

Peran Evaluasi Pembelajaran dalam Meningkatkan Kualitas Proses Belajar Siswa

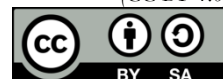
Robiah Ivadatun Nisa¹

¹Universitas Islam Negeri Sunan Ampel Surabaya, Indonesia

robiahivadatunnisa@gmail.com

Informasi Artikel	Abstrak (9 poin)
Riwayat artikel: Diterima: 2026-06-04 Direvisi: 2026-06-10 Diterima: 2026-06-16	Evaluasi pembelajaran merupakan unsur pokok dalam sistem pendidikan yang menentukan arah dan kualitas proses belajar siswa. Fenomena yang menjadi perhatian dalam kajian ini adalah kecenderungan praktik evaluasi di sekolah yang masih menitikberatkan pada fungsi sumatif semata, sehingga potensi evaluasi diagnostic dan formatif sebagai instrument perbaikan pembelajaran belum dimanfaatkan secara optimal. Kesenjangan penelitian terletak pada minimnya kajian yang membahas ketiga jenis evaluasi, diagnostik, formatif, dan sumatif secara integrative dalam satu kerangka analisis yang utuh dan kohesif. Penelitian ini bertujuan menganalisis peran masing-masing jenis evaluasi tersebut serta kontribusinya secara sinergis terhadap peningkatan kualitas belajar siswa. Pertanyaan penelitian yang diajukan meliputi: (1) bagaimana evaluasi diagnostik mengidentifikasi kebutuhan belajar siswa; (2) bagaimana evaluasi formatif memperbaiki proses pembelajaran secara berkelanjutan; (3) bagaimana evaluasi sumatif mengukur pencapaian kompetensi; dan (4) bagaimana ketiga jenis evaluasi berkontribusi secara integrative terhadap kualitas proses belajar. Penelitian menggunakan metode Library Research dengan teknik dokumentasi dan penelusuran literatur ilmiah dari jurnal nasional dan internasional bereputasi, buku referensi, serta dokumen kebijakan pendidikan. Analisis dilakukan melalui content analysis dan analisis tematik terhadap literatur yang terpilih. Hasil kajian menunjukkan bahwa evaluasi diagnostik efektif memetakan kesenjangan pemahaman awal siswa; evaluasi formatif terbukti meningkatkan motivasi, regulasi diri, dan kualitas umpan balik; sedangkan evaluasi sumatif berperan dalam akuntabilitas capaian kompetensi. Secara integrative ketiga jenis evaluasi yang diterapkan secara sinergis dalam satu sistem yang kohesif mampu meningkatkan kualitas proses belajar siswa secara bermakna.
Kata kunci: Evaluasi Diagnostik; Evaluasi Formatif; Evaluasi Sumatif; Kualitas Proses Belajar; Library Research.	
Jenis artikel: Artikel penelitian	

Karya ini dilisensikan di bawah Lisensi Creative Commons Attribution 4.0 International (CC BY 4.0).



1. Pendahuluan

1.1. Informasi latar belakang

Mutu pendidikan suatu bangsa tidak dapat dilepaskan dari kualitas sistem evaluasi yang diterapkan di dalamnya. Evaluasi pembelajaran bukan sekadar alat ukur pencapaian akhir, melainkan merupakan proses yang melekat dalam setiap tahapan kegiatan belajar mengajar dari perencanaan hingga refleksi pascapembelajaran. (Stufflebeam & Shinkfield, 2007) menegaskan bahwa evaluasi pendidikan yang komprehensif mencakup penilaian terhadap konteks, masukan, proses, dan produk pembelajaran. Pandangan ini mengarahkan pada kesadaran bahwa evaluasi sejatinya harus hadir di seluruh siklus pembelajaran, bukan hanya di penghujungnya. Relevansi kerangka CIPP ini semakin dikukuhkan oleh kajian-kajian terbaru yang membuktikan aplikabilitasnya dalam konteks pendidikan abad ke-21 yang dinamis (Mishra, 2022).

Kenyataan di lapangan memperlihatkan potret yang berbeda. Sejumlah kajian konseptual dan empiris menunjukkan bahwa praktik evaluasi di banyak sekolah masih berporos pada fungsi sumatif, yakni mengukur sejauh mana siswa telah menguasai materi setelah pembelajaran selesai (Popham, 2008). Orientasi ini menghasilkan budaya 'belajar untuk ujian' yang menempatkan siswa sebagai penerima pasif informasi, bukan sebagai agen aktif yang membangun pemahaman secara bertahap. Akibatnya, informasi berharga yang seharusnya dapat digali melalui evaluasi diagnostik dan formatif tidak pernah menjadi bagian dari siklus perbaikan pembelajaran.

Hubungan antara evaluasi dan kualitas proses belajar bersifat resiprokal dan dinamis. Evaluasi yang dirancang dengan cermat tidak hanya mencerminkan seberapa jauh siswa telah berkembang, tetapi juga secara aktif mendorong perkembangan itu sendiri (Black & Wiliam, 1998). Ketika guru menggunakan informasi dari evaluasi untuk menyesuaikan pendekatan mengajarnya, dan ketika siswa menggunakan umpan balik untuk memperbaiki strategi belajarnya, siklus peningkatan yang berkelanjutan dapat terwujud. Inilah yang oleh (Hattie & Timperley, 2007) disebut sebagai 'the feedback loop' lingkaran umpan balik yang menjadi mesin penggerak kualitas pembelajaran. Temuan ini mendapat konfirmasi empiris terbaru dari (Foster, 2024) yang melalui meta-analisis terhadap studi-studi antara tahun 2000 hingga 2023 menyimpulkan bahwa kualitas umpan balik formatif, keterlibatan siswa, dan kompetensi guru merupakan tiga variable yang paling determinan dalam efektivitas evaluasi terhadap prestasi belajar.

Dalam konteks Pendidikan Agama Islam (PAI), dimensi evaluasi memiliki makna yang melampaui sekadar pengukuran kognitif. Tujuan pembelajaran PAI mencakup pembentukan akhlak, internalisasi nilai-nilai keislaman, dan pengembangan kecerdasan spiritual yang tidak cukup hanya diukur melalui tes tertulis di akhir semester. Oleh karena itu, kebutuhan akan sistem evaluasi yang komprehensif yang mampu mendeteksi kebutuhan belajar awal, memantau perkembangan selama proses, sekaligus mengukur capaian kompetensi secara valid menjadi semakin mendesak dan relevan dalam konteks pendidikan Islam.

