

Problematika Guru dalam Penggunaan Instrumen Evaluasi PAI pada Ranah Kognitif, Afektif, dan Psikomotorik

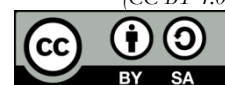
Nafisa Dzatinnuha¹, Syaifuddin²

¹Universitas Islam Negeri Sunan Ampel Surabaya, Indonesia

²Universitas Islam Negeri Sunan Ampel Surabaya, Indonesia

nafisadzatinnuha@gmail.com, syaifuddin.tarbiyah@uinsa.ac.id.

Informasi Artikel	Abstrak
Riwayat artikel: Diterima: 2026-06-05 Direvisi: 2026-06-22 Diterima: 2026-06-26	Evaluasi pembelajaran Pendidikan Agama Islam (PAI) memiliki kompleksitas yang lebih tinggi dibandingkan evaluasi mata pelajaran umum karena tidak hanya berorientasi pada penguasaan pengetahuan, tetapi juga mencakup dimensi moralitas, spiritualitas, karakter, dan keterampilan praktik keagamaan peserta didik. Dalam implementasi Kurikulum Merdeka, guru dituntut melaksanakan penilaian autentik pada ranah kognitif, afektif, dan psikomotorik secara simultan. Akan tetapi, berbagai penelitian menunjukkan bahwa implementasi evaluasi PAI masih menghadapi problem konseptual, teknis, dan pedagogis yang cukup serius. Penelitian ini bertujuan untuk menganalisis problematika penggunaan instrumen evaluasi PAI pada ketiga ranah tersebut, faktor penyebabnya, serta implikasinya terhadap kualitas pembelajaran pendidikan Islam. Penelitian menggunakan pendekatan metode <i>library research</i> dengan teknik dokumentasi terhadap berbagai jurnal ilmiah, buku akademik, dokumen kurikulum, dan penelitian relevan. Analisis data dilakukan menggunakan teknik <i>content analysis</i> berbasis sintesis literatur dengan model analisis interaktif melalui tahapan reduksi data, kategorisasi tematik, triangulasi sumber literatur, serta interpretasi kritis. Hasil penelitian memperlihatkan bahwa evaluasi PAI masih didominasi instrumen berbasis hafalan (<i>Lower Order Thinking Skills</i>), subjektivitas penilaian afektif, dan lemahnya validitas evaluasi praktik ibadah. Berbagai problem tersebut tidak hanya bersifat teknis, tetapi juga berkaitan dengan krisis epistemologis dalam mengukur religiusitas dan karakter peserta didik melalui instrumen formal sekolah. Eksistensi problem tersebut mengonfirmasi perlunya rekonstruksi paradigma evaluasi PAI melalui penguatan literasi evaluasi guru, penggunaan penilaian autentik berbasis digital, serta pengembangan instrumen yang lebih kontekstual, reflektif, dan berkeadilan. <i>Karya ini dilisensikan di bawah Lisensi Creative Commons Attribution 4.0 International (CC BY 4.0).</i>
Kata kunci: Evaluasi PAI Penilaian Autentik Kognitif Afektif Psikomotorik Kurikulum Merdeka	
Jenis artikel: Artikel penelitian	



1. Pendahuluan

1.1. Informasi latar belakang

Evaluasi pembelajaran merupakan bagian integral dalam sistem pendidikan karena berfungsi mengukur ketercapaian tujuan pembelajaran sekaligus menjadi dasar refleksi pedagogis dalam meningkatkan kualitas proses pendidikan. Dalam konteks Pendidikan Agama Islam (PAI), evaluasi tidak hanya diarahkan untuk mengukur penguasaan konsep keagamaan peserta didik, tetapi juga mencakup pembentukan karakter, internalisasi nilai Islam, kesadaran spiritual, serta keterampilan praktik ibadah. Kompleksitas tersebut menyebabkan evaluasi PAI memiliki dimensi yang lebih luas dibandingkan evaluasi pada mata pelajaran umum. Secara filosofis, pendidikan Islam tidak berhenti pada transfer pengetahuan agama, melainkan berupaya membentuk manusia yang memiliki keseimbangan intelektual, moral, spiritual, dan sosial. Konsekuensinya, evaluasi PAI tidak cukup dilakukan melalui tes tertulis yang hanya mengukur kemampuan mengingat materi, tetapi juga

harus mampu menilai perubahan perilaku, kualitas akhlak, dan implementasi nilai-nilai Islam dalam kehidupan sehari-hari. Persoalan muncul ketika aspek-aspek tersebut bersifat abstrak, dinamis, dan sulit diukur secara objektif.

Berbagai penelitian menunjukkan bahwa praktik evaluasi PAI di sekolah masih didominasi penggunaan instrumen tes tertulis berbasis hafalan. Guru cenderung menggunakan soal pilihan ganda yang lebih mudah disusun dan dikoreksi dibandingkan instrumen autentik yang mampu mengukur kemampuan berpikir kritis, refleksi moral, dan keterampilan praktik keagamaan peserta didik (Hanafi et al., 2025; Fadholi et al., 2022). Dominasi instrumen evaluasi yang terjebak pada level *Lower Order Thinking Skills* (LOTS) memperlihatkan bahwa pembelajaran PAI masih berorientasi pada penguasaan materi tekstual dibandingkan pengembangan kesadaran keagamaan yang reflektif dan kontekstual. Dalam implementasi Kurikulum Merdeka, guru dituntut melaksanakan penilaian autentik yang mencakup ranah kognitif, afektif, dan psikomotorik secara integratif. Namun, berbagai kajian memperlihatkan adanya kontradiksi antara tuntutan kurikulum dan kapasitas evaluatif guru di lapangan. Penelitian mengenai implementasi penilaian autentik menemukan bahwa banyak guru mengalami kesulitan menyusun instrumen evaluasi yang valid, reliabel, dan sesuai dengan karakteristik pembelajaran PAI (Hanafi et al., 2025; Marzoan et al., 2025). Kondisi tersebut memperlihatkan bahwa transformasi kurikulum tidak selalu diikuti oleh penguatan kompetensi evaluasi guru secara memadai.

Persoalan evaluasi menjadi semakin kompleks ketika menyentuh ranah afektif. Penilaian karakter dan religiusitas peserta didik sering kali dilakukan melalui observasi singkat yang rentan terhadap subjektivitas guru. Dalam banyak kasus, penilaian sikap berubah menjadi formalitas administratif tanpa indikator perilaku yang jelas dan terukur. Kesenjangan tersebut mengindikasikan bahwa evaluasi afektif dalam PAI masih menghadapi problem validitas dan reliabilitas yang serius. Dalam perspektif psikologi pendidikan, penilaian afektif sangat rentan terhadap *halo effect*, yaitu kecenderungan guru memberikan penilaian berdasarkan kesan umum terhadap peserta didik tertentu. Kerentanan tersebut menjadi lebih kompleks dalam PAI karena objek evaluasinya berkaitan langsung dengan moralitas dan simbol kesalehan. Peserta didik dapat menampilkan perilaku religius secara situasional ketika berada di bawah pengawasan guru, tetapi belum tentu merepresentasikan karakter autentik dalam kehidupan sehari-hari. Fenomena tersebut dikenal sebagai *performative religiosity* atau kesalehan performatif.

Selain itu, evaluasi psikomotorik dalam PAI juga menghadapi berbagai kendala teknis dan pedagogis. Penilaian praktik ibadah seperti salat, wudhu, membaca Al-Qur'an, dan praktik keagamaan lainnya membutuhkan observasi detail serta rubrik analitik yang sistematis. Akan tetapi, berbagai penelitian menunjukkan bahwa implementasi penilaian praktik masih cenderung dilakukan secara global tanpa indikator yang jelas karena keterbatasan waktu, rasio siswa yang besar, dan rendahnya literasi evaluasi guru (Marzoan et al., 2025). Dalam perspektif pendidikan Islam, evaluasi sesungguhnya memiliki dimensi moral dan spiritual yang sangat kuat. Prinsip keadilan dan objektivitas dalam penilaian tercermin dalam QS. Al-Zalzalah ayat 7–8 yang menjelaskan bahwa seluruh amal manusia akan diperhitungkan secara adil tanpa ada pengurangan sedikit pun. Konsep hisab dan mizan dalam Islam memperlihatkan bahwa evaluasi bukan sekadar aktivitas administratif, tetapi juga bentuk amanah moral yang harus dilakukan secara jujur, objektif, dan bertanggung jawab.

