

INTEGRASI EVALUASI PENDIDIKAN DAN SUPERVISI AKADEMIK BERBASIS BUKTI: KAJIAN LITERATUR UNTUK PENINGKATAN MUTU PEMBELAJARAN

Eka Yuliani¹, Muhammad Isnaini², Rahmad Azazi Rhomantoro³

^{1,2,3}Universitas Islam Negeri Sultan Aji Muhammad Idris Samarinda, Indonesia

Email: yulianieka452@gmail.com

Abstrak

Evaluasi pendidikan memiliki fungsi strategis ketika hasilnya tidak berhenti sebagai skor, tetapi diterjemahkan menjadi bukti untuk pengambilan keputusan dan perbaikan pembelajaran. Artikel ini bertujuan menganalisis fungsi komplementer evaluasi penempatan, formatif, diagnostik, dan sumatif serta mensintesis bagaimana bukti evaluatif dapat diintegrasikan ke dalam supervisi akademik. Kajian menggunakan pendekatan narrative literature review terstruktur. Literatur ditelusuri secara terarah melalui ERIC, Google Scholar, laman penerbit jurnal, dan repositori resmi pemerintah dengan fokus publikasi 2020–2026; sumber klasik tetap digunakan untuk konsep seminal. Literatur diseleksi berdasarkan relevansi terhadap evaluasi pembelajaran, kualitas bukti asesmen, supervisi akademik, pengembangan profesional guru, dan kebijakan asesmen di Indonesia. Hasil kajian menunjukkan bahwa tiap jenis evaluasi menghasilkan informasi yang berbeda: evaluasi penempatan memetakan kesiapan awal, formatif menyediakan umpan balik selama pembelajaran, diagnostik menelusuri akar kesulitan, dan sumatif menilai capaian akhir. Namun, nilai evaluatif baru bermakna bagi supervisi ketika bukti ditafsirkan melalui prinsip validitas, triangulasi, konteks, dan dialog reflektif. Bukti mutakhir mendukung manfaat formative assessment, tetapi juga menunjukkan variasi efek dan keterbatasan kepastian bukti, sehingga supervisor perlu menghindari keputusan berbasis satu indikator. Artikel ini menawarkan model konseptual lima tahap: bukti evaluatif, interpretasi, penetapan fokus supervisi, tindak lanjut profesional, dan evaluasi ulang. Model tersebut menempatkan evaluasi dan supervisi sebagai satu siklus perbaikan berkelanjutan yang berorientasi pada mutu pembelajaran, bukan sekadar kepatuhan administratif.

Kata Kunci: Evaluasi Pendidikan, Asesmen Formatif, Supervisi Akademik, Bukti Evaluatif, Mutu Pembelajaran.

Abstract

Educational evaluation becomes strategically meaningful when its results do not end as scores but are translated into evidence for decision-making and instructional improvement. This article analyzes the complementary functions of placement, formative, diagnostic, and summative evaluation and synthesizes how evaluative evidence can be integrated into academic supervision. A structured narrative literature review was employed. Literature was purposively identified through ERIC, Google Scholar, journal publisher websites, and official government repositories, with emphasis on publications from 2020–2026; seminal older works were retained for foundational concepts. Sources were selected for their relevance to classroom evaluation, assessment evidence quality, academic supervision, teacher professional development, and Indonesian assessment policy. The review indicates that each evaluation type generates different information: placement evaluation maps initial readiness, formative evaluation provides evidence during learning, diagnostic evaluation investigates

the roots of learning difficulties, and summative evaluation establishes final attainment. Yet evidence becomes useful for supervision only when interpreted through validity, triangulation, context, and reflective professional dialogue. Recent evidence supports the benefits of formative assessment but also points to heterogeneous effects and limitations in certainty, cautioning against decisions based on a single indicator. This article proposes a five-stage conceptual model consisting of evaluative evidence, interpretation, supervision focus-setting, professional follow-up, and re-evaluation. The model positions evaluation and academic supervision as an integrated cycle of continuous improvement oriented toward learning quality rather than administrative compliance.

Keywords: Educational Evaluation, Formative Assessment, Academic Supervision, Evaluative Evidence, Learning Quality.

A. PENDAHULUAN

Evaluasi merupakan komponen inheren dalam pendidikan karena setiap keputusan pembelajaran membutuhkan bukti mengenai kondisi awal peserta didik, proses belajar, kesulitan yang muncul, dan capaian akhir. Dalam praktiknya, pengukuran, asesmen, penilaian, dan evaluasi sering dipertukarkan, padahal masing-masing memiliki fokus yang berbeda. Pengukuran berorientasi pada kuantifikasi atribut atau performa; asesmen mengumpulkan dan menafsirkan bukti; penilaian memberi makna terhadap bukti; sedangkan evaluasi menggunakan bukti tersebut untuk menetapkan nilai atau keberhargaan dan menentukan tindak lanjut. Dengan demikian, tes tertulis hanyalah salah satu instrumen, bukan keseluruhan proses evaluasi.