Pergeseran paradigma evaluasi dari 'assessment of learning' menuju 'assessment for learning' dan bahkan 'assessment as learning' merupakan salah satu perkembangan paling signifikan dalam ilmu pendidikan dua dekade terakhir (Stiggins, 2007). Paradigma baru ini menempatkan evaluasi bukan sebagai ancaman yang memicu kecemasan, melainkan sebagai mitra pembelajaran yang membantu siswa memahami posisi mereka dalam perjalanan belajar. Perkembangan ini semakin menguat dalam literatur kontemporer: penelitian terbaru konsisten menunjukkan bahwa ketika siswa dilibatkan aktif dalam proses evaluative melalui asesmen diri dan asesmen sejawat, kapasitas regulasi diri mereka berkembang secara signifikan dan berkelanjutan (Yan et al., 2022; Fang & Xu, 2023; Junaštiková, 2024) sebuah capaian yang jauh melampaui sekadar perbaikan nilai ujian.

1.2. Kesenjangan penelitian

Kajian-kajian terdahulu dalam bidang evaluasi pembelajaran telah menghasilkan pemahaman yang cukup mendalam tentang masing-masing jenis evaluasi secara terpisah. (Black & Wiliam, 1998) telah memberi landasan empiris yang kuat tentang efektivitas asesmen formatif. (Hattie & Timperley, 2007) telah mengembangkan model umpan balik yang berpengaruh luas. (Heritage, 2010) telah merinci bagaimana evaluasi diagnostik bekerja dalam praktik kelas sehari-hari. Namun, kajian-kajian tersebut umumnya berdiri sendiri dan belum secara eksplisit menghubungkan ketiga jenis evaluasi dalam satu kerangka integratif.

Kesenjangan kedua terletak pada konteks kajian. Sebagian besar literatur yang ada berasal dari tradisi pendidikan Barat yang karakteristik sistem, budaya, dan kurikulumnya berbeda dari konteks pendidikan Indonesia, khususnya madrasah dan sekolah berbasis agama Islam. Implikasi teoretis dan praktis dari temuan internasional tersebut belum banyak dieksplorasi dalam konteks lokal yang

memiliki kekhasan tersendiri, terutama dalam kaitannya dengan pembelajaran PAI yang multidimensi.

Ketiga, kajian yang melakukan sintesis literatur secara komprehensif terhadap ketiga jenis evaluasi—dengan menelusuri persamaan, perbedaan, dan ketegangan antartemuan dari berbagai sumber—masih relatif terbatas dalam literatur pendidikan berbahasa Indonesia. Artikel ini hadir untuk mengisi kesenjangan tersebut dengan menyajikan analisis integratif yang dapat menjadi rujukan bagi akademisi, praktisi, dan pengambil kebijakan pendidikan, khususnya dalam konteks pendidikan Islam di Indonesia.

1.3. Tujuan penelitian

Penelitian ini bertujuan untuk: (1) menganalisis peran evaluasi diagnostik dalam mengidentifikasi kebutuhan belajar siswa; (2) mengkaji peran evaluasi formatif dalam memperbaiki proses pembelajaran secara berkelanjutan; (3) mendeskripsikan peran evaluasi sumatif dalam mengukur pencapaian kompetensi siswa; dan (4) menjelaskan kontribusi integratif ketiga jenis evaluasi terhadap peningkatan kualitas proses belajar siswa berdasarkan sintesis literatur ilmiah.

1.4. Pertanyaan penelitian

Berdasarkan tujuan di atas, pertanyaan penelitian dirumuskan sebagai berikut: (1) Bagaimana evaluasi diagnostik berperan dalam mengidentifikasi kebutuhan belajar dan kesenjangan pemahaman awal siswa? (2) Bagaimana evaluasi formatif berfungsi sebagai instrumen perbaikan proses pembelajaran yang berkelanjutan? (3) Bagaimana evaluasi sumatif mengukur pencapaian kompetensi dan mendukung akuntabilitas pendidikan? (4) Bagaimana ketiga jenis evaluasi dapat bekerja secara integratif dan sinergis untuk meningkatkan kualitas proses belajar siswa?

1.5. Signifikansi penelitian

Secara teoretis, artikel ini berkontribusi pada pengembangan kerangka konseptual evaluasi pembelajaran yang integratif, dengan menjembatani literatur tentang evaluasi diagnostik, formatif, dan sumatif dalam satu narasi analisis yang kohesif. Kerangka ini dapat menjadi landasan bagi penelitian-penelitian selanjutnya yang ingin mengkaji implementasi sistem evaluasi terpadu di berbagai jenjang dan konteks pendidikan.

Secara praktis, artikel ini memberikan sintesis pengetahuan yang dapat menjadi acuan bagi guru PAI dan pendidik Islam dalam merancang sistem evaluasi yang lebih komprehensif dan berorientasi pada perbaikan mutu pembelajaran. Bagi kepala madrasah dan sekolah Islam, temuan ini dapat menjadi dasar pengembangan kebijakan evaluasi internal yang lebih bermakna. Bagi pengambil kebijakan, artikel ini menawarkan perspektif konseptual yang relevan dalam konteks implementasi Kurikulum Merdeka dan reformasi asesmen nasional.

2. Metodologi

2.1. Jenis Penelitian

Penelitian ini menggunakan metode Library Research (penelitian kepustakaan), yaitu pendekatan penelitian yang menjadikan bahan-bahan tertulis berupa buku, jurnal ilmiah, prosiding, dan dokumen resmi sebagai sumber data utama (Zed, 2008). Metode ini dipilih karena tujuan penelitian adalah menghasilkan sintesis pengetahuan yang komprehensif berdasarkan kajian terhadap literatur ilmiah yang telah tersedia, bukan untuk mengumpulkan data baru dari pengalaman empiris di lapangan. Pendekatan ini lazim digunakan dalam riset-riset tinjauan (review research) yang bertujuan memetakan, mensintesis, dan mengkritisi capaian pengetahuan dalam suatu bidang (Snyder, 2019).

Penelitian kepustakaan dalam kajian ini tidak sekadar merangkum literatur, melainkan melakukan analisis kritis dan sintesis interpretatif terhadap temuan-temuan dari berbagai sumber. Melalui proses ini, dihasilkan pemahaman baru yang melampaui apa yang tersedia di sumber-sumber individual, yakni sebuah kerangka integratif tentang peran evaluasi pembelajaran dalam meningkatkan kualitas proses belajar siswa.

2.2. Sumber Data

Sumber data primer dalam penelitian ini adalah artikel-artikel jurnal ilmiah bereputasi yang secara langsung membahas evaluasi diagnostik, evaluasi formatif, evaluasi sumatif, atau hubungan antara praktik evaluasi dan kualitas pembelajaran. Jurnal-jurnal yang menjadi sumber utama antara lain *Assessment in Education: Principles, Policy & Practice*, *Educational Measurement: Issues and Practice*, *Review of Educational Research*, *Educational Researcher*, dan *Practical Assessment, Research, and Evaluation*. Prioritas diberikan kepada artikel yang diterbitkan dalam lima tahun terakhir, dengan pengecualian untuk karya-karya klasik yang memiliki dampak besar terhadap perkembangan teori evaluasi pendidikan.

Sumber data sekunder terdiri dari buku teks dan monograf dalam bidang evaluasi pendidikan, asesmen kelas, dan psikologi pembelajaran, serta dokumen kebijakan pendidikan nasional yang relevan, termasuk panduan asesmen dari Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia. Seluruh sumber ditelusuri melalui basis data akademik Google Scholar, ERIC, dan Scopus menggunakan kata kunci yang telah ditentukan.

