

PENGARUH KESALAHAN PENGUKURAN TERHADAP KREDIBILITAS EVALUASI PENDIDIKAN AGAMA ISLAM (PAI): STUDI PADA INSTRUMEN TES DAN NON- TES di SDN 023 TENGGARONG

Khairudin¹, Sugeng²

Universitas Islam Negeri Sultan Aji Muhammad Idris Samarinda¹²

khairudin023tgr@gmail.com

ABSTRAK

Kesalahan pengukuran merupakan faktor yang dapat menurunkan ketepatan dan mutu hasil evaluasi pembelajaran. Dalam konteks evaluasi Pendidikan Agama Islam (PAI) di SDN 023 Tenggarong, kerentanan kesalahan muncul pada penilaian kognitif, afektif, dan psikomotor, baik melalui instrumen tes maupun non-tes. Penelitian ini menggunakan metode studi pustaka dengan menelaah teori evaluasi, prinsip psikometri, serta penelitian terkait pelaksanaan penilaian di sekolah dasar. Hasil kajian menunjukkan bahwa kesalahan pengukuran terutama disebabkan oleh instrumen yang tidak memenuhi kriteria validitas dan reliabilitas, perumusan butir soal yang kurang tepat, subjektivitas guru dalam observasi, bias penilaian, dan kendala administratif. Faktor internal peserta didik seperti kondisi emosional, motivasi, dan pemahaman instruksi turut memengaruhi skor sehingga hasil tidak selalu mencerminkan kemampuan aktual. Penelitian ini menyimpulkan bahwa peningkatan kredibilitas evaluasi memerlukan ketelitian profesional guru serta budaya evaluasi berorientasi mutu melalui pelatihan berkelanjutan, praktik reflektif, dan supervisi akademik. Penerapan prosedur evaluasi yang terstandar diharapkan menghasilkan data yang lebih akurat dan mendukung pengambilan keputusan pembelajaran.

Kata Kunci: Kesalahan Pengukuran, Evaluasi PAI, Instrumen Tes, Instrumen Non-Tes

ABSTRACT

Measurement errors are factors that can reduce the accuracy and quality of learning evaluation results. In the context of Islamic Education (PAI) evaluation at SDN 023 Tenggarong, such errors may occur in cognitive, affective, and psychomotor assessments, both in test and non-test instruments. This study employed a literature review method, examining evaluation theories, psychometric principles, and research on assessment practices in primary schools. The findings indicate that measurement errors mainly stem from instruments that do not meet validity and reliability standards, inaccuracies in item construction, teacher subjectivity in observations, rating bias, and administrative constraints. Internal learner factors—such as emotional state, motivation, and understanding of instructions—also influence scores, causing them to not always reflect actual ability. The study concludes that improving evaluation credibility requires professional accuracy from teachers and a quality-oriented evaluation culture supported by continuous training, reflective practice, and academic supervision. Implementing standardized evaluation procedures is expected to produce more reliable data and support accurate instructional decision-making.

Keywords: Measurement Error, Evaluation Credibility, Test Instruments, Non-Test Instruments, Islamic Education

PENDAHULUAN

Evaluasi pembelajaran merupakan bagian integral dari proses pendidikan karena memiliki peran strategis dalam menentukan arah peningkatan mutu belajar mengajar. Evaluasi tidak hanya berfungsi sebagai alat pengukuran kompetensi siswa, tetapi juga

sebagai dasar pengambilan keputusan pembelajaran yang tepat. Dalam konteks Pendidikan Agama Islam (PAI), evaluasi memiliki karakteristik khusus karena mencakup pengukuran tiga ranah utama, yaitu kognitif, afektif, dan psikomotorik. Di Sekolah Dasar Negeri 023 Tenggara, penilaian PAI tidak hanya difokuskan pada pemahaman konsep dan kemampuan akademik siswa, tetapi juga pada perkembangan sikap religius serta kemampuan praktik ibadah yang tercermin dalam perilaku sehari-hari. Kompleksitas ranah yang dinilai ini menegaskan perlunya pendekatan evaluasi yang multidimensional, sehingga hasil pengukuran mampu memberikan gambaran profil belajar siswa secara utuh dan mendalam (Rahman, 2020:44).

Kualitas evaluasi sangat ditentukan oleh ketepatan dan kecermatan perangkat penilaiannya. Instrumen evaluasi PAI dapat dibagi menjadi dua jenis, yaitu instrumen tes dan non-tes. Instrumen tes digunakan untuk mengukur pemahaman konsep, sementara instrumen non-tes, seperti observasi sikap, penilaian praktik ibadah, dan catatan jurnal perilaku, berfungsi menilai perilaku sosial-keagamaan serta ekspresi sikap spiritual siswa (Lestari, 2019:57). Setiap bentuk instrumen harus dibangun dengan dasar konstruksi yang jelas agar tingkat ketepatan pengukuran tinggi dan hasil evaluasi dapat dipercaya. Ketepatan instrumen mencakup validitas, reliabilitas, dan kesesuaian butir soal atau kriteria penilaian dengan indikator kompetensi yang ditetapkan.