Landasan klasik evaluasi pendidikan memperlihatkan pergeseran dari orientasi ketercapaian tujuan menuju pengambilan keputusan. Tyler (1949) menempatkan tujuan pendidikan sebagai rujukan evaluasi, sementara Cronbach (1963) memperluas fungsi evaluasi sebagai sumber informasi untuk menentukan tindakan. Scriven (1967) kemudian membedakan fungsi formatif dan sumatif, yang menegaskan bahwa waktu penggunaan bukti dan tujuan keputusan sama pentingnya dengan bentuk instrumen. Kerangka ini tetap relevan, tetapi kebutuhan pendidikan kontemporer menuntut integrasi yang lebih kuat antara bukti asesmen, refleksi profesional, dan keputusan perbaikan.

Persoalan utama bukan ketiadaan instrumen evaluasi, melainkan kecenderungan mereduksi evaluasi menjadi skor, nilai rapor, dan keputusan akhir. Ketika data digunakan hanya untuk klasifikasi, fungsi diagnostik dan formatif menjadi lemah. Nilai rendah dapat menunjukkan bahwa peserta didik belum mencapai standar, tetapi tidak dengan sendirinya menjelaskan apakah penyebabnya adalah miskonsepsi, lemahnya pengetahuan prasyarat, strategi pembelajaran yang tidak sesuai, kendala bahasa, keterbatasan akses, atau faktor nonakademik. Karena itu, evaluasi yang bermakna harus mampu mengubah data menjadi informasi dan informasi menjadi keputusan yang dapat diuji kembali.

Bukti empiris mutakhir menguatkan pentingnya penggunaan asesmen selama pembelajaran, tetapi sekaligus mengingatkan bahwa dampaknya tidak seragam. Yao et al. (2024) menganalisis 258 effect sizes dari 118 studi primer dan menemukan efek keseluruhan formative assessment sebesar Hedges' $g = 0,25$ terhadap capaian akademik K-12. Sementara itu, umbrella review oleh Sortwell et al. (2024) terhadap 13 meta-analisis menunjukkan kecenderungan efek positif, namun kepastian bukti pada banyak meta-analisis dinilai rendah atau sangat rendah. Temuan tersebut penting karena menunjukkan dua hal: formative assessment berpotensi memperbaiki pembelajaran, tetapi kualitas keputusan tetap bergantung pada desain asesmen, konteks implementasi, dan cara bukti ditafsirkan.

Dalam konteks Indonesia, Panduan Pembelajaran dan Asesmen Edisi Revisi 2025 menempatkan pembelajaran dan asesmen sebagai satu siklus. Asesmen digunakan untuk

mencari bukti ketercapaian tujuan pembelajaran sekaligus untuk mengecek efektivitas pembelajaran yang berlangsung, dengan penekanan kuat pada asesmen formatif (Anggraena et al., 2025). Arah kebijakan ini membuka ruang untuk menghubungkan evaluasi dengan supervisi akademik secara lebih substantif. Supervisi tidak seharusnya berhenti pada pemeriksaan administrasi, tetapi memanfaatkan bukti tentang belajar peserta didik dan praktik mengajar guru untuk merancang dialog profesional dan tindak lanjut.

Kajian supervisi di Indonesia menunjukkan bahwa fungsi perbaikan tersebut belum selalu berjalan optimal. Setiyadi et al. (2025), melalui sintesis penelitian 2021–2024, menemukan bahwa program supervisi akademik pada jenjang dasar dan menengah pada umumnya diarahkan untuk meningkatkan kinerja guru, tetapi masih menghadapi persoalan pemahaman konsep, pelaksanaan, dan tindak lanjut. Penelitian Maisyaroh et al. (2021) juga menunjukkan bahwa teknik, prinsip, dan pendekatan supervisi berhubungan dengan variasi model dan bahan pembelajaran yang digunakan guru. Studi-studi korelasional di Indonesia selanjutnya memperlihatkan hubungan positif antara supervisi akademik dengan kinerja atau profesionalisme guru (Awam et al., 2023; Isriyati et al., 2024; Yuliarti et al., 2024), meskipun desain penelitian tersebut tidak serta-merta membuktikan hubungan kausal.

Dari sini tampak adanya ruang konseptual yang belum cukup terintegrasi. Literatur evaluasi banyak membahas fungsi penempatan, formatif, diagnostik, dan sumatif, sedangkan literatur supervisi banyak membahas observasi kelas, umpan balik, coaching, dan pengembangan profesional. Namun, hubungan antara jenis bukti evaluatif, cara menafsirkan bukti, penentuan fokus supervisi, dan evaluasi dampak tindak lanjut sering diperlakukan sebagai bagian yang terpisah. Padahal, kualitas supervisi sangat bergantung pada kualitas bukti yang digunakan dan kualitas penalaran dalam menerjemahkan bukti tersebut menjadi keputusan pembinaan.

