



Multidisciplinary: Journal of Education and Learning

Evaluasi Pendidikan: Antara Konsep, Urgensi dan Efektivitasnya dalam Meningkatkan Mutu Pendidikan

Mohamad Toha^{1*}, Munirah Munirah², Siti Mulkiyah³

Affiliation: Pascasarjana IPEBA Cirebon

Email: mtoha9247@gmail.com, munirahbae323@gmail.com, sitimulkiyah95@gmail.com

ABSTRACT

Educational evaluation is a systematic process of collecting, analyzing, and interpreting information to assess the achievement of educational goals and improve the learning process and management of educational units. This article examines the concept of educational evaluation, its urgency in quality governance, and its effectiveness in driving improvements in educational quality. The discussion emphasizes the differences between evaluation and measurement and assessment, the various approaches (formative–summative, internal–external), and the requirements for evaluation to have a real impact: improvement-oriented, based on valid data, stakeholder involvement, and planned follow-up. The article also examines implementation challenges such as bias, achievement reductionism, administrative burden, and a "just report" culture. In conclusion, evaluation is effective when positioned as an organizational learning instrument that guides decisions, not merely an accountability ritual.

Keywords: *educational evaluation, educational quality, effectiveness, quality assurance, formative–summative*

Evaluasi pendidikan merupakan komponen esensial dalam sistem pembelajaran karena berfungsi menilai ketercapaian tujuan, memotret kualitas proses, serta menjadi dasar pengambilan keputusan dan tindak lanjut perbaikan. Dalam praktiknya, evaluasi menyediakan informasi yang diperlukan pendidik dan lembaga untuk menentukan apakah strategi, metode, serta layanan pembelajaran telah berjalan efektif, sekaligus mengarahkan penyesuaian program secara berkelanjutan. Arikunto (2013) menegaskan bahwa melalui evaluasi, guru dan sekolah dapat mengumpulkan data secara menyeluruh sehingga ketercapaian tujuan pembelajaran dapat diketahui. Sejalan dengan itu, Mardapi (2012) memandang evaluasi sebagai proses sistematis dalam menafsirkan informasi hasil pengukuran untuk menentukan tingkat pencapaian tujuan pendidikan dan mendukung penetapan kebijakan pada satuan pendidikan. Dengan demikian, evaluasi pendidikan dapat dipahami sebagai proses terencana untuk mengukur, menafsirkan, dan menilai informasi pembelajaran secara sistematis agar diperoleh gambaran capaian, mutu proses, serta rekomendasi tindak lanjut yang tepat.

Seiring perkembangan teori dan praktik asesmen, paradigma evaluasi modern tidak lagi terbatas pada penilaian hasil akhir (sumatif), tetapi juga menekankan peran evaluasi formatif sebagai penggerak peningkatan pembelajaran. Temuan kajian berpengaruh menunjukkan bahwa asesmen formatif—khususnya umpan balik yang berkualitas dan digunakan untuk memperbaiki proses belajar-mengajar—berkontribusi signifikan terhadap peningkatan capaian belajar peserta didik (Black & Wiliam, 1998). Dalam kerangka "learning culture", asesmen diposisikan sebagai bagian dari proses belajar itu sendiri: membantu peserta didik memahami tujuan belajar, memonitor kemajuan, dan mengembangkan kemampuan refleksi melalui umpan balik dan perbaikan berkelanjutan (Shepard, 2000). Selain itu, evaluasi pada level program juga berkembang melalui pendekatan komprehensif seperti model CIPP (Context, Input, Process, Product) yang menempatkan evaluasi sebagai kerangka pengambilan keputusan untuk menilai konteks kebutuhan, kelayakan masukan/sumber daya, kualitas proses implementasi, serta hasil dan dampak program (Stufflebeam, 2003). Pergeseran ini memperlihatkan bahwa evaluasi tidak hanya berorientasi pada "membuktikan" capaian, tetapi juga pada "memperbaiki" proses dan hasil pendidikan.

Dalam pendidikan Islam, evaluasi memiliki kesamaan dengan evaluasi pendidikan pada umumnya—yakni sama-sama berfungsi menilai ketercapaian tujuan dan meningkatkan kualitas proses—namun memiliki kekhasan pada landasan dan fokus penilaian. Pendidikan Islam memandang tujuan pendidikan tidak berhenti pada penguasaan pengetahuan, tetapi juga pembentukan akhlak dan spiritualitas menuju manusia paripurna (insan

kāmil). Karena itu, evaluasi dalam pendidikan Islam cenderung menekankan ranah afektif dan psikomotorik, termasuk kualitas hubungan personal dengan Tuhan, relasi sosial kemasyarakatan, sikap terhadap diri dan lingkungan, serta kesadaran peran sebagai hamba dan khalifah (Jalaluddin, 1994). Kekhasan tersebut juga berimplikasi pada perbedaan penekanan dalam prinsip dan metode, karena indikator keberhasilan tidak semata-mata dinilai melalui skor kognitif, tetapi juga melalui internalisasi nilai, perilaku, dan pembiasaan yang mencerminkan karakter Islami.