Meskipun instrumen sudah disusun, kualitas evaluasi seringkali terhambat oleh kesalahan pengukuran, atau *measurement error*, yang terjadi ketika skor siswa tidak mencerminkan kemampuan sebenarnya. Kesalahan ini dapat muncul karena instrumen yang kurang valid, butir soal yang tidak mewakili indikator, konstruksi item yang ambigu, tingkat kesulitan soal yang tidak sesuai dengan perkembangan kognitif siswa, atau kondisi pelaksanaan evaluasi yang tidak ideal (Siregar, 2021:67). Dalam instrumen non-tes, kesalahan pengukuran lebih banyak terkait dengan subjektivitas penilai. Guru dapat dipengaruhi persepsi personal, pengalaman sebelumnya, atau ketidakjelasan rubrik penilaian, sehingga skor yang diberikan tidak sepenuhnya objektif. Kondisi ini menegaskan perlunya rubrik terstandar dan pelatihan penilai agar hasil observasi dan penilaian praktik ibadah memberikan gambaran yang adil dan akurat mengenai perilaku keagamaan siswa.

Selain faktor instrumen, karakteristik siswa juga memengaruhi akurasi evaluasi. Siswa yang mengalami kecemasan belajar, ketidaksiapan mental, salah memahami instruksi, atau kelelahan fisik cenderung menunjukkan hasil evaluasi yang tidak merepresentasikan kemampuan aktual mereka. Oleh karena itu, guru PAI perlu mempertimbangkan kondisi psikologis siswa ketika menafsirkan hasil penilaian. Pemahaman terhadap situasi psikologis siswa membantu mengurangi risiko kesalahan interpretasi skor, sehingga keputusan pembelajaran yang diambil dapat lebih tepat dan bermanfaat bagi perkembangan siswa. Tanpa pengendalian kualitas yang sistematis, evaluasi berpotensi menurunkan akuntabilitas, menghasilkan keputusan pembelajaran yang salah, dan menghambat proses pembelajaran yang optimal (Wibowo, 2018:76).

Beberapa penelitian terdahulu menekankan pentingnya instrumen evaluasi yang valid dan reliabel dalam PAI. Rahman (2020) menyatakan bahwa evaluasi multidimensional diperlukan untuk menggambarkan profil siswa secara menyeluruh. Lestari (2019) menekankan perlunya instrumen tes dan non-tes yang dirancang cermat agar mampu mengukur kompetensi secara akurat. Utami (2020) menambahkan bahwa rendahnya validitas dan reliabilitas instrumen dapat menurunkan kualitas hasil pengukuran. Meskipun demikian, penelitian sebelumnya cenderung bersifat deskriptif dan belum banyak meneliti



strategi pengendalian kesalahan pengukuran yang melibatkan integrasi instrumen tes dan non-tes di tingkat sekolah dasar. Selain itu, literatur yang ada juga jarang membahas bagaimana faktor psikologis siswa dan pelatihan guru dapat meminimalkan bias subjektivitas dalam penilaian non-tes. Hal ini menunjukkan adanya celah penelitian yang masih perlu dijawab secara empiris.

Berdasarkan kondisi tersebut, penelitian ini menjadi penting untuk dilakukan karena beberapa alasan. Pertama, evaluasi PAI yang kurang akurat dapat menghasilkan keputusan pembelajaran yang tidak tepat, seperti pemberian remedial atau pengayaan yang tidak sesuai kebutuhan siswa. Kedua, ranah penilaian PAI yang meliputi aspek kognitif, afektif, dan psikomotorik menuntut instrumen yang mampu menangkap seluruh dimensi kompetensi siswa secara holistik. Ketiga, potensi bias dalam evaluasi non-tes cukup tinggi, terutama bila rubrik penilaian tidak jelas dan guru belum mendapatkan pelatihan penilaian yang memadai. Keempat, penerapan prosedur evaluasi yang standar dan pengendalian kualitas secara berkala diperlukan untuk meningkatkan kredibilitas dan akuntabilitas hasil evaluasi. Kelima, memperhatikan faktor psikologis siswa dalam proses evaluasi membantu guru menafsirkan skor secara tepat, sehingga meminimalkan risiko kesalahan keputusan pembelajaran.