Artikel ini bertujuan: (1) menganalisis fungsi komplementer jenis-jenis evaluasi pendidikan; (2) menjelaskan prinsip agar bukti evaluatif layak digunakan dalam keputusan supervisi; dan (3) menyusun model konseptual integrasi evaluasi pendidikan dan supervisi akademik berbasis bukti. Kontribusi utama artikel bukan menawarkan jenis evaluasi baru, melainkan menyatukan fungsi evaluasi dengan siklus supervisi dalam kerangka yang dapat membantu guru, kepala sekolah, pengawas, dan pengelola pendidikan mengubah data pembelajaran menjadi tindakan perbaikan yang terarah dan dapat dievaluasi kembali.

B. METODE

Artikel ini menggunakan pendekatan narrative literature review terstruktur. Pendekatan ini dipilih karena tujuan kajian bukan menghitung efek gabungan secara statistik atau menghasilkan systematic review penuh, melainkan menyintesis konsep, bukti empiris, dan implikasi kebijakan untuk membangun kerangka konseptual yang koheren. Prinsip transparansi tetap diterapkan agar pembaca dapat memahami bagaimana literatur dipilih dan bagaimana sintesis dilakukan.

Penelusuran literatur dilakukan secara terarah melalui ERIC, Google Scholar, laman penerbit jurnal, dan repositori resmi pemerintah. Kata kunci digunakan dalam bahasa Indonesia dan Inggris, antara lain: evaluasi pendidikan, asesmen formatif, evaluasi diagnostik, academic supervision, instructional supervision, formative assessment, evidence-based supervision, teacher professional development, dan assessment policy. Literatur kontemporer diprioritaskan pada rentang 2020–2026. Sumber sebelum 2020 dipertahankan apabila memiliki posisi seminal, seperti Tyler, Cronbach, Scriven, Black dan Wiliam, Cogan, serta standar evaluasi AERA, APA, dan NCME.

Kriteria inklusi meliputi: (a) sumber yang secara langsung membahas fungsi evaluasi atau asesmen pembelajaran; (b) penelitian atau kajian mengenai supervisi akademik/instruksional dan pengembangan profesional guru; (c) sumber yang membahas

kualitas bukti asesmen, umpan balik, atau pengambilan keputusan; dan (d) dokumen kebijakan resmi yang relevan dengan konteks Indonesia. Sumber dikeluarkan apabila hanya menyebut evaluasi atau supervisi secara umum tanpa hubungan dengan pembelajaran, tidak menyediakan informasi bibliografis yang memadai, atau tidak memiliki relevansi substantif dengan pertanyaan kajian.

Analisis dilakukan melalui sintesis tematik. Pertama, konsep evaluasi dikelompokkan berdasarkan fungsi keputusan yang dilayani. Kedua, temuan tentang kualitas bukti dianalisis untuk mengidentifikasi syarat penggunaan data dalam supervisi. Ketiga, penelitian mengenai supervisi dipetakan untuk melihat bagaimana observasi, umpan balik, dialog reflektif, dan tindak lanjut digunakan. Keempat, hasil sintesis diterjemahkan menjadi model integratif yang menghubungkan bukti evaluatif dengan siklus supervisi akademik. Karena kajian ini bersifat naratif terstruktur dan bukan systematic review, artikel tidak mengklaim bahwa seluruh publikasi yang tersedia telah teridentifikasi. Keterbatasan ini dipertimbangkan dalam penarikan kesimpulan.

C. HASIL DAN PEMBAHASAN

1. Evaluasi sebagai Proses Pengambilan Keputusan Berbasis Bukti

Evaluasi pendidikan dapat dipahami sebagai proses berjenjang: bukti dikumpulkan, ditafsirkan, diberi makna, lalu digunakan untuk menentukan tindakan. Perspektif ini penting karena dua instrumen yang sama dapat memiliki fungsi berbeda tergantung tujuan penggunaannya. Sebuah kuis, misalnya, berfungsi formatif ketika hasilnya digunakan untuk memperbaiki strategi pembelajaran sebelum unit selesai, tetapi dapat berfungsi sumatif apabila skor digunakan sebagai bagian dari nilai akhir. Oleh sebab itu, klasifikasi evaluasi lebih tepat ditentukan oleh fungsi keputusan daripada semata-mata bentuk instrumen.

Konsekuensi dari pandangan berbasis bukti adalah bahwa kualitas evaluasi tidak dapat dinilai hanya dari banyaknya data. Data yang melimpah tidak otomatis menghasilkan keputusan yang baik. Bukti perlu relevan dengan konstruk yang ingin dinilai, cukup konsisten, dipahami dalam konteks, dan digunakan secara adil. Standards for Educational and Psychological Testing menekankan bahwa validitas berkaitan dengan dukungan bukti dan teori terhadap interpretasi skor untuk tujuan penggunaan tertentu (AERA, APA, & NCME, 2014). Dengan demikian, pertanyaan penting bagi supervisor bukan hanya 'berapa nilai peserta didik?', tetapi 'interpretasi apa yang sah dari nilai tersebut dan tindakan apa yang layak diambil berdasarkan interpretasi itu?'.

