agama Islam dalam Mengukur Pemahaman Keagamaan Peserta Didik

Salmaa Dewi Anjeli^{1*}, Helmi Fitriansyah², Annisa Sri Rahayu³, Hamida Farkhati⁴,
Amalia Muazzah Adawiah⁵, Nurhasanah Nurhasanah⁶

^{1, 3, 4, 6}STAI Siliwangi Bandung, Indonesia

²Institut Daarul Qur'an Jakarta, Indonesia

⁵Universitas Sultan Ageng Tirtayasa, Indonesia

*Correspondence: sdewianjeli@gmail.com

Abstrak: Penilaian Pendidikan Agama Islam (PAI) tidak hanya berfungsi untuk mengukur penguasaan pengetahuan keagamaan, tetapi juga seharusnya mampu merepresentasikan kemampuan peserta didik dalam memahami, memaknai, dan mengontekstualisasikan ajaran Islam dalam kehidupan. Penelitian ini bertujuan menganalisis validitas instrumen penilaian PAI dan menilai sejauh mana kelayakan psikometriknya dapat merepresentasikan pemahaman keagamaan peserta didik. Penelitian menggunakan jenis data kuantitatif dengan metode penelitian deskriptif-evaluatif. Instrumen berupa 30 butir tes objektif PAI diujicobakan pada peserta didik sekolah dasar dan dianalisis menggunakan ANATES. Hasil penelitian menunjukkan bahwa 20 butir (66,67%) memenuhi kriteria validitas, sedangkan 10 butir (33,33%) memerlukan revisi. Koefisien reliabilitas sebesar 0,68 berada pada kategori sedang. Sebanyak 19 butir (63,33%) memiliki daya pembeda yang memadai, sedangkan 11 butir (36,67%) masih memiliki daya pembeda yang rendah. Tingkat kesukaran terdiri atas 12 butir mudah (40,00%), 14 butir sedang (46,67%), dan 4 butir sukar (13,33%). Temuan tersebut menunjukkan kelayakan psikometrik yang cukup, tetapi validitas statistik belum sepenuhnya sejalan dengan validitas substantif dalam mengukur pemahaman keagamaan. Tes objektif lebih kuat merepresentasikan pengetahuan faktual-konseptual dibandingkan dengan kemampuan interpretatif, reflektif, kontekstual, dan internalisasi nilai. Penelitian ini menawarkan validitas integratif penilaian PAI melalui keterkaitan kualitas psikometrik, keterwakilan konstruk, kesesuaian inferensi, dan triangulasi bukti untuk menghasilkan representasi pemahaman keagamaan yang lebih utuh.

Kata kunci: *validitas instrumen; Pendidikan Agama Islam; pemahaman keagamaan; asesmen autentik; literasi keagamaan.*

Abstract: Islamic Religious Education (IRE) assessment should not only measure students' mastery of religious knowledge but also represent their ability to understand, interpret, and contextualise Islamic teachings in everyday life. This study aims to analyse the validity of IRE assessment instruments and

examine the extent to which their psychometric adequacy reflects students' religious understanding. The study employs a quantitative approach with a descriptive-evaluative design. The researchers administered an objective IRE test to elementary school students and analysed it using ANATES to determine item validity, reliability, discriminating power, and difficulty level. The findings indicate that most items demonstrate adequate validity, with 20 of 30 items (66.67%) considered suitable for use, while the instrument's reliability coefficient ($r = 0.68$) falls within the moderate category. Nineteen items (63.33%) show adequate discriminating power, whereas 11 items (36.67%) require revision. The difficulty distribution consists of 12 easy items (40.00%), 14 moderate items (46.67%), and 4 difficult items (13.33%). These findings indicate adequate psychometric properties for measuring learning outcomes; however, statistical validity does not fully correspond to substantive validity in assessing religious understanding. Objective tests are more effective in representing factual and conceptual knowledge than interpretive, reflective, contextual, and value-internalisation abilities. This study proposes an integrative validity framework that combines psychometric quality, construct representation, inference appropriateness, and evidence triangulation to provide a more comprehensive representation of students' religious understanding.

Keywords: *instrument validity; Islamic Religious Education; religious understanding; authentic assessment; religious literacy.*

Pendahuluan

Penilaian merupakan bagian integral dari pembelajaran Pendidikan Agama Islam (PAI) karena memberikan informasi mengenai ketercapaian tujuan pendidikan sekaligus menjadi dasar untuk memperbaiki proses pembelajaran. Kedudukan penilaian dalam PAI memiliki kompleksitas tersendiri karena objek yang hendak dikembangkan tidak hanya berkaitan dengan kemampuan peserta didik untuk mengetahui ajaran Islam, tetapi juga memahami maknanya, membangun kesadaran nilai, dan menghubungkan ajaran tersebut dengan kehidupan. Arfah menjelaskan bahwa evaluasi PAI memiliki fungsi yang lebih luas daripada sekadar mengukur hasil belajar karena berkaitan dengan ketercapaian tujuan pendidikan Islam secara menyeluruh.¹ Konsekuensinya, keberhasilan PAI tidak seharusnya direduksi menjadi skor yang hanya merepresentasikan jumlah jawaban benar peserta didik.

Kompleksitas penilaian tersebut berkaitan dengan karakter kompetensi yang dikembangkan dalam PAI. Pengetahuan mengenai akidah, ibadah, akhlak, Al-Qur'an, hadis, fikih, dan sejarah Islam merupakan fondasi yang harus dikuasai peserta didik, tetapi pengetahuan tersebut memperoleh signifikansi pendidikan ketika dapat dipahami, dimaknai, dan digunakan sebagai dasar pertimbangan dalam kehidupan. Prihatin dan Hamami menunjukkan bahwa model penilaian PAI perlu bergerak melampaui *assessment of learning* dengan mengintegrasikan *assessment for learning*, *assessment as learning*, dan asesmen autentik.² Orientasi tersebut menempatkan penilaian bukan hanya sebagai aktivitas untuk menentukan hasil akhir, melainkan sebagai bagian dari proses pembentukan kompetensi keagamaan.

Kebutuhan akan penilaian yang komprehensif semakin penting karena penggunaan satu jenis instrumen berpotensi menghasilkan gambaran yang parsial mengenai kemampuan peserta didik. Abdurahman dan Erihadiana membedakan fungsi tes dan non-tes dalam evaluasi pendidikan Islam. Tes efektif untuk memperoleh informasi mengenai kompetensi yang dapat direpresentasikan melalui respons terhadap pertanyaan atau tugas, sedangkan teknik non-tes seperti observasi, wawancara, skala sikap, dan laporan personal dapat

¹ M. Arfah Arfah, "Evaluasi Hasil Belajar Pendidikan Agama Islam (PAI)," *Jurnal Literasiologi* 7, no. 2 (October 13, 2021): 1–14, <https://doi.org/10.47783/literasiologi.v7i2.282>.