2.3. Teknik Pengumpulan Data

Teknik pengumpulan data yang digunakan adalah dokumentasi sistematis, yaitu proses mengidentifikasi, mencatat, dan mengorganisasikan informasi relevan dari berbagai sumber literatur. Penelusuran literatur dilakukan menggunakan kata kunci dalam dua bahasa: dalam bahasa Inggris digunakan kata kunci 'diagnostic assessment', 'formative assessment', 'summative assessment', 'assessment for learning', 'feedback in education', dan 'quality of learning'; dalam bahasa Indonesia digunakan 'evaluasi diagnostik', 'evaluasi formatif', 'evaluasi sumatif', 'asesmen pembelajaran', dan 'kualitas proses belajar'.

Kriteria inklusi yang diterapkan meliputi: (a) literatur yang secara eksplisit membahas salah satu atau lebih jenis evaluasi yang dikaji; (b) diterbitkan dalam jurnal akademik atau oleh penerbit akademik terpercaya; (c) memiliki relevansi konseptual atau empiris yang jelas terhadap pertanyaan penelitian; dan (d) dapat diverifikasi sumber dan penulisnya. Literatur yang tidak memenuhi kriteria ini dikeluarkan dari proses analisis.

2.4. Analisis data

Analisis data dilakukan menggunakan dua teknik yang saling melengkapi: analisis isi (content analysis) dan analisis tematik. Analisis isi diterapkan untuk mengidentifikasi dan mengkategorisasikan informasi relevan dari setiap sumber secara sistematis (Krippendorff, 2018). Analisis tematik digunakan untuk menemukan pola, tema, dan kesimpulan yang muncul secara konsisten atau yang memperlihatkan ketegangan produktif di antara berbagai sumber (Braun & Clarke, 2006).

Proses analisis berlangsung dalam lima tahap. Tahap pertama (pengumpulan literatur): seluruh literatur yang relevan dikumpulkan berdasarkan kriteria inklusi. Tahap kedua (seleksi literatur): literatur disaring berdasarkan kualitas, relevansi, dan signifikansi terhadap pertanyaan penelitian. Tahap ketiga (kategorisasi): literatur dikelompokkan ke dalam empat kategori tematik utama sesuai pertanyaan penelitian. Tahap keempat (analisis tematik): setiap kategori dianalisis secara mendalam untuk mengungkap pola, persamaan, perbedaan, dan hubungan antarkonsep. Tahap kelima

(sintesis): seluruh hasil analisis diintegrasikan menjadi narasi yang kohesif dan menjawab pertanyaan penelitian secara komprehensif.

3. Pembahasan

3.1. Peran Evaluasi Diagnostik dalam Mengidentifikasi Kebutuhan Belajar Siswa

Evaluasi diagnostik adalah proses terencana untuk mengumpulkan informasi tentang pengetahuan awal, pemahaman, dan keterampilan siswa sebelum atau pada awal sebuah unit pembelajaran. Tujuan pokoknya bukan menghasilkan nilai, melainkan memperoleh gambaran tentang apa yang sudah dan belum dikuasai siswa sebagai pijakan perencanaan instruksi yang responsif (Heritage, 2010). Sebuah tinjauan sistematis terbaru yang memanfaatkan data bibliometric dari scopus menemukan bahwa perkembangan kajian tentang *cognitive diagnostic assessment* mengalami akselerasi signifikansi sejak tahun 2015, dengan fokus utama beralih dari sekadar mengidentifikasi atribut kognitif menuju pengembangan model yang mampu menghasilkan umpan balik diagnostic yang dipersonalisasi secara real-time (Herliana, 2026). Temuan ini menegaskan bahwa bidang evaluasi diagnostik bukan hanya berkembang, bertransformasi secara fundamental dalam cara ia dioperasionalkan di dalam kelas.

Kajian literatur memperlihatkan setidaknya dua fungsi utama evaluasi diagnostik. Fungsi pertama adalah pemetaan pengetahuan awal. Ketika guru mengetahui skema kognitif yang sudah dimiliki siswa, mereka dapat merancang pembelajaran yang menghubungkan pengetahuan baru dengan pengetahuan lama secara bermakna. (Tomlinson & Moon, 2013) menegaskan bahwa tanpa pemetaan ini, diferensiasi instruksi tidak dapat dilakukan secara tepat sasaran guru hanya akan berjalan berdasarkan asumsi, bukan berdasarkan bukti tentang kebutuhan nyata setiap siswa. Penelitian terkini memperkuat pandangan ini dengan menunjukkan bahwa guru yang menggunakan data diagnostic secara sistematis dalam perencanaan pembelajaran menghasilkan desain instruksional yang lebih responsive terhadap keragaman kebutuhan belajar siswa, termasuk dalam konteks implementasi Kurikulum Merdeka di Indonesia (Iskak et al., 2023; Badrun Kholid et al., 2024).

Fungsi kedua adalah identifikasi miskonsepsi. Miskonsepsi adalah pemahaman yang secara konseptual keliru namun diyakini benar oleh siswa, bersifat persisten dan resisten terhadap instruksi konvensional. (Black & Wiliam, 1998) memperluas konsep evaluasi diagnostik melampaui penggunaan di awal pembelajaran informasi diagnostic yang dikumpulkan secara berkelanjutan selama proses pembelajaran melalui pertanyaan kelas, diskusi, dan observasi guru sama pentingnya dengan assesmen diagnostik formal di awal unit. Gagasan ini mendapat elaborasi lebih lanjut dalam kerangka *cognitive diagnostic assessment* yang menggabungkan teori respons butir dengan model atribut kognitif untuk menghasilkan profil diagnostik individual yang spesifik dan dapat ditindaklanjuti (Herliana, 2026). Implikasi kritis yang perlu mendapat perhatian adalah temuan bahwa banyak guru mengumpulkan informasi diagnostik tetapi tidak secara efektif memanfaatkannya untuk memodifikasi instruksi pengembangan profesionalisme guru dalam literasi data penilaian karenanya menjadi kebutuhan yang mendesak (Black & Wiliam, 1998; McMillan, 2014).

Tabel 1. Sintesis Literatur tentang Peran Evaluasi Pembelajaran

Penulis & Tahun	Jenis Evaluasi	Fokus Kajian	Temuan Pokok
Heritage (2010)	Diagnostik	Formative Assesment dalam kelas	Evaluasi diagnostik yang menjadi bagian rutinitas kelas membantu guru memahami kebutuhan belajar siswa

			secara berkelanjutan, bukan hanya di awal unit.
Black & Wiliam (1998,2021)	Diagnostik & Formatif	Asesmen dalam kelas dan perbaikan pembelajaran	Diagnostik informal melalui interaksi harian sama pentingnya dengan instrumen formal; pembaruan 2021 memperkuat argument tentang hubungan guru-siswa yang lebih holistik.
Herliana et al. (2026)	Diagnostik Kognitif	Tren global asesmen diagnostik kognitif	Model asesmen diagnostik kognitif berkembang menuju profil diagnostik yang dipersonalisasi dan dapat ditindaklanjuti secara real-time.
McMillan (2022)	Diagnostik & Sumatif	Asesmen kelas dan motivasi siswa	Guru membutuhkan pelatihan profesional berkelanjutan untuk mengoptimalkan penggunaan data penilaian dalam perencanaan pembelajaran.

