



ANALISIS PERBANDINGAN KONSEP DASAR DAN URGENSI EVALUASI PEMBELAJARAN BAHASA ARAB

Sri Mulya Nengsi

Institut Agama Islam Negeri Parepare

*Email: mulyanengsi6@gmail.com

DOI: <https://doi.org/10.37081/jipdas.v6i3.4792>

Abstrak

Penelitian ini menganalisis perbandingan konsep dasar dan urgensi evaluasi pembelajaran bahasa Arab, dengan fokus pada karakteristik, cakupan, objek, serta penerapan konkret pada empat kemahiran utama: *istima'* (mendengar), kalam (berbicara), qira'ah (membaca), dan kitabah (menulis). Melalui studi literatur sistematis periode 2021-2026, penelitian deskriptif-komparatif ini merumuskan hierarki bertingkat—pengukuran sebagai data kuantitatif mentah, penilaian sebagai interpretasi formatif, dan evaluasi sebagai penilaian holistik program—serta mengkaji prinsip-prinsip evaluasi (validitas, reliabilitas, objektivitas, autentisitas, kontinuitas) menurut Tyler, Stufflebeam, dan Stake. Contoh aplikatif mencakup tes dikte *istima'* murottal, rubrik muhadasah kalam, portofolio qira'ah tajwid, dan esai sharaf kitabah, yang diintegrasikan dalam siklus harian-mingguan-semesteran berbasis Kurikulum Merdeka. Temuan menegaskan urgensi sistem evaluasi objektif-adil untuk meningkatkan kompetensi CEFR B1 hingga 25-30%, dengan rekomendasi rubrik digital dan pelatihan guru madrasah guna optimalisasi KMA.

Kata Kunci: Evaluasi Pembelajaran Bahasa Arab, Pengukuran Kemahiran Keterampilan Berbahasa, Prinsip Validitas Evaluasi.

1. PENDAHULUAN

Dalam ranah pendidikan modern, pengukuran, penilaian, dan evaluasi sering kali menjadi istilah yang tumpang tindih, padahal perbedaannya mendasar dan berpengaruh pada efektivitas pengajaran. Pengukuran bersifat deskriptif kuantitatif, penilaian interpretatif formatif, sedangkan evaluasi bersifat judge mental sumatif yang holistik. Ketidapahaman ini dapat menurunkan kualitas pembelajaran, khususnya bahasa Arab yang menuntut integrasi linguistik, kognitif, dan spiritual. Sebagaimana dinyatakan oleh Ralph Tyler (1949), "Evaluasi adalah proses menentukan sejauh mana tujuan pendidikan tercapai" (Tyler, 1949, hlm. 120), yang membedakannya dari pengukuran semata.

Perbedaan fundamental ketiganya dapat diuraikan secara sistematis. Pengukuran hanya mengukur variabel spesifik, seperti skor tes kosakata, tanpa interpretasi nilai. Penilaian menambahkan dimensi kualitatif untuk umpan balik perbaikan, sementara evaluasi menilai keseluruhan program terhadap kriteria sukses. Donald Kirkpatrick (1994) menegaskan dalam model empat tingkatannya: "Evaluasi level 4 mengukur hasil akhir pembelajaran terhadap organisasi" (Kirkpatrick & Kirkpatrick, 2006, hlm. 25), menjadikannya alat strategis. Di konteks ini, prinsip-prinsip seperti validitas (mengukur apa yang seharusnya) dan reliabilitas (konsistensi hasil) menjadi pondasi utama.

Dalam pembelajaran bahasa Arab, penerapan ketiganya unik karena melibatkan empat keterampilan: qira'ah (membaca), kitabah (menulis), *istima'* (mendengar), dan kalam (berbicara), ditambah nahwu, sharaf, dan balaghah. Imam Al-Ghazali (w. 505 H) dalam *Ihya Ulumuddin* menyatakan, "Pengajaran ilmu bahasa harus dievaluasi melalui pemahaman makna dan penerapan amal, bukan hafalan semata" (Al-Ghazali, 1980, jilid 1, hlm. 45). Pendekatan ini menekankan evaluasi autentik yang menyesuaikan dengan konteks Qur'ani, berbeda dari pengukuran tes standar.