Implikasi ini menggeser posisi evaluasi dari aktivitas administratif menuju fungsi epistemik dan manajerial. Evaluasi menjadi cara organisasi pendidikan mengetahui kondisi nyata pembelajaran. Supervisi kemudian menjadi mekanisme profesional untuk menafsirkan pengetahuan tersebut bersama guru dan mengubahnya menjadi tindakan. Tanpa tahap interpretasi, supervisi berisiko menjadikan skor sebagai indikator tunggal kualitas mengajar; tanpa tindak lanjut, evaluasi hanya menghasilkan dokumentasi tanpa perbaikan.

2. Fungsi Komplementer Evaluasi Penempatan, Formatif, Diagnostik, dan Sumatif

Empat jenis evaluasi utama memiliki fungsi berbeda dan sebaiknya dipandang sebagai sistem yang saling melengkapi. Evaluasi penempatan menentukan titik awal, evaluasi formatif memandu perjalanan belajar, evaluasi diagnostik menelusuri penyebab hambatan, dan evaluasi sumatif memberikan informasi tentang capaian pada akhir periode. Integrasi keempatnya menghasilkan gambaran yang lebih utuh daripada ketergantungan pada satu jenis evaluasi.

Tabel 1. Fungsi Evaluasi dan Relevansinya Bagi Supervisi Akademik

Jenis Evaluasi	Pertanyaan Utama	Bukti yang Lazim	Keputusan Pembelajaran	Relevansi Supervisi
Penempatan	Dari titik mana peserta didik memulai?	Tes prasyarat, asesmen awal, wawancara, inventori kebutuhan	Diferensiasi, matrikulasi, pengelompokan fleksibel	Menilai kesesuaian perencanaan guru dengan kesiapan awal
Formatif	Apa yang sudah dipahami dan apa yang perlu diperbaiki?	Kuis, observasi, exit ticket, draf, rubrik, umpan balik	Penyesuaian strategi, revisi, pengayaan, remedial cepat	Mengamati kualitas pertanyaan, respons guru, umpan balik, dan adaptasi pembelajaran
Diagnostik	Mengapa kesulitan terjadi dan di mana hambatannya?	Analisis kesalahan, wawancara, tes subkompetensi, portofolio	Intervensi terarah, bantuan khusus, rujukan bila diperlukan	Memfokuskan supervisi pada sumber miskonsepsi dan hambatan belajar
Sumatif	Seberapa jauh capaian akhir telah diraih?	Ujian akhir, proyek final, praktik, portofolio akhir	Nilai, kelulusan, sertifikasi, akuntabilitas	Menilai pola capaian dan efektivitas keseluruhan, bukan menghakimi guru dari satu skor

Evaluasi penempatan memiliki nilai strategis karena membantu guru dan supervisor menguji apakah rancangan pembelajaran berangkat dari kondisi awal peserta didik. Jika hasil awal menunjukkan heterogenitas tinggi, supervisi dapat berfokus pada diferensiasi tugas, dukungan bertingkat, dan pemilihan sumber belajar. Kesalahan umum adalah menggunakan penempatan sebagai label kemampuan permanen; padahal, fungsinya seharusnya fleksibel dan mendukung keputusan awal yang dapat diperbarui ketika bukti baru muncul.

Evaluasi formatif memiliki hubungan paling langsung dengan perbaikan selama pembelajaran. Black dan Wiliam (1998) menempatkan formative assessment sebagai proses menggunakan bukti untuk memperkecil kesenjangan antara posisi belajar saat ini dan tujuan yang hendak dicapai. Bukti mutakhir mendukung manfaatnya, namun efeknya tidak seragam. Meta-analisis Yao et al. (2024) menunjukkan efek positif rata-rata yang relatif kecil sampai moderat, sedangkan Sortwell et al. (2024) menemukan variasi dari efek trivial hingga besar dan kepastian bukti yang beragam. Karena itu, supervisor sebaiknya tidak hanya menanyakan apakah guru 'melakukan asesmen formatif', tetapi bagaimana bukti dari asesmen tersebut memengaruhi keputusan pengajaran berikutnya.

Evaluasi diagnostik memberikan kedalaman yang tidak tersedia dari skor umum. Dua peserta didik dengan skor 60 dapat memerlukan intervensi yang sangat berbeda. Analisis pola kesalahan, wawancara singkat, atau telaah karya dapat membedakan apakah hambatan berasal dari konsep prasyarat, miskonsepsi, bahasa akademik, perhatian, atau faktor akses. Dalam supervisi, fungsi diagnostik mencegah simplifikasi masalah. Fokus pembinaan dapat

diarahkan pada desain penjelasan, representasi konsep, contoh yang digunakan, atau strategi umpan balik, bukan sekadar meminta guru 'meningkatkan nilai'.