Sumber: Diolah dari berbagai literatur oleh penulis, 2025.

3.2. Peran Evaluasi Formatif dalam Memperbaiki Proses Pembelajaran

Evaluasi formatif mencakup berbagai proses yang digunakan oleh guru dan siswa untuk mengenali dan merespons perkembangan belajar dalam rangka memperbaiki dan memajukan pembelajaran (Cowie & Bell, 1999). Karakter utama yang membedakan evaluasi formatif dari jenis evaluasi lainnya bukan terletak pada bentuk instrumennya, melainkan pada fungsi penggunaannya: apakah informasi yang dihasilkan segera dimanfaatkan untuk perbaikan, atau hanya disimpan sebagai catatan akhir.

(Hattie & Timperley, 2007) menunjukkan bahwa umpan balik komponen inti evaluasi formatif merupakan salah satu faktor paling berpengaruh terhadap prestasi belajar siswa. Umpan balik yang beroperasi pada level proses dan regulasi diri terbukti jauh lebih efektif daripada umpan balik yang bersifat evaluatif atau personal. Temuan ini mendapat konfirmasi dan perluasan dari meta-analisis terbaru: (Foster, 2024) menemukan bahwa efektivitas evaluasi formatif sangat ditentukan oleh tiga faktor utama yang saling berinteraksi kualitas umpan balik, keterlibatan aktif siswa dalam proses evaluatif, dan kompetensi guru dalam menginterpretasikan serta merespons data formatif secara tepat waktu. Lebih lanjut, (Rioux, 2023) melalui kajian terhadap praktik asesmen formatif guru di berbagai konteks menemukan bahwa variasi dalam cara guru memberikan umpan balik—apakah bersifat dialogis atau direktif berkontribusi signifikan terhadap perbedaan kualitas regulasi diri siswa.

(Black & Wiliam, 1998) dalam tinjauan mereka atas ratusan studi tentang asesmen kelas menyimpulkan bahwa evaluasi formatif yang diimplementasikan dengan baik secara konsisten meningkatkan prestasi belajar siswa, dengan keuntungan yang lebih besar dirasakan oleh siswa berprestasi rendah. Temuan ini secara implisit menyatakan bahwa evaluasi formatif bukan sekadar alat peningkatan mutu, melainkan juga instrumen keadilan pendidikan—ia memberi peluang lebih besar kepada siswa yang secara tradisional kurang berhasil dalam sistem evaluasi konvensional. (Wiliam, 2011) merumuskan lima strategi inti evaluasi formatif yang menjadi kerangka kerja paling berpengaruh di bidang ini: (a) mengklarifikasi tujuan pembelajaran dan kriteria keberhasilan; (b) merancang aktivitas yang menghasilkan bukti pembelajaran; (c) memberikan umpan balik yang dapat memajukan pembelajaran; (d) mengaktifkan siswa sebagai sumber belajar bagi satu sama lain; dan (e) mengaktifkan siswa sebagai pemilik dan pengelola proses belajar mereka sendiri. Lima

strategi ini saling berkaitan dan membentuk sebuah ekosistem evaluasi formatif yang komprehensif.

Dari perspektif motivasi belajar, evaluasi formatif yang dirancang dengan baik meningkatkan motivasi intrinsik dan kepercayaan diri siswa (Stiggins, 2007). Kaitannya dengan Self-Determination Theory (Deci & Ryan, 2000), evaluasi formatif yang efektif memenuhi tiga kebutuhan psikologis dasar yang mendorong motivasi intrinsik: kebutuhan kompetensi, kebutuhan otonomi, dan kebutuhan keterkaitan. Temuan terbaru memperkuat dimensi regulasi diri ini: penelitian (Fang & Xu, 2023) membuktikan bahwa umpan balik yang terdiferensiasi disesuaikan dengan level kemampuan dan gaya belajar siswa menghasilkan peningkatan regulasi diri yang lebih substansial dibanding umpan balik yang seragam. Dalam konteks pembelajaran PAI, nilai-nilai Islam tentang muhasabah (introspeksi diri), islah (perbaikan diri), dan tawadu (kerendahan hati dalam belajar) sejatinya selaras dengan prinsip-prinsip evaluasi formatif yang efektif.

3.3. Peran Evaluasi Sumatif dalam Mengukur Pencapaian Kompetensi

Evaluasi sumatif adalah penilaian yang dilakukan pada akhir sebuah periode pembelajaran—unit, semester, atau tahun ajaran—dengan tujuan utama membuat keputusan tentang status pencapaian siswa (Harlen, 2006). Evaluasi ini menghasilkan ringkasan tentang seberapa jauh siswa telah mencapai tujuan pembelajaran yang ditetapkan, dan informasinya digunakan untuk berbagai keperluan akuntabilitas: pemberian nilai, kenaikan kelas, sertifikasi, dan pelaporan kepada orang tua.

(Harlen, 2006) membuat argumen yang menarik bahwa evaluasi sumatif yang dirancang dengan baik tidak harus berseberangan dengan tujuan pembelajaran yang bermakna data sumatif sesungguhnya berfungsi sebagai titik tolak diagnostik untuk siklus pembelajaran berikutnya. Pandangan ini mendapat dukungan empiris terbaru dari kajian tentang pemanfaatan data asesmen untuk perbaikan mutu pendidikan. Penelitian terkini menegaskan bahwa evaluasi sumatif baru bermakna secara pedagogis ketika datanya digunakan untuk memperbaiki instruksi, bukan sekadar meranking siswa (Popham, 2008); temuan ini mendapat perluasan dalam konteks keberagaman budaya kritik terhadap evaluasi sumatif standar yang mengabaikan representasi budaya semakin menguat, dengan tuntutan yang jelas untuk asesmen yang lebih responsif terhadap latar belakang, keyakinan, dan pengalaman belajar siswa yang beragam (Walker et al., 2023).

(McMillan, 2014) secara kritis mengidentifikasi kelemahan umum dalam praktik evaluasi sumatif di kelas: nilai akhir sering kali mencerminkan campuran antara capaian akademis dan faktor-faktor non-akademis seperti kehadiran, perilaku, dan usaha pencampuran ini menghasilkan informasi yang tidak valid tentang kompetensi siswa yang sesungguhnya. McMillan menganjurkan pemisahan yang jelas antara penilaian kompetensi akademis dengan penilaian karakter dan perilaku. Dalam konteks pendidikan Islam, hal ini menjadi lebih kompleks karena pembelajaran PAI memang secara inheren mencakup dimensi karakter dan akhlak; tantangannya adalah merancang sistem evaluasi yang mampu mengukur kedua dimensi ini secara adil dan terpisah.