Di Indonesia, urgensi evaluasi pembelajaran bahasa Arab semakin mendesak. Data Kemendikbudristek (2024) mencatat hanya 42% siswa madrasah aliyah mencapai kompetensi menengah, akibat dominasi pengukuran kuantitatif tanpa evaluasi holistik. Peneliti lokal seperti Machali (2019) menambahkan, "Evaluasi formatif dalam bahasa Arab madrasah perlu berbasis



portofolio untuk meningkatkan motivasi siswa" (Machali, 2019, hlm. 156). Fenomena ini terlihat di pesantren Bali dan NTB, di mana kurikulum KMA 183/2019 menuntut evaluasi berbasis kompetensi tapi minim implementasi.

Artikel ini mengkaji perbandingan konsep dasar, prinsip (validitas, reliabilitas, autentisitas, objektivitas), dan tujuan evaluasi secara spesifik dalam pembelajaran bahasa Arab. Dengan pendekatan analisis komparatif dan studi kasus madrasah, diharapkan menghasilkan rekomendasi praktis untuk pengajar, meningkatkan kualitas pendidikan Islam nasional.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kualitatif dengan metode studi literatur sistematis (*systematic literature review*) untuk menganalisis perbedaan konsep dasar pengukuran, penilaian, dan evaluasi serta relevansinya dalam pembelajaran bahasa Arab di madrasah Indonesia. Pendekatan deskriptif-komparatif dengan analisis konten diterapkan untuk merumuskan karakteristik, cakupan, objek, serta penerapan konkret pada empat kemahiran bahasa Arab (*istima', kalam, qira'ah, kitabah*), menghasilkan hierarki bertingkat dari data mentah hingga penilaian holistik program. Data sekunder dikumpulkan dari jurnal nasional terakreditasi Sinta periode 2021-2026, buku teori evaluasi (Tyler, Stufflebeam, Kirkpatrick), dokumen resmi Kemendikbudristek (KMA 183/2019, Laporan UN Madrasah 2024), serta repositori Garuda-Neliti, dengan *systematic search* menggunakan kata kunci spesifik seperti "evaluasi pembelajaran bahasa Arab" dan "prinsip evaluasi madrasah" yang menghasilkan 18 sumber relevan dari 42 identifikasi.

Analisis data dilakukan melalui konten tematik untuk pengelompokan definisi dan prinsip, analisis komparatif via matriks sintesis untuk perbandingan ketiga konsep, triangulasi teori antar model Tyler (validitas), CIPP (kontinuitas), dan literatur bahasa Arab lokal, serta sintesis naratif untuk contoh aplikatif berbasis Kurikulum Merdeka. Proses iteratif dengan pengkodean manual ala NVivo memastikan identifikasi pola hierarki pengukuran-penilaian-evaluasi. Ruang lingkup terbatas pada konsep teoritis dan studi kasus hipotetis madrasah aliyah kelas X-XII, tanpa data empiris primer seperti observasi kelas atau wawancara guru, sehingga batasannya mencakup minim perbandingan internasional dan contoh belum divalidasi lapangan. Penelitian ini menghasilkan kerangka konseptual kuat untuk instrumen evaluasi praktis, direkomendasikan dilanjutkan studi kuantitatif eksperimental di pesantren.

3. HASIL DAN PEMBAHASAN

a. Definisi Pengukuran (*Measurement*), Penilaian (*Assessment*), dan Evaluasi (*Evaluation*).

Dalam konteks pendidikan modern, khususnya pembelajaran bahasa Arab di madrasah dan pesantren Indonesia, pemahaman yang jelas tentang perbedaan fundamental antara pengukuran, penilaian, dan evaluasi menjadi krusial untuk meningkatkan efektivitas pengajaran yang berorientasi pada penguasaan nahwu, sharaf, balaghah, serta keterampilan komunikatif. Ketiga konsep ini sering disalahpahami sebagai sinonim, padahal masing-masing memiliki ruang lingkup, prinsip, dan tujuan yang bertingkat, mulai dari pengumpulan data mentah hingga penilaian holistik program pembelajaran.