Evaluasi sumatif diperlukan untuk akuntabilitas dan penetapan capaian akhir, tetapi memiliki risiko ketika diperlakukan sebagai satu-satunya indikator mutu pembelajaran. Skor akhir sering merupakan hasil akumulasi banyak faktor dan tidak selalu mampu mengisolasi kontribusi praktik mengajar. Oleh sebab itu, data sumatif lebih bermakna jika digunakan untuk mengidentifikasi pola: kompetensi mana yang konsisten rendah, kelompok mana yang tertinggal, dan apakah kelemahan yang sama muncul lintas periode. Pola tersebut kemudian perlu diverifikasi dengan bukti formatif, diagnostik, observasi kelas, dan artefak pembelajaran.

3. Dari Data Menjadi Bukti: Validitas, Triangulasi, dan Konteks

Integrasi evaluasi dan supervisi hanya bermakna apabila data yang digunakan cukup layak untuk mendukung keputusan. Pertama, bukti harus relevan dengan tujuan. Jika supervisor ingin menilai kualitas argumentasi peserta didik, skor pilihan ganda semata mungkin kurang memadai. Dibutuhkan artefak yang menunjukkan alasan, struktur argumen, atau performa autentik. Kedua, interpretasi harus proporsional terhadap kualitas bukti. Satu kuis tidak seharusnya digunakan untuk membuat kesimpulan luas mengenai kualitas guru, sebagaimana satu observasi kelas tidak cukup untuk menyimpulkan efektivitas pembelajaran sepanjang semester.

Triangulasi menjadi prinsip penting. Bukti hasil belajar perlu dibaca bersama proses pembelajaran. Skor, karya peserta didik, observasi, portofolio, percakapan dengan guru, kehadiran, dan data konteks dapat saling mengonfirmasi atau justru menunjukkan kontradiksi yang perlu ditelusuri. Ketika skor rendah tetapi karya proyek menunjukkan pemahaman mendalam, supervisor perlu mempertanyakan kesesuaian instrumen. Sebaliknya, skor tinggi yang tidak diikuti kemampuan menjelaskan konsep dapat menunjukkan bahwa asesmen lebih banyak mengukur hafalan atau latihan pola soal.

Konteks juga menentukan makna bukti. Kesenjangan akses teknologi, variasi bahasa, ukuran kelas, waktu belajar, dan karakteristik sekolah dapat memengaruhi hasil. Karena itu, supervisi berbasis bukti berbeda dari supervisi berbasis angka. Supervisi berbasis bukti selalu menghubungkan data dengan pertanyaan interpretatif: bukti apa yang tersedia, seberapa kuat bukti itu, faktor apa yang mungkin memengaruhi, dan bukti tambahan apa yang dibutuhkan sebelum mengambil keputusan.

Prinsip keadilan dan transparansi harus menyertai proses tersebut. Guru perlu mengetahui fokus supervisi, sumber bukti yang digunakan, dan bagaimana kesimpulan dibuat. Pendekatan ini memperkuat posisi supervisi sebagai bantuan profesional, bukan mekanisme inspeksi yang sepihak. Pada saat yang sama, supervisor bertanggung jawab menjaga kerahasiaan data peserta didik dan menghindari penggunaan data asesmen di luar tujuan yang dapat dipertanggungjawabkan.

4. Supervisi Akademik sebagai Dialog Profesional Berbasis Bukti

Supervisi akademik pada dasarnya bertujuan membantu guru meningkatkan kemampuan mengelola pembelajaran. Dalam tradisi supervisi klinis, prosesnya meliputi konferensi awal, observasi, analisis data, konferensi balikan, dan tindak lanjut (Cogan, 1973; Goldhammer et al., 1993). Struktur tersebut menyediakan ruang alami untuk memasukkan bukti evaluatif. Konferensi awal dapat dimulai dengan analisis data belajar; observasi diarahkan pada hipotesis yang muncul dari data; konferensi balikan menguji interpretasi; dan tindak lanjut dirancang untuk memperbaiki praktik yang paling relevan.

Temuan penelitian mutakhir mendukung fungsi pengembangan tersebut. Maisyaroh et al. (2021) menunjukkan bahwa teknik, prinsip, dan pendekatan supervisi berpengaruh

terhadap variasi model pembelajaran dan bahan ajar yang digunakan guru di Indonesia dan Filipina. Awam et al. (2023) menemukan kontribusi supervisi akademik terhadap kinerja guru pada konteks SMP negeri, sementara Isriyati et al. (2024) menunjukkan hubungan supervisi akademik dengan profesionalisme guru melalui motivasi mengajar. Yuliarti et al. (2024) juga melaporkan pengaruh supervisi akademik terhadap kinerja guru honorer. Bukti-bukti ini perlu dibaca sesuai desainnya: sebagian bersifat korelasional atau ex-post facto, sehingga lebih tepat digunakan untuk menunjukkan hubungan dan relevansi, bukan membuktikan mekanisme kausal secara tunggal.

