

Melampaui Asesmen: Membangun Follow-Up Evaluation Improvement Model (FEIM) untuk Peningkatan Mutu Pembelajaran Pendidikan Agama Islam

Abiyyu Al Fatih¹, Selviyana², Reva Dwi Ariendra³, Much Febri Ardiyanto⁴, Mamun Hanif⁵

UIN Kh. Abdurrahman Wahid Pekalongan, Indonesia¹⁻⁵

Email Korespondensi: abiyyu.al.fatih@mhs.uingudur.ac.id, selviyana@mhs.uingudur.ac.id, reva.dwi.ariendra@mhs.uingudur.ac.id, much.febri.ardiyanto@mhs.uingudur.ac.id, mamun.hanif63@gmail.com

Article received: 05 Januari 2026, Review process: 28 Januari 2026

Article Accepted: 22 April 2026, Article published: 28 Juli 2026

ABSTRACT

Evaluation in Islamic Religious Education (PAI) plays a critical role in shaping students' character and academic competence, yet current pedagogical practices tend to reduce it to mere cognitive-summative assessment, neglecting the follow-up mechanisms essential for continuous quality improvement. This study aims to analyze existing PAI evaluation systems, identify patterns in follow-up programs, and construct a new conceptual framework termed the Follow-Up Evaluation Improvement Model (FEIM). A qualitative literature review was conducted following PRISMA 2020 guidelines, examining 42 reputable sources (predominantly published 2021–2026) from Scopus, Web of Science, DOAJ, and Sinta databases, analyzed using Braun and Clarke's six-stage thematic analysis. The findings reveal that current follow-up mechanisms in PAI evaluation remain fragmented, dominated by ad-hoc, student-centered remedial tasks, while overlooking systemic instructional weaknesses. Through the synthesis of the CIPP Model, the Continuous Quality Improvement (CQI) cycle via PDCA, and Islamic philosophical values (muhasabah, islah, itqan), this study constructs an eight-step FEIM framework as a conceptual blueprint. The model fundamentally links evaluation data aggregation with reflective analysis (muhasabah), systematic root-cause identification, targeted intervention design, implementation of corrective actions (islah), participatory monitoring, and cyclical re-evaluation to achieve pedagogical excellence (itqan). FEIM contributes as an integrative paradigm that shifts the evaluation stance from static assessment of learning toward a dynamic engine for sustained pedagogical excellence, although its operational validity requires further empirical testing.

Keywords: FEIM, CIPP Model, Islamic Religious Education, Evaluation Follow-Up, Muhasabah-Islah-Itqan.

ABSTRAK

Evaluasi dalam Pendidikan Agama Islam (PAI) memiliki peran penting dalam membentuk karakter dan kompetensi akademik siswa, namun praktik yang berjalan saat ini cenderung mereduksinya menjadi penilaian kognitif-sumatif belaka tanpa mekanisme tindak lanjut yang memadai untuk peningkatan mutu berkelanjutan. Penelitian ini bertujuan menganalisis sistem evaluasi PAI yang berjalan, mengidentifikasi pola program tindak lanjut, serta mengonstruksi kerangka konseptual baru bernama Follow-Up Evaluation

Improvement Model (FEIM). Penelitian menggunakan metode tinjauan literatur kualitatif berpedoman pada PRISMA 2020, dengan menelaah 42 literatur bereputasi (mayoritas terbitan 2021–2026) dari basis data Scopus, Web of Science, DOAJ, dan Sinta, yang dianalisis melalui enam tahap analisis tematik Braun dan Clarke. Hasil penelitian menunjukkan bahwa mekanisme tindak lanjut evaluasi PAI saat ini bersifat terfragmentasi, didominasi tugas remedial ad-hoc yang berorientasi individual siswa, serta mengabaikan kelemahan instruksional yang bersifat sistemik. Melalui sintesis Model CIPP, siklus Continuous Quality Improvement (CQI)/PDCA, dan nilai filosofis Islam (muhasabah, islah, itqan), penelitian ini merumuskan FEIM sebagai cetak biru konseptual yang terdiri atas delapan langkah. Model ini secara fundamental menghubungkan agregasi data evaluasi dengan analisis reflektif (muhasabah), identifikasi akar masalah secara sistematis, perancangan intervensi terarah, implementasi tindakan perbaikan (islah), pemantauan partisipatif, serta evaluasi ulang secara siklus guna mencapai keunggulan pedagogis (itqan). FEIM berkontribusi menggeser paradigma evaluasi dari sekadar penilaian statis menjadi mesin dinamis bagi keunggulan pedagogis yang berkelanjutan, meskipun validitas operasionalnya masih memerlukan uji empiris lebih lanjut.

Kata Kunci: FEIM, Model CIPP, Pendidikan Agama Islam, Tindak Lanjut Evaluasi, Muhasabah-Islah-Itqan.

PENDAHULUAN

Evaluasi merupakan pilar fundamental dalam sistem pendidikan yang berfungsi sebagai kompas pengarah bagi keefektifan strategi pedagogis, relevansi kurikulum, dan ketepatan kebijakan institusional. Dalam konteks Pendidikan Agama Islam (PAI), evaluasi tidak sekadar instrumen pengukur kognisi, melainkan memikul mandat holistik untuk memvalidasi dimensi afektif (spiritual-moral) dan psikomotorik yang melekat pada pembentukan karakter peserta didik (Lubis et al., 2025). Urgensi evaluasi komprehensif dalam PAI menjadi sangat krusial mengingat mata pelajaran ini bertanggung jawab atas internalisasi worldview Islam ke dalam kepribadian siswa, bukan sekadar transfer dogma teoretis (Yuniartin et al., 2024). Melalui evaluasi yang terstruktur, pendidik dapat mendeteksi sejauh mana nilai-nilai keagamaan mengkristal dalam perilaku sehari-hari peserta didik.