Reduksi makna evaluasi yang hanya berorientasi pada pemberian angka berimplikasi pada hilangnya substansi pendidikan Islam sebagai proses pembentukan manusia berkarakter dan berkepribadian Islami. Dalam konteks tersebut, problem evaluasi PAI tidak dapat dipahami semata sebagai persoalan teknis penyusunan instrumen, melainkan sebagai problem epistemologis mengenai bagaimana religiusitas, moralitas, dan spiritualitas manusia diukur melalui mekanisme evaluasi formal sekolah.

1.2. Kesenjangan penelitian

Kajian mengenai evaluasi pembelajaran PAI telah banyak dilakukan, tetapi mayoritas penelitian masih bersifat parsial dan lebih menitikberatkan pada aspek teknis penggunaan instrumen penilaian. Penelitian Hanafi et al. (2025) misalnya, lebih memfokuskan pembahasan pada implementasi penilaian autentik dalam pembelajaran PAI tanpa mengaitkannya dengan problem filosofis pengukuran karakter dan religiusitas peserta didik. Penelitian tersebut juga belum mengintegrasikan problem evaluasi pada ranah kognitif, afektif, dan psikomotorik secara simultan. Kajian Marzooan et al. (2025) lebih berfokus pada evaluasi keterampilan membaca Al-Qur'an dan praktik ibadah peserta didik sekolah dasar. Penelitian tersebut memberikan kontribusi penting dalam aspek psikomotorik, tetapi belum membahas hubungan antara problem evaluasi praktik dengan tekanan administratif dan kompetensi evaluasi guru.

Sementara itu, penelitian mengenai *Higher Order Thinking Skills* (HOTS) dalam pembelajaran PAI umumnya hanya menyoroti pengembangan soal berbasis kognitif tanpa mengaitkan budaya hafalan, orientasi nilai, dan problem literasi evaluasi guru sebagai akar persoalan evaluasi PAI secara lebih luas. Beberapa penelitian lain juga masih memosisikan problem evaluasi sebagai persoalan administratif semata, padahal realitas menunjukkan bahwa evaluasi PAI berkaitan erat dengan dilema objektivitas moral dan validitas pengukuran karakter. Berdasarkan kondisi tersebut, masih terdapat kekosongan kajian yang membahas problematika penggunaan instrumen evaluasi PAI pada ranah kognitif, afektif, dan psikomotorik secara integratif dengan pendekatan kritis-analitis. Artikel ini memosisikan problem evaluasi PAI tidak hanya sebagai problem pedagogis, tetapi juga sebagai krisis epistemologis dalam mengukur keberhasilan pendidikan Islam melalui instrumen formal pendidikan.

1.3. Tujuan penelitian

Penelitian ini bertujuan untuk:

1. Menganalisis problematika penggunaan instrumen evaluasi PAI pada ranah kognitif, afektif, dan psikomotorik.
2. Mengidentifikasi faktor-faktor internal dan eksternal yang memengaruhi rendahnya efektivitas evaluasi pembelajaran PAI.
3. Menganalisis implikasi problem evaluasi terhadap kualitas pembelajaran pendidikan Islam.
4. Merumuskan strategi pengembangan evaluasi PAI yang lebih autentik, objektif, dan kontekstual.

1.4. Pertanyaan penelitian

1. Bagaimana problematika penggunaan instrumen evaluasi PAI pada ranah kognitif, afektif, dan psikomotorik?
2. Mengapa implementasi penilaian autentik dalam pembelajaran PAI masih menghadapi berbagai kendala?
3. Faktor-faktor apa saja yang menyebabkan rendahnya validitas dan objektivitas evaluasi PAI?
4. Bagaimana strategi rekonstruksi evaluasi PAI yang lebih relevan dengan tuntutan Kurikulum Merdeka?.

1.5. Signifikansi penelitian

Secara teoritis, penelitian ini diharapkan memberikan kontribusi dalam pengembangan kajian evaluasi pendidikan Islam, khususnya terkait integrasi antara teori evaluasi modern dan paradigma evaluasi dalam Islam. Penelitian ini juga memperluas diskursus mengenai problem epistemologis penilaian karakter dan religiusitas peserta didik. Secara praktis, penelitian ini dapat menjadi bahan refleksi bagi guru PAI dalam mengembangkan instrumen evaluasi yang lebih autentik dan objektif. Selain itu, penelitian ini diharapkan dapat menjadi rujukan bagi sekolah dan pengembang kurikulum dalam memperkuat sistem evaluasi pembelajaran PAI yang lebih kontekstual dan berorientasi pada pembentukan karakter.

2. Metodologi

2.1. Desain Penelitian

Penelitian ini menggunakan pendekatan berbasis studi kepustakaan (*library research*). Pilihan metode tersebut didasarkan pada karakter persoalan yang dikaji, yakni problematika penggunaan instrumen evaluasi Pendidikan Agama Islam (PAI) pada ranah kognitif, afektif, dan psikomotorik yang telah banyak dibahas dalam berbagai penelitian sebelumnya, tetapi masih tersebar dan belum dibaca secara integratif. Fokus penelitian tidak diarahkan pada pengumpulan data lapangan melalui observasi atau wawancara, melainkan pada penelaahan kritis terhadap dokumen akademik yang memotret praktik evaluasi PAI di sekolah. Pendekatan ini dipilih karena problem evaluasi PAI sesungguhnya bukan sekadar persoalan teknis penyusunan instrumen, tetapi juga berkaitan dengan persoalan pedagogis, filosofis, dan epistemologis dalam mengukur religiusitas, moralitas, dan keterampilan praktik keagamaan peserta didik.

Penelitian kepustakaan dalam artikel ini tidak diposisikan sebagai aktivitas merangkum teori secara mekanis. Pembacaan literatur dilakukan secara analitis untuk menemukan pola problem yang terus berulang dalam praktik evaluasi pembelajaran PAI. Dari berbagai penelitian yang ditelaah, terlihat adanya kecenderungan serupa: dominasi instrumen berbasis hafalan, subjektivitas penilaian sikap, lemahnya validitas evaluasi praktik ibadah, serta tekanan administratif yang memengaruhi kualitas asesmen guru. Pendekatan deskriptif analitis digunakan untuk membaca hubungan antarproblem secara lebih mendalam. Literatur dipahami bukan hanya sebagai sumber informasi teoritis, tetapi juga sebagai representasi realitas pendidikan yang memperlihatkan ketegangan antara idealitas penilaian autentik dengan praktik evaluasi di sekolah.

2.2. Sumber Data dan Parameter Seleksi Pustaka

Data penelitian sepenuhnya bersifat sekunder berupa dokumen akademik yang berkaitan dengan evaluasi pembelajaran Pendidikan Agama Islam. Sumber data meliputi artikel jurnal nasional terakreditasi, buku evaluasi pendidikan, dokumen kebijakan kurikulum, serta penelitian ilmiah yang membahas asesmen pembelajaran agama Islam di tingkat dasar maupun menengah.

Pemilihan bahan pustaka dilakukan menggunakan teknik seleksi berbasis kriteria (*criterion-based selection*). Langkah ini ditempuh agar literatur yang digunakan benar-benar relevan dengan fokus penelitian. Tidak seluruh kajian tentang evaluasi pendidikan dimasukkan ke dalam analisis karena banyak penelitian hanya membahas evaluasi secara umum tanpa mengaitkannya dengan karakteristik khas pembelajaran PAI.

Beberapa parameter digunakan dalam menentukan kelayakan literatur, yaitu:

1. Literatur membahas evaluasi pembelajaran Pendidikan Agama Islam.
2. Kajian memuat pembahasan mengenai penilaian kognitif, afektif, atau psikomotorik.
3. Artikel berasal dari jurnal ilmiah kredibel, buku akademik, atau dokumen resmi pendidikan.
4. Publikasi berada dalam rentang tahun 2020–2026 agar tetap relevan dengan perkembangan Kurikulum Merdeka.
5. Literatur memuat problem praksis evaluasi di sekolah, bukan sekadar uraian konseptual normatif.

Penelusuran pustaka dilakukan melalui Google Scholar, Garuda, DOAJ, serta beberapa repository perguruan tinggi. Dari hasil pencarian awal ditemukan beberapa artikel dengan tema evaluasi pendidikan Islam. Namun setelah melalui proses penyaringan bertahap, hanya sumber yang memiliki keterkaitan langsung dengan problem instrumen evaluasi PAI yang digunakan sebagai bahan sintesis utama dalam penelitian ini.

2.3. Instrumen

Dalam penelitian berbasis studi kepustakaan, peneliti berfungsi sebagai instrumen utama (*human instrument*). Peneliti secara langsung melakukan pembacaan, interpretasi, klasifikasi, dan sintesis

terhadap berbagai dokumen yang digunakan sebagai sumber data penelitian. Proses analisis tetap membutuhkan perangkat bantu agar pembacaan data tidak kehilangan fokus. Karena itu, penelitian ini menggunakan lembar dokumentasi dan matriks sintesis literatur. Lembar dokumentasi digunakan untuk mencatat identitas sumber, fokus kajian, argumentasi utama, serta relevansi temuan penelitian terhadap problem evaluasi PAI.