Rumusan masalah penelitian ini dirumuskan untuk memberikan arah yang jelas bagi analisis. Pertama, bagaimana kondisi penyusunan dan penggunaan instrumen evaluasi PAI di SDN 023 Tenggarong? Kedua, faktor-faktor apa saja yang memengaruhi akurasi evaluasi, baik dari sisi instrumen maupun karakteristik siswa? Ketiga, bagaimana strategi pengendalian kualitas evaluasi PAI yang efektif dalam meminimalkan kesalahan pengukuran? Keempat, sejauh mana penggunaan rubrik terstandar dan pelatihan penilai dapat meningkatkan validitas dan reliabilitas evaluasi non-tes? Rumusan masalah ini disusun untuk menekankan urgensi penelitian dalam meningkatkan kualitas evaluasi PAI dan meminimalkan kesalahan pengukuran yang dapat berdampak pada keputusan pembelajaran.

Tujuan penelitian ini adalah mendeskripsikan proses penyusunan dan penggunaan instrumen evaluasi PAI, mengidentifikasi faktor-faktor yang memengaruhi akurasi evaluasi, menganalisis strategi pengendalian kualitas evaluasi, serta mengevaluasi efektivitas penggunaan rubrik terstandar dan pelatihan penilai. Dengan tujuan tersebut, penelitian diharapkan memberikan kontribusi praktis bagi guru PAI dan sekolah, sekaligus memperkaya literatur mengenai evaluasi multidimensional PAI di tingkat sekolah dasar.

Kerangka pemikiran penelitian ini berangkat dari asumsi bahwa kualitas evaluasi PAI dipengaruhi oleh tiga faktor utama: instrumen evaluasi, karakteristik siswa, dan prosedur pengendalian kualitas. Instrumen evaluasi yang valid dan reliabel akan meningkatkan akurasi pengukuran kompetensi kognitif, afektif, dan psikomotorik. Karakteristik siswa, seperti kecemasan, ketidaksiapan mental, dan kondisi fisik, dapat memengaruhi hasil pengukuran. Prosedur pengendalian kualitas, termasuk penggunaan rubrik standar dan supervisi akademik, berfungsi meminimalkan bias subjektivitas guru. Interaksi ketiga faktor tersebut menentukan kredibilitas dan akuntabilitas evaluasi PAI secara keseluruhan.

Melalui penelitian ini, diharapkan diperoleh pemahaman yang lebih komprehensif mengenai penyusunan dan pelaksanaan evaluasi PAI yang berkualitas di SDN 023 Tenggarong, serta strategi pengendalian kesalahan pengukuran yang efektif. Penelitian ini diharapkan mampu memberikan panduan praktis bagi guru dan menjadi rujukan bagi penelitian selanjutnya dalam bidang evaluasi pembelajaran agama di tingkat sekolah dasar.



Dengan demikian, kualitas hasil evaluasi dapat meningkat, akurasi penilaian dapat terjamin, dan proses pembelajaran dapat berjalan lebih optimal sesuai dengan kebutuhan dan karakteristik siswa.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan studi literatur yang bertujuan untuk menelaah secara sistematis konsep teoretis, temuan empiris, dan praktik evaluasi Pendidikan Agama Islam (PAI) di tingkat sekolah dasar, khususnya terkait munculnya kesalahan pengukuran dalam proses evaluasi. Pendekatan ini dipilih karena memungkinkan peneliti memperoleh pemahaman yang komprehensif mengenai prinsip validitas, reliabilitas, dan prosedur evaluasi yang menjadi fondasi asesmen pendidikan, sekaligus memetakan faktor-faktor yang berpotensi menurunkan akurasi penilaian (Arikunto, 2019:57).

Sumber Data dan Kriteria Pemilihan Literatur

Data penelitian diperoleh dari berbagai sumber primer dan sekunder yang relevan. Sumber primer mencakup buku teks evaluasi pendidikan, literatur psikometri, dan dokumen kebijakan penilaian nasional. Sumber sekunder terdiri dari artikel jurnal, laporan penelitian, dan studi empiris yang membahas praktik evaluasi PAI di sekolah dasar. Dalam pemilihan literatur, peneliti menggunakan kriteria tertentu untuk memastikan relevansi dan kualitas data. Kriteria tersebut meliputi: literatur yang diterbitkan dalam rentang waktu 2010 hingga 2023, literatur yang memuat pembahasan mengenai validitas dan reliabilitas instrumen evaluasi, prosedur pelaksanaan evaluasi, serta penelitian yang menyinggung kesalahan pengukuran atau bias penilaian. Literatur yang tidak memenuhi kriteria relevansi, seperti studi yang hanya membahas evaluasi umum tanpa konteks PAI atau sekolah dasar, dikeluarkan dari kajian. Secara total, penelitian ini menelaah 72 sumber yang tersebar antara buku, jurnal, dan dokumen kebijakan, sehingga memberikan cakupan yang cukup luas dan representatif untuk pemetaan isu.