Berdasarkan uraian tersebut, artikel ini mengkaji konsep dasar, urgensi, prinsip, serta paradigma modern evaluasi pendidikan dan relevansinya dalam pendidikan Islam. Kajian ini menegaskan bahwa evaluasi berfungsi tidak hanya untuk mengukur hasil belajar, tetapi juga sebagai instrumen pembinaan yang mendukung pembentukan karakter dan spiritualitas peserta didik secara holistik menuju insan kāmil.

METHOD

Penelitian ini menggunakan pendekatan penelitian kepustakaan (library research), yaitu penelitian yang berfokus pada penelusuran, pemilihan, dan pengkajian sistematis terhadap sumber-sumber tertulis untuk membangun argumentasi ilmiah berbasis literatur. Dalam penelitian kepustakaan, data utama berasal dari bahan pustaka yang relevan dan otoritatif, sehingga peneliti dapat memetakan konsep dan kerangka teoretis secara mendalam (Zed, 2004). Pendekatan ini dipilih karena topik evaluasi pendidikan Islam membutuhkan pembacaan konseptual-filosofis yang komprehensif atas gagasan tokoh dan literatur akademik.

Teknik pengumpulan data dilakukan melalui dokumentasi (documentary method), yaitu pengumpulan data dengan menginventarisasi dan menelaah dokumen yang relevan seperti buku ilmiah, artikel jurnal, prosiding, dokumen kebijakan pendidikan, dan sumber akademik lain. Metode dokumentasi dipandang sesuai untuk penelitian kepustakaan karena memungkinkan peneliti memperoleh data konseptual, historis, dan argumentatif secara mendalam dari teks (Bowen, 2009). Analisis data menggunakan analisis isi (content analysis) dengan pendekatan deskriptif-analitis. Analisis isi digunakan untuk mengidentifikasi, mengodekan, mengkategorikan, dan menafsirkan makna dari teks secara sistematis agar diperoleh pola konsep dan hubungan antar gagasan (Krippendorff, 2018).

HASIL DAN PEMBAHASAN

Konsep Dasar Evaluasi Pendidikan

Evaluasi pendidikan merupakan proses sistematis dan terencana untuk menentukan nilai (value) suatu program, kebijakan, kinerja pembelajaran, maupun capaian peserta didik berdasarkan kriteria tertentu, dengan tujuan utama menghasilkan keputusan yang berdampak pada perbaikan mutu. Dalam literatur kontemporer, evaluasi dipahami bukan sekadar aktivitas teknis, melainkan praktik reflektif berbasis bukti yang mengintegrasikan data kuantitatif dan kualitatif untuk mendukung pengambilan keputusan yang adil dan efektif. Evaluasi pendidikan berfungsi sebagai mekanisme akuntabilitas sekaligus sebagai alat peningkatan mutu berkelanjutan. Secara konseptual, evaluasi menekankan pada penggunaan data yang sah (valid), andal (reliable), serta relevan dengan konteks pembelajaran dan kebijakan yang dievaluasi (Black & Wiliam, 2018; Hattie, 2017; Stufflebeam & Zhang, 2017). Dengan demikian, evaluasi tidak hanya menilai hasil, tetapi juga memeriksa proses dan konteks yang memengaruhi hasil tersebut.

Ruang lingkup evaluasi pendidikan sangat luas dan mencakup berbagai level sistem pendidikan. Pada tingkat mikro, evaluasi pembelajaran berfokus pada strategi mengajar, ketercapaian kompetensi, serta praktik asesmen di kelas. Penelitian menunjukkan bahwa praktik evaluasi yang terintegrasi dalam pembelajaran (assessment for learning) dapat meningkatkan motivasi dan hasil belajar peserta didik secara signifikan (Panadero et al., 2018; Hattie, 2017). Pada tingkat meso, evaluasi program mencakup penilaian kurikulum, pelatihan guru, program literasi, serta intervensi remedial. Evaluasi program bertujuan menilai efektivitas dan efisiensi suatu intervensi pendidikan sebelum diperluas atau direvisi (Darling-Hammond et al., 2017; OECD, 2020). Sementara itu, pada tingkat makro, evaluasi institusi dan kebijakan mencakup manajemen sekolah, budaya mutu, layanan peserta didik, iklim belajar, standar nasional, sistem ujian, pendanaan, hingga kebijakan zonasi. Studi komparatif internasional menunjukkan bahwa sistem evaluasi yang koheren dan berbasis data berkontribusi terhadap peningkatan kinerja sistem pendidikan secara keseluruhan (OECD, 2020; Chapman & Sammons, 2016).