Kajian Setiyadi et al. (2025) memperlihatkan bahwa masalah penting supervisi di Indonesia bukan hanya pelaksanaan observasi, melainkan kesinambungan dari pemahaman konsep, perencanaan, implementasi, hingga tindak lanjut. Temuan ini konsisten dengan argumen bahwa supervisi akan kehilangan makna bila berhenti pada pengumpulan instrumen atau pemberian skor observasi. Tindak lanjut merupakan titik di mana bukti berubah menjadi pembelajaran profesional. Tanpa tindak lanjut, supervisi menghasilkan arsip; dengan tindak lanjut yang tepat, supervisi menghasilkan perubahan praktik yang kemudian dapat diuji kembali melalui bukti baru.

Karena itu, umpan balik dalam supervisi sebaiknya bersifat spesifik, berbasis kriteria, dan berorientasi tindakan. Pernyataan seperti 'pengelolaan kelas perlu ditingkatkan' terlalu umum. Umpan balik yang lebih bermakna menghubungkan bukti dengan tindakan, misalnya: 'Pada tiga kesempatan, pertanyaan tingkat tinggi dijawab oleh siswa yang sama; pada siklus berikutnya, strategi distribusi respons dapat diubah dan dampaknya dilihat melalui kualitas jawaban lebih banyak siswa.' Struktur tersebut membuat supervisi dapat diamati, ditindaklanjuti, dan dievaluasi.

5. Model Integrasi Evaluasi Pendidikan dan Supervisi Akademik Berbasis Bukti

Berdasarkan sintesis literatur, artikel ini mengusulkan model lima tahap yang menempatkan evaluasi dan supervisi sebagai satu siklus perbaikan. Model tidak dimaksudkan sebagai prosedur baku yang menggantikan kebijakan sekolah, melainkan sebagai kerangka berpikir untuk memastikan bahwa setiap tahap supervisi mempunyai dasar bukti dan setiap tindak lanjut dievaluasi dampaknya.



Gambar 1. Model Konseptual Integrasi Evaluasi dan Supervisi Akademik Berbasis Bukti

Tahap pertama adalah pengumpulan bukti evaluatif. Sumber bukti dipilih sesuai pertanyaan yang ingin dijawab. Evaluasi penempatan cocok untuk memetakan kesiapan awal, formatif untuk memantau perkembangan, diagnostik untuk menelusuri hambatan, dan sumatif untuk melihat capaian akhir. Penggunaan beberapa sumber lebih dianjurkan ketika keputusan memiliki konsekuensi besar.

Tahap kedua adalah interpretasi. Data diperiksa dari sisi kesesuaian dengan tujuan, pola, konsistensi, konteks, dan kebutuhan bukti tambahan. Pada tahap ini guru dan supervisor

menghindari kesimpulan cepat. Tujuannya bukan mencari siapa yang salah, melainkan membangun penjelasan yang paling masuk akal terhadap masalah pembelajaran.

Tahap ketiga adalah penetapan fokus supervisi. Fokus dipilih sesempit mungkin agar dapat diamati dan ditindaklanjuti, misalnya kualitas pertanyaan guru, kejelasan representasi konsep, pola umpan balik, penggunaan asesmen formatif, diferensiasi, atau keterlibatan peserta didik. Fokus yang terlalu luas akan menghasilkan umpan balik yang umum dan sulit dievaluasi.

Tahap keempat adalah tindak lanjut profesional. Bentuknya dapat berupa coaching, peer observation, lesson study, perancangan ulang asesmen, revisi modul ajar, demonstrasi praktik, atau pendampingan individual. Tindak lanjut perlu disesuaikan dengan penyebab masalah. Jika hambatannya adalah miskonsepsi peserta didik, intervensi berbeda dari situasi ketika masalah berasal dari desain instrumen yang tidak selaras dengan tujuan.

Tahap kelima adalah evaluasi ulang. Setelah perbaikan diterapkan, bukti baru dikumpulkan untuk menjawab apakah perubahan praktik menghasilkan perbaikan yang diharapkan. Jika belum, diagnosis diperbarui. Siklus ini mencegah supervisi menjadi kegiatan periodik yang terputus-putus dan mengubahnya menjadi proses pembelajaran organisasi.

Tabel 2. Contoh transformasi bukti evaluatif menjadi keputusan supervisi

Temuan bukti	Interpretasi awal	Fokus supervisi	Tindak lanjut yang dapat diuji
Banyak siswa salah pada konsep yang sama	Ada kemungkinan miskonsepsi atau representasi konsep kurang efektif	Cara guru merepresentasikan konsep dan memeriksa pemahaman	Gunakan representasi alternatif dan exit ticket; bandingkan pola kesalahan berikutnya
Nilai akhir rendah tetapi proyek menunjukkan pemahaman baik	Instrumen sumatif mungkin kurang selaras dengan kompetensi	Kesesuaian asesmen dengan tujuan dan rubrik	Revisi blueprint/rubrik; uji konsistensi hasil dari dua bentuk asesmen
Hanya sedikit siswa merespons pertanyaan tingkat tinggi	Partisipasi dan distribusi kesempatan berpikir mungkin tidak merata	Teknik bertanya dan distribusi respons	Gunakan wait time dan strategi random response; ukur keterlibatan lebih luas
Perbedaan capaian awal sangat lebar	Kelas membutuhkan diferensiasi dukungan	Pemanfaatan data awal dalam perencanaan	Kelompok fleksibel/tugas bertingkat; evaluasi perubahan setelah beberapa pertemuan