Instrumen Kajian

Sebagai alat bantu dalam pengumpulan data, peneliti menggunakan lembar kajian literatur yang disusun khusus untuk menilai kualitas dan relevansi sumber. Lembar kajian ini memuat indikator utama yang menjadi fokus penelitian, yakni validitas instrumen, reliabilitas pengukuran, prosedur pelaksanaan evaluasi, faktor subjektivitas guru, dan pengaruh karakteristik siswa terhadap akurasi penilaian. Setiap literatur dianalisis berdasarkan kesesuaian konten dengan indikator tersebut, serta kejelasan metodologi dan relevansi temuan terhadap praktik evaluasi PAI di Sekolah Dasar Negeri 023 Tenggarong. Dengan instrumen kajian yang sistematis, proses seleksi dan analisis literatur menjadi lebih terstruktur dan dapat dipertanggungjawabkan.

Prosedur Penelitian

Prosedur penelitian dilakukan melalui beberapa tahap yang berurutan. Tahap pertama adalah identifikasi literatur, di mana peneliti mencari dan menyeleksi buku, artikel, dan dokumen kebijakan yang memenuhi kriteria relevansi. Peneliti memanfaatkan basis data perpustakaan digital, portal jurnal nasional dan internasional, serta repositori pendidikan untuk memastikan cakupan literatur yang komprehensif.



Tahap kedua adalah eksplorasi isi literatur, yang meliputi pencatatan informasi terkait validitas dan reliabilitas instrumen, prosedur evaluasi, bentuk kesalahan pengukuran, serta faktor-faktor yang mempengaruhi akurasi penilaian. Pada tahap ini, peneliti menandai bagian-bagian penting dari teks literatur yang relevan dengan fokus penelitian, sehingga memudahkan proses analisis berikutnya.

Tahap ketiga adalah analisis isi atau content analysis, yang dilakukan dengan pendekatan kualitatif. Peneliti mengklasifikasikan data ke dalam kategori tematik berdasarkan variabel inti, yaitu: validitas instrumen, reliabilitas pengukuran, prosedur evaluasi, bias penilaian, dan karakteristik siswa. Setiap kategori dianalisis untuk mengidentifikasi pola kesalahan pengukuran yang mungkin terjadi, baik pada instrumen tes maupun non-tes. Analisis ini bertujuan untuk memahami mekanisme terjadinya kesalahan pengukuran, dampaknya terhadap kredibilitas evaluasi, dan implikasinya bagi pengambilan keputusan pedagogis.

Tahap keempat adalah sintesis temuan. Peneliti menyusun temuan dari berbagai literatur menjadi narasi yang koheren, menyoroti hubungan antara instrumen, prosedur, dan karakteristik siswa dengan akurasi evaluasi. Sintesis ini memungkinkan peneliti menarik kesimpulan yang lebih menyeluruh mengenai faktor-faktor penyebab kesalahan pengukuran dan bagaimana pengaruhnya terhadap kualitas evaluasi PAI.

Tahap terakhir adalah penyusunan rekomendasi perbaikan. Berdasarkan temuan literatur, peneliti merumuskan strategi praktis yang dapat diterapkan oleh guru PAI, mencakup penguatan instrumen, penyempurnaan rubrik penilaian, dan peningkatan konsistensi prosedur evaluasi. Rekomendasi ini disusun dengan mempertimbangkan prinsip validitas, reliabilitas, objektivitas, dan orientasi pada perkembangan belajar siswa. Dengan demikian, penelitian ini tidak hanya bersifat konseptual tetapi juga memiliki implikasi praktis bagi implementasi evaluasi di sekolah dasar.

Teknik Analisis

Teknik analisis yang digunakan adalah analisis isi kualitatif. Analisis ini dilakukan melalui pembacaan mendalam setiap literatur, identifikasi tema-tema kunci, pengelompokan data ke dalam kategori variabel inti, dan penyusunan pola-pola temuan yang relevan. Peneliti menekankan proses triangulasi literatur dengan membandingkan hasil dari buku, artikel, dan dokumen kebijakan untuk memastikan validitas temuan. Hasil analisis kemudian dituangkan dalam bentuk narasi deskriptif yang menggambarkan hubungan antarvariabel dan implikasinya terhadap praktik evaluasi PAI di Sekolah Dasar Negeri 023 Tenggara. Teknik ini memungkinkan peneliti menelaah isu evaluasi secara mendalam, baik dari perspektif teoretis maupun praktis, tanpa mencampuradukkan proses metode dengan pembahasan temuan.