Evaluasi pembelajaran merupakan salah satu komponen penting dalam praktik pendidikan modern. Evaluasi tidak hanya dilakukan pada akhir pembelajaran (summative evaluation), tetapi juga selama proses berlangsung (formative evaluation). Evaluasi formatif berperan sebagai umpan balik berkelanjutan yang membantu guru menyesuaikan strategi pembelajaran sesuai kebutuhan siswa. Bukti empiris menunjukkan bahwa umpan balik yang spesifik dan berorientasi pada tugas memiliki dampak positif terhadap pencapaian akademik

dan regulasi diri peserta didik (Wisniewski et al., 2020; Panadero et al., 2018). Di sisi lain, evaluasi sumatif berfungsi untuk mengukur pencapaian akhir terhadap standar tertentu, seperti kelulusan atau sertifikasi kompetensi. Integrasi kedua pendekatan ini menciptakan sistem evaluasi yang komprehensif dan seimbang antara fungsi pengembangan (improvement) dan fungsi pertanggungjawaban (accountability) (Black & Wiliam, 2018; Hattie, 2017).

Pada tataran program dan institusi, evaluasi berfungsi sebagai instrumen penjaminan mutu (quality assurance). Evaluasi kurikulum, misalnya, bertujuan memastikan bahwa isi, strategi, dan asesmen selaras dengan kebutuhan peserta didik serta perkembangan zaman. Penelitian menunjukkan bahwa evaluasi kurikulum berbasis data dan partisipatif mampu meningkatkan relevansi pembelajaran dan kinerja guru (Darling-Hammond et al., 2017). Selain itu, evaluasi pelatihan guru yang sistematis terbukti meningkatkan kompetensi pedagogik dan hasil belajar siswa (Cordingley et al., 2018). Evaluasi institusi juga mencakup analisis budaya sekolah dan kepemimpinan pendidikan. Sekolah dengan budaya reflektif dan kolaboratif cenderung memiliki capaian akademik yang lebih baik dan lingkungan belajar yang lebih positif (Chapman & Sammons, 2016). Dengan demikian, evaluasi berperan sebagai instrumen transformasi kelembagaan yang berkelanjutan.

Pada level kebijakan, evaluasi pendidikan memiliki fungsi strategis dalam menentukan arah reformasi sistem pendidikan. Evaluasi kebijakan mencakup analisis dampak standar nasional, sistem ujian, mekanisme pendanaan, serta kebijakan pemerataan akses seperti zonasi sekolah. Pendekatan berbasis bukti (evidence-based policy) menjadi paradigma dominan dalam evaluasi kebijakan pendidikan modern (OECD, 2020). Studi lintas negara menunjukkan bahwa kebijakan yang dirancang dan direvisi berdasarkan hasil evaluasi yang sistematis cenderung lebih efektif dan berkelanjutan (Darling-Hammond et al., 2017). Oleh karena itu, evaluasi kebijakan tidak hanya bersifat administratif, tetapi juga menjadi sarana refleksi kritis terhadap kesenjangan dan tantangan sistem pendidikan.

Pemahaman yang jelas mengenai perbedaan antara pengukuran (measurement), penilaian (assessment), dan evaluasi (evaluation) sangat penting dalam praktik pendidikan. Pengukuran merujuk pada proses kuantifikasi karakteristik tertentu melalui instrumen yang terstandar, seperti skor tes, angka kehadiran, atau indeks prestasi. Fokus utama pengukuran adalah menghasilkan data numerik yang objektif dan dapat dibandingkan. Literatur psikometri menekankan bahwa kualitas pengukuran ditentukan oleh validitas dan reliabilitas instrumen yang digunakan (Brown & Abdulnabi, 2017; Kane, 2016). Tanpa validitas yang memadai, angka yang dihasilkan tidak memiliki makna substantif bagi pengambilan keputusan pendidikan.

Penilaian (assessment) memiliki cakupan yang lebih luas dibandingkan pengukuran. Penilaian merupakan proses sistematis dalam mengumpulkan, menganalisis, dan menafsirkan bukti belajar untuk memahami perkembangan peserta didik. Penilaian dapat bersifat kuantitatif maupun kualitatif, seperti portofolio, observasi, proyek, atau refleksi diri. Penelitian menunjukkan bahwa pendekatan penilaian autentik dan berbasis kinerja mampu menggambarkan kompetensi siswa secara lebih komprehensif dibandingkan tes pilihan ganda semata (Panadero et al., 2018; Andrade & Brookhart, 2019). Selain itu, praktik penilaian formatif yang melibatkan siswa dalam refleksi diri terbukti meningkatkan motivasi dan regulasi diri (Wisniewski et al., 2020). Dengan demikian, penilaian tidak hanya berfungsi sebagai alat pengukuran, tetapi juga sebagai strategi pembelajaran itu sendiri.