Matriks sintesis literatur dipakai untuk membaca pola problem yang terus muncul dalam berbagai penelitian. Dari proses pemetaan tersebut terlihat kecenderungan yang cukup konsisten: evaluasi PAI masih didominasi instrumen berbasis hafalan, implementasi *Higher Order Thinking Skills* (HOTS) berjalan terbatas, dan penilaian praktik ibadah sering berakhir sebagai formalitas administratif. Ada satu problem yang cukup sering muncul dalam penelitian kepustakaan, yakni kecenderungan menjadikan artikel hanya sebagai kumpulan kutipan teori tanpa arah analisis yang jelas. Karena itu, pemetaan data dalam penelitian ini tidak diarahkan sekadar merangkum isi literatur, tetapi digunakan untuk membaca hubungan antar temuan secara lebih reflektif dan kritis.

2.4. Prosedur Pengumpulan Data

Pengumpulan data dilakukan melalui teknik dokumentasi terhadap berbagai sumber pustaka yang relevan dengan fokus penelitian. Penelusuran dokumen diawali dengan pelacakan kata kunci pada basis data akademik menggunakan kombinasi istilah seperti “Evaluasi PAI”, “Asesmen Autentik Pendidikan Islam”, “HOTS Pendidikan Agama Islam”, “Problem Penilaian Afektif”, dan “Evaluasi Psikomotorik PAI”. Dokumen yang ditemukan tidak seluruhnya langsung digunakan. Peneliti melakukan proses penyaringan (*screening*) secara bertahap untuk memisahkan artikel yang benar-benar membahas problem evaluasi PAI dengan artikel yang hanya menyinggung evaluasi pendidikan secara umum.

Tahap berikutnya dilakukan melalui pembacaan intensif terhadap dokumen yang lolos seleksi. Pada proses ini peneliti mulai melakukan anotasi data dengan mencatat bagian-bagian yang menunjukkan letak problem penggunaan instrumen evaluasi oleh guru PAI. Sejumlah penelitian memperlihatkan bahwa guru cenderung menggunakan soal berbasis hafalan karena tekanan administrasi dan keterbatasan waktu. Penelitian lain menunjukkan tingginya subjektivitas penilaian afektif akibat absennya indikator yang benar-benar terukur.

Beberapa hasil penelitian juga memperlihatkan kontradiksi tertentu. Ada penelitian yang menempatkan rendahnya kompetensi guru sebagai akar problem evaluasi, tetapi penelitian lain menunjukkan bahwa sistem administrasi sekolah dan tuntutan kurikulum justru menjadi faktor dominan yang membatasi ruang reflektif guru dalam menyusun asesmen autentik. Perbedaan-perbedaan temuan tersebut tidak dieliminasi, tetapi dibaca sebagai bagian penting untuk memahami kompleksitas evaluasi PAI secara lebih utuh. Dari proses itulah mulai terlihat bahwa problem evaluasi PAI tidak berdiri sendiri, melainkan saling berkaitan dengan budaya pedagogi, sistem sekolah, serta orientasi pendidikan yang terlalu administratif.

2.5. Analisis data

Analisis data dilakukan menggunakan teknik *content analysis* berbasis sintesis literatur. Teknik ini digunakan untuk mengidentifikasi pola, kecenderungan, dan kontradiksi problem evaluasi pembelajaran PAI yang muncul dalam berbagai penelitian terdahulu.

Tahap pertama dimulai melalui reduksi data. Peneliti memilah berbagai hasil penelitian yang benar-benar relevan dengan fokus kajian. Literatur yang terlalu umum atau tidak memiliki hubungan langsung dengan evaluasi pembelajaran PAI dieliminasi agar analisis tetap terarah. Setelah proses reduksi, data kemudian dikelompokkan berdasarkan tema-tema problematik tertentu. Pada tahap ini mulai terlihat pola yang cukup konsisten, seperti dominasi instrumen *Lower Order Thinking Skills* (LOTS) dalam evaluasi kognitif, subjektivitas penilaian sikap pada ranah afektif, serta lemahnya validitas evaluasi praktik ibadah pada ranah psikomotorik.

Analisis tidak berhenti pada pengelompokan tema. Penelitian ini juga berupaya membaca hubungan antarproblem secara lebih mendalam. Dominasi soal hafalan, misalnya, ternyata tidak hanya dipengaruhi rendahnya kemampuan guru menyusun soal HOTS. Ada faktor lain yang lebih struktural, seperti tekanan administratif sekolah, keterbatasan waktu pembelajaran, dan budaya evaluasi yang terlalu berorientasi pada angka. Persoalan serupa muncul dalam penilaian afektif. Secara normatif sekolah ingin membentuk karakter religius peserta didik, tetapi praktik evaluasinya justru sering terjebak pada simbol kesalehan yang mudah dimanipulasi. Akibatnya, penilaian afektif berpotensi berubah menjadi formalitas administratif yang sulit menangkap kualitas moral peserta didik secara autentik. Karena itu, analisis dalam penelitian ini tidak diarahkan untuk mencari problem guru semata. Fokus utamanya adalah membaca bagaimana sistem evaluasi pendidikan modern sering kali memaksa sesuatu yang abstrak seperti religiusitas, moralitas, dan spiritualitas kemudian masuk ke dalam format penilaian kuantitatif yang seragam dan administratif.

3. Hasil

3.1. Problematika Evaluasi Ranah Kognitif

Evaluasi ranah kognitif dalam pembelajaran PAI masih bergerak di sekitar wilayah hafalan dan reproduksi teks. Potret tersebut muncul berulang dalam berbagai penelitian yang menelaah implementasi asesmen PAI di sekolah. Instrumen evaluasi yang digunakan guru sebagian besar belum diarahkan untuk membaca kemampuan berpikir kritis peserta didik, melainkan sekadar memastikan apakah siswa mampu mengingat kembali materi yang sudah diajarkan. Berhentinya implementasi evaluasi autentik pada *level Lower Order Thinking Skills* (LOTS) tercatat dalam temuan Hanafi et al. (2025), serta Fadholi et al. (2022). Guru masih sangat bergantung pada soal definisional dan pertanyaan normatif yang jawabannya tersedia langsung di buku ajar. Kecenderungan ini tampak dari dominasi penggunaan Kata Kerja Operasional (KKO) seperti menyebutkan, menghafal, menjelaskan, atau menuliskan kembali.

Jejak problem tersebut terlihat cukup jelas pada contoh instrumen evaluasi yang ditemukan dalam berbagai penelitian terdahulu. Pertanyaan seperti Jelaskan pengertian zakat menurut bahasa dan istilah atau Sebutkan syarat wajib puasa masih mendominasi lembar evaluasi semester. Instrumen semacam ini memang relevan untuk mengukur penguasaan konsep dasar, tetapi gagal membaca kemampuan peserta didik ketika dihadapkan pada problem sosial keagamaan yang lebih kompleks. Kondisi di lapangan belum menunjukkan perubahan yang signifikan meskipun Kurikulum Merdeka menekankan pentingnya asesmen autentik dan penguatan *Higher Order Thinking Skills* (HOTS). Kajian Iqbal et al. (2024) menemukan bahwa penggunaan stimulus berbasis kasus dalam pembelajaran fikih masih sangat terbatas. Guru mengalami kesulitan menyusun soal kontekstual karena belum terbiasa membangun pertanyaan yang mendorong kemampuan interpretasi dan argumentasi.

Menariknya, sebagian guru memahami soal HOTS hanya sebagai soal yang panjang dan sulit secara bahasa. Problem utamanya ternyata bukan sekadar kemampuan teknis membuat soal, tetapi cara pandang terhadap evaluasi itu sendiri. Pembelajaran agama masih dipahami sebagai proses transfer materi normatif yang harus dihafal siswa. Akibatnya, evaluasi lebih banyak diarahkan pada pengukuran daya ingat dibanding kemampuan refleksi keagamaan. Tekanan administratif sekolah ikut memperkuat pola tersebut. Instrumen berbasis hafalan dianggap lebih praktis, mudah dikoreksi, dan aman digunakan untuk memenuhi target penilaian. Orientasi evaluasi yang terlalu berfokus pada capaian angka dan penyelesaian administratif cenderung mempersempit ruang refleksi pedagogis dalam pembelajaran PAI. Akibatnya, evaluasi lebih diarahkan pada penguasaan informasi tekstual dibanding pengembangan kemampuan berpikir reflektif peserta didik. Pengetahuan Islam direduksi menjadi kumpulan informasi tekstual, bukan sarana membangun kesadaran kritis peserta didik terhadap realitas sosial yang terus berubah.