3.4. Kontribusi Integratif Evaluasi Pembelajaran terhadap Kualitas Proses Belajar Siswa

Sintesis literatur memperlihatkan bahwa dampak evaluasi terhadap kualitas proses belajar mencapai titik optimalnya ketika ketiga jenis evaluasi—diagnostik, formatif, dan sumatif—diintegrasikan dalam satu sistem yang kohesif dan saling menguatkan. (Wiggins & McTighe, 2005) melalui kerangka Understanding by Design (UbD) menyediakan arsitektur konseptual untuk integrasi ini: evaluasi sumatif mendefinisikan tujuan akhir yang hendak dicapai, evaluasi diagnostik memetakan posisi awal siswa, dan evaluasi formatif memandu perjalanan dari titik awal menuju tujuan.

Dimensi keterlibatan siswa dalam proses evaluasi mendapat perhatian yang semakin besar dalam literatur kontemporer. (Yan et al., 2022) melalui meta-analisis terhadap 57 studi menemukan bahwa intervensi asesmen diri dan asesmen sejawat secara konsisten meningkatkan kinerja akademis

siswa, dengan efek yang lebih kuat pada keterampilan berpikir tingkat tinggi. Meta-analisis susulannya (Yan et al., 2022) lebih lanjut menegaskan bahwa eksplisitnya instruksi tentang cara melakukan asesmen diri bukan sekadar memberikan kesempatan untuk melakukannya merupakan kunci efektivitasnya. Temuan ini memperkuat argumentasi (Weeden et al., 2002) tentang pentingnya self-assessment dan peer assessment yang terbimbing untuk pengembangan keterampilan metakognitif.

Dimensi teknologi dalam regulasi diri semakin menjadi perhatian dalam konteks evaluasi integratif. (Junaštíková, 2024) dalam ulasan sistematisnya menemukan bahwa pemanfaatan teknologi untuk mendukung regulasi diri belajar telah berkembang pesat, termasuk melalui implementasi kecerdasan buatan untuk mengukur dan mendukung regulasi diri menggunakan data jejak aktivitas belajar (*trace data*) secara real-time. Implikasi bagi guru adalah urgensi pengembangan kompetensi literasi data penilaian agar potensi teknologi dalam mendukung siklus evaluasi integratif dapat dimanfaatkan secara optimal.

(Gardner, 2011)mengingatkan bahwa sistem evaluasi yang homogen secara struktural menguntungkan siswa dengan kecerdasan linguistik-logis dan merugikan siswa dengan kecerdasan lain yang sama-sama valid. Temuan terbaru memperkuat perspektif ini melalui konteks diferensiasi instruksi: pembelajaran yang berdiferensiasi berbasis kecerdasan majemuk terbukti menciptakan pengalaman belajar yang lebih bermakna dan menyenangkan bagi semua kelompok siswa (Anggoro, 2024). Dalam konteks pendidikan Islam, integrasi ketiga jenis evaluasi sejalan dengan konsep tafakkur (perenungan mendalam), tadabur (refleksi sistematis), dan tazkiyah (pemurnian diri melalui proses belajar) evaluasi diagnostik mencerminkan proses mengenali diri (*ma'rifat al-nafs*), evaluasi formatif mencerminkan proses perbaikan diri yang berkelanjutan (*islah al-nafs*), dan evaluasi sumatif mencerminkan proses pertanggungjawaban atas capaian (*hisab*).

Tabel 2. Sintesis Literatur tentang Kontribusi Integratif Evaluasi terhadap Kualitas Proses Belajar (Diperbarui)

Penulis & Tahun	Aspek yang Dikaji	Kontribusi terhadap Kualitas Proses Belajar
Wiggins & McTighe (2005)	Backward design: keselarasan tujuan, asesmen, dan instruksi	Evaluasi (diagnostik–formatif–sumatif) yang dirancang sebelum instruksi meningkatkan keselarasan antara tujuan dan capaian secara signifikan.
Black & Wiliam (1998)	Asesmen dalam kelas dan peningkatan pembelajaran	Evaluasi formatif meningkatkan prestasi belajar, dengan manfaat lebih besar bagi siswa berprestasi rendah; efek ekuitas pendidikan yang signifikan.
Foster (2024)	Meta-analisis dampak asesmen formatif pada hasil belajar	Kualitas umpan balik, keterlibatan siswa, dan kompetensi guru adalah tiga faktor paling determinan dalam efektivitas evaluasi formatif.
Yan, Lao, Panadero et al. (2022)	Meta-analisis intervensi self-assessment dan peer-assessment	Asesmen diri dan sejawat secara konsisten meningkatkan kinerja akademis, terutama pada keterampilan berpikir tingkat tinggi.
Yan, Wang, Boud et al. (2023)	Efek self-assessment pada kinerja akademis dan peran eksplisitnya	Instruksi eksplisit tentang cara melakukan asesmen diri adalah kunci efektivitasnya—bukan sekadar memberi kesempatan melakukannya.

Penulis & Tahun	Aspek yang Dikaji	Kontribusi terhadap Kualitas Proses Belajar
Fang & Xu (2023)	Self-regulated learning melalui umpan balik diferensiasi	Umpan balik yang terdiferensiasi sesuai kemampuan siswa menghasilkan peningkatan regulasi diri yang lebih substansial dibanding umpan balik seragam.
Junaštíková (2024)	Regulasi diri dalam konteks teknologi modern	AI dan teknologi pendidikan dapat mengukur dan mendukung regulasi diri menggunakan trace data real-time, memperluas kapasitas evaluasi formatif.
Anggoro et al. (2024)	Diferensiasi berbasis kecerdasan majemuk	Evaluasi yang mengakomodasi kecerdasan majemuk menciptakan pengalaman belajar yang lebih bermakna dan inklusif bagi semua kelompok siswa.

Sumber: Diolah dari berbagai literatur oleh penulis, 2025.

4. Pembahasan

4.1 Evaluasi Diagnostik dan Identifikasi Kebutuhan Belajar

Temuan dari sintesis literatur tentang evaluasi diagnostik menyingkap sebuah ketegangan konseptual yang produktif: di satu sisi, tradisi psikometrik menempatkan evaluasi diagnostik sebagai tes terstandar yang dilaksanakan sebelum pembelajaran; di sisi lain, tradisi asesmen kelas kontemporer memandangnya sebagai proses yang berkelanjutan dan mengalir sepanjang interaksi pembelajaran (Heritage, 2010; Black & Wiliam, 1998). Ketegangan ini sesungguhnya tidak perlu diselesaikan dengan memilih salah satunya, melainkan dengan memahami keduanya sebagai moda diagnostik yang saling melengkapi pada level yang berbeda.