Sementara itu, evaluasi (evaluation) merupakan proses memberikan pertimbangan nilai terhadap hasil pengukuran dan penilaian dengan membandingkannya pada kriteria atau standar tertentu untuk menghasilkan keputusan. Evaluasi melibatkan interpretasi data serta pertimbangan kontekstual, seperti tujuan program, kebutuhan peserta didik, dan dampak jangka panjang. Model evaluasi modern, seperti pendekatan CIPP (Context, Input, Process, Product), menekankan pentingnya menilai tidak hanya hasil akhir, tetapi juga konteks dan proses pelaksanaan program (Stufflebeam & Zhang, 2017). Dalam praktik kebijakan, evaluasi digunakan untuk menentukan apakah suatu program perlu dilanjutkan, direvisi, dihentikan, atau diperluas (OECD, 2020).

Pembedaan konseptual ini penting agar praktik evaluasi tidak berhenti pada angka atau laporan statistik semata. Tanpa proses evaluasi yang komprehensif, data pengukuran dan penilaian tidak akan menghasilkan rekomendasi yang dapat ditindaklanjuti. Penelitian menunjukkan bahwa penggunaan data secara reflektif dan kolaboratif oleh guru dan pemimpin sekolah berkontribusi pada peningkatan mutu pembelajaran (Chapman & Sammons, 2016; Cordingley et al., 2018). Oleh karena itu, integrasi pengukuran yang valid, penilaian yang bermakna, dan evaluasi yang reflektif menjadi fondasi utama dalam sistem pendidikan yang berorientasi pada perbaikan berkelanjutan.

Efektivitas Evaluasi: Kapan Evaluasi Benar-benar Meningkatkan Mutu?

Evaluasi pendidikan sering dipandang sebagai instrumen utama peningkatan mutu, tetapi berbagai penelitian menunjukkan bahwa evaluasi tidak secara otomatis menghasilkan perbaikan. Evaluasi hanya efektif ketika dirancang dan diimplementasikan sebagai bagian dari siklus perbaikan berkelanjutan yang terintegrasi dengan tujuan pembelajaran dan pengambilan keputusan berbasis bukti. Studi empiris menunjukkan bahwa penggunaan data evaluasi yang reflektif dan kolaboratif di tingkat sekolah berkorelasi dengan peningkatan hasil belajar dan profesionalisme guru (Datnow & Hubbard, 2016; Schildkamp et al., 2019). Namun, evaluasi yang bersifat administratif semata—misalnya hanya untuk pelaporan atau akuntabilitas—cenderung tidak berdampak signifikan terhadap mutu pembelajaran (Hopfenbeck et al., 2015; Brown et al., 2018). Oleh karena itu, efektivitas evaluasi sangat ditentukan oleh kualitas desain, instrumen, penggunaan data, dan tindak lanjutnya dalam praktik nyata di kelas dan institusi pendidikan.

Evaluasi yang efektif harus berakar pada tujuan pembelajaran atau tujuan program yang spesifik, terukur, dan selaras dengan standar kompetensi. Tanpa tujuan yang jelas, evaluasi hanya menjadi pengumpulan data tanpa arah yang bermakna. Penelitian menunjukkan bahwa kejelasan tujuan pembelajaran dan kriteria keberhasilan meningkatkan pemahaman siswa tentang ekspektasi akademik serta memperbaiki capaian belajar mereka (Hattie, 2017; Panadero et al., 2018). Ketika guru mengkomunikasikan tujuan dan kriteria secara eksplisit, siswa lebih mampu melakukan regulasi diri dan refleksi terhadap kemajuan mereka.

Selain itu, evaluasi program pendidikan yang efektif juga menuntut indikator keberhasilan yang terdefinisi dengan baik agar dapat menilai dampak secara objektif. Studi tentang implementasi kebijakan pendidikan menunjukkan bahwa program dengan indikator yang kabur cenderung menghasilkan interpretasi yang subjektif dan inkonsisten antar evaluator (Schildkamp et al., 2019; Brown et al., 2018). Kejelasan kriteria memungkinkan proses evaluasi menjadi transparan, adil, dan dapat dipertanggungjawabkan. Tanpa kriteria yang disepakati, evaluasi berisiko menjadi opini personal atau sekadar kompilasi angka tanpa makna substantif bagi peningkatan mutu. Dengan demikian, penetapan tujuan dan kriteria merupakan fondasi utama efektivitas evaluasi pendidikan.