² Rani Putri Prihatin and Tasman Hamami, "Learning Assessment Model for Islamic Religious Education," *Nusantara: Jurnal Pendidikan Indonesia* 2, no. 2 (May 31, 2022): 373–90, <https://doi.org/10.14421/njpi.2022.v2i2-10>.

memberikan informasi mengenai aspek yang tidak mudah direpresentasikan melalui tes tertulis.³ Kajian tersebut memperkuat argumen bahwa penggunaan tes dan nontes perlu ditempatkan secara komplementer, bukan saling bertentangan. Artikel mereka memang secara eksplisit membahas keseimbangan antara kedua pendekatan tersebut dalam evaluasi pendidikan Islam. *eJournal of Sunan Gunung Djati Praktik evaluasi PAI* pada kenyataannya masih banyak bertumpu pada tes tertulis. Atika et al. menunjukkan bahwa analisis tes formatif PAI tetap memerlukan pemeriksaan kualitas butir agar hasil pengukuran dapat dipertanggungjawabkan.⁴ Penggunaan tes memberikan keuntungan berupa efisiensi, objektivitas dalam penskoran, dan kemungkinan melakukan analisis statistik terhadap karakteristik butir. Febrianti et al. juga menunjukkan bahwa tes harian tetap memiliki kedudukan penting dalam evaluasi PAI dan berkaitan dengan dinamika pembelajaran peserta didik.⁵ Tes karena itu tidak perlu ditinggalkan, tetapi interpretasi terhadap skor yang dihasilkannya perlu ditempatkan secara proporsional.

Persoalan muncul ketika keberhasilan statistik instrumen diperlakukan sebagai bukti bahwa instrumen tersebut mampu mengukur keseluruhan pemahaman keagamaan. Instrumen dapat memiliki validitas butir dan reliabilitas yang memadai, tetapi tetap hanya merepresentasikan sebagian dari konstruk yang seharusnya diukur. Kondisi tersebut relevan dengan instrumen penelitian ini, yaitu tes objektif yang dianalisis menggunakan validitas, reliabilitas, daya pembeda, dan tingkat kesukaran. Sebagian besar butir dapat digunakan, tetapi pendekatan tersebut pada dasarnya lebih kuat dalam menjelaskan bagaimana instrumen bekerja secara statistik daripada seberapa lengkap instrumen tersebut merepresentasikan pemahaman keagamaan. Perbedaan tersebut menjadi penting ketika pemahaman keagamaan dipahami sebagai konstruk multidimensional. Diskursus mengenai religious and interreligious literacy dan diskursus asesmen hasil pendidikan agama selama ini sering berkembang secara terpisah, padahal keduanya membutuhkan landasan teoretis dan bukti empiris yang saling terkait. Fondasi empiris mengenai hubungan antara literasi keagamaan dan asesmen pendidikan agama masih relatif terbatas. Temuan tersebut memberikan dasar teoretis yang penting untuk mempertanyakan apakah penilaian PAI yang valid secara psikometrik secara otomatis valid dalam merepresentasikan pemahaman keagamaan.

Konsep religious literacy memperluas pengertian kompetensi keagamaan dari sekadar mengetahui fakta agama menjadi kemampuan untuk memahami agama dalam konteksnya. Amina dan Lakhania mengembangkan kembali konsep literasi keagamaan dengan menekankan perlunya kemampuan untuk memahami agama secara lebih kompleks dalam konteks pendidikan.⁶ BinTaleb juga menunjukkan bahwa literasi keagamaan relevan dalam pendidikan formal karena memberikan ruang bagi peserta didik untuk memahami agama

³ Asep Abdurahman and Mohamad Erihadiana, "Comprehensive Assessment: Balancing Test and Non-Test Methods in Islamic Education Evaluation," *Jurnal Inovasi Pendidikan Agama Islam (JIPAI)* 4, no. 2 (December 26, 2024): 95–108, <https://doi.org/10.15575/jipai.v4i2.40752>.

⁴ Yova Atika et al., "Formative Test Analysis Of Islamic Religious Education Learning Evaluation Practices (Study At SMP Negeri 1 Muara Kemumu)," *Indonesian Journal of Pedagogy and Teacher Education* 1, no. 2 (December 31, 2023): 38–42, <https://doi.org/10.58723/ijopate.v1i2.113>.

⁵ Evi Febrianti et al., "Implementation of Islamic Religious Education Learning Evaluation Through Daily Tests on Students' Learning Motivation UPTD SMP Negeri 4 Parepare," *Indonesian Journal of Interdisciplinary Research in Science and Technology* 1, no. 9 (October 31, 2023): 823–38, <https://doi.org/10.55927/marcopolo.v1i9.6588>.

⁶ Anar Amin and Anisha Lakhani, "Redefining Religious Literacy in Secondary Mainstream United Kingdom Classrooms," *Journal of Beliefs & Values* 45, no. 2 (April 2, 2024): 180–97, <https://doi.org/10.1080/13617672.2024.2309746>.

secara lebih kontekstual.⁷ Dalam konteks Indonesia, Hidayah et al. menghubungkan literasi keagamaan peserta didik dengan kemampuan mereka untuk memahami keberagaman dalam kehidupan sosial.⁸ Perspektif ini menunjukkan bahwa pemahaman keagamaan tidak identik dengan jumlah informasi keagamaan yang dapat diingat.

Masalah literasi keagamaan semakin kompleks di masyarakat digital. Ashari, Faizin, dan Shiddiq menunjukkan bahwa religious digital literacy menjadi kompetensi penting bagi peserta didik di Indonesia dan Malaysia karena sumber informasi agama semakin tersebar melalui media digital.⁹ Noprijon et al. juga menunjukkan bahwa digitalisasi pendidikan memiliki konsekuensi terhadap pengembangan literasi keagamaan.¹⁰ Peserta didik sekarang tidak hanya membutuhkan kemampuan untuk mengetahui ajaran agama, tetapi juga kemampuan untuk memahami, menilai, dan menempatkan informasi keagamaan yang diperoleh dari berbagai sumber. Sehingga, instrumen PAI yang hanya mengukur reproduksi informasi memiliki keterbatasan dalam merepresentasikan kompetensi keagamaan kontemporer.

Pengembangan asesmen autentik merupakan salah satu respons terhadap keterbatasan tersebut. Baroroh dan Hamami menegaskan bahwa asesmen autentik PAI di sekolah dasar perlu mencakup ranah kognitif, afektif, dan psikomotorik.¹¹ Penelitian tersebut menunjukkan bahwa asesmen autentik berusaha menghasilkan bukti yang lebih luas mengenai hasil pembelajaran dibandingkan dengan tes kognitif saja. Imanudin juga menunjukkan bahwa asesmen autentik memberikan kesempatan untuk menilai kemampuan peserta didik melalui tugas yang lebih dekat dengan konteks penggunaan kompetensi.¹² Namun, asesmen autentik tidak selalu mudah diterapkan. Prihantoro menemukan adanya kegagalan dalam pelaksanaan asesmen autentik PAI di tingkat pendidikan dasar dan menengah, yang antara lain berkaitan dengan kompleksitas implementasi penilaian.¹³ Temuan tersebut menunjukkan bahwa perluasan bentuk asesmen harus disertai dengan kejelasan konstruk, indikator, rubrik, dan prosedur pengumpulan bukti. Penggunaan instrumen yang semakin beragam tanpa konstruk yang jelas justru dapat menimbulkan persoalan validitas baru.