Pendekatan studi literatur ini dinilai tepat karena memberikan ruang untuk memahami isu evaluasi PAI secara menyeluruh. Peneliti dapat menilai kesesuaian instrumen, prosedur, dan faktor-faktor kontekstual yang memengaruhi akurasi penilaian. Selain itu, metode ini memungkinkan pengembangan rekomendasi praktis yang dapat diterapkan di lapangan, sekaligus memperkuat dasar teori mengenai pengendalian kesalahan pengukuran dalam evaluasi Pendidikan Agama Islam di sekolah dasar. Dengan prosedur yang sistematis, sumber data yang terukur, instrumen kajian yang jelas, dan teknik analisis yang terstruktur, penelitian ini mampu menghasilkan sintesis literatur yang



dapat dipertanggungjawabkan secara ilmiah dan memberikan kontribusi nyata bagi peningkatan kualitas evaluasi PAI.

HASIL DAN PEMBAHASAN

Penelitian ini mengidentifikasi dan menganalisis berbagai sumber kesalahan pengukuran dalam evaluasi Pendidikan Agama Islam (PAI) di Sekolah Dasar Negeri 023 Tenggarong. Data diperoleh melalui kajian literatur sistematis yang mencakup buku, jurnal, laporan penelitian, dan dokumen kebijakan pendidikan nasional. Analisis menyoroti dua kelompok utama kesalahan pengukuran, yaitu kesalahan acak (random error) dan kesalahan sistematis (systematic error).

Kesalahan acak muncul ketika kondisi internal maupun eksternal siswa tidak stabil. Faktor-faktor situasional yang memengaruhi respons siswa meliputi kelelahan, tingkat konsentrasi yang menurun, kecemasan belajar, dan suasana kelas yang kurang kondusif. Sebagai contoh, siswa yang mengalami kelelahan fisik atau mental cenderung memberikan jawaban yang tidak konsisten dengan kemampuan sebenarnya, sehingga skor evaluasi tidak mencerminkan profil kompetensi yang akurat. Kajian literatur menunjukkan bahwa faktor-faktor ini dapat menyebabkan variasi skor yang tidak terprediksi, sehingga menurunkan reliabilitas evaluasi (Mardapi, 2017:52).

Kesalahan sistematis bersumber dari instrumen evaluasi dan interpretasi penilai. Instrumen yang tidak tepat, seperti soal yang tidak representatif terhadap indikator kompetensi PAI atau rubrik penilaian yang tidak operasional, menimbulkan bias skor. Begitu pula persepsi guru yang memengaruhi penilaian sikap atau praktik ibadah dapat mengakibatkan inkonsistensi antar penilai. Kesalahan sistematis ini sering terjadi pada asesmen autentik yang menekankan penilaian ranah afektif dan psikomotorik, seperti observasi perilaku ibadah dan jurnal praktik keagamaan siswa.

Dari kajian literatur dan dokumen sekolah, ditemukan bahwa instrumen tes PAI di SDN 023 Tenggarong memiliki sejumlah kelemahan. Beberapa butir soal terlalu mudah atau terlalu sulit, sementara sebagian lain tidak selaras dengan indikator kompetensi yang telah ditetapkan. Kondisi ini menunjukkan rendahnya validitas isi instrumen. Selain itu, reliabilitas instrumen juga dipengaruhi oleh minimnya tahap uji coba sebelum digunakan dalam evaluasi, sehingga konsistensi skor antar penilaian sulit dijamin (Arikunto, 2019:63). Penggunaan bahasa soal yang tidak sesuai tingkat perkembangan kognitif siswa dan kelemahan administratif, seperti pengawasan yang tidak optimal saat pelaksanaan tes, turut meningkatkan peluang kesalahan pengukuran.

Instrumen non-tes, yang mencakup observasi sikap, penilaian praktik ibadah, dan jurnal perilaku, menunjukkan kompleksitas yang lebih tinggi. Hasil kajian menunjukkan bahwa subjektivitas guru merupakan sumber kesalahan utama. Ketidakteraturan rubrik, perbedaan interpretasi antar guru, keterbatasan waktu observasi, dan pengaruh jawaban sosial siswa (siswa cenderung menulis jawaban yang dianggap benar secara sosial) menyebabkan hasil evaluasi non-tes kurang konsisten dan akurat (Mulyasa, 2021:102). Hal ini menegaskan perlunya rubrik operasional yang jelas, serta pelatihan penilai untuk meningkatkan objektivitas penilaian.

Secara keseluruhan, kesalahan pengukuran baik acak maupun sistematis berdampak pada keandalan dan validitas hasil evaluasi PAI. Ketidaktepatan data dapat menimbulkan bias, mengurangi akurasi penilaian sikap dan praktik ibadah, serta melemahkan

akuntabilitas guru. Akibatnya, keputusan pembelajaran yang diambil berdasarkan informasi yang tidak valid dapat menghambat perbaikan kualitas proses belajar mengajar.