6. Implikasi untuk Praktik Pendidikan Indonesia dan Era AI Generatif

Panduan Pembelajaran dan Asesmen 2025 menegaskan bahwa asesmen merupakan bagian dari siklus pembelajaran, bukan aktivitas terpisah (Anggraena et al., 2025). Implikasi bagi supervisi adalah bahwa kepala sekolah dan pengawas perlu memperhatikan bagaimana guru menggunakan hasil asesmen untuk menyesuaikan pembelajaran. Pemeriksaan terhadap dokumen asesmen saja tidak cukup. Bukti yang lebih bermakna adalah jejak keputusan: apa yang diketahui guru dari asesmen, perubahan apa yang dilakukan, dan apakah perubahan itu memberikan dampak.

Pada level sekolah, model integratif dapat diterapkan melalui rapat data yang sederhana tetapi terarah. Guru membawa satu atau dua jenis bukti, menjelaskan pola yang

terlihat, mendiskusikan kemungkinan penyebab, memilih satu fokus perbaikan, lalu menentukan bukti apa yang akan dikumpulkan setelah intervensi. Dengan demikian, supervisi tidak bergantung sepenuhnya pada kunjungan formal, tetapi diperkuat oleh budaya inquiry profesional.

Era AI generatif menambah alasan untuk memperkuat asesmen autentik dan bukti proses. Tugas tertulis yang dapat dihasilkan dengan bantuan AI tidak selalu menunjukkan proses berpikir peserta didik. Respons yang memadai bukan hanya meningkatkan pengawasan, melainkan merancang bukti yang lebih kaya: presentasi lisan, demonstrasi, portofolio dengan riwayat revisi, refleksi metodologis, studi kasus kontekstual, atau proyek yang memiliki akuntabilitas individual. Dalam konteks supervisi, fokus bergeser dari 'apakah siswa menggunakan AI' menuju 'apakah desain asesmen memungkinkan guru memperoleh bukti yang sah tentang kompetensi siswa'.

Digitalisasi supervisi juga perlu memperhatikan etika data. Platform dapat mempermudah dokumentasi, agregasi skor, dan pemantauan tindak lanjut, tetapi data pendidikan tidak boleh diperlakukan sebagai komoditas administratif. Akses perlu dibatasi sesuai peran, interpretasi otomatis perlu diverifikasi, dan data sensitif peserta didik harus dilindungi. Prinsip ini sejalan dengan tuntutan evaluasi modern yang menempatkan validitas, keadilan, transparansi, dan etika sebagai satu kesatuan.

7. Keterbatasan Kajian dan Agenda Penelitian Lanjutan

Kajian ini memiliki keterbatasan. Pertama, pendekatan yang digunakan adalah narrative literature review terstruktur, bukan systematic review, sehingga tidak dimaksudkan untuk memetakan seluruh studi yang tersedia atau menghasilkan estimasi efek baru. Kedua, literatur supervisi di Indonesia masih didominasi studi korelasional, studi kasus, dan kajian pada konteks tertentu, sehingga generalisasi perlu dilakukan dengan hati-hati. Ketiga, istilah supervisi akademik, supervisi instruksional, coaching, dan instructional leadership kadang digunakan dengan batas yang berbeda antarpublikasi.

Penelitian berikutnya dapat menguji model integratif ini secara empiris. Studi longitudinal dapat menilai apakah supervisi yang dimulai dari analisis bukti evaluatif menghasilkan perubahan praktik mengajar dan capaian belajar yang lebih konsisten dibandingkan supervisi administratif. Penelitian juga perlu mengembangkan instrumen yang mampu menilai kualitas penggunaan data oleh guru dan supervisor, bukan hanya frekuensi pelaksanaan supervisi. Selain itu, kajian pada sekolah dengan kondisi sumber daya berbeda penting untuk melihat apakah model dapat diterapkan secara adil dalam konteks kesenjangan digital dan heterogenitas kapasitas sekolah.

Agenda lain adalah menguji kualitas keputusan supervisi. Banyak penelitian menilai keberhasilan melalui persepsi guru atau skor kinerja, tetapi belum banyak yang menelusuri rantai bukti secara lengkap: bukti awal apa yang digunakan, interpretasi apa yang dibuat, intervensi apa yang dipilih, dan bukti apa yang menunjukkan perubahan. Penelitian yang menelusuri rantai tersebut akan memperkuat pemahaman tentang mekanisme, bukan hanya hubungan statistik antara supervisi dan kinerja.

D. KESIMPULAN

Evaluasi pendidikan memiliki nilai strategis ketika digunakan sebagai dasar pengambilan keputusan, bukan hanya sebagai mekanisme pemberian skor. Evaluasi penempatan, formatif, diagnostik, dan sumatif menghasilkan informasi yang berbeda tetapi saling melengkapi. Penempatan membantu memetakan titik awal, formatif menyediakan bukti untuk penyesuaian selama pembelajaran, diagnostik menelusuri akar kesulitan, sedangkan sumatif memberikan gambaran capaian akhir dan akuntabilitas.