Efektivitas evaluasi juga bergantung pada kualitas instrumen yang digunakan. Instrumen dikatakan valid apabila benar-benar mengukur konstruk yang dimaksud, bukan aspek lain yang tidak relevan. Validitas menjadi isu sentral dalam asesmen pendidikan karena kesalahan interpretasi dapat berdampak serius terhadap keputusan akademik maupun kebijakan (Kane, 2016; Zumbo, 2017). Misalnya, tes literasi yang hanya mengukur hafalan fakta tidak dapat diklaim sebagai pengukur kemampuan berpikir kritis atau pemahaman mendalam. Oleh sebab itu, pengembangan instrumen harus melalui proses uji validitas isi, konstruk, dan konsekuensi.

Selain valid, instrumen evaluasi harus andal (reliabel), yaitu menghasilkan skor yang konsisten dalam kondisi yang relatif sama. Reliabilitas penting untuk memastikan bahwa perbedaan skor mencerminkan perbedaan kemampuan, bukan kesalahan pengukuran (Brown et al., 2018; Kane, 2016). Instrumen yang tidak reliabel dapat menghasilkan diagnosis yang keliru dan berdampak pada intervensi yang tidak tepat sasaran. Lebih jauh, prinsip keadilan (fairness) dalam asesmen juga menjadi perhatian utama dalam literatur mutakhir. Evaluasi harus mempertimbangkan keragaman latar belakang sosial, budaya, dan bahasa peserta didik agar tidak menimbulkan bias sistemik (AERA, APA, & NCME, 2018; Zumbo, 2017). Instrumen yang adil membantu memastikan bahwa setiap peserta didik memiliki kesempatan yang setara untuk menunjukkan kompetensinya. Dengan demikian, validitas, reliabilitas, dan keadilan merupakan syarat fundamental agar evaluasi benar-benar berkontribusi pada peningkatan mutu.

Mutu pendidikan tidak dapat direpresentasikan hanya melalui skor tes. Evaluasi yang efektif memerlukan pendekatan triangulatif yang menggabungkan berbagai sumber data untuk memperoleh gambaran yang komprehensif. Penelitian tentang data-driven decision making menunjukkan bahwa kombinasi data kuantitatif dan kualitatif meningkatkan akurasi diagnosis serta kualitas keputusan pendidikan (Datnow & Hubbard, 2016; Schildkamp et al., 2019). Data hasil belajar seperti skor tes dan portofolio memberikan informasi tentang capaian kognitif, tetapi belum tentu menjelaskan faktor penyebab di balik capaian tersebut.

Oleh karena itu, evaluasi perlu memadukan data proses seperti observasi kelas, rekaman praktik mengajar, serta jurnal refleksi guru untuk memahami dinamika pembelajaran secara mendalam. Selain itu, data konteks seperti tingkat kehadiran, dukungan keluarga, fasilitas belajar, dan iklim sekolah berpengaruh signifikan terhadap hasil belajar siswa (Hopfenbeck et al., 2015; OECD, 2016). Suara pemangku kepentingan—melalui wawancara atau survei terhadap siswa, guru, dan orang tua—juga memberikan perspektif penting tentang pengalaman belajar dan persepsi terhadap mutu layanan pendidikan. Pendekatan triangulatif memungkinkan evaluator menghindari kesimpulan yang terlalu sederhana serta menghasilkan rekomendasi yang lebih akurat dan

kontekstual. Dengan demikian, integrasi berbagai jenis data menjadi ciri utama evaluasi yang komprehensif dan bermakna.

Evaluasi formatif memiliki dampak paling besar terhadap peningkatan mutu ketika disertai umpan balik yang spesifik, tepat waktu, dan dapat ditindaklanjuti. Meta-analisis menunjukkan bahwa umpan balik yang menjelaskan kesenjangan antara performa aktual dan standar yang diharapkan secara signifikan meningkatkan hasil belajar siswa (Wisniewski et al., 2020; Hattie, 2017). Namun, umpan balik yang bersifat umum atau terlambat cenderung tidak efektif bahkan dapat menurunkan motivasi belajar. Oleh karena itu, kecepatan dan kualitas respons menjadi faktor kunci dalam efektivitas evaluasi formatif.