Perkembangan penelitian berikutnya bergerak ke arah integrasi beberapa metode asesmen. Fadlillah dan Kusaeri, melalui systematic literature review, menunjukkan bahwa asesmen autentik dan diagnostik dapat digunakan untuk mengoptimalkan assessment for

⁷ Abdulaziz BinTaleb, "Exploring the Potential for Religious Literacy in Saudi Secondary Education," *Religious Education* 119, no. 1 (January 28, 2024): 17–30, <https://doi.org/10.1080/00344087.2023.2294607>.

⁸ Noer Hidayah et al., "Assessing Religious Literacy in Indonesian Secondary Education and Its Implications for Religious Moderation," *EDUKASI: Jurnal Penelitian Pendidikan Agama Dan Keagamaan* 23, no. 3 (December 31, 2025): 627–38, <https://doi.org/10.32729/edukasi.v23i3.2152>.

⁹ Muhammad Khakim Ashari, Moh Faizin, and Jamaluddin Shiddiq, "Religious Digital Literacy of Students in Indonesia and Malaysia," *TADRIS: Jurnal Pendidikan Islam* 18, no. 1 (June 30, 2023): 189–210, <https://doi.org/10.19105/tjpi.v18i1.8794>.

¹⁰ Noprijon Noprijon et al., "The Digitalization of Islamic Education and Its Impact on Improving Students' Religious Literacy," *Ahlussunnah: Journal of Islamic Education* 3, no. 2 (August 28, 2024): 93–100, <https://doi.org/10.58485/jie.v3i2.320>.

¹¹ Umi Baroroh and Tasman Hamani, "Development of Authentic Assessment in Islamic Religious Education in Elementary School," *Nazhruna: Jurnal Pendidikan Islam* 5, no. 3 (July 18, 2022): 940–55, <https://doi.org/10.31538/nzh.v5i3.2380>.

¹² Bachtiar Annas Imanudin, "Penilaian Autentik Terhadap Pembelajaran Pendidikan Agama Islam," *Raudhah Proud To Be Professionals: Jurnal Tarbiyah Islamiyah* 8, no. 2 (2023): 527–39, <https://doi.org/10.48094/raudhah.v8i2.310>.

¹³ Agung Prihantoro, "Kegagalan Pelaksanaan Asesmen Autentik Pada Pelajaran Pendidikan Agama Islam Di Sekolah Dasar Dan Menengah," *Ulumuddin : Jurnal Ilmu-Ilmu Keislaman* 11, no. 1 (June 12, 2021): 29–54, <https://doi.org/10.47200/ulumuddin.v11i1.688>.

learning dalam pendidikan Islam.¹⁴ Mukmin dan Nuraini secara lebih spesifik menawarkan integrasi tes dan nontes untuk menghasilkan evaluasi PAI yang lebih holistik.¹⁵ Muharrozi, Anwar, dan Romli kemudian menegaskan perlunya penguatan asesmen PAI melalui keseimbangan ranah kognitif, psikomotorik, dan afektif.¹⁶ Keseluruhan temuan tersebut menunjukkan pergeseran paradigma dari asesmen sebagai pengukuran tunggal menuju asesmen sebagai pengumpulan berbagai bukti pembelajaran. Pada ranah kognitif, terjadi perkembangan yang penting. Rahmah et al. menunjukkan perlunya asesmen PAI yang berorientasi pada Higher-Order Thinking Skills agar pertanyaan tidak berhenti pada kemampuan mengingat dan memahami pada tingkat yang sederhana.¹⁷ Bahsoan et al. bahkan menggunakan Rasch-based cognitive diagnostic assessment untuk mengukur kemampuan berpikir kritis dalam PAI.¹⁸ Perkembangan tersebut menunjukkan bahwa tes kognitif dapat diperkuat untuk mengidentifikasi struktur kemampuan yang lebih kompleks. Dengan kata lain, problem tes bukan terletak pada bentuk tertulisnya semata, tetapi pada kedalaman konstruk dan proses kognitif yang dituntut oleh butir.

Dimensi lain yang juga penting adalah karakter dan internalisasi nilai. Zainuddin, Putra, dan Nahadi mengembangkan instrumen self-assessment berbasis Al-Qur'an untuk mengukur sekaligus menginternalisasi nilai-nilai karakter religius.¹⁹ Pendekatan tersebut memperluas sumber bukti penilaian dari jawaban tes hingga refleksi peserta didik. Meskipun demikian, penggunaan self-assessment juga tidak lepas dari masalah validitas. Falahain dan Munip menunjukkan adanya social desirability bias dalam penilaian diri terhadap karakter keagamaan, sehingga bukti dari self-assessment perlu ditriangulasikan dengan bukti performa.²⁰ Temuan ini menunjukkan bahwa validitas tidak otomatis meningkat hanya dengan mengganti tes dengan instrumen non-tes.

Penelitian terdahulu telah menunjukkan tiga kecenderungan. Pertama, penelitian mengenai tes PAI menekankan kualitas teknis instrumen melalui validitas, reliabilitas, kesukaran, dan daya pembeda. Kedua, penelitian asesmen autentik mendorong integrasi ranah kognitif, afektif, dan psikomotorik. Ketiga, penelitian religious literacy menempatkan pemahaman agama sebagai kompetensi kontekstual dan interpretatif yang melampaui sekadar penguasaan informasi. Ketiga tradisi tersebut telah berkembang, tetapi belum sepenuhnya disatukan dalam satu kerangka analitis.

¹⁴ Nilna Fadlillah and Kusaeri Kusaeri, "Optimizing Assessment for Learning in Islamic Education through Authentic and Diagnostic Assessment: A Systematic Literature Review," *Jurnal Kependidikan: Jurnal Hasil Penelitian Dan Kajian Kepustakaan Di Bidang Pendidikan, Pengajaran Dan Pembelajaran* 10, no. 2 (June 22, 2024): 654, <https://doi.org/10.33394/jk.v10i2.11555>.

¹⁵ Mukmin Mukmin and Nuraini Nuraini, "Integrasi Penilaian Tes Dan Non-Tes Dalam Pendidikan Agama Islam: Menuju Evaluasi Holistik Untuk Pembelajaran Berkelanjutan," *Journal of Instructional and Development Researches* 4, no. 5 (October 31, 2024): 370–79, <https://doi.org/10.53621/jider.v4i5.384>.

¹⁶ Marprilyan Muharrozi, Saepul Anwar, and Usup Romli, "Strengthening Holistic Assessment in Islamic Religious Education: Cognitive, Psychomotor, and Affective Domains," *At-Tarbawi: Jurnal Kajian Kependidikan Islam* 10, no. 2 (December 31, 2025): 174–90, <https://doi.org/10.22515/attarbawi.v10i2.13364>.

¹⁷ Syarifah Rahmah et al., "Implementing Higher-Order Thinking Skills Oriented Assessment in Islamic Religious Education," *Al-Hayat: Journal of Islamic Education* 10, no. 1 (March 26, 2026): 237–53, <https://doi.org/10.35723/ajie.v10i1.152>.

¹⁸ Agil Bahsoan et al., "Rasch-Based Cognitive Diagnostic Assessment to Measure Critical Thinking Skills in Islamic Religious Education," *Jurnal Iqra': Kajian Ilmu Pendidikan* 10, no. 3 (November 16, 2025): 121–37, <https://doi.org/10.25217/ji.v10i3.6648>.