Dampaknya tidak kecil. Peserta didik mungkin mampu menjelaskan definisi akhlak, zakat, atau toleransi secara teoritis, tetapi kebingungan ketika diminta membaca konflik sosial keagamaan di ruang publik menggunakan perspektif nilai Islam yang substantif. Problem evaluasi kognitif PAI pada akhirnya bukan hanya persoalan instrumen, melainkan persoalan paradigma pedagogi yang masih terlalu tekstual dan berorientasi hafalan.

3.2. Problematika Evaluasi Ranah Afektif

Ranah afektif menjadi wilayah evaluasi yang paling problematik dalam pembelajaran PAI. Literatur yang dianalisis memperlihatkan bahwa penilaian sikap religius peserta didik masih sangat rentan terhadap subjektivitas dan bias personal guru. Masalahnya bermula dari penggunaan indikator penilaian yang terlalu abstrak. Guru sering menggunakan kategori seperti religius, jujur, bertanggung jawab, atau sopan tanpa parameter perilaku yang benar-benar operasional. Akibatnya, standar observasi menjadi kabur. Penilaian akhirnya lebih banyak ditentukan oleh kesan personal dibanding pengamatan sistematis. Penilaian sikap keagamaan di sekolah juga rentan terdistorsi oleh kedekatan emosional, sebuah gejala yang erat kaitannya dengan *halo effect*. Dari dokumen riset yang ditelaah, guru cenderung memberikan skor afektif tinggi kepada siswa yang aktif mengikuti kegiatan keagamaan sekolah atau mereka yang secara visual menampilkan atribut religius tertentu. Sebaliknya, siswa yang lebih pendiam dan tidak terlalu ekspresif dalam menunjukkan simbol kesalehan sering dipersepsikan memiliki kualitas afektif yang rendah (Halimah 2021).

Bias penilaian tersebut diperparah oleh munculnya fenomena *social desirability* bias. Demi mengamankan nilai rapor, peserta didik kerap “menyutradarai” perilaku mereka ketika berada di bawah pengawasan guru. Mereka belajar bagaimana tampil religius di ruang sekolah: rajin memberi salam, aktif mengikuti kegiatan rohani, atau menunjukkan kepatuhan simbolik tertentu ketika proses observasi berlangsung.

Dampaknya cukup mengkhawatirkan. Sistem evaluasi afektif yang terlalu administratif berpotensi memelihara budaya kesalehan performatif (*performative religiosity*). Alih-alih membangun kesadaran moral yang autentik, sekolah justru tanpa sadar melatih peserta didik untuk mahir memanipulasi simbol kesalehan demi kebutuhan nilai akademik (Dona et al. 2024). Di titik inilah evaluasi afektif PAI menghadapi dilema filosofis yang serius. Sekolah ingin mengukur karakter religius siswa, tetapi instrumen yang digunakan hanya mampu menangkap perilaku yang tampak di permukaan. Pertanyaan mendasarnya menjadi sulit dihindari: apakah sistem evaluasi sekolah benar-benar sedang membentuk moralitas peserta didik, atau hanya membentuk kemampuan mereka memainkan citra moral di depan guru?

Sejumlah penelitian juga memperlihatkan praktik penilaian afektif yang dilakukan sekadar untuk memenuhi administrasi rapor. Guru terkadang memberikan nilai sikap secara kolektif karena keterbatasan waktu observasi dan tingginya beban kerja administratif. Dalam situasi seperti ini, penilaian afektif kehilangan fungsi pedagogisnya dan berubah menjadi rutinitas birokratis. Problem tersebut memperlihatkan bahwa evaluasi afektif dalam PAI tidak dapat dipahami sekadar sebagai kelemahan teknis penyusunan rubrik (Pribadi et al. 2022). Ada problem epistemologis yang jauh lebih dalam: moralitas dan religiusitas manusia dipaksa masuk ke dalam format penilaian administratif yang seragam dan kuantitatif.

3.3. Problematika Evaluasi Ranah Psikomotorik

Evaluasi psikomotorik dalam pembelajaran PAI umumnya berkaitan dengan praktik ibadah seperti salat, wudhu, membaca Al-Qur'an, khutbah, hingga praktik penyelenggaraan jenazah. Secara teoritis, ranah ini terlihat lebih mudah dievaluasi karena guru dapat mengamati performa siswa secara langsung. Akan tetapi, berbagai penelitian justru memperlihatkan banyak persoalan muncul pada tahap implementasinya. Kajian Marzoan et al. (2025) menunjukkan bahwa penilaian praktik salat dan membaca Al-Qur'an di sejumlah sekolah masih dilakukan secara global tanpa rubrik analitik yang rinci. Guru umumnya memberikan skor berdasarkan impresi umum terhadap penampilan siswa ketika praktik berlangsung. Situasi ini menyebabkan standar penilaian menjadi

tidak konsisten antar guru. Instrumen evaluasi praktik yang ditemukan dalam beberapa penelitian juga masih sangat sederhana. Sebagian guru hanya menggunakan checklist umum dengan kategori baik, cukup, atau kurang tanpa indikator operasional yang jelas. Akibatnya, penilaian praktik ibadah sulit dipastikan validitas dan reliabilitasnya.

Kesulitan paling besar muncul ketika guru harus menilai aspek ibadah yang sifatnya abstrak. Gerakan salat memang dapat diamati secara langsung, tetapi kualitas kekhayalan, penghayatan doa, dan kesadaran spiritual peserta didik tidak mudah diterjemahkan menjadi skor angka. Konfirmasi ini mempertegas batas evaluasi psikomotorik PAI tidak bisa mentah-mentah direduksi menjadi sekadar pengukuran perilaku motorik biasa. Persoalan lain datang dari keterbatasan waktu pembelajaran. Dalam kelas dengan jumlah siswa besar, guru hampir tidak memiliki cukup waktu melakukan observasi detail terhadap seluruh peserta didik. Akibatnya, evaluasi praktik sering dilakukan secara cepat agar seluruh siswa tetap memperoleh nilai.

Sebagian penelitian juga memperlihatkan rendahnya pemahaman guru terhadap konsep *performance assessment*. Evaluasi praktik dipahami hanya sebagai observasi perilaku sederhana tanpa penyusunan rubrik performa yang sistematis. Padahal, asesmen praktik membutuhkan indikator rinci agar hasil penilaian lebih konsisten dan dapat dipertanggungjawabkan secara akademik. Keterbatasan fasilitas sekolah ikut memperumit keadaan. Tidak semua sekolah memiliki ruang praktik atau sarana pembelajaran ibadah yang memadai. Pada beberapa kasus, praktik ibadah dilakukan secara bergantian dengan waktu yang sangat terbatas sehingga guru tidak memiliki ruang cukup untuk melakukan observasi mendalam.

Keseluruhan temuan tersebut memperlihatkan bahwa problem evaluasi psikomotorik PAI bukan semata persoalan kompetensi guru. Ada problem struktural yang ikut bekerja di belakangnya: tekanan administrasi sekolah, rasio siswa yang besar, keterbatasan waktu pembelajaran, dan paradigma evaluasi yang terlalu berorientasi pada angka kuantitatif.

3.4. Faktor Penyebab Problematika Evaluasi PAI

Problematika evaluasi PAI tidak lahir dari satu sumber tunggal. Literatur yang dianalisis memperlihatkan adanya tumpang tindih antara persoalan kompetensi guru, kultur pedagogi sekolah, tekanan administratif, hingga cara pandang ideologis terhadap pembelajaran agama itu sendiri. Karena itu, reduksi masalah evaluasi hanya pada aspek kurangnya pelatihan sesungguhnya terlalu menyederhanakan persoalan. Beban administrasi memang menjadi faktor yang paling sering muncul dalam berbagai penelitian. Guru PAI harus menyelesaikan perangkat ajar, penginputan nilai digital, laporan pembelajaran, asesmen diagnostik, hingga tuntutan dokumentasi Kurikulum Merdeka dalam waktu yang terbatas. Situasi ini membuat evaluasi lebih diarahkan untuk memenuhi kebutuhan birokrasi sekolah dibanding kebutuhan reflektif peserta didik. Tidak mengherankan jika sebagian guru memilih menggunakan soal hafalan dan penilaian instan yang dianggap lebih aman secara administratif.