Namun, praktik pendidikan kontemporer menunjukkan bahwa proses evaluasi sering kali berhenti secara prematur begitu nilai akhir didistribusikan. Evaluasi cenderung direduksi menjadi vonis administratif terhadap kapasitas siswa (assessment of learning), alih-alih difungsikan sebagai instrumen diagnostik untuk memodifikasi dan meningkatkan kualitas pengajaran secara berkelanjutan (assessment for learning) (Halimah et al., 2026). Ketiadaan mekanisme sistematis yang menghubungkan hasil penilaian dengan tindakan korektif pedagogis menyebabkan data evaluasi menjadi artefak administratif yang mati, tidak memberikan umpan balik bagi perbaikan mutu instruksional institusi.

Penelitian terdahulu yang relevan menunjukkan bahwa diskursus evaluasi PAI masih terfokus pada aspek hulu dan tengah. (Halimah et al., 2026) mengkaji pentingnya implementasi evaluasi formatif dalam pembelajaran PAI dan menemukan bahwa evaluasi formatif masih terhambat oleh keterbatasan waktu dan pemahaman guru, sehingga jarang digunakan untuk perbaikan pembelajaran

berkelanjutan. (Fauzi & Nurfuadi, 2026) meneliti evaluasi pembelajaran PAI berbasis nilai-nilai Islam dan menyimpulkan bahwa praktik evaluasi masih dominan pada ranah kognitif, sementara dimensi afektif dan psikomotorik terabaikan karena kurangnya instrumen yang terstandar. (Julhadi et al., 2026) mengkaji tantangan dan praktik penilaian autentik dalam PAI, mengungkap bahwa guru kesulitan menerapkan penilaian autentik karena beban administratif dan kurangnya pelatihan, sehingga penilaian cenderung kembali pada tes tertulis. (Yuniartin et al., 2024) mengembangkan model evaluasi PAI yang mengintegrasikan aspek kognitif, moral, dan spiritual berdasarkan Al-Qur'an dan Hadis, namun model tersebut masih bersifat konseptual dan belum dilengkapi mekanisme tindak lanjut yang terukur. Sementara itu, (Puspita et al., 2026) mengusulkan kerangka Adaptive Quality Cycle (AQC) yang mengintegrasikan siklus PDCA dengan prinsip Islam (muhasabah, tahsin, itqan) untuk perbaikan mutu berkelanjutan, tetapi kerangka ini masih bersifat umum dan belum dirancang secara spesifik untuk fase pasca-evaluasi pembelajaran.

Berdasarkan penelusuran terhadap state of the art tersebut, terdapat kesenjangan penelitian (research gap) yang mencolok: diskursus akademis mengenai fase pasca-evaluasi yakni bagaimana hasil asesmen tersebut secara metodologis diterjemahkan menjadi program tindak lanjut (follow-up) untuk perbaikan mutu berkelanjutan dalam hal ini masih sangat minim disentuh. Penelitian tentang evaluasi PAI masih terpisah-pisah antara kajian instrumen, proses penilaian, dan manajemen mutu, tanpa menyediakan kerangka utuh yang menghubungkan hasil evaluasi dengan tindakan perbaikan yang terstruktur (Ahmadi & Khozin, 2025). Belum ada model yang secara komprehensif mengintegrasikan analisis data evaluasi, identifikasi akar masalah, desain intervensi, dan pemantauan implementasi dalam satu siklus yang koheren dan aplikatif bagi guru dan pengelola PAI.

Penelitian ini bertujuan untuk menjembatani kesenjangan tersebut dengan: (1) menganalisis secara kritis sistem evaluasi pembelajaran PAI dalam literatur kontemporer; (2) mengidentifikasi bentuk-bentuk program tindak lanjut evaluasi yang eksis; dan (3) mengonstruksi model konseptual baru yang dinamakan Follow-Up Evaluation Improvement Model (FEIM). Orisinalitas dan kebaruan (novelty) dari riset ini terletak pada pengembangan kerangka FEIM yang secara eklektik menyintesis Model Evaluasi CIPP (Stufflebeam), Continuous Quality Improvement melalui Siklus PDCA (Deming, 1986), Teori Mutu Pendidikan (Sallis, 2014), serta epistemologi evaluasi Islam (muhasabah, islah, itqan) yang telah diangkat oleh (Puspita et al., 2026) dan (Fithri & Ali, 2026). Model ini diproyeksikan memberikan kontribusi teoretis strategis dalam menggeser paradigma evaluasi PAI dari orientasi pemenuhan birokrasi nilai menuju mesin penggerak peningkatan mutu yang sirkular dan transformatif.

METODE

Penelitian ini menggunakan desain kajian kepustakaan kualitatif (qualitative literature review) untuk mengidentifikasi, mengevaluasi, dan menyintesis publikasi akademis terkini mengenai evaluasi pembelajaran Pendidikan Agama Islam (PAI)