¹⁹ Zainuddin Zainuddin et al., "Developing Qur'an-Based Self-Assessment Instruments for Measuring and Internalizing Religious Character Values," *Tadris: Jurnal Keguruan Dan Ilmu Tarbiyah* 9, no. 2 (December 22, 2024): 487–97, <https://doi.org/10.24042/tadris.v9i2.25027>.

²⁰ Muchtar Falahain and Abdul Munip, "Addressing Social Bias and Enhancing the Validity of Student Self-Assessment with a Performance Evidence-Based Triangulation Model in Islamic Religious Education Learning," *Journal of Islamic Education and Ethics* 4, no. 2 (2026): 161–75, <https://doi.org/10.18196/jjee.v4i2.170>.

Research gap penelitian ini adalah belum kuatnya hubungan antara validitas psikometrik instrumen PAI dengan validitas substantifnya dalam merepresentasikan pemahaman keagamaan. Penelitian terdahulu cenderung menentukan kelayakan instrumen berdasarkan karakteristik statistik atau, sebaliknya, menawarkan asesmen holistik tanpa menjadikan hasil analisis butir sebagai titik awal untuk menentukan batas inferensi instrumen. Penelitian ini mempertemukan kedua pendekatan tersebut dengan menjadikan hasil analisis empiris instrumen sebagai dasar untuk menentukan sejauh mana skor tes dapat diinterpretasikan sebagai bukti pemahaman keagamaan. Sehingga, penelitian ini bertujuan untuk menganalisis kualitas psikometrik instrumen penilaian PAI berdasarkan validitas, reliabilitas, daya pembeda, dan tingkat kesukaran; mengidentifikasi batas kemampuan tes objektif dalam merepresentasikan pemahaman keagamaan; serta merumuskan model validitas instrumen yang mengintegrasikan kualitas psikometrik dengan validitas substantif-keagamaan. Kebaruan penelitian terletak pada konsep validitas integratif penilaian PAI, yaitu bahwa instrumen yang valid untuk mengukur pemahaman keagamaan harus memenuhi kualitas teknis pengukuran sekaligus memiliki keterwakilan konstruk yang mencakup pengetahuan, pemahaman konseptual, kemampuan kontekstual-interpretatif, refleksi nilai, dan penerapan ajaran.

Metodologi

Penelitian ini menggunakan pendekatan kuantitatif dengan desain deskriptif-evaluatif untuk menganalisis kualitas psikometrik instrumen penilaian Pendidikan Agama Islam (PAI) sekaligus mengidentifikasi batas kemampuannya dalam merepresentasikan pemahaman keagamaan peserta didik. Subjek penelitian adalah peserta didik sekolah dasar yang mengikuti uji coba instrumen, sedangkan objek penelitian berupa 30 butir tes objektif PAI yang dikembangkan berdasarkan kompetensi pembelajaran. Data diperoleh melalui tes tertulis dan dianalisis menggunakan aplikasi ANATES berdasarkan empat parameter, yaitu validitas butir berdasarkan hubungan skor setiap butir dengan skor total, reliabilitas untuk menentukan konsistensi internal instrumen, daya pembeda untuk mengidentifikasi kemampuan butir dalam membedakan peserta didik berdasarkan tingkat penguasaan, serta indeks kesukaran untuk mengklasifikasikan butir ke dalam kategori mudah, sedang, dan sukar. Analisis dilakukan dalam dua tahap. Tahap pertama merupakan analisis psikometrik terhadap karakteristik empiris instrumen, sedangkan tahap kedua merupakan analisis interpretatif terhadap keterwakilan konstruk berdasarkan sintesis kajian asesmen PAI dan literasi religius. Pada tahap kedua, pemahaman keagamaan dipetakan ke dalam lima dimensi, yaitu pengetahuan faktual-konseptual, pemahaman eksplanatif, kemampuan kontekstual-interpretatif, refleksi nilai, dan kemampuan aplikatif. Kelima dimensi tersebut digunakan sebagai kerangka untuk menentukan batas interpretasi skor, bukan sebagai konstruk yang sepenuhnya telah diuji secara empiris melalui instrumen penelitian. Dengan demikian, penelitian membedakan validitas psikometrik, yang menunjukkan kualitas teknis instrumen, dari validitas substantif-keagamaan, yang menunjukkan kecukupan keterwakilan konstruk pemahaman keagamaan.

Hasil Penelitian dan Pembahasan

Kualitas Psikometrik Instrumen Penilaian Pendidikan Agama Islam

Hasil analisis menunjukkan bahwa 20 dari 30 butir soal (66,67%) memenuhi kriteria validitas dan dapat dipertahankan, sedangkan 10 butir (33,33%) memerlukan revisi. Koefisien reliabilitas instrumen sebesar 0,68 menunjukkan konsistensi internal pada kategori sedang. Hasil ini menunjukkan bahwa instrumen ini memiliki dasar psikometrik yang cukup untuk

digunakan sebagai alat evaluasi, tetapi kualitas antarbutirnya belum sepenuhnya homogen. Karena itu, kelayakan instrumen perlu dipahami sebagai kelayakan teknis untuk fungsi pengukuran yang direpresentasikan oleh tes, bukan sebagai bukti bahwa seluruh dimensi pemahaman keagamaan dapat diukur.

Validitas butir merupakan tahap penting dalam evaluasi PAI karena skor yang dihasilkan instrumen harus berasal dari pertanyaan yang benar-benar berkontribusi terhadap kompetensi yang hendak diukur. Atika et al. menegaskan pentingnya analisis tes formatif PAI untuk memastikan kualitas instrumen sebelum hasilnya digunakan sebagai dasar evaluasi pembelajaran.²¹ Temuan penelitian ini memperkuat argumen tersebut. Butir yang tampak sesuai secara materi tidak selalu memiliki kemampuan empiris yang sama dalam membedakan tingkat penguasaan peserta didik. Reliabilitas instrumen berada pada kategori sedang. Temuan tersebut menunjukkan bahwa konsistensi internal instrumen telah terbentuk, tetapi masih dapat ditingkatkan. Beberapa butir dengan validitas dan daya pembeda rendah, serta distribusi kesukaran yang kurang seimbang, dapat berkontribusi terhadap reliabilitas yang rendah. Validitas dan reliabilitas harus dibaca secara bersamaan: validitas menunjukkan kualitas hubungan antara butir dan konstruk yang diukur, sedangkan reliabilitas menunjukkan konsistensi bukti yang dihasilkan oleh instrumen.

Hasil analisis daya pembeda menunjukkan bahwa 19 butir (63,33%) memiliki kemampuan yang memadai dalam membedakan peserta didik berdasarkan tingkat penguasaan, sedangkan 11 butir (36,67%) masih menunjukkan fungsi diskriminatif yang rendah dan perlu direvisi. Analisis tingkat kesukaran menunjukkan 12 butir (40,00%) berada pada kategori mudah, 14 butir (46,67%) pada kategori sedang, dan 4 butir (13,33%) pada kategori sukar. Distribusi tersebut menunjukkan dominasi butir mudah dan sedang, sedangkan proporsi butir sukar masih relatif kecil. Komposisi ini perlu diperbaiki agar instrumen tidak hanya efektif dalam mengidentifikasi penguasaan dasar, tetapi juga memberikan ruang yang memadai untuk mengungkap kemampuan analitis dan pemahaman yang lebih mendalam.