Namun akar problemnya tidak berhenti di sana. Sejumlah penelitian memperlihatkan adanya hambatan epistemologis yang khas dalam pendidikan agama. Dalam kultur sebagian pendidik PAI, penguasaan teks keagamaan secara literal dan hafalan dalil masih dipahami sebagai indikator utama keberhasilan pembelajaran. Cara pandang ini lahir dari kekhawatiran bahwa pembelajaran yang terlalu dialogis dan kritis dapat menggeser kemurnian ajaran agama. Di titik tertentu, muncul semacam sakralisasi terhadap model evaluasi tekstual. Hafalan dianggap identik dengan kesungguhan beragama, sedangkan pertanyaan berbasis studi kasus kadang dicurigai membuka ruang relativisme moral. Akibatnya, guru lebih nyaman menggunakan instrumen definisional yang jawabannya tunggal dan pasti dibanding pertanyaan reflektif yang memungkinkan munculnya argumentasi beragam dari siswa.

Situasi tersebut menjelaskan mengapa transformasi evaluasi PAI sering berjalan lambat meskipun kurikulum telah berubah. Problematika ternyata bukan semata kurangnya keterampilan teknis menyusun soal HOTS, tetapi juga benturan paradigma mengenai apa yang dianggap sebagai pengetahuan agama yang benar. Dalam kultur pedagogi yang sangat normatif, evaluasi kritis sering

dipersepsikan berisiko mengaburkan otoritas teks keagamaan. Masalah lain muncul pada rendahnya literasi evaluasi guru. Banyak guru masih memahami evaluasi hanya sebagai aktivitas pemberian skor akhir, bukan proses pedagogis untuk membaca perkembangan belajar peserta didik secara komprehensif. Dampaknya terlihat pada instrumen yang monoton, minim stimulus kontekstual, dan tidak memiliki indikator performa yang jelas.

Persoalan tersebut menjadi semakin rumit ketika dikaitkan dengan rasio siswa yang besar. Guru kesulitan melakukan observasi afektif dan evaluasi praktik secara mendalam dalam kelas yang padat. Penilaian akhirnya dilakukan secara global agar seluruh siswa tetap memperoleh nilai rapor tepat waktu. Dalam kondisi semacam ini, objektivitas evaluasi hampir mustahil dijaga secara maksimal. Keterbatasan sarana pembelajaran juga ikut memengaruhi kualitas asesmen praktik ibadah. Tidak semua sekolah memiliki fasilitas yang mendukung pelaksanaan evaluasi psikomotorik secara ideal. Pada beberapa kasus, praktik salat atau membaca Al-Qur'an dilakukan dalam waktu yang sangat singkat sehingga guru tidak memiliki ruang cukup untuk mengamati detail performa peserta didik. Yang sering luput dibicarakan adalah kenyataan bahwa sebagian guru sendiri mengalami kebingungan menghadapi perubahan paradigma evaluasi. Mereka diminta menerapkan asesmen autentik, tetapi pada saat yang sama masih dibebani target ketuntasan kurikulum dan tuntutan angka kuantitatif rapor. Situasi ini memunculkan dilema praktis: guru didorong melakukan evaluasi reflektif, tetapi sistem sekolah tetap menuntut kepastian angka yang cepat dan seragam.

3.5. Tren Rekonstruksi dan Inovasi Evaluasi PAI dalam Literatur Kontemporer

Literatur mutakhir mulai mendorong perubahan yang cukup radikal dalam desain evaluasi PAI. Fokusnya tidak lagi sekadar mempercantik format instrumen, melainkan menggeser orientasi evaluasi dari budaya hafalan menuju pembelajaran reflektif yang lebih kontekstual. Pergeseran paradigma tersebut tampak dalam berbagai kajian asesmen autentik PAI yang dilakukan oleh Hanafi et al. (2025), Fadholi et al. (2022), serta Iqbal et al. (2024). Kecenderungan tersebut terlihat dari semakin banyaknya penelitian yang mengkritik dominasi soal definisional dalam pembelajaran agama. Instrumen evaluasi dinilai tidak lagi memadai apabila hanya mengukur kemampuan siswa mengulang isi buku teks atau dalil normatif. Arifin (2021) dan Hanafi et al. (2025) menegaskan bahwa pembelajaran PAI perlu diarahkan pada kemampuan membaca problem sosial keagamaan secara reflektif dan argumentatif, bukan berhenti pada penguasaan konsep verbal semata.

Karena itu, sejumlah penelitian mulai menawarkan penggunaan *project-based assessment* dan studi kasus kontekstual sebagai alternatif evaluasi. Iqbal et al. (2024), misalnya, merekomendasikan pengembangan instrumen berbasis problem sosial kontemporer agar peserta didik mampu menghubungkan konsep fikih dengan realitas empiris. Model evaluasi semacam ini tampak pada tugas analisis fenomena pinjaman online dari perspektif fikih muamalah, kajian etika digital berbasis akhlakul karimah, atau proyek kampanye literasi anti hoaks yang dikaitkan dengan prinsip tabayyun dalam Al-Qur'an. Transformasi tersebut sekaligus memperlihatkan bahwa evaluasi PAI tidak cukup lagi disusun melalui pola soal satu jawaban tunggal. Instrumen yang terlalu tekstual justru berisiko menjauhkan pembelajaran agama dari dinamika kehidupan sosial yang terus berubah. Marzoan et al. (2025) bahkan mengingatkan bahwa dominasi evaluasi hafalan dalam pendidikan Islam berpotensi melahirkan apa yang mereka sebut sebagai *textual religiosity*, yakni keberagamaan yang kuat secara verbal tetapi lemah dalam kemampuan membaca realitas.

Perombakan besar juga mulai diarahkan pada evaluasi afektif. Sejumlah penelitian menilai bahwa lembar observasi singkat milik guru tidak lagi memadai untuk membaca perkembangan moral peserta didik secara autentik Hanafi et al. (2025), Fadholi et al. (2022). Sebagai alternatif, muncul rekomendasi penggunaan peer assessment, jurnal refleksi diri, dan observasi berkelanjutan yang memungkinkan peserta didik merefleksikan perkembangan karakter mereka secara lebih jujur dan gradual. Pendekatan tersebut mengubah posisi siswa dalam proses evaluasi. Mereka tidak lagi sekadar menjadi objek yang diamati kesalahannya dari luar, melainkan didorong menjadi subjek reflektif yang mampu mengevaluasi perkembangan spiritual dan perilaku sosialnya sendiri.

Pergeseran ini penting karena problem terbesar evaluasi afektif justru terletak pada budaya kepura-puraan moral demi kebutuhan nilai rapor.

Pada evaluasi psikomotorik, literatur terbaru mulai menekankan penggunaan rubrik analitik yang lebih rinci dan berbasis performa nyata. Kajian Marzoan et al. (2025) menunjukkan bahwa penilaian praktik ibadah yang terlalu global menyebabkan rendahnya reliabilitas skor antarguru. Karena itu, penyusunan indikator performa secara spesifik menjadi kebutuhan penting agar asesmen praktik salat, membaca Al-Qur'an, atau praktik ibadah lainnya tidak hanya bergantung pada impresi umum guru. Digitalisasi evaluasi juga mulai banyak dibahas dalam penelitian kontemporer. Namun sejumlah peneliti memberikan catatan kritis bahwa penggunaan teknologi tidak otomatis memperbaiki kualitas asesmen PAI (Arifin, 2021; Iqbal et al., 2024). Google Form, platform daring, atau aplikasi evaluasi digital justru dapat memperparah budaya LOTS apabila hanya digunakan untuk memindahkan soal hafalan dari lembar kertas ke layar gawai tanpa perubahan struktur stimulus dan pola berpikir.

Karena itu, inti rekonstruksi evaluasi PAI sesungguhnya bukan terletak pada pergantian media, melainkan perubahan paradigma pedagogi. Teknologi hanya menjadi alat bantu. Yang jauh lebih menentukan adalah keberanian mengubah orientasi pembelajaran agama: dari budaya transmisi hafalan menuju proses pembentukan kesadaran etis, refleksi kritis, dan kemampuan membaca realitas sosial melalui perspektif nilai-nilai Islam secara substantif.

4. Pembahasan

4.1. Analisis Problematika Ranah Kognitif

Evaluasi ranah kognitif dalam pembelajaran Pendidikan Agama Islam (PAI) semestinya tidak direduksi hanya sebagai aktivitas menguji kemampuan menghafal materi keagamaan. Dalam kerangka pendidikan Islam, dimensi kognitif justru diarahkan untuk membangun kesadaran intelektual religius agar peserta didik mampu memahami, merefleksikan, serta mengontekstualisasikan nilai-nilai Islam dalam realitas sosial yang terus berubah. Persoalannya, orientasi tersebut belum sepenuhnya tercermin dalam praktik evaluasi di sekolah. Sejumlah penelitian memperlihatkan bahwa instrumen evaluasi PAI masih berkatut pada hafalan konsep normatif dan penguasaan teks baku (Hanafi et al., 2025; Fadholi et al., 2022). Guru lebih sering menggunakan soal pilihan ganda dengan kata kerja operasional seperti menyebutkan, menghafal, atau menjelaskan kembali. Pola ini membuat evaluasi berhenti pada level remembering dan understanding dalam Taksonomi Bloom yang direvisi oleh Anderson dan Krathwohl. Ruang bagi kemampuan analisis, refleksi moral, atau pemecahan masalah keagamaan justru menjadi sangat sempit.