dan program tindak lanjutnya. Pendekatan ini dipilih karena memungkinkan peneliti untuk menganalisis secara mendalam fenomena yang kompleks dengan mensintesis temuan-temuan dari berbagai sumber terpercaya (Puspita et al., 2026), (Fithri & Ali, 2026), dan (Ahmadi & Khozin, 2025) Guna menjamin akuntabilitas dan replikabilitas metodologis, pengumpulan serta penyeleksian artikel dipandu secara ketat oleh protokol PRISMA 2020 (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) (Page et al., 2021). (A) Lokasi Penelitian dan Kehadiran Peneliti. Penelitian ini tidak memiliki lokasi fisik karena bersifat kepustakaan, namun objek kajiannya adalah literatur ilmiah yang terindeks dalam basis data bereputasi yang diakses secara daring melalui jaringan internet. Kehadiran peneliti bersifat virtual dan berperan sebagai instrumen kunci (key instrument) yang melakukan penelusuran, seleksi, analisis, dan sintesis literatur secara sistematis. Peneliti terlibat penuh dalam seluruh tahapan penelitian, mulai dari perumusan masalah, penelusuran literatur, seleksi artikel, hingga interpretasi dan penarikan kesimpulan. (B) Populasi dan Sampel (Sasaran Penelitian). Populasi dalam penelitian ini adalah seluruh artikel ilmiah yang membahas evaluasi pembelajaran, asesmen, tindak lanjut, dan perbaikan mutu dalam konteks PAI yang terbit dalam rentang waktu lima tahun terakhir (2021–2026). Sampel penelitian ditentukan melalui prosedur purposive sampling dengan kriteria inklusi yang ketat, menghasilkan 42 artikel ilmiah yang paling relevan dan berkualitas tinggi (Ahmadi & Khozin, 2025); (Halimah et al., 2026); (Julhadi et al., 2026); (Yuniartin et al., 2024). (C) Subjek Penelitian dan Informan. Subjek penelitian dalam kajian kepustakaan ini adalah teks-teks akademis dari artikel-artikel terpilih yang memuat data dan informasi terkait praktik evaluasi dan tindak lanjut PAI. Adapun informan penelitian adalah para penulis artikel yang menjadi sumber data sekunder, yang gagasan dan temuannya dianalisis secara kritis melalui karya tulis mereka. Karena penelitian ini bersifat kepustakaan, tidak ada informan langsung yang diwawancarai, melainkan seluruh data diperoleh dari dokumen publikasi ilmiah yang terindeks (Lubis et al., 2025). (D) Teknik Pengumpulan Data. Teknik pengumpulan data dilakukan melalui penelusuran literatur komprehensif pada kuartal pertama tahun 2026 secara daring dengan mengakses empat basis data bereputasi: Scopus, Web of Science (WoS), Directory of Open Access Journals (DOAJ), dan Sinta (Science and Technology Index) (Puspita et al., 2026); (Fithri & Ali, 2026). Kata kunci disesuaikan menggunakan Boolean Operators: ("Pendidikan Agama Islam" OR "Islamic Religious Education") AND ("evaluasi pembelajaran" OR "learning evaluation" OR "assessment") AND ("tindak lanjut" OR "follow-up" OR "PDCA" OR "quality improvement"). Kriteria inklusi meliputi: (1) Artikel orisinal atau review yang terbit di jurnal peer-reviewed (berkualifikasi Q1-Q4 Scopus/WoS atau Sinta 1-3); (2) Spesifik membahas evaluasi, asesmen, tindak lanjut, dan perbaikan mutu dalam konteks PAI di level pendidikan dasar hingga menengah; (3) Dipublikasikan dalam rentang waktu lima tahun terakhir (2021–2026) guna menjaga kemutakhiran data; dan (4) Berbahasa Inggris atau Indonesia. Adapun kriteria eksklusi adalah: (1) Artikel yang hanya melakukan uji validitas/reliabilitas instrumen statistik murni tanpa menyinggung follow-up sama sekali; (2) Publikasi non-jurnal seperti opini,

book chapter, atau prosiding yang tidak peer-reviewed; dan (3) Kajian evaluasi di luar rumpun ilmu agama Islam (Lubis et al., 2025). Proses filterisasi mengikuti empat alur sekuensial sesuai protokol PRISMA 2020. Pertama (Identifikasi), pencarian awal menghasilkan 285 dokumen. Kedua (Skrining), setelah duplikasi dihapus dengan bantuan reference manager Mendeley, tersisa 192 dokumen. Skrining berbasis kesesuaian judul dan abstrak mengeliminasi 112 dokumen. Ketiga (Kelayakan), 80 naskah lengkap dinilai (full-text review), mengakibatkan eksklusi 38 artikel yang metodologinya lemah atau bias konteks di luar PAI. Keempat (Inklusi), didapatkan 42 artikel ilmiah paling relevan yang dijadikan data primer untuk disintesis dalam studi ini (Ahmadi & Khozin, 2025); (Fithri & Ali, 2026); (Puspita et al., 2026). (E) Teknik Analisis Data. Seluruh literatur yang terinklusi dianalisis menggunakan pendekatan Thematic Analysis yang dirumuskan oleh (Braun & Clarke, 2024). Tahapannya meliputi: (1) Familiarization yakni membaca dan memahami teks artikel secara mendalam untuk memperoleh gambaran menyeluruh tentang temuan-temuan terkait evaluasi dan tindak lanjut PAI (Halimah et al., 2026) dan (Julhadi et al., 2026); (2) memberi label konseptual pada data yang menyinggung fungsi evaluasi dan tindak lanjut, seperti "remedial ad-hoc", "asesmen afektif terabaikan", "siklus PDCA terputus", dan "muhasabah" (Fauzi & Nurfuadi, 2026) dan (Yuniartin et al., 2024); (3) Generating Themes yakni mengelompokkan kode-kode ke dalam pola-pola bermakna yang merepresentasikan temuan utama, seperti dominansi evaluasi kognitif, fragmentasi asesmen afektif, reduksi tindak lanjut, dan terputusnya rantai pemeliharaan mutu pedagogis (Ahmadi & Khozin, 2025) dan (Aulia & Yuliyanti, 2024); (4) Reviewing Themes – mengonfrontasi draf tema dengan teori CIPP dan CQI untuk memastikan koherensi teoretis (Puspita et al., 2026) dan (Fithri & Ali, 2026); (5) Defining and Naming Themes – menetapkan definisi final dari setiap tema yang ditemukan; dan (6) Writing Up – menarasikan temuan menjadi konstruksi model FEIM yang utuh dan sistematis (Lubis et al., 2025).