Tes diagnostik terstandar memberikan peta pemahaman yang sistematis dan komprehensif, khususnya berguna saat guru menghadapi kelompok siswa baru yang karakteristiknya belum dikenal. Sementara itu, teknik diagnostik informal—pertanyaan kelas yang dirancang dengan cermat, diskusi terbuka, dan observasi aktif—memberikan informasi yang lebih dinamis dan kontekstual yang tidak dapat ditangkap oleh instrumen terstandar. (Tomlinson & Moon, 2013) menegaskan bahwa kombinasi keduanya menghasilkan pemahaman diagnostik yang paling kaya tentang kebutuhan belajar siswa.

Dari perspektif teori belajar konstruktivisme, evaluasi diagnostik yang efektif memungkinkan guru untuk bekerja dalam zona perkembangan proksimal setiap siswa sebagaimana dikonseptualisasikan oleh Vygotsky. Ketika instruksi dirancang persis di batas antara apa yang sudah dikuasai dan apa yang belum, dengan scaffolding yang tepat, pembelajaran menjadi paling bermakna dan efisien. Ini adalah argumen teoretis yang paling kuat untuk investasi waktu dan energi dalam evaluasi diagnostik.

Implikasi kritis yang perlu mendapat perhatian adalah temuan (Black & Wiliam, 1998) bahwa banyak guru mengumpulkan informasi diagnostik tetapi tidak secara efektif memanfaatkannya untuk memodifikasi instruksi. Kesenjangan antara pengumpulan data dan penggunaannya ini mengindikasikan bahwa masalah utama bukan pada kualitas instrumen diagnostik, melainkan pada kemampuan guru dalam menginterpretasikan data dan membuat keputusan instruksional berdasarkan data tersebut. Pengembangan profesional guru dalam literasi data penilaian karenanya menjadi kebutuhan yang mendesak.

4.2 Evaluasi Formatif sebagai Alat Perbaikan Pembelajaran

Posisi sentral evaluasi formatif dalam literatur asesmen kontemporer telah mapan sejak tinjauan landmark (Black & Wiliam, 1998). Namun, meta-analisis (Kingston & Nash, 2011) yang lebih ketat menemukan efek yang lebih moderat dibanding klaim awal. Perbedaan ini bukan melemahkan argumen untuk evaluasi formatif, melainkan mempertegas bahwa efektivitasnya sangat bergantung pada kualitas implementasi. Evaluasi formatif yang dilaksanakan secara mekanis—sekadar menambahkan kuis di tengah pembelajaran tanpa mengubah cara instruksi berdasarkan hasilnya—tidak akan menghasilkan dampak yang bermakna.

Model umpan balik (Hattie & Timperley, 2007) menyediakan kerangka yang sangat berguna untuk memahami mengapa tidak semua umpan balik formatif memiliki efek yang sama. Umpan balik pada level tugas (apakah jawaban benar atau salah) memiliki efek yang terbatas. Umpan balik pada level proses (bagaimana cara berpikir yang lebih efektif untuk menyelesaikan tugas) memiliki efek yang lebih kuat. Umpan balik pada level regulasi diri (bagaimana siswa dapat memantau dan mengelola belajarnya sendiri) memiliki efek yang paling kuat dan tahan lama. Model ini memberi arah yang jelas bagi guru tentang bagaimana meningkatkan kualitas umpan balik yang mereka berikan.

Kaitannya dengan Self-Determination Theory (Deci & Ryan, 2000), evaluasi formatif yang efektif memenuhi tiga kebutuhan psikologis dasar yang mendorong motivasi intrinsik: kebutuhan kompetensi (terpenuhi melalui umpan balik yang membantu siswa merasa semakin mampu), kebutuhan otonomi (terpenuhi melalui keterlibatan siswa dalam self-assessment dan penetapan tujuan belajar), dan kebutuhan keterkaitan (terpenuhi melalui interaksi evaluatif yang konstruktif antara guru dan siswa). Ketika evaluasi formatif dirancang untuk memenuhi ketiga kebutuhan ini, ia bukan hanya meningkatkan prestasi akademis, tetapi juga membangun fondasi motivasi belajar yang intrinsik dan berkelanjutan.

Dalam konteks pembelajaran PAI secara khusus, evaluasi formatif memiliki dimensi spiritual yang tidak dapat diabaikan. Nilai-nilai Islam tentang muhasabah (introspeksi diri), islah (perbaikan diri), dan tawadu (kerendahan hati dalam belajar) sejatinya selaras dengan prinsip-prinsip evaluasi formatif yang efektif. Self-assessment dan refleksi berkelanjutan atas proses belajar merupakan praktik yang tidak hanya bernilai pedagogis, tetapi juga bernilai spiritual dalam tradisi keilmuan Islam.

4.3 Evaluasi Sumatif sebagai Pengukur Capaian Belajar

Evaluasi sumatif kerap mendapat stigma negatif dalam diskusi reformasi evaluasi sebagai penyebab 'teaching to the test' dan kecemasan akademis yang kontraproduktif. Namun, sintesis literatur ini menunjukkan bahwa masalah sesungguhnya bukan pada evaluasi sumatif itu sendiri, melainkan pada bagaimana ia dirancang, dikomunikasikan, dan digunakan.

(Harlen, 2006) membuat distinsi yang penting antara evaluasi sumatif yang berfungsi sebagai 'asesmen terhadap standar' dan yang berfungsi sebagai 'asesmen untuk akuntabilitas eksternal'. Ketika evaluasi sumatif dirancang dan digunakan dengan bijak oleh guru sebagai alat untuk memahami capaian siswa secara menyeluruh—dan kemudian direfleksikan untuk perbaikan unit pembelajaran berikutnya—ia tidak bertentangan dengan tujuan formatif. Kesalahpahaman ini perlu diluruskan dalam program pengembangan profesional guru.

(Popham, 2008) memberikan kontribusi konseptual yang sangat relevan dengan konteks Indonesia melalui konsep 'asesmen transformatif'. Asesmen Nasional yang saat ini diterapkan oleh (Kemendikbudristek, 2022) dirancang berdasarkan prinsip serupa: data AN seharusnya digunakan oleh sekolah sebagai cermin mutu pembelajaran, bukan sebagai alat seleksi. Namun, efektivitas

prinsip ini bergantung pada kemampuan guru dan kepala sekolah dalam menginterpretasikan dan memanfaatkan data AN untuk perbaikan praktik pembelajaran yang nyata.

(McMillan, 2014) memberikan kritik konstruktif tentang praktik penilaian yang mencampur capaian akademis dengan faktor non-akademis dalam nilai akhir. Dalam konteks pendidikan Islam, hal ini menjadi lebih kompleks karena pembelajaran PAI memang secara inheren mencakup dimensi karakter dan akhlak yang tidak dapat dipisahkan dari dimensi kognitif. Tantangannya adalah merancang sistem evaluasi yang mampu mengukur kedua dimensi ini secara adil dan terpisah, sehingga informasi yang dihasilkan tentang kompetensi akademis tidak terdistorsi oleh penilaian karakter, dan sebaliknya.