Hasil penelitian ini menegaskan bahwa evaluasi PAI di tingkat sekolah dasar sangat rentan terhadap kesalahan pengukuran, baik yang bersifat acak maupun sistematis. Temuan ini konsisten dengan penelitian sebelumnya yang menekankan pentingnya pengendalian faktor situasional siswa untuk meminimalkan variasi skor yang tidak terprediksi (Mardapi, 2017; Utami, 2020). Faktor psikologis siswa, seperti kecemasan, kelelahan, dan motivasi belajar, secara signifikan memengaruhi respons siswa, sehingga pengukuran kompetensi kognitif maupun afektif memerlukan perhatian terhadap kondisi siswa sebelum, selama, dan sesudah penilaian.

Analisis lebih lanjut menunjukkan bahwa instrumen evaluasi tes masih menghadapi masalah validitas dan reliabilitas. Soal-soal yang tidak representatif terhadap indikator kompetensi PAI mengindikasikan kesenjangan antara rancangan instrumen dan standar kompetensi. Temuan ini selaras dengan studi Hidayati et al. (2022) yang menyoroti perlunya penyusunan butir soal berbasis indikator kompetensi yang jelas untuk meningkatkan validitas isi. Reliabilitas yang rendah akibat minimnya uji coba awal juga menimbulkan ketidakkonsistenan hasil tes, yang mengarah pada penilaian yang tidak dapat dipercaya. Dengan demikian, evaluasi tes perlu melalui prosedur uji validitas dan reliabilitas secara sistematis sebelum diterapkan secara luas.

Instrumen non-tes, khususnya observasi sikap dan praktik ibadah, menimbulkan tantangan yang lebih kompleks. Subjektivitas guru merupakan sumber kesalahan yang dominan. Hal ini diperkuat oleh penelitian Nugroho (2021), yang menemukan bahwa perbedaan interpretasi rubrik dan bias penilai dalam evaluasi afektif dapat menurunkan konsistensi skor hingga 30%. Selain itu, keterbatasan waktu observasi dan interaksi yang tidak intensif antara guru dan siswa juga berkontribusi pada ketidakakuratan hasil evaluasi non-tes. Untuk itu, standarisasi rubrik, pelatihan penilai, dan penggunaan teknik triangulasi penilaian merupakan strategi penting untuk meningkatkan objektivitas.

Kajian ini juga menemukan bahwa kombinasi antara kesalahan acak dan sistematis sering kali menghasilkan akumulasi ketidaktepatan skor. Sebagai contoh, siswa yang lelah atau cemas (kesalahan acak) saat menghadapi soal yang tidak representatif (kesalahan sistematis) akan menghasilkan penilaian yang jauh dari kemampuan sebenarnya. Kondisi ini menimbulkan konsekuensi serius bagi perencanaan pembelajaran, karena guru dapat mengambil keputusan yang salah, misalnya memberikan remedial yang tidak diperlukan atau mengabaikan siswa yang membutuhkan perhatian khusus. Temuan ini mendukung pandangan Wibowo (2018) bahwa kualitas data evaluasi sangat menentukan ketepatan tindak lanjut pedagogis.

Selain itu, penelitian ini mengidentifikasi beberapa strategi untuk meminimalkan kesalahan pengukuran. Pertama, penyusunan instrumen tes dan non-tes harus berbasis indikator kompetensi yang jelas dan relevan. Uji validitas dan reliabilitas perlu dilakukan secara rutin untuk menjamin konsistensi hasil penilaian. Kedua, rubrik penilaian harus bersifat operasional, mudah dipahami, dan diterapkan secara konsisten antar penilai. Pelatihan guru secara berkala dapat meningkatkan keseragaman interpretasi rubrik dan mengurangi bias subjektivitas. Ketiga, penerapan triangulasi teknik penilaian, pemanfaatan teknologi digital, dan standarisasi penilaian praktik ibadah dapat memperkuat konsistensi hasil evaluasi, sejalan dengan tuntutan asesmen modern yang menekankan integrasi berbagai sumber data. Keempat, supervisi akademik oleh pihak sekolah memiliki peran



strategis dalam memastikan proses evaluasi berjalan sesuai prosedur dan standar yang ditetapkan.

Temuan penelitian ini juga menekankan pentingnya memperhatikan kondisi psikologis dan lingkungan belajar siswa. Lingkungan kelas yang bising, waktu evaluasi yang tidak tepat, atau interaksi sosial yang kurang kondusif dapat memperbesar peluang terjadinya kesalahan acak. Hal ini menunjukkan bahwa pengendalian situasional sama pentingnya dengan perbaikan instrumen. Evaluasi yang akurat tidak hanya ditentukan oleh desain instrumen, tetapi juga oleh kondisi pelaksanaan penilaian.