Seluruh proses analisis dilakukan dengan tetap merujuk pada kerangka teoretis Model Evaluasi CIPP (Stufflebeam), Continuous Quality Improvement melalui siklus PDCA (Deming), dan nilai filosofis Islam (muhasabah, islah, itqan) sebagaimana diintegrasikan oleh (Puspita et al., 2026) dan (Fithri & Ali, 2026). Validitas temuan dijaga melalui triangulasi sumber dengan membandingkan hasil analisis antar-literatur yang memiliki perspektif berbeda, serta diskusi antarpeneiti untuk menghindari bias interpretasi.

HASIL DAN PEMBAHASAN

Dominansi Evaluasi Kognitif dan Fragmentasi Asesmen Afektif dalam Pembelajaran PAI

Temuan penelitian ini mengindikasikan bahwa praktik evaluasi Pendidikan Agama Islam (PAI) masih terpusat pada penilaian kognitif tertulis. Sementara itu, asesmen afektif dan psikomotorik yang menjadi esensi pendidikan agama sering dinilai secara artifisial tanpa instrumen yang terstandar. Fauzi dan Nurfuadi (2026) menegaskan bahwa dimensi afektif dan psikomotorik terabaikan karena keterbatasan instrumen yang valid dan terukur. Akibatnya, proses evaluasi

kehilangan fungsi diagnostiknya dan hanya digunakan untuk memenuhi tuntutan administratif kelulusan, bukan untuk memetakan internalisasi nilai-nilai moral dan spiritual peserta didik (Yuniartin et al., 2024).

Kesulitan guru dalam mengimplementasikan penilaian autentik turut memperburuk kondisi ini. (Julhadi et al., 2026) mengungkap bahwa beban administratif dan minimnya pelatihan menyebabkan guru cenderung kembali pada tes tertulis yang lebih praktis. (Halimah et al., 2026) juga menemukan bahwa evaluasi formatif untuk memantau perkembangan afektif siswa terhambat oleh keterbatasan waktu dan pemahaman guru tentang pentingnya penilaian holistik. Kondisi ini menyebabkan pengukuran sikap spiritual, kejujuran, dan tanggung jawab sosial hanya dilakukan secara subjektif tanpa rubrik yang jelas.

Dalam perspektif Model CIPP, dominansi evaluasi kognitif mencerminkan kegagalan pada fase Product, di mana data capaian belajar tidak merepresentasikan seluruh dimensi pendidikan yang seharusnya dievaluasi. Padahal (Yuniartin et al., 2024) menekankan bahwa evaluasi PAI idealnya mengintegrasikan aspek kognitif, moral, dan spiritual berdasarkan Al-Qur'an dan Hadis. Fragmentasi ini bertentangan dengan prinsip muhasabah dalam evaluasi Islam yang menuntut introspeksi komprehensif atas seluruh dimensi kinerja (Majid & Abdullah, 2022; Zarkasyi, 2020). Dengan mengabaikan asesmen afektif yang terstruktur, institusi kehilangan kemampuan memetakan kristalisasi nilai-nilai keagamaan dalam perilaku siswa.

Implikasi lebih jauh dari dominansi ini adalah terabaikannya pengembangan karakter yang menjadi mandat utama PAI (Aulia & Yuliyanti, 2024); (Lubis et al., 2025) Dalam manajemen mutu, Sallis (2014) menyatakan bahwa evaluasi parsial menghasilkan gambaran yang menyesatkan tentang kualitas pendidikan. Ahmadi dan Khozin (2025) menambahkan bahwa penjaminan mutu di lembaga pendidikan Islam masih bersifat prosedural dan belum menyentuh evaluasi afektif secara mendalam. Hal ini mencerminkan kelemahan sistemik di mana penilaian dipandang sebagai kewajiban birokrasi, bukan sebagai instrumen peningkatan mutu berkelanjutan (Puspita et al., 2026). Kegagalan mengukur dimensi afektif secara akurat tidak hanya menghambat pemetaan karakter siswa tetapi juga memutus umpan balik bagi perbaikan pedagogis, sehingga menjadi pijakan penting bagi pengembangan model FEIM yang menempatkan analisis reflektif (muhasabah) sebagai langkah awal identifikasi kelemahan sistemik.

Reduksi Tindak Lanjut Menjadi Remedial Ad-Hoc yang Berpusat pada Siswa

Temuan kedua penelitian ini mengungkap bahwa program tindak lanjut (follow-up) evaluasi PAI di lapangan sangat terbatas pada kegiatan remedial dan pengayaan yang bersifat dangkal dan instan. Remedial diterapkan dengan paradigma yang berpusat pada kegagalan siswa (student-centric), di mana peserta didik diminta mengerjakan ulang soal yang sama atau diberi tugas tambahan tanpa adanya perubahan strategi mengajar dari pendidik. (Halimah et al., 2026) menegaskan bahwa evaluasi formatif yang seharusnya menjadi dasar perbaikan pembelajaran sering terhambat dan tidak diikuti dengan tindakan korektif yang

terstruktur. Mekanisme ini mengasumsikan bahwa kesalahan mutlak berada pada siswa, sementara kelemahan pada metodologi guru atau desain kurikulum diabaikan sama sekali.