Tabel 1.
Interpretasi Hasil Analisis Instrumen PAI

Parameter	Hasil Empiris	Interpretasi Psikometrik	Keputusan
Butir valid	20 butir (66,67%)	Memenuhi kriteria validitas	Dipertahankan
Butir perlu revisi	10 butir (33,33%)	Validitas belum optimal	Direvisi
Reliabilitas	$r = 0,68$	Sedang	Perlu penguatan
Daya pembeda baik	19 butir (63,33%)	Mampu membedakan kemampuan peserta didik	Dipertahankan
Daya pembeda rendah	11 butir (36,67%)	Fungsi diskriminatif terbatas	Direvisi
Mudah	12 butir (40,00%)	Proporsi relatif tinggi	Evaluasi distribusi
Sedang	14 butir (46,67%)	Proporsi memadai	Dipertahankan
Sukar	4 butir (13,33%)	Proporsi relatif rendah	Evaluasi distribusi

Sumber: Hasil pengolahan data penelitian menggunakan ANATES, 2026.

Tabel 1 menunjukkan bahwa kelayakan psikometrik instrumen memberikan bukti mengenai kualitas teknis pengukuran, tetapi belum tentu menunjukkan keluasan konstruk pemahaman keagamaan yang direpresentasikan. Validitas butir, reliabilitas, daya pembeda,

²¹ Atika et al., "Formative Test Analysis Of Islamic Religious Education Learning Evaluation Practices (Study At SMP Negeri 1 Muara Kemumu)."

dan tingkat kesukaran menjelaskan sejauh mana instrumen bekerja secara konsisten dan mampu membedakan kemampuan peserta didik. Namun, indikator tersebut belum cukup untuk memastikan bahwa skor yang dihasilkan merepresentasikan dimensi pemahaman keagamaan secara utuh, terutama kemampuan kontekstual-interpretatif, refleksi nilai, dan kemampuan aplikatif. Oleh karena itu, hasil analisis psikometrik perlu dijadikan dasar awal untuk menentukan kelayakan instrumen sekaligus batas inferensi terhadap skor yang dihasilkan.

Pemahaman Keagamaan sebagai Konstruk Penilaian PAI

Pemahaman keagamaan merupakan konstruk yang lebih luas daripada pengetahuan keagamaan. Pengetahuan dapat ditunjukkan melalui kemampuan mengenali konsep, mengingat ketentuan, menyebutkan dalil, atau menentukan jawaban yang tepat. Pemahaman menuntut kemampuan untuk menjelaskan hubungan antarkonsep, memberikan alasan, menangkap nilai yang mendasari ajaran, dan menerapkan prinsip keagamaan dalam menghadapi situasi yang berbeda-beda. Amina dan Lakhania mengembangkan religious literacy sebagai kemampuan yang tidak berhenti pada pengetahuan deklaratif tentang agama.²² BinTaleb juga menunjukkan potensinya dalam pendidikan menengah sebagai dasar untuk memahami agama secara lebih kontekstual.²³ Dalam konteks Indonesia, Suyanto et al. menunjukkan bahwa literasi keagamaan masih menghadapi permasalahan dalam sistem pendidikan.²⁴ Konsep ini relevan bagi PAI karena peserta didik hidup dalam masyarakat di mana pemahaman agama berinteraksi dengan budaya, perkembangan sosial, dan informasi digital.

Dimensi digital memperluas tuntutan tersebut. Ashari, Faizin, dan Shiddiq menunjukkan pentingnya literasi digital keagamaan bagi peserta didik di Indonesia dan Malaysia.²⁵ Informasi agama saat ini tidak lagi diperoleh hanya dari guru dan buku pelajaran, tetapi juga dari video, media sosial, mesin pencari, dan berbagai otoritas keagamaan di dunia digital. Kemampuan peserta didik untuk memahami informasi tersebut tidak dapat dinilai hanya dengan pertanyaan yang menuntut reproduksi informasi. Berdasarkan hal tersebut, penelitian ini memetakan pemahaman keagamaan ke dalam lima dimensi operasional.

Tabel 2.
Kerangka Konseptual Dimensi Pemahaman Keagamaan dalam Asesmen PAI

Dimensi	Kompetensi yang Direpresentasikan	Contoh Bukti Asesmen
Faktual-konseptual	Mengetahui konsep, ketentuan, dalil, istilah	Tes objektif
Eksplanatif	Menjelaskan makna, alasan, hubungan dan hikmah	Uraian
Kontekstual-interpretatif	Menggunakan prinsip Islam untuk membaca kasus	Studi kasus/HOTS
Reflektif-nilai	Memberikan pertimbangan berdasarkan nilai Islam	Refleksi/portofolio
Aplikatif	Menunjukkan penggunaan kompetensi dalam tugas nyata	Kinerja/proyek/observasi

Sumber: Diolah oleh penulis berdasarkan referensi dari Amin dan Lakhani (2024), BinTaleb (2024), Suyanto et al. (2024), Hidayah et al. (2025), serta Ashari, Faizin, dan Shiddiq (2023).

²² Amin and Lakhani, "Redefining Religious Literacy in Secondary Mainstream United Kingdom Classrooms."

²³ BinTaleb, "Exploring the Potential for Religious Literacy in Saudi Secondary Education."

²⁴ Bagong Suyanto et al., "Problems of Religious Literacy in Indonesian Education," *Journal of Religious Education* 72, no. 2 (July 17, 2024): 165–81, <https://doi.org/10.1007/s40839-024-00228-1>.

²⁵ Ashari, Faizin, and Shiddiq, "Religious Digital Literacy of Students in Indonesia and Malaysia."

Pemetaan tersebut juga menetapkan batas epistemologis. Penilaian PAI tidak dimaksudkan untuk mengukur kualitas iman peserta didik secara absolut karena wilayah batin tidak dapat direduksi menjadi skor pendidikan. Yang dapat dinilai adalah bukti belajar yang dapat diobservasi dan dipertanggungjawabkan sebagai representasi pengetahuan, penalaran, refleksi, dan kemampuan menerapkan nilai. Lima dimensi tersebut merupakan kerangka konseptual-interpretatif yang disusun melalui sintesis literatur asesmen PAI dan religious literacy, bukan hasil analisis faktor dari instrumen yang diuji. Karena instrumen empiris penelitian berbentuk tes objektif, bukti yang dihasilkan terutama mencakup dimensi faktual-konseptual dan sebagian kemampuan eksplanatif atau analitis, bergantung pada konstruksi setiap butir. Dimensi kontekstual-interpretatif, reflektif-nilai, dan aplikatif ditempatkan sebagai wilayah konstruk yang memerlukan jenis bukti tambahan. Penegasan ini penting untuk mencegah klaim berlebihan bahwa satu jenis tes telah mengukur seluruh pemahaman keagamaan.