Lembar evaluasi semesteran yang dikaji dalam beberapa penelitian menunjukkan pola pertanyaan yang hampir seragam. Misalnya Jelaskan pengertian zakat menurut bahasa dan istilah atau Sebutkan rukun wudhu. Instrumen seperti ini memang efektif mengukur daya ingat jangka pendek, tetapi gagal membaca kemampuan peserta didik ketika berhadapan dengan problem sosial keagamaan yang lebih kompleks. Siswa mungkin hafal definisi zakat, tetapi belum tentu mampu menjelaskan bagaimana konsep zakat produktif dapat digunakan untuk mengurangi ketimpangan ekonomi di tengah maraknya pinjaman daring dan budaya konsumtif digital. Masalahnya tidak berhenti pada kelemahan teknis penyusunan soal. Ada faktor pedagogis dan bahkan ideologis yang ikut mempertahankan budaya evaluasi hafalan. Dalam sebagian kultur pembelajaran PAI, penguasaan teks agama secara literal masih dipandang sebagai simbol utama keberhasilan belajar agama. Hafalan dalil dianggap identik dengan kedalaman religiusitas. Cara pandang ini tanpa sadar menjadikan dialog kritis atau analisis kontekstual sebagai sesuatu yang kurang aman bagi stabilitas pemahaman agama siswa. Akibatnya, evaluasi berbasis studi kasus sering dihindari karena dikhawatirkan membuka ruang relativisme interpretasi.

Padahal, tuntutan Kurikulum Merdeka justru bergerak ke arah sebaliknya. Peserta didik didorong untuk mampu berpikir reflektif, memecahkan persoalan nyata, dan menghubungkan pengetahuan dengan kehidupan sosial. Kontradiksi inilah yang membuat implementasi evaluasi autentik dalam PAI berjalan lambat. Guru berada di tengah tarik menarik antara tuntutan pedagogi modern dengan kultur pembelajaran agama yang masih sangat tekstual. Kondisi tersebut melahirkan implikasi yang tidak sederhana. Pembelajaran agama berisiko menghasilkan *textual religiosity* keberagamaan yang kuat secara hafalan, tetapi lemah dalam kemampuan membaca realitas sosial. Siswa dapat memperoleh nilai tinggi pada ujian PAI, namun belum tentu memiliki sensitivitas moral ketika menghadapi persoalan intoleransi, etika digital, korupsi, atau ketidakadilan sosial. Persoalan lain muncul pada implementasi soal berbasis *Higher Order Thinking Skills* (HOTS). Banyak guru menganggap soal sulit otomatis termasuk kategori HOTS. Padahal keduanya tidak selalu identik. Iqbal et al. (2024) menemukan bahwa sebagian instrumen yang diklaim berbasis HOTS ternyata hanya mempersulit redaksi soal tanpa mengubah level berpikir peserta didik. Artinya, siswa tetap hanya diminta mengingat informasi, hanya saja dengan bahasa yang lebih rumit. Fenomena ini memperlihatkan adanya krisis literasi evaluasi di kalangan guru PAI. Sebagian guru belum memperoleh pelatihan yang cukup mengenai desain asesmen berbasis refleksi dan pemecahan masalah. Akibatnya, evaluasi lebih diarahkan pada instrumen yang cepat dibuat dan mudah dikoreksi dibanding asesmen kontekstual yang menuntut analisis lebih mendalam. Situasi tersebut menjadi semakin problematik ketika sekolah tetap menjadikan nilai numerik sebagai ukuran utama keberhasilan belajar agama. Pembelajaran akhirnya terjebak pada logika administratif: yang penting nilai tuntas dan rapor selesai. Dimensi reflektif agama perlahan kehilangan ruangnya. Perbandingan karakteristik instrumen evaluasi konvensional dan instrumen rekonstruktif berbasis HOTS dapat dilihat pada tabel berikut.

Tabel 1 Perbandingan karakteristik instrumen evaluasi konvensional dan instrumen rekonstruktif berbasis HOTS

No.	Dimensi Evaluasi	Instrumen Konvensional (LOTS)	Kelemahan Pedagogis	Instrumen Rekonstruktif (HOTS)
1	Zakat	Jelaskan pengertian zakat!	Hanya mengukur hafalan definisi normatif.	Analisis bagaimana zakat produktif dapat mengurangi kesenjangan ekonomi masyarakat digital.
2	Akhlak	Sebutkan macam-macam akhlak terpuji!	Tidak mengukur refleksi moral peserta didik.	Bagaimana prinsip akhlak Islam dapat diterapkan dalam etika bermedia sosial?
3	Fikih Muamalah	Apa hukum jual beli?	Bersifat tekstual dan tidak kontekstual.	Analisis praktik transaksi daring berdasarkan prinsip fikih muamalah.

Tabel tersebut memperlihatkan perbedaan mendasar antara evaluasi berbasis hafalan dan evaluasi berbasis refleksi. Instrumen HOTS menuntut peserta didik menghubungkan konsep Islam dengan problem kehidupan nyata. Akan tetapi, rekonstruksi evaluasi seperti ini membutuhkan perubahan paradigma pembelajaran secara menyeluruh, bukan sekadar mengganti format soal.

4.2. Analisis Problematika Ranah Afektif

Ranah afektif merupakan wilayah evaluasi yang paling rumit dalam pembelajaran PAI. Jika dimensi kognitif masih dapat diukur melalui tes tertulis, maka penilaian afektif menyentuh wilayah moralitas, sikap religius, kesadaran spiritual, dan internalisasi nilai yang sifatnya jauh lebih abstrak. Di titik inilah problem objektivitas mulai muncul. Dokumen penelitian yang dianalisis menunjukkan bahwa penilaian afektif di banyak sekolah masih dilakukan secara administratif dan

seremonial. Guru biasanya menggunakan lembar observasi sederhana dengan kategori umum seperti baik, cukup, atau kurang tanpa indikator perilaku yang benar-benar operasional. Bahkan terdapat praktik merapel nilai sikap, yakni memberikan nilai afektif secara massal demi mempermudah administrasi rapor. Persoalan ini memunculkan dilema serius. Sekolah dituntut mengukur karakter religius peserta didik secara objektif, sementara karakter manusia sendiri tidak pernah sepenuhnya bisa direduksi menjadi angka. Moralitas tidak bekerja seperti rumus matematika.

Fakta lain yang cukup mengkhawatirkan adalah munculnya gejala *social desirability* bias. Sejumlah penelitian menemukan bahwa siswa cenderung menampilkan perilaku religius tertentu ketika berada di bawah pengawasan guru. Mereka belajar cepat membaca ekspektasi sekolah: kapan harus tampak santun, kapan harus aktif mengikuti kegiatan ibadah, dan kapan harus menunjukkan atribut kesalehan formal demi menjaga citra di hadapan guru. Praktik evaluasi afektif yang terlalu administratif berpotensi mendorong munculnya perilaku performatif dalam menampilkan simbol kesalehan di lingkungan sekolah. Kondisi ini dapat terjadi ketika penilaian lebih berorientasi pada tampilan perilaku religius yang terlihat secara formal dibanding proses internalisasi nilai moral peserta didik secara mendalam. Pertanyaannya menjadi sangat mendasar: apakah sekolah sedang membentuk kesadaran moral yang autentik, atau justru melatih siswa memainkan peran religius demi keamanan nilai rapor?

Gejala tersebut lebih rentan terjadi dalam PAI dibanding mata pelajaran umum karena evaluasi agama membawa tekanan moral dan simbolik yang jauh lebih kuat. Dalam konteks pendidikan Islam, siswa tidak hanya dinilai berdasarkan kemampuan akademik, tetapi juga berdasarkan citra kesalehan. Akibatnya, atribut religius sering kali memperoleh bobot sosial yang lebih besar dibanding kualitas moral yang sesungguhnya tidak tampak secara visual. Bias penilaian juga muncul melalui halo effect. Guru cenderung memberikan kesan positif kepada siswa yang aktif mengikuti kegiatan keagamaan sekolah, berpakaian religius, atau dikenal santun secara simbolik. Sebaliknya, siswa yang pendiam atau kurang ekspresif sering memperoleh penilaian afektif yang lebih rendah meskipun sebenarnya memiliki perkembangan karakter yang baik.