Kelemahan paradigma remedial ad-hoc ini semakin diperkuat oleh temuan (Julhadi et al., 2026). yang menyatakan bahwa guru PAI kesulitan menerapkan penilaian autentik dan cenderung mengulangi pola remedial yang sama tanpa evaluasi terhadap efektivitasnya. (Fauzi & Nurfuadi, 2026) juga menyoroti bahwa tindak lanjut evaluasi sering direduksi menjadi kegiatan administratif tanpa analisis mendalam terhadap akar permasalahan pembelajaran. Akibatnya, program remedial tidak mampu mengatasi kesulitan belajar siswa secara bermakna karena tidak didasarkan pada diagnosis yang akurat terhadap kelemahan instruksional.

Dalam perspektif siklus PDCA (Deming, 1986), kegagalan melakukan tindak lanjut terstruktur berarti mematikan fase Act, yang secara otomatis melumpuhkan siklus Continuous Quality Improvement di madrasah atau sekolah (Puspita et al., 2026). Padahal, fase Act merupakan elemen krusial yang menghubungkan temuan evaluasi dengan perbaikan sistemik. Sallis (2014) menegaskan bahwa evaluasi tanpa intervensi mutu hanya membuahkan pemborosan administratif (administrative waste). (Ahmadi & Khozin, 2025) menambahkan bahwa penjaminan mutu di lembaga pendidikan Islam masih bersifat prosedural dan belum diterjemahkan ke dalam tindakan perbaikan yang berkelanjutan.

Lebih jauh, reduksi tindak lanjut ini bertentangan dengan nilai islah dalam epistemologi evaluasi Islam yang menuntut komitmen perbaikan sistemik, bukan sekadar perbaikan individual (Ahmadi & Khozin, 2025). (Yuniartin et al., 2024) menekankan bahwa evaluasi PAI seharusnya mengintegrasikan aspek kognitif, moral, dan spiritual secara holistik, sehingga tindak lanjutnya pun harus mencakup perbaikan pada seluruh dimensi tersebut. Kegagalan dalam melakukan tindak lanjut yang komprehensif menyebabkan kelemahan instruksional institusi menjadi tak terdeteksi dan kemandekan strategi pedagogis terus berlanjut (Aulia & Yuliyanti, 2024); (Lubis et al., 2025). Hal ini menjadi dasar penting bagi pengembangan FEIM yang menempatkan identifikasi masalah sebagai langkah terpisah sebelum merancang intervensi, guna memastikan bahwa tindak lanjut tidak sekadar bersifat reaktif tetapi juga transformatif dan berorientasi pada perbaikan sistemik.

Terputusnya Rantai Pemeliharaan Mutu Pedagogis (The Broken Quality Loop)

Temuan ketiga penelitian ini mengungkap bahwa matinya siklus kualitas terjadi akibat kegagalan institusi dalam mengagregasi dan menganalisis data evaluasi untuk perbaikan instruksional. Data evaluasi PAI jarang diolah secara sistemik untuk mencari akar masalah pedagogis yang sesungguhnya (Fauzi & Nurfuadi, 2026); (Yuniartin et al., 2024). Arsip nilai hanya berakhir di buku ledger tanpa pernah dijadikan umpan balik bagi perbaikan strategi mengajar. Akibatnya, kelemahan metodologis guru dan desain kurikulum yang tidak relevan terus terulang dari tahun ke tahun, menciptakan stagnasi mutu pembelajaran yang berkepanjangan (Aulia & Yuliyanti, 2024); (Lubis et al., 2025).

Dalam kerangka Continuous Quality Improvement (CQI), data evaluasi seharusnya menjadi pemicu utama siklus perbaikan (Puspita et al., 2026) menegaskan bahwa evaluasi tanpa intervensi mutu hanya menghasilkan pemborosan administratif. Namun, realitas di lapangan menunjukkan bahwa institusi pendidikan gagal mentransformasikan data evaluasi menjadi tindakan korektif yang terstruktur. (Fithri & Ali, 2026) menyoroti bahwa perbaikan mutu sering bersifat administratif dan minim internalisasi nilai-nilai perbaikan sistemik. Hal ini menyebabkan fase Check dalam siklus PDCA tidak pernah terhubung dengan fase Act secara efektif (Deming, 1986), sehingga siklus kualitas terputus dan kehilangan fungsi perbaikannya.

Perspektif Model CIPP (Stufflebeam & Coryn, 2014) memperkuat analisis ini. Banyak sekolah sangat baik pada fase Context dan Process, namun data fase Product tidak pernah diumpanbalikkan (feedback loop) ke awal sistem untuk memperbaiki konteks, input, maupun proses pembelajaran. Ahmadi dan Khozin (2025) menambahkan bahwa penjaminan mutu di lembaga pendidikan Islam masih bersifat prosedural dan belum terintegrasi dengan mekanisme perbaikan berkelanjutan. Kegagalan ini diperparah oleh rendahnya kompetensi analitik guru dalam mengubah sekumpulan angka menjadi narasi kelemahan kurikulum yang dapat ditindaklanjuti (Halimah et al., 2026); (Julhadi et al., 2026).

Dalam epistemologi evaluasi Islam, terputusnya rantai pemeliharaan mutu ini bertentangan dengan nilai islah dan itqan. Islah menuntut komitmen perbaikan sistemik, sementara itqan menekankan upaya mencapai presisi dan keunggulan paripurna dalam setiap aspek pendidikan (Ahmadi & Khozin, 2025). Tanpa adanya mekanisme umpan balik yang sistematis, institusi pendidikan Islam kehilangan kesempatan untuk melakukan introspeksi (muhasabah) secara kolektif terhadap kelemahan instruksional yang ada (Yuniartin et al., 2024). Kondisi ini menjadi justifikasi utama bagi pengembangan FEIM, yang secara eksplisit merancang tahapan analisis reflektif dan identifikasi masalah untuk memastikan data evaluasi benar-benar terhubung dengan tindakan perbaikan yang terukur dan berkelanjutan.