Selain spesifik dan tepat waktu, umpan balik harus disertai kesempatan perbaikan, seperti remedial, pengayaan, atau revisi tugas. Penelitian menunjukkan bahwa siswa yang diberi kesempatan merevisi pekerjaannya berdasarkan umpan balik menunjukkan peningkatan signifikan dalam pemahaman konsep dan keterampilan metakognitif (Panadero et al., 2018; Andrade & Brookhart, 2020). Evaluasi yang berhenti pada pemberian nilai tanpa tindak lanjut kehilangan potensi transformasinya. Dengan demikian, evaluasi benar-benar meningkatkan mutu ketika menjadi bagian dari siklus pembelajaran yang responsif: mengidentifikasi kebutuhan, memberikan umpan balik bermakna, dan menyediakan ruang perbaikan berkelanjutan.

Tantangan Implementasi Evaluasi Pendidikan

Implementasi evaluasi pendidikan di berbagai sistem sekolah menghadapi beragam tantangan konseptual dan praktis. Meskipun evaluasi dirancang untuk meningkatkan mutu, dalam praktiknya ia dapat menghasilkan konsekuensi yang tidak diinginkan ketika desain, indikator, dan pemanfaatannya tidak tepat. Literatur internasional menunjukkan bahwa tekanan akuntabilitas berbasis tes, beban administratif, bias konteks sosial-ekonomi, serta rendahnya literasi data di kalangan pendidik menjadi hambatan utama efektivitas evaluasi (Hopfenbeck et al., 2015; Datnow & Hubbard, 2016). Tantangan-tantangan ini menegaskan bahwa evaluasi bukan sekadar persoalan teknis, melainkan juga persoalan etis, struktural, dan kapasitas profesional. Tanpa perhatian terhadap faktor-faktor tersebut, evaluasi berisiko berubah dari alat peningkatan mutu menjadi mekanisme kontrol yang kontraproduktif (Schildkamp et al., 2019; Lingard et al., 2016).

Salah satu tantangan terbesar dalam evaluasi pendidikan adalah kecenderungan mereduksi mutu menjadi angka-angka hasil tes. Sistem evaluasi berbasis skor sering kali menyederhanakan kompleksitas proses belajar menjadi indikator kuantitatif tunggal, seperti nilai ujian standar. Penelitian menunjukkan bahwa tekanan terhadap capaian numerik dapat mendorong praktik "teaching to the test", yaitu pengajaran yang berfokus semata-mata pada materi yang diujikan, bukan pada pemahaman mendalam atau pengembangan kompetensi holistik (Au, 2016; Koretz, 2017). Akibatnya, dimensi penting seperti kreativitas, karakter, kolaborasi, dan kesejahteraan siswa terabaikan karena tidak mudah diukur melalui tes standar.

Selain itu, studi lintas negara menemukan bahwa kebijakan akuntabilitas berbasis tes dapat meningkatkan stres akademik dan menurunkan motivasi intrinsik siswa (Hopfenbeck et al., 2015; Ryan & Weinstein, 2019). Ketika mutu dipersempit menjadi skor, sekolah cenderung memprioritaskan strategi jangka pendek demi peningkatan nilai, bukan penguatan kapasitas belajar jangka panjang. Pendekatan reduksionistik ini juga dapat menciptakan distorsi kurikulum, di mana mata pelajaran yang tidak diujikan memperoleh perhatian lebih rendah (Lingard et al., 2016). Oleh karena itu, evaluasi yang terlalu menekankan angka berisiko mengabaikan tujuan pendidikan yang lebih luas, termasuk pembentukan karakter, kesehatan mental, dan iklim belajar yang positif.

Tantangan kedua adalah meningkatnya beban administratif akibat praktik evaluasi yang menumpuk instrumen tanpa fokus yang jelas. Di banyak sistem pendidikan, guru diwajibkan mengisi berbagai format pelaporan, mulai dari rencana pembelajaran, instrumen asesmen, hingga laporan refleksi, yang sering kali lebih berorientasi pada kepatuhan administratif daripada analisis substantif. Penelitian menunjukkan bahwa beban dokumentasi yang berlebihan dapat mengurangi waktu guru untuk merancang pembelajaran bermakna dan berinteraksi dengan siswa (Schildkamp et al., 2019; Datnow & Hubbard, 2016).

Fenomena ini dikenal sebagai ritualisme pelaporan, yaitu ketika evaluasi menjadi aktivitas formalitas tanpa tindak lanjut nyata. Studi tentang implementasi kebijakan evaluasi sekolah menemukan bahwa laporan yang dihasilkan sering tidak dianalisis secara mendalam atau tidak diikuti dengan dukungan profesional yang memadai (Brown et al., 2018; Perryman et al., 2017). Akibatnya, evaluasi kehilangan fungsi reflektif dan transformasionalnya. Guru mungkin memenuhi tuntutan administrasi, tetapi tidak memperoleh umpan balik yang membantu memperbaiki praktik pembelajaran. Tanpa dukungan analitis dan kepemimpinan yang kuat, evaluasi berisiko menjadi beban birokratis yang menguras energi profesional pendidik.