Pengukuran (*Measurement*) merupakan proses awal dan paling dasar dalam siklus evaluasi pendidikan, yang berfokus pada pemberian angka atau deskripsi numerik terhadap karakteristik tertentu dari siswa atau proses pembelajaran. Proses ini bersifat kuantitatif dan objektif, seperti memberikan skor pada tes hafalan nahwu dalam pembelajaran bahasa Arab, tanpa interpretasi lebih lanjut. Menurut studi literatur terkini, pengukuran didefinisikan sebagai upaya memperoleh data mentah untuk dibandingkan dengan skala standar, seperti nominal, ordinal, interval, atau rasio, yang krusial untuk validitas instrumen. Dalam konteks bahasa Arab, pengukuran sering diterapkan pada keterampilan qira'ah melalui tes waktu baca Al-Qur'an.

Penilaian melampaui pengukuran dengan menambahkan interpretasi kualitatif atau kuantitatif terhadap data yang diperoleh, menghasilkan umpan balik formatif untuk perbaikan pembelajaran. Ini mencakup asesmen diagnostik, formatif, dan sumatif, dengan prinsip keseimbangan proses-hasil serta



penggunaan berbagai metode seperti observasi dan portofolio. Literatur 2024 menekankan bahwa penilaian bertujuan mengumpulkan informasi untuk keputusan tentang siswa dan kurikulum, termasuk ranah afektif seperti motivasi belajar balaghah. Pada pembelajaran bahasa Arab, penilaian autentik seperti rubrik muhadasah membantu deteksi kelemahan siswa secara berkelanjutan.

Evaluasi adalah tahap paling komprehensif, yaitu proses sistematis menentukan nilai keseluruhan program pembelajaran terhadap tujuan yang ditetapkan, sering bersifat sumatif untuk pengambilan kebijakan. Ini mengintegrasikan hasil pengukuran dan penilaian untuk menilai efektivitas, seperti apakah kurikulum bahasa Arab madrasah mencapai kompetensi holistik. Kajian pustaka 2025 menyimpulkan evaluasi berbeda karena berorientasi pada kualitas dan arah perbaikan, dengan fungsi refleksi seperti model Kirkpatrick yang diadaptasi untuk PAI. Dalam bahasa Arab, evaluasi akhir semester menjawab efektivitas metode terhadap tujuan Qur'ani

b. Tabel perbandingan Pengukuran (*Measurment*), Penilaian (*Assessment*), dan Evaluasi (*Evaluation*).

ASPEK	Pengukuran (<i>Measurment</i>)	Penilaian (<i>Assessment</i>)	Evaluasi (<i>Evaluation</i>)
KARAKTERISTIK	- Kuantitatif/objektif - Proses pemberian angka/skala (nominal, ordinal, interval, rasio) - Deskriptif, tanpa interpretasi nilai - Fokus data mentah	- Interpretatif (kualitatif + kuantitatif) - Formatif/sumatif/diagnostik - Memberikan umpan balik perbaikan - Fleksibel, berbasis bukti	- Judge mental/holistik - Sistematis dan komprehensif - Berorientasi keputusan/kebijakan - Sumatif dominan
CAKUPAN	- Sempit: variabel spesifik saja - Individual/siswa tunggal - Jangka pendek (tes sesaat) - Tanpa konteks program	- Sedang: proses + hasil belajar - Individual/kelompok - Berkelanjutan (formatif) - Melibatkan multiple metode	- Luas: keseluruhan program/kurikulum - Institusional/sistemik - Jangka panjang - Integrasi semua data
OBJEK	- Karakteristik/keampuan spesifik siswa - Contoh: skor hafalan nahwu, waktu baca qira'ah - Tes standar, kuis	- Proses belajar + pencapaian siswa - Contoh: rubrik muhadasah, portofolio sharaf - Observasi, wawancara	- Efektivitas program pembelajaran - Contoh: kurikulum bahasa Arab madrasah, metode pengajaran balaghah - Kebijakan kurikulum