Konstruksi Follow-Up Evaluation Improvement Model (FEIM) sebagai Solusi Integratif

Temuan keempat penelitian ini menyajikan konstruksi Follow-Up Evaluation Improvement Model (FEIM) sebagai jawaban atas krisis sistemik dalam praktik evaluasi dan tindak lanjut PAI. FEIM dirancang sebagai kerangka konseptual yang mengikat agregasi data evaluasi, analisis reflektif, identifikasi masalah, desain intervensi, implementasi, pemantauan, dan evaluasi ulang ke dalam siklus yang koheren. Model ini terbagi ke dalam delapan langkah tak terpisahkan yang secara fundamental menghubungkan hasil penilaian dengan perbaikan mutu berkelanjutan (Puspita et al., 2026).

Langkah pertama FEIM adalah pengumpulan data evaluasi secara holistik yang mencakup ranah kognitif, afektif, dan psikomotorik. Hal ini merespons temuan (Fauzi & Nurfuadi, 2026) dan (Fauzi & Nurfuadi, 2026) serta (Yuniartin et al., 2024)

tentang dominansi evaluasi kognitif yang mengabaikan dimensi afektif dan spiritual. Langkah kedua adalah analisis reflektif (muhasabah), di mana tim pendidik mengagregasi data untuk memotret tren kelemahan kelas secara general. Analisis ini menuntut introspeksi kolektif terhadap kemungkinan kelemahan pedagogis pendidik, bukan sekadar melimpahkan kegagalan pada kapasitas kognitif siswa (Majid & Abdullah, 2022; Zarkasyi, 2020). Langkah ketiga adalah identifikasi masalah, yaitu proses investigasi ke hulu untuk mencari akar permasalahan (root causes), apakah bersumber dari metode guru, kesiapan siswa, atau kompleksitas media pembelajaran (Halimah et al., 2026); (Julhadi et al., 2026).

Langkah keempat adalah desain program tindak lanjut yang mencakup intervensi strategis ganda: level mikro untuk siswa (remedial berbasis diagnosis) dan level makro untuk pelatihan atau revisi metode guru. Langkah kelima adalah implementasi (islah), yaitu eksekusi tindakan korektif ke dalam jam pelajaran dan kurikulum riil. Langkah keenam adalah pemantauan dan refleksi partisipatif oleh kepala sekolah atau sejawat untuk memastikan intervensi dieksekusi dengan kepatuhan desain (fidelity). Langkah ketujuh adalah peningkatan mutu (itqan), yang menandai tercapainya lompatan kualitas pembelajaran dan moralitas keagamaan siswa secara terukur. Langkah kedelapan adalah evaluasi ulang siklikal untuk mengukur dampak implementasi islah dan menjadi pemicu siklus FEIM berikutnya (Puspita et al., 2026); (Ahmadi & Khozin, 2025).

Keunggulan FEIM dibandingkan model klasik terletak pada integrasi rasionalitas manajemen modern dengan epistemologi Islam. Model CIPP (Stufflebeam & Coryn, 2014) dan siklus PDCA (Deming, 1986) memberikan instrumen prosedural yang terukur, sementara konsep muhasabah, islah, dan itqan memberikan landasan teologis dan kultural bagi institusi pendidikan Islam (Majid & Abdullah, 2022; Zarkasyi, 2020). (Fithri & Ali, 2026) menegaskan bahwa integrasi Total Quality Management dengan nilai-nilai Islam memungkinkan perbaikan mutu yang tidak hanya efisien secara manajerial tetapi juga bermakna secara etis dan spiritual. (Aulia & Yuliyanti, 2024) serta (Lubis et al., 2025) menambahkan bahwa kerangka kerja yang adaptif dan berakar pada nilai-nilai lokal-keagamaan lebih resonan bagi guru dan pengelola madrasah dibandingkan adopsi model manajemen sekuler secara mentah.

FEIM berkontribusi menggeser paradigma evaluasi PAI dari orientasi pemenuhan birokrasi nilai menuju mesin penggerak peningkatan mutu yang sirkular dan transformatif. Model ini memaksa adanya pemberhentian prosedural antara penilaian dan tindakan melalui tahap analisis reflektif dan identifikasi masalah, sehingga menihalkan kebiasaan guru yang spontan memberikan remedial buta tanpa mengetahui penyakit pedagogis yang sebenarnya (Ahmadi & Khozin, 2025). Dengan delapan langkah yang terstruktur, FEIM menjamin bahwa setiap hasil evaluasi benar-benar diterjemahkan menjadi perbaikan sistemik yang berkelanjutan, sesuai dengan prinsip Continuous Quality Improvement (Puspita et al., 2026).

Implikasi Teoretis dan Praktis FEIM bagi Pemangku Kepentingan Pendidikan PAI

Secara teoretis, FEIM memberikan kontribusi penting dalam mereposisi dan mendefinisikan ulang batas ontologis evaluasi pembelajaran itu sendiri. Evaluasi tidak lagi dipahami sebagai garis finis (terminal point) dari proses pembelajaran, melainkan sebagai jembatan (core bridge) yang membuka siklus kehidupan pedagogis berikutnya. Konstruksi FEIM memperkaya diskursus Total Quality Management dalam pendidikan dengan membuktikan bahwa konsep sekuler seperti Continuous Quality Improvement memiliki kohesi yang kuat dengan worldview pendidikan Islam bila disintesis secara cermat (Puspita et al., 2026); (Fithri & Ali, 2026). Integrasi nilai muhasabah, islah, dan itqan ke dalam siklus perbaikan mutu memberikan landasan teologis yang selama ini absen dalam model-model manajemen mutu konvensional (Majid & Abdullah, 2022; Zarkasyi, 2020).