4.4 Evaluasi Pembelajaran dan Peningkatan Kualitas Proses Belajar Siswa

Pembahasan ketiga subbab sebelumnya mengarah pada satu kesimpulan yang kokoh: kualitas proses belajar siswa meningkat secara optimal ketika evaluasi diagnostik, formatif, dan sumatif tidak dipraktikkan sebagai entitas yang terpisah, melainkan sebagai komponen yang saling terhubung dalam satu ekosistem evaluasi yang dirancang secara sadar dan kohesif.

Kerangka UbD dari (Wiggins & McTighe, 2005) menyediakan arsitektur yang paling komprehensif untuk integrasi ini. Prinsip 'mulai dari akhir'—menentukan tujuan dan bukti penilaian sebelum merancang aktivitas pembelajaran—memastikan keselarasan antara apa yang hendak dicapai, bagaimana pencapaian itu diukur, dan aktivitas apa yang dirancang untuk mencapainya. Ketika keselarasan ini terwujud, proses belajar siswa menjadi lebih terarah, bermakna, dan produktif.

(Weeden et al., 2002) menyoroti dimensi yang sering terabaikan dalam diskusi evaluasi: keterlibatan siswa. Ketika siswa hanya menjadi objek yang dievaluasi, mereka cenderung mengembangkan orientasi kinerja yang ekstrinsik—belajar untuk mendapatkan nilai, bukan untuk memahami. Namun ketika siswa dilibatkan sebagai agen aktif dalam proses evaluasi—menilai karya sendiri, memberikan umpan balik kepada teman, dan menetapkan tujuan belajar—mereka mengembangkan keterampilan metakognitif dan motivasi intrinsik yang jauh lebih berharga dari sekadar nilai yang baik.

(Gardner, 2011) mengingatkan bahwa sistem evaluasi yang homogen—yang hanya mengakui satu cara mengetahui dan satu cara menunjukkan pemahaman—secara struktural menguntungkan siswa dengan kecerdasan linguistik-logis dan merugikan siswa dengan kecerdasan lain yang sama-sama valid. Evaluasi yang komprehensif dan berkualitas harus mampu mengakomodasi keragaman ini, tidak hanya demi keadilan, tetapi juga demi akurasi penilaian kompetensi yang sesungguhnya.

Dalam konteks pendidikan Islam, integrasi ketiga jenis evaluasi sejalan dengan konsep tafakkur (perenungan mendalam), tadabur (refleksi sistematis), dan tazkiyah (purnian diri melalui proses belajar). Evaluasi diagnostik mencerminkan proses mengenali diri (ma'rifat al-nafs), evaluasi formatif mencerminkan proses perbaikan diri yang berkelanjutan (islah al-nafs), dan evaluasi sumatif mencerminkan proses pertanggungjawaban atas capaian (hisab). Ketika ketiga dimensi ini hadir dalam sistem evaluasi pembelajaran PAI, evaluasi tidak sekadar menjadi alat pedagogis, tetapi juga menjadi wahana pembentukan karakter dan kepribadian Islami yang autentik.

5. Kesimpulan

5.1 Kesimpulan

Berdasarkan sintesis literatur yang telah dilakukan, artikel ini menghasilkan empat kesimpulan utama yang menjawab pertanyaan penelitian.

Pertama, evaluasi diagnostik berperan fundamental dalam mengidentifikasi kebutuhan belajar siswa melalui dua fungsi pokok: pemetaan pengetahuan awal dan identifikasi miskonsepsi. Evaluasi

diagnostik yang efektif bukan hanya dilaksanakan di awal pembelajaran, tetapi juga berlangsung secara informal dan berkelanjutan melalui interaksi sehari-hari di kelas. Kualitas evaluasi diagnostik tidak hanya ditentukan oleh instrumennya, melainkan juga oleh kemampuan guru dalam menginterpretasikan dan memanfaatkan data yang dihasilkan untuk perbaikan instruksi.

Kedua, evaluasi formatif terbukti memiliki kontribusi signifikan terhadap peningkatan proses pembelajaran melalui umpan balik yang efektif, peningkatan motivasi intrinsik, dan pengembangan kemampuan regulasi diri siswa. Efektivitasnya sangat bergantung pada kualitas implementasi: umpan balik yang beroperasi pada level proses dan regulasi diri terbukti lebih efektif dibanding umpan balik yang bersifat evaluatif atau personal. Evaluasi formatif juga berkontribusi pada kesetaraan pendidikan dengan memberikan manfaat yang lebih besar bagi siswa berprestasi rendah.

Ketiga, evaluasi sumatif memiliki peran yang tidak tergantikan dalam mengukur pencapaian kompetensi dan menyediakan informasi akuntabilitas yang valid. Evaluasi sumatif yang dirancang dengan baik tidak bertentangan dengan tujuan formatif—data sumatif dapat berfungsi sebagai titik tolak diagnostik untuk siklus pembelajaran berikutnya. Masalah bukan pada evaluasi sumatif itu sendiri, melainkan pada bagaimana ia dirancang, dikomunikasikan, dan digunakan.

Keempat, dampak optimal terhadap kualitas proses belajar siswa hanya dapat dicapai ketika ketiga jenis evaluasi diintegrasikan dalam satu sistem yang kohesif dan saling menguatkan. Integrasi ini memerlukan keselarasan antara tujuan pembelajaran, proses evaluasi, dan aktivitas instruksional; keterlibatan aktif siswa dalam proses evaluasi; serta keragaman format evaluasi yang mengakomodasi keberagaman cara belajar dan cara mengetahui siswa. Dalam konteks pendidikan Islam, integrasi ini juga memiliki dimensi spiritual yang sejalan dengan nilai-nilai tafakkur, islah, dan hisab dalam tradisi keilmuan Islam.

5.2 Saran

Bagi guru, khususnya guru Pendidikan Agama Islam: disarankan untuk (a) membiasakan penggunaan evaluasi diagnostik informal—melalui pertanyaan terbuka, jurnal refleksi, dan diskusi awal—sebelum memulai unit pembelajaran baru; (b) menggeser fokus umpan balik dari evaluatif (nilai dan peringkat) menuju deskriptif (apa yang sudah dikuasai, apa yang perlu diperbaiki, dan bagaimana cara memperbaikinya); (c) melibatkan siswa secara aktif dalam proses self-assessment dan peer assessment yang terbimbing sebagai bagian dari pembelajaran, bukan sebagai tambahan di luarnya; dan (d) memperlakukan data sumatif sebagai bahan refleksi untuk perbaikan desain pembelajaran ke depan, bukan hanya sebagai angka akhir.

Bagi sekolah dan madrasah: disarankan untuk (a) mengembangkan kebijakan evaluasi internal yang menekankan fungsi formatif dan diagnostik, tidak hanya akuntabilitas sumatif; (b) menyediakan waktu terjadwal untuk kolaborasi guru dalam menganalisis data evaluasi dan merancang respons instruksional bersama; (c) menyelenggarakan program pengembangan profesional yang berfokus pada literasi data penilaian dan teknik umpan balik yang efektif; dan (d) membangun budaya evaluasi yang positif di mana evaluasi dipandang sebagai instrumen pertumbuhan, bukan hukuman.