Masalah ini menunjukkan bahwa evaluasi afektif sangat mudah dipengaruhi impresi personal. Guru tidak sepenuhnya menilai perilaku, melainkan juga citra sosial yang melekat pada peserta didik. Di sinilah objektivitas evaluasi mulai goyah. Kerumitan lain muncul ketika sekolah memaksa aspek spiritualitas untuk dikonversi menjadi skor numerik. Guru diminta memberikan angka terhadap indikator seperti kejujuran, tanggung jawab, kedisiplinan, bahkan religiusitas. Padahal indikator tersebut memiliki batas interpretasi yang sangat cair. Apa parameter konkret siswa disebut religius? Apakah rajin salat berjamaah cukup? Bagaimana dengan kejujuran yang tidak terlihat di ruang kelas?

Pertanyaan-pertanyaan tersebut memperlihatkan bahwa evaluasi afektif sesungguhnya berada di persimpangan antara tuntutan administrasi pendidikan modern dan hakikat moralitas manusia yang sangat personal. Sekolah membutuhkan angka untuk kepentingan birokrasi akademik, sementara perkembangan karakter berlangsung melalui proses sosial dan spiritual yang jauh lebih kompleks. Karena itu, evaluasi afektif tidak bisa dilakukan melalui observasi sesaat. Penilaian karakter membutuhkan pendekatan longitudinal, reflektif, dan kolaboratif. Guru, keluarga, dan lingkungan sosial harus dilibatkan agar perkembangan moral peserta didik dapat dipahami secara lebih utuh.

4.3. Analisis Problematika Ranah Psikomotorik

Evaluasi psikomotorik dalam PAI berkaitan dengan praktik ibadah dan keterampilan keagamaan, mulai dari wudhu, salat, membaca Al-Qur'an, khutbah, hingga praktik penyelenggaraan jenazah. Dalam paradigma pendidikan Islam, dimensi ini sangat penting karena keberagamaan tidak cukup dipahami sebagai pengetahuan teoritis, tetapi juga praktik nyata dalam kehidupan sehari-hari. Meski demikian, sejumlah penelitian memperlihatkan bahwa evaluasi praktik ibadah masih menghadapi problem validitas dan reliabilitas yang serius (Marzoan et al., 2025). Guru sering memberikan nilai

praktik secara global tanpa menggunakan rubrik analitik yang jelas. Penilaian akhirnya lebih bergantung pada impresi umum ketika siswa tampil di depan kelas.

Dalam praktik penilaian salat, misalnya, guru kerap hanya menggunakan kategori baik, cukup, atau kurang tanpa indikator rinci mengenai ketepatan gerakan, kelancaran bacaan, urutan pelaksanaan, ataupun kesesuaian tajwid. Akibatnya, standar penilaian menjadi sangat subjektif dan berbeda antara satu guru dengan guru lain. Masalah menjadi lebih pelik ketika evaluasi menyentuh aspek yang sangat abstrak seperti kekhusyukan atau penghayatan ibadah. Bagaimana sebenarnya cara mengukur kekhusyukan? Tidak ada parameter yang benar-benar pasti. Guru dapat memiliki interpretasi berbeda mengenai siswa yang dianggap fokus atau sungguh-sungguh dalam ibadah.

Konfirmasi ini mempertegas batas evaluasi psikomotorik PAI: spiritualitas tidak bisa sepenuhnya diterjemahkan menjadi angka kuantitatif. Pengalaman religius manusia memiliki dimensi personal yang sering kali melampaui kemampuan instrumen evaluasi formal sekolah. Keterbatasan struktural ikut memperparah keadaan. Guru sering harus menilai praktik ibadah puluhan siswa dalam waktu yang sangat terbatas. Situasi ini membuat observasi mendalam hampir mustahil dilakukan. Akhirnya, penilaian berubah menjadi aktivitas administratif untuk mengejar penyelesaian nilai. Di banyak sekolah, fasilitas praktik juga belum memadai. Ruang ibadah sempit, alat praktik terbatas, dan waktu pembelajaran terlalu singkat. Ketika evaluasi autentik menuntut observasi individual dan berkelanjutan, kapasitas sistem pendidikan justru belum mampu mendukung kebutuhan tersebut secara optimal. Kontradiksi ini memperlihatkan kesenjangan antara idealitas Kurikulum Merdeka dan realitas implementasi di lapangan. Kurikulum mendorong asesmen berbasis praktik dan proyek, tetapi dukungan teknis maupun kapasitas evaluatif guru belum sepenuhnya siap.

4.4. Analisis Faktor Penyebab Internal dan Eksternal

Problematika evaluasi PAI tidak dapat dipahami semata sebagai kelemahan teknis individual guru. Persoalan ini terbentuk melalui kombinasi faktor internal dan eksternal yang saling bertaut. Faktor internal paling dominan berkaitan dengan rendahnya literasi evaluasi guru. Banyak guru PAI belum memperoleh pelatihan yang memadai mengenai desain asesmen autentik, penyusunan rubrik HOTS, maupun performance assessment. Akibatnya, guru lebih nyaman menggunakan instrumen evaluasi konvensional yang dianggap praktis dan aman secara administratif.

Namun akar masalahnya tidak sesederhana kekurangan pelatihan. Ada dimensi epistemologis religius yang ikut memengaruhi cara guru memandang evaluasi PAI. Dalam sebagian kultur pendidikan agama, hafalan dalil dan penguasaan teks literal masih dianggap sebagai bentuk utama keberhasilan belajar agama. Cara pandang ini membuat evaluasi reflektif dan kontekstual sering dipandang kurang penting, bahkan kadang dicurigai dapat mengaburkan kemurnian ajaran. Dilema tersebut menjelaskan mengapa budaya evaluasi hafalan tetap bertahan kuat di sekolah. Guru tidak hanya menghadapi keterbatasan teknis, tetapi juga tekanan ideologis mengenai bagaimana agama seharusnya diajarkan dan diukur.

Faktor eksternal muncul dari sistem pendidikan yang sangat administratif. Guru PAI tidak hanya mengajar, tetapi juga dibebani penyusunan modul ajar, asesmen diagnostik, pelaporan capaian pembelajaran, hingga administrasi kurikulum yang terus berubah. Energi guru banyak tersita untuk penyelesaian dokumen dibanding refleksi pedagogis. Rasio siswa yang besar ikut memperburuk kualitas evaluasi. Penilaian autentik membutuhkan observasi individual dan berkelanjutan, sedangkan guru sering harus menangani puluhan siswa dalam satu kelas. Situasi seperti ini membuat evaluasi afektif dan psikomotorik lebih mudah berubah menjadi formalitas administratif. Budaya sekolah yang terlalu menekankan capaian angka juga mempersempit ruang evaluasi reflektif. Keberhasilan pembelajaran agama sering diukur berdasarkan nilai rapor dan tingkat ketuntasan belajar. Akibatnya, evaluasi diarahkan untuk memenuhi target administratif, bukan membangun kesadaran moral dan spiritual peserta didik secara mendalam.

4.5. Rekonstruksi Evaluasi PAI Berbasis Kurikulum Merdeka

Literatur kontemporer memperlihatkan adanya dorongan kuat untuk merekonstruksi evaluasi PAI agar lebih relevan dengan paradigma Kurikulum Merdeka. Rekonstruksi ini tidak cukup dilakukan dengan mengganti format soal, tetapi harus menyentuh perubahan paradigma evaluasi secara menyeluruh. Di wilayah kognitif, sejumlah penelitian menawarkan pergeseran dari evaluasi definisional menuju asesmen berbasis studi kasus dan proyek sosial-keagamaan. Peserta didik tidak lagi hanya diminta menghafal konsep Islam, melainkan menguji relevansi nilai tersebut dalam problem kehidupan nyata. Misalnya, siswa diminta menganalisis fenomena pinjaman daring menggunakan perspektif fikih muamalah atau merancang kampanye etika digital berbasis akhlakul karimah.

Pergeseran ini penting karena pembelajaran agama tidak lagi cukup menghasilkan siswa yang hafal teks, tetapi juga mampu membaca realitas sosial secara kritis dan etis. Meski demikian, digitalisasi evaluasi tidak otomatis menyelesaikan persoalan pedagogis. Penggunaan Google Form atau platform digital lainnya hanya akan memindahkan budaya hafalan dari lembar kertas ke layar gawai apabila struktur stimulus soal tetap berbasis LOTS. Karena itu, rekonstruksi evaluasi harus dimulai dari perubahan logika asesmen, bukan sekadar perubahan medianya. Di ranah afektif, banyak penelitian mulai merekomendasikan penggunaan peer assessment, jurnal refleksi, dan evaluasi longitudinal (Imamuddin & Khuriyah, 2023). Melalui pendekatan ini, siswa tidak semata menjadi objek penilaian guru, tetapi diajak melakukan refleksi terhadap perkembangan moral dan spiritual mereka sendiri.