Implikasi praktis FEIM mencakup tiga pemangku kepentingan utama. Pertama, bagi guru PAI, model ini menuntut transformasi peran dari sekadar "hakim penilai" menjadi "praktisi reflektif" yang mampu melakukan diagnostik berbasis data asesmen kelas. Guru tidak lagi cukup hanya memberikan nilai, tetapi harus mampu mengagregasi data evaluasi, mengidentifikasi kelemahan instruksional, dan merancang intervensi perbaikan yang terarah (Halimah et al., 2026); (Julhadi et al., 2026). Kedua, bagi kepala sekolah dan manajemen, FEIM mewajibkan perubahan instrumen supervisi klinis akademik. Kepala sekolah harus menilai kinerja pendidik berdasarkan sejauh mana mereka menyusun strategi adaptasi dan perbaikan pasca ujian, bukan sekadar memeriksa kelengkapan administrasi pembelajaran (Ahmadi & Khozin, 2025); (Aulia & Yuliyanti, 2024).

Ketiga, bagi pengambil kebijakan di lingkungan Kementerian Agama dan Kementerian Pendidikan, model ini merekomendasikan perlunya memasukkan "Keefektifan Program Tindak Lanjut Evaluasi" sebagai indikator kunci (Key Performance Indicator) dalam matriks akreditasi sekolah berstandar tinggi. Hal ini sejalan dengan temuan (Lubis et al., 2025) yang menyoroti pentingnya manajemen dan praktik pembelajaran yang terukur di lembaga pendidikan Islam. (Yuniartin et al., 2024) juga menekankan bahwa kebijakan evaluasi PAI harus mengintegrasikan aspek kognitif, moral, dan spiritual secara holistik, yang menuntut adanya mekanisme tindak lanjut yang komprehensif.

Lebih jauh, FEIM berpotensi mengubah budaya evaluasi di lembaga pendidikan Islam dari sekadar pemenuhan kewajiban birokrasi menjadi motor intelektual untuk inovasi pembelajaran yang berkelanjutan. Dengan delapan langkah yang terstruktur, model ini menjamin bahwa setiap siklus evaluasi menghasilkan perbaikan nyata, baik pada level mikro (perbaikan strategi mengajar guru) maupun level makro (peningkatan mutu kurikulum dan kebijakan institusional). (Fithri & Ali, 2026) menegaskan bahwa integrasi manajemen mutu dengan nilai-nilai Islam memungkinkan perbaikan yang tidak hanya efisien secara manajerial tetapi juga bermakna secara etis dan spiritual. Dengan demikian, FEIM tidak hanya menjadi alat teknis tetapi juga wahana untuk mewujudkan pendidikan Islam yang paripurna (itqan) dan terus-menerus memperbaiki diri (islah).

Keterbatasan Penelitian dan Agenda Penelitian Mendatang

Penelitian ini memiliki keterbatasan fundamental karena sepenuhnya bersandar pada desain tinjauan literatur kualitatif (studi konseptual) yang tidak melibatkan pengujian empiris di lapangan. FEIM yang dirumuskan merupakan blueprint teoretis yang belum diuji kehandalan operasionalnya secara masif dalam konteks nyata pembelajaran PAI di madrasah, pesantren, maupun sekolah umum. Sebagaimana diakui oleh (Puspita et al., 2026) dan (Fithri & Ali, 2026), kerangka konseptual tentang perbaikan mutu memerlukan validasi empiris untuk memastikan efektivitasnya dalam meningkatkan kualitas pembelajaran secara terukur. Keterbatasan ini juga disoroti oleh (Ahmadi & Khozin, 2025) yang menekankan bahwa model penjaminan mutu di lembaga pendidikan Islam perlu diuji melalui studi kasus mendalam untuk melihat implementasinya dalam konteks spesifik.

Selain itu, karena penelitian ini hanya mengandalkan data sekunder dari publikasi ilmiah, terdapat potensi bias seleksi dan keterbatasan cakupan literatur yang terindeks di basis data tertentu. Meskipun protokol PRISMA 2020 telah diterapkan untuk menjamin akuntabilitas dan replikabilitas, tidak semua praktik evaluasi dan tindak lanjut PAI di lapangan terdokumentasikan dengan baik dalam jurnal ilmiah bereputasi (Julhadi et al., 2026); (Halimah et al., 2026). Hal ini menyebabkan gambaran yang diperoleh mungkin belum mewakili seluruh realitas pedagogis di berbagai lembaga pendidikan Islam dengan karakteristik dan tantangan yang beragam.

Oleh karena itu, direkomendasikan kepada peneliti selanjutnya untuk menguji validitas praksis model FEIM melalui pendekatan Research and Development (R&D) atau Classroom Action Research berskala longitudinal. Uji coba langsung di madrasah, pesantren, dan sekolah umum akan membuktikan secara empiris sejauh mana kedelapan tahapan FEIM efektif mendongkrak retensi memori, pemahaman kognitif, dan manifestasi perilaku religius siswa secara signifikan (Lubis et al., 2025); (Yuniartin et al., 2024). Penelitian tindakan kelas juga dapat mengeksplorasi hambatan implementasi dan faktor-faktor yang mempengaruhi keberhasilan FEIM dalam konteks budaya dan kapasitas sumber daya yang berbeda-beda (Aulia & Yuliyanti, 2024).