Bagi pengambil kebijakan pendidikan: disarankan untuk (a) merancang regulasi evaluasi yang mendorong keseimbangan antara fungsi formatif dan sumatif dalam sistem penilaian kelas; (b) memasukkan kompetensi evaluasi formatif dan diagnostik secara eksplisit dalam standar kompetensi guru dan kurikulum program pendidikan profesi guru; (c) mengevaluasi dan memperkuat implementasi asesmen diagnostik dalam Kurikulum Merdeka agar tidak hanya menjadi formalitas administratif; dan (d) mengembangkan platform digital yang membantu guru melaksanakan evaluasi formatif secara efisien tanpa menambah beban administratif yang berlebihan.

Bagi peneliti selanjutnya: disarankan untuk (a) mengembangkan penelitian empiris tentang implementasi sistem evaluasi terpadu dalam konteks madrasah dan sekolah Islam di Indonesia; (b) mengkaji faktor-faktor yang memediasi efektivitas evaluasi formatif dalam konteks pembelajaran PAI yang multidimensi (kognitif, afektif, spiritual); (c) menginvestigasi potensi teknologi artificial intelligence dalam mendukung personalisasi umpan balik diagnostik dan formatif secara real-time; dan (d) mengembangkan instrumen evaluasi holistik yang mampu mengukur kompetensi PAI secara valid dan reliabel, termasuk dimensi akhlak dan kecerdasan spiritual.

Referensi

- Anggoro, S. (2024). Differentiated instruction based on multiple intelligences as promising joyful and meaningful learning. *Journal of Education*. <https://doi.org/10.1155/2024/7853205>
- Badrun Kholid, Arif Rahman, & Lalu Ari Irawan. (2024). Implementing Diagnostic Assessment in Designing Differentiated Learning for English Language Learning at the Junior High Schools. *Journal of Language and Literature Studies*, 4(2), 445–458. <https://doi.org/10.36312/jolls.v4i2.1934>
- Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Cowie, B., & Bell, B. (1999). A Model of Formative Assessment in Science Education. *Assessment in Education: Principles, Policy & Practice*, 6(1), 101–116. <https://doi.org/10.1080/09695949993026>
- Deci, E. L., & Ryan, R. M. (2000). The “What” and “Why” of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry*, 11(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01
- Fang, Z., & Xu, X. (2023). Exploring self-regulated learning through differentiated feedback. *Cogent Education*, 10(1), 2212113. <https://doi.org/10.1080/2331186X.2023.2212113>
- Foster, H. (2024). The impact of formative assessment on student learning outcomes: A meta-analytical review. *Academy of Educational Leadership Journal*, 28(S1), 1–3.
- Gardner, H. (2011). *Multiple Intelligences: New Horizons in Theory and Practice*. Basic Books.
- Harlen, W. (2006). On the Relationship Between Assessment for Formative and Summative Purposes. In *Assessment and Learning* (pp. 103–118). SAGE Publications.
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Heritage, M. (2010). *Formative Assessment: Making It Happen in the Classroom*. Corwin Press.

- Herliana, F. (2026). Cognitive diagnostic assessment in educational contexts: Global trends, methodological innovations, and emerging directions. *Eurasia Journal of Mathematics, Science and Technology Education*, 22(2), em2775. <https://doi.org/10.29333/ejmste/17827>
- Iskak, K. N. N., A.G Thamrin, & Budi Tri Cahyono. (2023). The IMPLEMENTATION OF DIAGNOSTIC ASSESSMENT AS ONE OF THE STEPS TO IMPROVE LEARNING IN THE IMPLEMENTATION OF THE INDEPENDENT CURRICULUM. *JISAE: Journal of Indonesian Student Assessment and Evaluation*, 9(1), 15–25. <https://doi.org/10.21009/jisae.v9i1.32714>
- Junaštíková, J. (2024). Self-regulation of learning in the context of modern technology: A review of empirical studies. *Interactive Technology and Smart Education*, 21(2), 270–291. <https://doi.org/10.1108/ITSE-02-2023-0030>
- Kemendikbudristek. (2022). *Kurikulum Merdeka: Panduan pembelajaran dan asesmen Pendidikan Agama Islam dan Budi Pekerti*. Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi.
- Kingston, N., & Nash, B. (2011). Formative Assessment: A Meta-Analysis and a Call for Research. *Educational Measurement: Issues and Practice*, 30(4), 28–37. <https://doi.org/10.1111/j.1745-3992.2011.00220.x>
- Krippendorff, K. (2018). *Content analysis: An introduction to its methodology*. SAGE Publications.
- McMillan, J. H. (2014). *Classroom Assessment: Principles and Practice for Effective Standards-Based Instruction* (6th ed.). Pearson.
- Mishra, S. (2022). Evaluating learning effectiveness using CIPP model. *Education Sciences*, 12(9), 605. <https://doi.org/10.3390/educsci12090605>
- Popham, W. J. (2008). *Transformative Assessment*. ASCD.
- Rioux, C. (2023). Exploring formative assessment practices and teacher feedback. *Assessment in Education: Principles, Policy & Practice*, 30(1), 1–20. <https://doi.org/10.1080/0969594X.2023.2175038>
- 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>
- Stiggins, R. J. (2007). Assessment for Learning: An Essential Foundation of Productive Instruction. In *Ahead of the Curve: The Power of Assessment to Transform Teaching and Learning* (pp. 59–76). Solution Tree.
- Stufflebeam, D. L., & Shinkfield, A. J. (2007). *Evaluation Theory, Models, and Applications*. Jossey-Bass.
- Tomlinson, C. A., & Moon, T. R. (2013). *Assessment and Student Success in a Differentiated Classroom*. ASCD.

- Walker, M. E., Olivera-Aguilar, M., Lehman, B., Laitusis, C., Guzman-Orth, D., & Gholson, M. (2023). Culturally Responsive Assessment: Provisional Principles. *ETS Research Report Series*, 2023(1), 1–24.
<https://doi.org/10.1002/ets2.12374>
- Weeden, P., Winter, J., & Broadfoot, P. (2002). *Assessment: What's in It for Schools?* RoutledgeFalmer.
- Wiggins, G., & McTighe, J. (2005). *Understanding by Design* (Expanded 2). ASCD.
- Wiliam, D. (2011). *Embedded Formative Assessment*. Solution Tree Press.
- Yan, Z., Lao, H., & Panadero, E. (2022). Effects of self-assessment and peer-assessment interventions on academic performance: A meta-analysis. *Educational Research Review*, 37, 100484.
<https://doi.org/10.1016/j.edurev.2022.100484>
- Zed, M. (2008). *Metode Penelitian Kepustakaan*. Yayasan Obor Indonesia.