Pendekatan semacam ini jauh lebih relevan dibanding observasi sesaat yang administratif. Penilaian karakter tidak cukup dilakukan melalui kesan umum guru di ruang kelas. Ia membutuhkan proses refleksi yang terus-menerus. Sementara itu, rekonstruksi evaluasi psikomotorik diarahkan pada penguatan *performance assessment* berbasis rubrik analitik digital. Guru didorong menggunakan indikator yang lebih spesifik, terukur, dan operasional agar reliabilitas penilaian praktik ibadah dapat meningkat. Beberapa penelitian juga menawarkan penggunaan penilaian praktik berbasis performa autentik sebagai alternatif evaluasi berkelanjutan (Syamsiah & Prastowo, 2022). Strategi ini memungkinkan guru melakukan observasi lebih mendalam tanpa sepenuhnya bergantung pada keterbatasan waktu tatap muka di kelas. Pada akhirnya, rekonstruksi evaluasi PAI membutuhkan perubahan cara pandang terhadap makna penilaian itu sendiri. Evaluasi tidak boleh berhenti sebagai mekanisme pemberian angka rapor, tetapi harus diposisikan sebagai instrumen pembinaan kesadaran intelektual, moral, dan spiritual peserta didik secara lebih autentik.

5. Kesimpulan dan saran

Berdasarkan sintesis literatur kontemporer yang telah dilakukan, problematika evaluasi Pendidikan Agama Islam (PAI) tidak hanya berkaitan dengan aspek teknis penyusunan instrumen, tetapi juga mengindikasikan adanya problem paradigmatis dalam praktik evaluasi, khususnya dalam upaya mengukur moralitas dan religiusitas peserta didik melalui instrumen formal sekolah. Evaluasi PAI saat ini masih terjebak dalam nalar transmisi pengetahuan normatif dan formalitas administratif, yang pada akhirnya mereduksi hakikat pendidikan agama itu sendiri. Temuan penelitian menunjukkan bahwa problem evaluasi PAI muncul pada ketiga ranah penilaian, baik kognitif, afektif, maupun psikomotorik, dengan karakteristik dan tantangan yang berbeda pada masing-masing ranah. Pada ranah kognitif, asesmen masih didominasi budaya hafalan tekstual (LOTS) yang gagap menghadapi realitas sosial. Pada ranah afektif, penilaian terjebak dalam subjektivitas guru (*halo effect*) dan rentan memproduksi kesalehan performatif (*social desirability bias*) demi tuntutan nilai rapor. Sementara pada ranah psikomotorik, penilaian praktik ibadah kehilangan validitas dan reliabilitasnya akibat absennya rubrik analitik yang operasional, diperparah oleh hambatan struktural seperti keterbatasan waktu dan besarnya rasio kelas. Akar dari seluruh problem ini adalah benturan struktural antara tuntutan birokrasi pendidikan yang berbasis angka kuantitatif dengan hakikat spiritualitas Islam yang bersifat abstrak dan longitudinal. Rekonstruksi evaluasi PAI dalam konteks Kurikulum Merdeka menuntut transformasi paradigma yang radikal, bukan sekadar

digitalisasi instrumen. Solusi strategis yang ditawarkan oleh literatur mutakhir meliputi penggunaan project-based assessment untuk nalar kritis, penilaian sejawat (*peer assessment*) dan jurnal refleksi untuk mengikis kepura-puraan moral, serta rubrik analitik berbasis performa nyata. Teknologi dan digitalisasi hanya akan menjadi alat bantu yang bermakna jika diletakkan di atas fondasi paradigma pedagogi yang berorientasi pada kesadaran etis substantif peserta didik.

Saran

Berdasarkan kesimpulan di atas, beberapa langkah strategis direkomendasikan sebagai berikut:

1. Bagi Pendidik (Guru PAI): Diharapkan meningkatkan literasi evaluasi melalui pelatihan kedisiplinan asesmen autentik yang menyentuh dimensi filosofis, sehingga mampu menyusun rubrik analitik dan stimulus berbasis kasus kontemporer.
2. Bagi Pengelola Lembaga Pendidikan (Sekolah): Perlu melakukan dekonstruksi terhadap budaya sekolah yang mendewakan capaian angka formal, serta memberikan kelonggaran administratif agar guru memiliki ruang refleksi pedagogis dalam mengamati perkembangan karakter siswa secara longitudinal.
3. Bagi Pengembang Kebijakan: Perlu merumuskan regulasi evaluasi yang lebih realistis dan adaptif terhadap kendala struktural di lapangan, khususnya terkait rasio ideal siswa dalam kelas dan ketersediaan fasilitas praktik keagamaan.
4. Bagi Peneliti Selanjutnya: Mengingat penelitian ini berbasis studi kepustakaan, disarankan bagi peneliti berikutnya untuk melakukan riset empiris di lapangan guna memetakan dinamika, resistensi, dan efektivitas implementasi rekonstruksi evaluasi ini secara kontekstual di berbagai jenjang sekolah.

Referensi

- Arifin, S. (2021). Kebijakan merdeka belajar dan implikasinya terhadap pengembangan desain evaluasi pembelajaran PAI. *Jurnal Manajemen dan Pendidikan Islam*, 7(1).
<https://journal.um-surabaya.ac.id/Tadarus/article/view/16847>
- Dona, R., Puspa, P., Rahmayanti, R., & Arifmiboy, A. (2024). Evaluasi pembelajaran PAI: Tinjauan ranah kognitif, afektif dan psikomotorik di SDIT Cahaya Hati Bukittinggi. *Jurnal Pendidikan Tambusai*, 8(3), 43041–43052.
<https://jptam.org/index.php/jptam/article/view/23385>
- Fadholi, M. I., Sukiman, S., & Fadilah, L. (2022). Implementasi pendekatan saintifik dalam pengembangan high order thinking skills (HOTS) pada pembelajaran Pendidikan Agama Islam (PAI) kelas X SMA N 1 Kalasan tahun pelajaran 2021/2022. *Berkala Ilmiah Pendidikan*, 2(2). <https://doi.org/10.51214/bip.v2i2.444>
- Halimah, S. (2021). Implementasi pendekatan HOTS (higher order thinking skills) dalam pembelajaran PAI. *Evaluasi: Jurnal Manajemen Pendidikan Islam*, 5(2), 342–362.
<https://doi.org/10.32478/evaluasi.v5i2.762>
- Hanafi, S. S. A. N., Fadhila, A. S., & Bashith, A. (2025). Pendekatan penilaian autentik dalam evaluasi pembelajaran Pendidikan Agama Islam (PAI). *Educan*.
<https://educan.journal.unida.gontor.ac.id/index.php/educan/article/view/38>
- Imamuddin, M., & Khuriyah. (2023). Penilaian autentik pembelajaran PAI dengan blended learning. *JiIP - Jurnal Ilmiah Ilmu Pendidikan*, 6(3).
<https://doi.org/10.54371/jiip.v6i3.1428>
- Iqbal, M. G., Khasanah, S. U., & Luthfi, A. (2024). Strategi dan evaluasi high order thinking skill (HOTS) dalam pembelajaran PAI (Studi kasus di SMAN 114 Jakarta). *Ideguru: Jurnal Karya Ilmiah Guru*, 10(1), 23–29. <https://doi.org/10.51169/ideguru.v10i1.1428>

- Isnaini, N., & Jannah, N. (2024). Islamic religious education (PAI) learning assessment based on higher order thinking skills (HOTS) at SMK Muhammadiyah 5 Jember. *AL-WIJDÁN Journal of Islamic Education Studies*, 9(3), 334–353. <https://doi.org/10.58788/alwijdn.v9i3.4647>
- Marzoan, M., Sumiadi, R., Ramzi, M., Sukarto, & Arum F, M. (2025). Analisis tingkat penguasaan keterampilan sholat dan membaca Al-Qur'an pada siswa sekolah dasar. *Ta'dib: Jurnal Pendidikan Islam dan Isu-Isu Sosial*, 23(1), 20–32. <https://doi.org/10.37216/tadib.v23i1.2239>
- Pribadi, R. A., Hidayah, W. N., & Afriyani, N. (2022). Proses pelaksanaan penilaian autentik dalam rangka pencapaian pembelajaran berkualitas. *JS (Jurnal Sekolah)*, 6(2), 58–66. <https://doi.org/10.24114/js.v6i2.33429>
- Syamsiah, P. N., & Prastowo, A. (2022). Implementasi penilaian autentik pada ranah psikomotorik dalam pembelajaran PAI di sekolah dasar. *Journal of Research and Thought on Islamic Education (JRTIE)*, 5(2). <https://doi.org/10.24260/jrtie.v5i2.2204>