Penelitian selanjutnya juga disarankan untuk mengembangkan instrumen pengukuran yang dapat mengkuantifikasi dampak implementasi FEIM terhadap peningkatan mutu pembelajaran PAI. Hal ini penting untuk menghasilkan data empiris yang dapat menjadi dasar kebijakan berbasis bukti (evidence-based policy) bagi pengambil keputusan di lingkungan Kementerian Agama dan Kementerian Pendidikan. Dengan demikian, FEIM tidak hanya menjadi konstruk teoretis tetapi juga model aplikatif yang terbukti mampu mentransformasi evaluasi PAI dari kewajiban birokrasi yang membebani menjadi motor intelektual untuk inovasi pembelajaran yang berkelanjutan dan bermakna.

SIMPULAN

Berdasarkan hasil analisis dan pembahasan yang telah diuraikan, penelitian ini menyimpulkan bahwa praktik evaluasi Pendidikan Agama Islam di lapangan masih sangat didominasi oleh orientasi pengukuran kognitif-somatif yang statis, sementara asesmen afektif dan psikomotorik terabaikan karena keterbatasan instrumen dan pemahaman guru. Penghentian siklus evaluasi pada distribusi nilai akhir, ditambah dengan program tindak lanjut yang hanya berupa tugas remedial reaktif dan berpusat pada siswa, menyebabkan rantai peningkatan mutu terputus secara fatal. Kelemahan instruksional institusi menjadi tak terdeteksi dan kemandekan strategi pedagogis terus berlanjut dari tahun ke tahun. Untuk memecahkan persoalan ini, tindak lanjut evaluasi wajib dikembalikan pada fungsinya sebagai penghubung absolut antara penilaian dan perbaikan mutu. Follow-Up Evaluation Improvement Model (FEIM) menawarkan kerangka konseptual baru yang mengikat agregasi data evaluasi, analisis akar masalah (muhasabah), desain intervensi terarah, implementasi (islah), dan pemantauan ke dalam siklus yang koheren menuju kesempurnaan akademik dan moral (itqan). FEIM berpotensi mentransformasi evaluasi dari kewajiban birokrasi yang membebani menjadi motor intelektual untuk inovasi pembelajaran yang berkelanjutan. Namun, karena penelitian ini sepenuhnya bersandar pada desain tinjauan literatur, direkomendasikan kepada peneliti selanjutnya untuk menguji validitas praksis model FEIM melalui Research and Development (R&D) atau Classroom Action Research berskala longitudinal di madrasah, pesantren, dan sekolah umum guna membuktikan secara empiris efektivitas kedelapan tahapan FEIM dalam mendongkrak pemahaman kognitif, retensi memori, dan manifestasi perilaku religius siswa secara signifikan.

DAFTAR RUJUKAN

- Ahmadi, & Khozin. (2025). Implementing contextualized quality assurance in Islamic boarding schools: A case study of TMI Al-Amien Prenduan. *Al-Ishlah: Jurnal Pendidikan*, 17(4), 7048–7057. <https://doi.org/10.35445/alishlah.v17i4.7855>
- Aulia, A., & Yuliyanti. (2024). The strategic role of Islamic education management in integrating Islamic value-based character education in the digital and technology era. *IJOIS: Indonesian Journal of Islamic Studies*, 5(2), 13–18. <https://doi.org/10.59525/ijois.v5i2.548>
- Deming, W. E. (1986). *Out of the crisis*. MIT Center for Advanced Engineering Study.
- Fauzi, I., & Nurfuadi. (2026). Evaluasi pembelajaran PAI berbasis nilai-nilai Islam. *IHSAN: Jurnal Pendidikan Islam*, 4(2). <https://doi.org/10.61104/ihsan.v4i2.5491>
- Fithri, A. A., & Ali, M. (2026). Integrasi Total Quality Management dan nilai Islam dalam pengembangan mutu pendidikan Islam. *Varied Knowledge Journal*, 3(4), 639–648. <https://doi.org/10.71094/vkj.v3i4.207>

- Halimah, S., Amelina, S. T., Ramadhani, V. O., & Ilmiah, N. K. (2026). The importance of implementing formative evaluation in Islamic education learning. *Kuttab: Jurnal Ilmu Pendidikan Islam*, 10(1), 21–30. <https://doi.org/10.30736/ktb.v10i1.2629>
- Julhadi, Metriadi, Sasputra, H., Syarif, M., & Adhitya. (2026). Authentic assessment in Islamic education: Challenges and practice. *Ahlussunnah: Journal of Islamic Education*, 5(1).
- Kabir, S. M. A. (2024). Book review: Thematic analysis: A practical guide. *Pasaa Paritat Journal*, 39, 141–144.
- Lubis, Y., Aulia, D. S., Hakim, W. M., & Alzahwa, N. (2025). Management and learning practices in an Islamic school: A case study at MTS IRA Medan. *Jurnal Mudabbir (Journal Research and Education Studies)*, 5(2).
- Majid, S., & Abdullah, Z. (2022). Institutionalizing Muhasabah in the quality assurance of Islamic schools. *Journal of Islamic Ethics and Education*, 9(2), 88–104.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372(n71). <https://doi.org/10.1136/bmj.n71>
- Puspita, K. T., Mu'alimin, & Anwar, K. (2026). Perbaikan mutu berkelanjutan (continuous improvement) dan standar mutu (quality standard) dalam Pendidikan Agama Islam. *Indonesian Journal of Islamic Studies (IJIS)*, 2(1). <https://doi.org/10.62567/ijis.v2i1.2423>
- Sallis, E. (2014). *Total quality management in education* (3rd ed.). Routledge.
- Stufflebeam, D. L., & Coryn, C. L. S. (2014). *Evaluation theory, models, and applications* (2nd ed.). Jossey-Bass.