Batas Validitas Tes Objektif dalam Mengukur Pemahaman Keagamaan

Temuan penelitian menunjukkan bahwa tes objektif masih memiliki fungsi penting. Instrumen ini efisien, memiliki prosedur penskoran yang relatif objektif, memungkinkan cakupan materi yang luas, serta dapat dianalisis menggunakan teknik psikometrik. Kualitas tersebut menjelaskan mengapa tes tertulis tetap digunakan dalam evaluasi PAI. Arfah menempatkan evaluasi hasil belajar sebagai bagian penting dari proses pembelajaran PAI, sedangkan Atika et al. menunjukkan relevansi analisis kualitas tes formatif.²⁶

Masalahnya bukan terletak pada penggunaan tes objektif, melainkan pada inferensi yang dibuat berdasarkan skor tes. Seorang peserta didik yang mampu menjawab pertanyaan tentang pengertian zakat telah menunjukkan pemahaman tentang zakat. Skor tersebut belum cukup untuk menyimpulkan bahwa ia mampu menjelaskan dimensi keadilan sosial zakat, menganalisis persoalan distribusi zakat kontemporer, atau memberikan pertimbangan terhadap suatu kasus berdasarkan nilai-nilai yang terkandung dalam zakat. Kondisi tersebut menjelaskan mengapa asesmen autentik berkembang dalam PAI. Baroroh dan Hamami menunjukkan bahwa asesmen autentik memungkinkan penilaian mencakup aspek kognitif, afektif, dan psikomotorik.²⁷ Penelitian ketiga ranah tersebut merupakan komponen penting dalam asesmen pendidikan. Ismail et al. secara khusus menunjukkan kompleksitas asesmen ranah psikomotorik PAI, yang tidak dapat sepenuhnya direpresentasikan melalui tes pengetahuan.²⁸

Integrasi tes dan non-tes menjadi lebih relevan dibandingkan dengan mengganti seluruh tes dengan instrumen lain. Abdurahman dan Erihadiana menunjukkan bahwa tes dan non-tes memiliki fungsi yang berbeda tetapi saling melengkapi; teknik non-tes dapat memperoleh informasi mengenai sikap, perilaku, karakteristik sosial, dan aspek personal yang tidak dapat ditangkap melalui tes.²⁹ Mukmin dan Nuraini juga mengarahkan integrasi kedua bentuk

²⁶ Atika et al., "Formative Test Analysis Of Islamic Religious Education Learning Evaluation Practices (Study At SMP Negeri 1 Muara Kemumu)."

²⁷ Baroroh and Hamami, "Development of Authentic Assessment in Islamic Religious Education in Elementary School."

²⁸ Fajri Ismail et al., "Assessment Model of Islamic Religious Education in the Psychomotor Domain during the COVID-19 Pandemic in Indonesia." *Eurasian Journal of Educational Research* 101 (2022): 190–203., " *Eurasian Journal of Educational Research* 101 (2022): 190–203, <https://doi.org/10.14689/ejer.2022.101.011>.

²⁹ Abdurahman and Erihadiana, "Comprehensive Assessment: Balancing Test and Non-Test Methods in Islamic Education Evaluation."

penilaian tersebut menuju evaluasi PAI yang lebih holistik. Penggunaan instrumen non-tes tetap harus memperhatikan validitas. Zainuddin, Putra, dan Nahadi menawarkan self-assessment berbasis Al-Qur'an untuk mengukur nilai-nilai karakter religius.³⁰ Pendekatan tersebut memperluas bukti evaluasi, tetapi jawaban peserta didik dalam self-assessment dapat dipengaruhi oleh keinginan untuk menampilkan citra diri yang dianggap baik. Falahain dan Munip mengusulkan triangulasi dengan bukti performa untuk mengurangi bias social desirability.³¹ Hal ini menunjukkan bahwa multimetode tidak identik dengan validitas jika setiap metode tidak memiliki dasar konstruk dan bukti yang jelas.

Fadlillah dan Kusaeri memberikan arah yang lebih sistematis dengan mengintegrasikan asesmen autentik dan diagnostik sebagai assessment for learning.³² Pendekatan tersebut relevan karena hasil penilaian tidak berhenti pada kategorisasi kemampuan, tetapi digunakan untuk menentukan kebutuhan pembelajaran selanjutnya. Berdasarkan hasil penelitian dan sintesis literatur, tes objektif dalam PAI mempunyai validity boundary. Instrumen tersebut dapat memberikan bukti kuat mengenai pengetahuan faktual-konseptual dan, apabila dirancang sebagai HOTS, sebagian kemampuan analitis. Instrumen tersebut menjadi semakin lemah apabila skor digunakan untuk membuat inferensi langsung mengenai refleksi nilai, internalisasi, atau kemampuan aplikatif peserta didik.

Validitas Integratif sebagai Rekonstruksi Penilaian PAI

Temuan utama penelitian ini mengarah pada perlunya merekonstruksi validitas penilaian PAI melalui dua lapisan yang saling berhubungan: validitas psikometrik dan validitas substantif-keagamaan. Validitas psikometrik menentukan apakah instrumen bekerja secara teknis. Validitas butir, reliabilitas, daya pembeda, dan tingkat kesukaran tetap menjadi syarat yang tidak dapat diabaikan. Instrumen dengan reliabilitas rendah atau butir yang tidak mampu membedakan kemampuan peserta didik tidak dapat menghasilkan bukti yang kuat. Validitas substantif-keagamaan menentukan apakah konstruk yang direpresentasikan oleh instrumen sesuai dengan pemahaman keagamaan yang menjadi tujuan pembelajaran. Dimensi ini mempertanyakan kesesuaian antara tujuan PAI, indikator, tugas asesmen, proses respons peserta didik, dan interpretasi skor. Muharrozi, Anwar, dan Romli menekankan pentingnya penguatan asesmen holistik melalui integrasi ranah kognitif, psikomotorik, dan afektif.³³ Fadlillah dan Kusaeri menunjukkan perlunya menggabungkan asesmen autentik dengan asesmen diagnostik.³⁴ Literatur tersebut mendukung konstruksi bahwa validitas penilaian PAI sebaiknya dibangun melalui berbagai jenis bukti.

Validitas integratif yang ditawarkan penelitian ini berbeda dari asesmen holistik maupun asesmen autentik. Asesmen holistik terutama menekankan keluasan domain yang dinilai, sedangkan asesmen autentik menekankan penggunaan tugas yang merepresentasikan situasi atau kinerja yang bermakna. Validitas integratif ditempatkan pada tingkat interpretasi bukti, yaitu untuk memastikan bahwa kesimpulan yang dibuat berdasarkan skor tidak

³⁰ Zainuddin et al., "Developing Qur'an-Based Self-Assessment Instruments for Measuring and Internalizing Religious Character Values."

³¹ Falahain and Munip, "Addressing Social Bias and Enhancing the Validity of Student Self-Assessment with a Performance Evidence-Based Triangulation Model in Islamic Religious Education Learning."

³² Fadlillah and Kusaeri, "Optimizing Assessment for Learning in Islamic Education through Authentic and Diagnostic Assessment: A Systematic Literature Review."

³³ Muharrozi, Anwar, and Romli, "Strengthening Holistic Assessment in Islamic Religious Education: Cognitive, Psychomotor, and Affective Domains."