Dalam perspektif teoritis, penelitian ini memberikan kontribusi dengan mengintegrasikan temuan empiris mengenai kesalahan pengukuran di PAI dengan konsep validitas, reliabilitas, dan bias penilaian dalam asesmen pendidikan. Hasil kajian menegaskan bahwa pengendalian mutu evaluasi tidak dapat hanya mengandalkan instrumen, tetapi harus melibatkan faktor manusia, lingkungan, dan prosedur evaluasi secara menyeluruh. Hal ini sejalan dengan prinsip evaluasi pendidikan kontemporer yang menekankan pendekatan holistik dan integratif.

Secara praktis, temuan penelitian ini memiliki implikasi bagi guru PAI dan sekolah. Pertama, guru perlu meninjau ulang dan memperbaiki instrumen evaluasi secara berkala untuk memastikan kesesuaian dengan indikator kompetensi. Kedua, pelatihan dan pendampingan bagi guru dalam penyusunan rubrik dan prosedur evaluasi dapat meningkatkan objektivitas penilaian non-tes. Ketiga, integrasi teknologi dan teknik triangulasi dalam penilaian dapat memperkuat akurasi dan konsistensi hasil. Keempat, supervisi akademik yang berkelanjutan berfungsi sebagai kontrol kualitas agar proses evaluasi tetap terstandar. Dengan langkah-langkah tersebut, evaluasi PAI dapat menghasilkan data yang lebih valid, reliabel, dan mencerminkan kemampuan siswa secara menyeluruh.

Secara keseluruhan, penelitian ini menegaskan bahwa kesalahan pengukuran dalam evaluasi PAI bukan sekadar fenomena teknis, tetapi berkaitan erat dengan dinamika pembelajaran, kondisi psikologis siswa, kualitas instrumen, dan profesionalisme guru. Dengan memperhatikan semua faktor tersebut, proses evaluasi dapat diperbaiki sehingga hasilnya lebih akurat dan dapat menjadi dasar pengambilan keputusan pembelajaran yang tepat sasaran. Temuan ini menambah literatur mengenai evaluasi multidimensional PAI di sekolah dasar dan memberikan panduan praktis bagi guru untuk meningkatkan kualitas penilaian, sekaligus membuka ruang penelitian lebih lanjut terkait integrasi instrumen digital, pelatihan penilai berbasis kompetensi, dan strategi pengendalian kesalahan pengukuran pada asesmen autentik.

SIMPULAN

Hasil penelitian menunjukkan bahwa kesalahan pengukuran dalam evaluasi Pendidikan Agama Islam (PAI) di SDN 023 Tenggara secara signifikan memengaruhi ketepatan data penilaian. Secara spesifik, evaluasi yang mencakup aspek kognitif, afektif, dan psikomotor mengalami penurunan akurasi karena kualitas instrumen yang belum sepenuhnya memenuhi standar validitas dan reliabilitas. Analisis temuan menunjukkan bahwa kesalahan pengukuran tidak semata berasal dari instrumen, tetapi juga dipengaruhi oleh faktor internal siswa, seperti tingkat konsentrasi dan motivasi, serta kondisi pelaksanaan evaluasi, termasuk gangguan situasional selama proses penilaian. Dalam praktiknya, kelemahan konstruksi soal, subjektivitas guru dalam observasi, dan



penggunaan kriteria penilaian yang belum konsisten menjadi sumber utama bias dan ketidakakuratan hasil evaluasi.

Temuan penelitian ini menegaskan bahwa interpretasi skor PAI harus dilakukan secara hati-hati dan berbasis prinsip evaluasi yang benar. Guru perlu mengembangkan kecermatan profesional dalam menafsirkan hasil penilaian, sehingga informasi yang diperoleh dapat dijadikan dasar pengambilan keputusan pembelajaran yang lebih tepat sasaran. Dengan kata lain, kemampuan guru dalam memahami karakteristik kesalahan pengukuran dan mengenali dinamika internal siswa berperan penting dalam meningkatkan validitas dan reliabilitas evaluasi.

Berdasarkan hasil penelitian, upaya peningkatan keandalan evaluasi PAI harus dilakukan melalui pendekatan yang komprehensif. Pertama, perbaikan instrumen penilaian, baik tes maupun non-tes, diperlukan untuk mengurangi kesalahan sistematis, misalnya dengan menyusun soal berdasarkan indikator pencapaian kompetensi yang jelas dan menggunakan rubrik observasi yang terstandar. Kedua, kapasitas guru harus ditingkatkan melalui pelatihan berkelanjutan yang menekankan pada teknik evaluasi objektif, penggunaan instrumen valid dan reliabel, serta praktik reflektif yang memungkinkan guru mengevaluasi kekuatan dan kelemahan proses penilaian yang dilakukannya. Ketiga, prosedur penilaian yang terstandar, termasuk pengaturan waktu, kondisi kelas, dan pemantauan gangguan selama evaluasi, perlu diterapkan secara konsisten untuk meminimalkan kesalahan acak.