Evaluasi pendidikan juga menghadapi tantangan serius terkait bias dan ketidaksetaraan konteks. Sekolah yang beroperasi di lingkungan dengan sumber daya terbatas—misalnya daerah terpencil atau komunitas berpenghasilan rendah—sering kali menunjukkan skor akademik lebih rendah dibandingkan sekolah di lingkungan lebih mapan. Namun, skor tersebut tidak selalu mencerminkan kualitas pengajaran, melainkan dipengaruhi oleh faktor eksternal seperti latar belakang sosial-ekonomi, dukungan keluarga, dan akses fasilitas belajar (OECD, 2016; Reardon, 2016).

Ketika indikator evaluasi tidak sensitif terhadap konteks, sekolah dengan tantangan struktural dapat dicap sebagai “bermutu rendah”, yang pada gilirannya memperburuk stigma dan mengurangi kepercayaan publik. Penelitian menunjukkan bahwa kebijakan berbasis peringkat sekolah dapat memperlebar kesenjangan karena sumber daya tambahan sering dialokasikan pada sekolah yang sudah berprestasi (Lingard et al., 2016; Au, 2016). Selain itu, bias budaya dalam instrumen asesmen dapat merugikan kelompok minoritas atau siswa dengan latar belakang bahasa berbeda (Koretz, 2017; Ryan & Weinstein, 2019). Oleh karena itu, evaluasi yang adil harus mempertimbangkan konteks sosial dan budaya serta menggunakan pendekatan komparatif yang proporsional, bukan absolut. Sensitivitas terhadap konteks menjadi kunci agar evaluasi tidak memperkuat ketimpangan yang sudah ada.

Tantangan lain yang signifikan adalah keterbatasan kapasitas analitik di tingkat sekolah. Meskipun data pendidikan semakin melimpah—baik dari hasil tes, survei, maupun sistem manajemen sekolah—tidak semua guru dan pemimpin sekolah memiliki literasi data yang memadai untuk menafsirkan dan menggunakannya secara efektif. Penelitian menunjukkan bahwa tanpa pelatihan yang memadai, data sering disalahartikan atau tidak dimanfaatkan dalam perencanaan pembelajaran (Datnow & Hubbard, 2016; Schildkamp et al., 2019).

Literasi data mencakup kemampuan memahami konsep statistik dasar, membaca tren, mengidentifikasi pola, serta mengaitkan temuan dengan strategi pedagogis. Studi menunjukkan bahwa sekolah yang memiliki tim data kolaboratif dengan dukungan kepemimpinan kuat lebih berhasil menerjemahkan temuan evaluasi menjadi tindakan nyata (Brown et al., 2018; Mandinach & Gummer, 2016). Sebaliknya, ketika kapasitas analitik lemah, evaluasi hanya menghasilkan laporan tanpa perubahan praktik. Data tersedia, tetapi tidak menggerakkan keputusan. Oleh karena itu, investasi dalam pengembangan profesional guru dan pemimpin sekolah menjadi prasyarat agar evaluasi benar-benar berfungsi sebagai motor peningkatan mutu.

KESIMPULAN

Evaluasi pendidikan adalah instrumen strategis dalam penjaminan dan peningkatan mutu apabila dipahami sebagai proses pemberian nilai berbasis kriteria untuk pengambilan keputusan perbaikan. Urgensinya terletak pada fungsi umpan balik, pemetaan kesenjangan, dan penguatan akuntabilitas. Namun efektivitas evaluasi bergantung pada kejelasan tujuan, kualitas instrumen, penggunaan data yang triangulatif, umpan balik yang dapat ditindaklanjuti, tindak lanjut terencana, serta budaya mutu yang mendukung. Tanpa prasyarat tersebut, evaluasi berisiko menjadi formalitas administratif yang minim dampak. Dengan menempatkan evaluasi sebagai proses belajar institusional yang berkelanjutan, mutu pendidikan dapat meningkat secara lebih terarah, adil, dan berkelanjutan.