³⁴ Fadlillah and Kusaeri, "Optimizing Assessment for Learning in Islamic Education through Authentic and Diagnostic Assessment: A Systematic Literature Review."

melampaui konstruk yang benar-benar direpresentasikan oleh instrumen. Dengan demikian, validitas integratif terdiri atas empat komponen yang saling berkaitan: kualitas psikometrik, keterwakilan konstruk, kesesuaian inferensi, dan triangulasi bukti.

Tabel 3.
Model Validitas Integratif Penilaian PAI

Lapisan	Pertanyaan Validitas	Bukti yang Dibutuhkan
Psikometrik	Apakah instrumen bekerja secara konsisten?	Validitas butir, reliabilitas, kesukaran, daya pembeda
Isi	Apakah materi dan kompetensi terwakili?	Blueprint, CP/TP, validasi ahli
Konstruk	Apakah tugas mengukur pemahaman yang dimaksud?	Analisis indikator dan proses respons
Kontekstual	Apakah peserta didik dapat menerapkan pemahaman pada kasus baru?	HOTS/studi kasus
Reflektif	Apakah peserta didik mampu memberikan argumentasi berbasis nilai?	Refleksi/portofolio
Aplikatif	Apakah kompetensi dapat ditunjukkan melalui tugas autentik?	Kinerja/proyek/observasi
Triangulatif	Apakah bukti dari berbagai instrumen konsisten?	Perbandingan tes dan non-tes

Sumber: Dikembangkan oleh penulis berdasarkan hasil penelitian dan sintesis literatur tentang asesmen PAI, asesmen autentik, serta religious literacy, 2026.

Model tersebut tidak mengharuskan semua dimensi diukur dalam satu kesempatan. Penilaian dapat dilakukan secara berkelanjutan. Tes objektif digunakan untuk mengukur fondasi pengetahuan; soal uraian dan kasus digunakan untuk mengidentifikasi penalaran; refleksi dan portofolio digunakan untuk memperoleh bukti proses pemaknaan; sedangkan tugas kinerja dan observasi digunakan ketika kompetensi membutuhkan demonstrasi tindakan.

Kerangka tersebut juga menjawab masalah social desirability. Sikap atau karakter tidak cukup dinilai hanya melalui pernyataan diri, sedangkan perilaku yang diamati pun tidak cukup untuk menentukan apakah nilai tersebut telah terinternalisasi. Bukti harus diperoleh dari beberapa situasi dan sumber sebelum digunakan untuk menarik kesimpulan yang lebih luas. Perkembangan asesmen digital membuka peluang untuk menerapkan model tersebut. Pendekatan digital seharusnya tidak sekadar memindahkan soal pilihan ganda dari kertas ke layar, tetapi juga memungkinkan diversifikasi bukti pembelajaran. Kontribusi konseptual penelitian ini dapat dirumuskan dalam hubungan berikut: Validitas Penilaian PAI = Kualitas Psikometrik + Keterwakilan Konstruk + Kesesuaian Inferensi + Triangulasi Bukti.

Formula tersebut juga bukan persamaan statistik, melainkan model konseptual. Kualitas psikometrik memastikan bukti pengukuran dapat dipercaya; keterwakilan konstruk memastikan dimensi pemahaman keagamaan tercakup; kesesuaian inferensi mencegah interpretasi skor melampaui kemampuan instrumen; sedangkan triangulasi memperkuat kesimpulan melalui beberapa sumber bukti. Temuannya memperjelas novelty penelitian. Penelitian ini tidak menolak tes objektif dan tidak pula menempatkan asesmen autentik sebagai satu-satunya pengganti. Penelitian ini menawarkan validitas integratif sebagai kerangka yang menghubungkan analisis psikometrik dengan konsep pemahaman keagamaan. Formula ini bukan persamaan matematis dan tidak dimaksudkan untuk

menghasilkan indeks numerik baru, melainkan sebagai representasi konseptual tentang syarat agar interpretasi hasil penilaian PAI dapat dipertanggungjawabkan secara proporsional.

Simpulan

Hasil penelitian menunjukkan bahwa instrumen penilaian Pendidikan Agama Islam yang diuji memiliki kualitas psikometrik yang memadai. Dari 30 butir, 20 butir (66,67%) memenuhi kriteria validitas dan 10 butir (33,33%) memerlukan revisi; koefisien reliabilitas sebesar 0,68 berada pada kategori sedang; 19 butir (63,33%) memiliki daya pembeda yang memadai dan 11 butir (36,67%) masih rendah; sedangkan tingkat kesukaran terdiri atas 12 butir mudah (40,00%), 14 butir sedang (46,67%), dan 4 butir sukar (13,33%). Temuan tersebut memberikan dasar untuk menggunakan instrumen dalam mengukur penguasaan pengetahuan PAI, tetapi tidak cukup untuk menyimpulkan bahwa tes objektif telah merepresentasikan pemahaman keagamaan secara utuh. Analisis konseptual menunjukkan bahwa pemahaman keagamaan mencakup pengetahuan faktual-konseptual, kemampuan eksplanatif, interpretasi kontekstual, refleksi nilai, dan kemampuan aplikatif. Lima dimensi tersebut merupakan kerangka interpretatif yang disintesis dari literatur dan belum sepenuhnya diuji secara empiris melalui instrumen penelitian. Berdasarkan temuan tersebut, penelitian ini menawarkan validitas integratif penilaian PAI yang menghubungkan kualitas psikometrik, keterwakilan konstruk, kesesuaian inferensi, dan triangulasi bukti. Kerangka ini berbeda dari asesmen holistik karena berfokus pada ketepatan hubungan antara bukti asesmen dan kesimpulan yang ditarik berdasarkan bukti tersebut. Implikasinya, tes objektif tetap relevan untuk mengukur pengetahuan dan sebagian kemampuan analitis, sedangkan pemahaman kontekstual, reflektif, dan aplikatif membutuhkan bukti tambahan melalui studi kasus, refleksi, portofolio, observasi, atau penilaian kinerja. Penelitian berikutnya perlu menguji secara empiris struktur multidimensional pemahaman keagamaan melalui validasi ahli, analisis faktor atau pemodelan Rasch, serta triangulasi antartenes asesmen.

Referensi

- Abdurahman, Asep, and Mohamad Erihadiana. "Comprehensive Assessment: Balancing Test and Non-Test Methods in Islamic Education Evaluation." *Jurnal Inovasi Pendidikan Agama Islam (JIPAI)* 4, no. 2 (December 26, 2024): 95–108. <https://doi.org/10.15575/jipai.v4i2.40752>.
- Amin, Anar, and Anisha Lakhani. "Redefining Religious Literacy in Secondary Mainstream United Kingdom Classrooms." *Journal of Beliefs & Values* 45, no. 2 (April 2, 2024): 180–97. <https://doi.org/10.1080/13617672.2024.2309746>.
- Arfah, M. Arfah. "Evaluasi Hasil Belajar Pendidikan Agama Islam (PAI)." *Jurnal Literasiologi* 7, no. 2 (October 13, 2021): 1–14. <https://doi.org/10.47783/literasiologi.v7i2.282>.
- Ashari, Muhammad Khakim, Moh Faizin, and Jamaluddin Shiddiq. "Religious Digital Literacy of Students in Indonesia and Malaysia." *TADRIS: Jurnal Pendidikan Islam* 18, no. 1 (June 30, 2023): 189–210. <https://doi.org/10.19105/tjpi.v18i1.8794>.
- Atika, Yova, Yessy Rahmani, Muslim, Deri Wanto, and Mirzon Daheri. "Formative Test Analysis Of Islamic Religious Education Learning Evaluation Practices (Study At SMP Negeri 1 Muara Kemumu)." *Indonesian Journal of Pedagogy and Teacher Education* 1, no. 2 (December 31, 2023): 38–42. <https://doi.org/10.58723/ijopate.v1i2.113>.
- Bahsoan, Agil, Nurhayati Tine, Ivan Rahmat Santoso, Neva Lionitha Ibrahim, and Branislav Pupala. "Rasch-Based Cognitive Diagnostic Assessment to Measure Critical Thinking