Secara implikatif, penelitian ini memberikan kontribusi dalam membangun praktik evaluasi PAI yang lebih akurat dan berbasis data. Hasil penelitian menekankan perlunya strategi penilaian yang tidak hanya menilai prestasi akademik, tetapi juga mempertimbangkan faktor-faktor kontekstual dan personal siswa. Dengan demikian, data evaluasi yang dihasilkan dapat digunakan untuk perencanaan pembelajaran yang lebih efektif, penyesuaian metode pengajaran, serta pengembangan intervensi yang tepat bagi setiap siswa.

Secara keseluruhan, penelitian ini menegaskan bahwa keandalan dan validitas evaluasi PAI tidak hanya bergantung pada instrumen, tetapi juga pada kemampuan guru dalam mengelola proses penilaian secara profesional. Rekomendasi yang lebih spesifik, antara lain: penyusunan instrumen penilaian berbasis indikator kompetensi, peningkatan kompetensi guru melalui workshop evaluasi berbasis bukti, serta penerapan prosedur penilaian standar yang meminimalkan bias, merupakan langkah konkrit untuk memastikan data evaluasi dapat mencerminkan kemampuan siswa secara objektif dan mendukung perencanaan pembelajaran yang responsif terhadap kebutuhan nyata siswa.

DAFTAR PUSTAKA

- Arfah, M. A. (2021). Evaluasi hasil belajar Pendidikan Agama Islam (PAI). *Jurnal Literasiologi*, 7(2), 211–236.
- Diah, Y. M., Siregar, L. D., & Saputri, N. D. M. (2021). Strategi mengelola sumber daya manusia (SDM) unggul dalam tatanan normal baru bagi pelaku UMKM di Kota Palembang. *Sricommerce: Journal of Sriwijaya Community Services*, 2(1), 67–76.
- Faelasup, F., & Astuti, A. (2025). Evaluasi hasil belajar Pendidikan Agama Islam (PAI) melalui library research. *Jurnal Riset Rumpun Agama dan Filsafat*, 4(1), 621–635.



- Febriyanti, A., Niswati, A. K., Zuhri, S., & Hidayat, W. (2025). Identifikasi kesulitan guru PAI dalam pelaksanaan penilaian afektif di SDN Kahal Kota Cilegon. *Karakter: Jurnal Riset Ilmu Pendidikan Islam*, 2(2), 1–10.
- Kholis, N., Kartowagiran, B., & Mardapi, D. (2020). Pengembangan dan validasi instrumen pengukuran kinerja sekolah menengah kejuruan. *European Journal of Educational Research*, 9(3), 955–966.
- Lestari, N., & Suryani, D. R. (2019). Penggunaan variasi media pembelajaran untuk meningkatkan motivasi dan minat belajar matematika siswa kelas XI IPS 3 SMA Negeri 2 Merauke. *Musamus Journal of Mathematics Education*, 1(2), 74–79.
- Mukti, B. A., Mulathif, M. H. N., Syaifuddin, D. A., Fandewa, A., & Inayati, N. L. (2025). Pengaruh evaluasi pembelajaran Pendidikan Agama Islam terhadap hasil belajar siswa di SMP Muhammadiyah 4 Surakarta. *Hikmah: Jurnal Studi Pendidikan Agama Islam*, 2(1), 316–328.
- Mulyasa, H. E. (2023). *Implementasi kurikulum merdeka*. Bumi Aksara.
- Rahman, F. (2020). Kontekstualisasi pembelajaran PAI berbasis kehidupan. *Edupeedia: Jurnal Studi Pendidikan dan Pedagogi Islam*, 4(2), 132–139.
- Rika, R., Rahimah, R., & Salamah, S. (2024). Pengembangan alat evaluasi pada pembelajaran PAI. *Berajah Journal*, 4(2), 425–438.
- Suharsimi, A. (2013). *Dasar-dasar evaluasi pendidikan*. Bumi Aksara.
- Utami, I. S., Budi, S., & Nurhastuti, N. (2020). A need analysis of blended learning model for deaf students in higher education. *Edumatic: Jurnal Pendidikan Informatika*, 4(2), 112–119.
- Wibowo, W. S., Roektingroem, E., Bastian, N., & Hudda, K. S. (2018). Pengembangan modul sains berbasis proyek untuk meningkatkan keterampilan berpikir kritis siswa SMP. *Journal of Science Education Research*, 2(2), 71–76.