DAFTAR PUSTAKA

- AERA, APA, & NCME. (2018). *Standards for educational and psychological testing*. American Educational Research Association.
- Andrade, H. L., & Brookhart, S. M. (2020). Classroom assessment as the co-regulation of learning. *Assessment in Education: Principles, Policy & Practice*, 27(4), 350–372. <https://doi.org/10.1080/0969594X.2019.1571992>
- Arikunto, Suharsimi, *Dasar-dasar Evaluasi Pendidikan*, Jakarta: Bumi Aksara, 2013.
- Au, W. (2016). Meritocracy 2.0: High-stakes, standardized testing as a racial project of neoliberal multiculturalism. *Educational Policy*, 30(1), 39–62.
- 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>
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27–40. <https://doi.org/10.3316/QRJ0902027>
- Brown, G. T. L., & Abdulnabi, H. H. A. (2017). Evaluating the quality of higher education instructor-constructed multiple-choice tests: Impact on student grades. *Frontiers in Education*, 2, Article 24. <https://doi.org/10.3389/educ.2017.00024>

- Datnow, A., & Hubbard, L. (2016). Teacher capacity for and beliefs about data-driven decision making: A literature review of international research. *Journal of Educational Change*, 17(1), 7–28.
- Harris, A., & Jones, M. (2019). Leading professional learning with impact. *School Leadership & Management*, 39(1), 1–4. <https://doi.org/10.1080/13632434.2018.1530892>
- Hattie, J. (2017). *Visible learning for teachers: Maximizing impact on learning*. Routledge.
- Hopfenbeck, T. N., et al. (2015). Balancing accountability and improvement: Learning from international experiences in test-based accountability. *Assessment in Education*, 22(3), 309–323.
- Jalaluddin. (1994). *Filsafat pendidikan Islam*. RajaGrafindo Persada.
- Kane, M. T. (2016). Explicating validity. *Assessment in Education: Principles, Policy & Practice*, 23(2), 198–211. <https://doi.org/10.1080/0969594X.2015.1060192>
- Koretz, D. (2017). The testing charade: Pretending to make schools better. *Educational Measurement: Issues and Practice*, 36(3), 7–15.
- Krippendorff, K. (2018). *Content analysis: An introduction to its methodology* (4th ed.). SAGE Publications.
- Kurniawan, Syamsul. *Ilmu Pendidikan Islam Sebuah Kajian Komprehensif*. Yogyakarta: Penerbit Ombak, 2016.
- Lingard, B., Sellar, S., & Lewis, S. (2016). Accountabilities in schools and the problem of numbers. *Educational Philosophy and Theory*, 48(6), 543–556.
- Makhsin, M. (2023). The Role of Muhasabah in Enhancing Community Solidarity in Islamic Society. *International Journal of Academic Research in Progressive Education and Development*, 12(2), 1390–1393.
- Mandinach, E. B., & Gummer, E. S. (2016). What does it mean for teachers to be data literate? *Teachers College Record*, 118(11), 1–40.
- Mardapi, D. (2012). *Pengukuran, penilaian, dan evaluasi pendidikan*. Nuha Medika.
- Masykur, A., Sunanto, S., & Surya, A. (2020). Social Engagement and Its Effect on Community Development in Islamic Context. *Journal of Islamic Management Studies*, 7(1), 22–35. <https://doi.org/10.34105/jims.v7i1.543>
- Nata, Abudin. *Filsafat Pendidikan Islam*, Jakarta: Gaya Media Pratama, 2005.
- OECD. (2020). *Education at a glance 2020: OECD indicators*. OECD Publishing. <https://doi.org/10.1787/69096873-en>
- Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, 22, 74–98. <https://doi.org/10.1016/j.edurev.2017.08.004>
- Perryman, J., Ball, S. J., Braun, A., & Maguire, M. (2017). Translating policy: Governmentality and the reflective teacher. *Journal of Education Policy*, 32(6), 745–761.
- Reardon, S. F. (2016). School segregation and racial academic achievement gaps. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 2(5), 34–57.
- Riinawati, *Pengantar Evaluasi Pendidikan* (Yogyakarta: Thema Publishing, 2021), 91.
- Ryan, R. M., & Weinstein, N. (2019). Undermining quality teaching and learning: A self-determination theory perspective on high-stakes testing. *Theory and Research in Education*, 17(2), 101–122.
- Salleh, M. (2023). Reflection as a Tool for Teacher Self-Assessment in Islamic Education.
- Schildkamp, K., Poortman, C., Ebbeler, J., & Pieters, J. (2019). How school leaders can build effective data teams: Five building blocks for a new wave of data-informed decision making. *Journal of Educational Change*, 20, 283–325.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4–14. <https://doi.org/10.3102/0013189X029007004>
- 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>
- Stufflebeam, D. L. (2003, October). *The CIPP model for evaluation: An update, a review of the model's development, a checklist to guide implementation* [Conference paper]. Oregon Program Evaluators Network (OPEN), Portland, OR.
- Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*, 10, Article 3087. <https://doi.org/10.3389/fpsyg.2019.03087>
- Zed, M. (2004). *Metode penelitian kepustakaan*. Yayasan Obor Indonesia.