Integrasi dengan supervisi akademik menuntut satu tahap yang sering terabaikan, yaitu interpretasi bukti. Data perlu dibaca melalui prinsip validitas, triangulasi, konteks, keadilan, dan transparansi sebelum diterjemahkan menjadi fokus pembinaan. Bukti mutakhir menunjukkan bahwa formative assessment pada umumnya bermanfaat bagi pembelajaran, tetapi variasi efek dan keterbatasan kepastian bukti menguatkan alasan untuk tidak menjadikan satu indikator sebagai dasar keputusan supervisi.

Kontribusi konseptual artikel ini adalah model lima tahap: pengumpulan bukti evaluatif, interpretasi, penetapan fokus supervisi, tindak lanjut profesional, dan evaluasi ulang. Model tersebut menempatkan evaluasi dan supervisi dalam satu siklus perbaikan berkelanjutan. Dengan orientasi ini, supervisi bergeser dari kegiatan administratif menuju proses inquiry profesional yang membantu guru memahami masalah pembelajaran, mencoba perbaikan, dan menilai dampaknya melalui bukti baru. Bagi praktik pendidikan Indonesia, pendekatan ini sejalan dengan arah kebijakan pembelajaran dan asesmen yang menempatkan asesmen sebagai bagian dari siklus pembelajaran serta menekankan penggunaan bukti untuk perkembangan kompetensi peserta didik.

DAFTAR PUSTAKA

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Anggraena, Y., Ginanto, D. E., Kesuma, A. T., & Setiyowati, D. (2025). *Panduan pembelajaran dan asesmen pendidikan anak usia dini, jenjang pendidikan dasar, dan jenjang pendidikan menengah: Edisi revisi tahun 2025*. Badan Standar, Kurikulum, dan Asesmen Pendidikan.
- Awam, R., Rifma, Hadiyanto, & Sabandi, A. (2023). The contribution of principal's academic supervision and pedagogical competence to teacher performance. *Indonesian Research Journal in Education*, 7(2), 529–544. <https://doi.org/10.22437/irje.v7i2.31476>
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.
- Cogan, M. L. (1973). *Clinical supervision*. Houghton Mifflin.
- Cronbach, L. J. (1963). Course improvement through evaluation. *Teachers College Record*, 64(8), 672–683.
- Goldhammer, R. J., Anderson, R. H., & Krajewski, R. A. (1993). *Clinical supervision: Special methods for the supervision of teachers* (3rd ed.). Harcourt Brace Jovanovich.
- Isriyati, Suminar, T., & Formen, A. (2024). The influence of principal's academic supervision on teachers' professionalism: Teaching motivation as a mediator for kindergarten teachers. *Indonesian Journal of Early Childhood Education Studies*, 13(1), 21–29. <https://doi.org/10.15294/ijeces.v13i1.13721>
- Maisyaroh, Wiyono, B. B., Hardika, Valdez, A. V., Mangorsi, S. B., & Canapi, S. P. T. (2021). The implementation of instructional supervision in Indonesia and the Philippines, and its effect on the variation of teacher learning models and materials. *Cogent Education*, 8(1), Article 1962232. <https://doi.org/10.1080/2331186X.2021.1962232>
- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagné, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (pp. 39–83). Rand McNally.
- Setiyadi, B., Pratama, M. A. N., Zamaludin, R., Komariah, A., & Ghani, M. F. A. (2025). Factual study of academic supervision at primary and secondary educational levels in

- Indonesia. *Jurnal Iqra': Kajian Ilmu Pendidikan*, 10(1), 453–476.
<https://doi.org/10.25217/ji.v10i1.6310>
- Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333–339.
<https://doi.org/10.1016/j.jbusres.2019.07.039>
- Sortwell, A., Trimble, K., Ferraz, R., Geelan, D. R., Hine, G., Ramirez-Campillo, R., Carter-Thuiller, B., Gkintoni, E., & Xuan, Q. (2024). A systematic review of meta-analyses on the impact of formative assessment on K-12 students' learning: Toward sustainable quality education. *Sustainability*, 16(17), Article 7826.
<https://doi.org/10.3390/su16177826>
- Tyler, R. W. (1949). *Basic principles of curriculum and instruction*. University of Chicago Press.
- Yao, Y., Amos, M., Snider, K., & Brown, T. (2024). The impact of formative assessment on K-12 learning: A meta-analysis. *Educational Research and Evaluation*, 29(7–8), 452–475. <https://doi.org/10.1080/13803611.2024.2363831>
- Yuliarti, D. P., Sari, E., & Fuad, N. (2024). Honorary teachers performance: Academic supervision and organizational culture in public junior high school. *Improvement: Jurnal Ilmiah untuk Peningkatan Mutu Manajemen Pendidikan*, 11(1), 81–89.
<https://doi.org/10.21009/improvement.v11i1.45037>