- Skills in Islamic Religious Education.” *Jurnal Iqra’: Kajian Ilmu Pendidikan* 10, no. 3 (November 16, 2025): 121–37. <https://doi.org/10.25217/ji.v10i3.6648>.
- Baroroh, Umi, and Tasman Hamani. “Development of Authentic Assessment in Islamic Religious Education in Elementary School.” *Nazhruna: Jurnal Pendidikan Islam* 5, no. 3 (July 18, 2022): 940–55. <https://doi.org/10.31538/nzh.v5i3.2380>.
- BinTaleb, Abdulaziz. “Exploring the Potential for Religious Literacy in Saudi Secondary Education.” *Religious Education* 119, no. 1 (January 28, 2024): 17–30. <https://doi.org/10.1080/00344087.2023.2294607>.
- Fadlillah, Nilna, and Kusaeri Kusaeri. “Optimizing Assessment for Learning in Islamic Education through Authentic and Diagnostic Assessment: A Systematic Literature Review.” *Jurnal Kependidikan: Jurnal Hasil Penelitian Dan Kajian Kepustakaan Di Bidang Pendidikan, Pengajaran Dan Pembelajaran* 10, no. 2 (June 22, 2024): 654. <https://doi.org/10.33394/jk.v10i2.11555>.
- Falahain, Muchtar, and Abdul Munip. “Addressing Social Bias and Enhancing the Validity of Student Self-Assessment with a Performance Evidence-Based Triangulation Model in Islamic Religious Education Learning.” *Journal of Islamic Education and Ethics* 4, no. 2 (2026): 161–75. <https://doi.org/10.18196/jiee.v4i2.170>.
- Febrianti, Evi, Makki, Rosmiati Ramli, Andi Fitriani Djollong, and Muhammad Nur Maallah. “Implementation of Islamic Religious Education Learning Evaluation Through Daily Tests on Students’ Learning Motivation UPTD SMP Negeri 4 Parepare.” *Indonesian Journal of Interdisciplinary Research in Science and Technology* 1, no. 9 (October 31, 2023): 823–38. <https://doi.org/10.55927/marcopolo.v1i9.6588>.
- Hidayah, Noer, Akhilarul Waro Wardani, Wahidul Anam, and Mona Addeq. “Assessing Religious Literacy in Indonesian Secondary Education and Its Implications for Religious Moderation.” *EDUKASI: Jurnal Penelitian Pendidikan Agama Dan Keagamaan* 23, no. 3 (December 31, 2025): 627–38. <https://doi.org/10.32729/edukasi.v23i3.2152>.
- Imanudin, Bachtiar Annas. “Penilaian Autentik Terhadap Pembelajaran Pendidikan Agama Islam.” *Raudhah Proud To Be Professionals: Jurnal Tarbiyah Islamiyah* 8, no. 2 (2023): 527–39. <https://doi.org/10.48094/raudhah.v8i2.310>.
- Ismail, Fajri, Mardiah Astuti, Dindin Nasrudin, Indah Wigati, and Zuhdiyah. “Assessment Model of Islamic Religious Education in the Psychomotor Domain during the COVID-19 Pandemic in Indonesia.” *Eurasian Journal of Educational Research* 101 (2022): 190–203. *Eurasian Journal of Educational Research* 101 (2022): 190–203. <https://doi.org/10.14689/ejer.2022.101.011>.
- Muharrozi, Marpriyan, Saepul Anwar, and Usup Romli. “Strengthening Holistic Assessment in Islamic Religious Education: Cognitive, Psychomotor, and Affective Domains.” *At-Tarbawi: Jurnal Kajian Kependidikan Islam* 10, no. 2 (December 31, 2025): 174–90. <https://doi.org/10.22515/attarbawi.v10i2.13364>.
- Mukmin, Mukmin, and Nuraini Nuraini. “Integrasi Penilaian Tes Dan Non-Tes Dalam Pendidikan Agama Islam: Menuju Evaluasi Holistik Untuk Pembelajaran Berkelanjutan.” *Journal of Instructional and Development Researches* 4, no. 5 (October 31, 2024): 370–79. <https://doi.org/10.53621/jider.v4i5.384>.
- Noprijon, Noprijon, Apriana Nofriastuti Rasdiany, Mona Yulia Zulfa, Mayang Belia Sameto, and Yusriandi Yusriandi. “The Digitalization of Islamic Education and Its Impact on Improving Students’ Religious Literacy.” *Ahlussunnah: Journal of Islamic Education* 3, no. 2 (August 28, 2024): 93–100. <https://doi.org/10.58485/jie.v3i2.320>.

- Prihantoro, Agung. "Kegagalan Pelaksanaan Asesmen Autentik Pada Pelajaran Pendidikan Agama Islam Di Sekolah Dasar Dan Menengah." *Ulumuddin: Jurnal Ilmu-Ilmu Keislaman* 11, no. 1 (June 12, 2021): 29–54. <https://doi.org/10.47200/ulumuddin.v11i1.688>.
- Prihatin, Rani Putri, and Tasman Hamami. "Learning Assessment Model for Islamic Religious Education." *Nusantara: Jurnal Pendidikan Indonesia* 2, no. 2 (May 31, 2022): 373–90. <https://doi.org/10.14421/njpi.2022.v2i2-10>.
- Rahmah, Syarifah, Ramlan, Nia Wardhani, Syarifah Alwiyah, and Syarbaini. "Implementing Higher-Order Thinking Skills Oriented Assessment in Islamic Religious Education." *Al-Hayat: Journal of Islamic Education* 10, no. 1 (March 26, 2026): 237–53. <https://doi.org/10.35723/ajie.v10i1.152>.
- Suyanto, Bagong, Mun'im Sirry, Rahma Sugihartati, Drajat Tri Kartono, and Muhammad Turhan Yani. "Problems of Religious Literacy in Indonesian Education." *Journal of Religious Education* 72, no. 2 (July 17, 2024): 165–81. <https://doi.org/10.1007/s40839-024-00228-1>.
- Zainuddin, Zainuddin, Syihabuddin Syihabuddin, Aminuddin Prahutama Putra, and Nahadi Nahadi. "Developing Qur'an-Based Self-Assessment Instruments for Measuring and Internalizing Religious Character Values." *Tadris: Jurnal Keguruan Dan Ilmu Tarbiyah* 9, no. 2 (December 22, 2024): 487–97. <https://doi.org/10.24042/tadris.v9i2.25027>.