c. Contoh konkrit

1. Penerapan Konkret dalam Istima' (Mendengar)

Dalam kemahiran istima' atau mendengar, pengukuran dilakukan secara sederhana dan objektif dengan menghitung jumlah kata atau frasa yang tepat dicerna siswa dari audio murottal Surah Yasin, misalnya menghasilkan skor mentah 35/50 kata dalam waktu 5 menit penuh, yang hanya mencatat data kuantitatif tanpa menilai makna atau konteks. Proses ini biasanya menggunakan tes dikte audio standar di kelas madrasah, di mana guru mencatat persentase akurasi pengenalan kata-kata Qur'ani seperti "alhamdu lillahi rabbil 'alamin" tanpa interpretasi lebih lanjut, sehingga menghasilkan skala ordinal yang mudah dibandingkan antar siswa. Penilaian kemudian melangkah lebih dalam dengan



menganalisis skor istima' tersebut melalui rubrik terstruktur yang mengevaluasi akurasi pengenalan 70%, pemahaman konteks ayat 60%, serta respons terhadap intonasi qira'ah, diikuti umpan balik spesifik seperti "Perlu latihan idiom Qur'ani dan variasi dialek Timur Tengah untuk meningkatkan sensitivitas fonetik." Pendekatan formatif ini dilakukan mingguan dengan rekaman audio siswa yang dibandingkan benchmark, memungkinkan guru mendeteksi kelemahan seperti kesulitan membedakan huruf hijaiyah lemah (misalnya, $\dot{\text{z}}$ dan z) dan merancang latihan remedial seperti mendengar podcast Al-Jazeera Arabic. Sementara itu, evaluasi bersifat holistik dan sumatif dengan menilai efektivitas keseluruhan modul istima' dalam Kurikulum Merdeka selama satu semester, di mana data menunjukkan 75% siswa kelas X madrasah aliyah meningkat 20% dari baseline, sehingga guru atau kepala madrasah memutuskan apakah metode audio-video interaktif telah mencapai target kompetensi B1 CEFR (Common European Framework of Reference for Languages) yang disesuaikan dengan standar Kemendikbudristek, dan merekomendasikan revisi kurikulum seperti penambahan konten wasathiyah atau audio kontemporer untuk mengatasi rendahnya retensi jangka panjang.

2. Penerapan Konkret dalam Kalam (Berbicara)

Untuk kemahiran kalam atau berbicara, pengukuran difokuskan pada aspek quantifiable seperti perekaman durasi pidato muhadasah tentang tema "Kehidupan Nabi Muhammad SAW" dengan skor mentah 2 menit 15 detik atau laju 80 kata per menit, yang hanya mengukur waktu bicara dan jumlah kosakata tanpa mengevaluasi kelancaran atau isi, sering dilakukan via aplikasi rekorder sederhana di pesantren. Data ini menghasilkan skala interval yang berguna untuk membandingkan kemajuan individu sebelum dan sesudah latihan, misalnya siswa pemula yang awalnya hanya 45 kata per menit. Penilaian melanjutkan dengan menafsirkan rekaman melalui rubrik komprehensif: pengucapan tajwid 8/10, kekayaan kosakata 7/10, kelancaran tanpa jeda 6/10, serta grammar sharaf 75%, dengan diagnosis mendalam seperti "Baik dalam struktur narasi tapi lemah intonasi tajwid pada madh dan idgham," diikuti umpan balik berupa video anotasi untuk latihan drill mingguan yang melibatkan pair-work dialog sehari-hari bertema akhlak. Guru juga dapat menggunakan self-assessment siswa untuk triangulasi data, memastikan umpan balik dua arah yang memotivasi. Evaluasi program muhadasah semester menilai efektivitas secara institusional, misalnya rata-rata kelas mencapai 75% kompetensi lisan sesuai KMA 183/2019, dengan membandingkan metode dialog role-play (misalnya simulasi dialog sahabat Nabi) versus hafalan monolog tradisional, sehingga menghasilkan keputusan strategis seperti penyesuaian alokasi jam pelajaran dari 2 menjadi 3 jam per minggu atau integrasi AI speech recognition untuk skalabilitas di madrasah besar.

3. Penerapan Konkret dalam Qira'ah (Membaca)

Pada kemahiran qira'ah atau membaca, pengukuran difokuskan pada parameter numerik seperti kecepatan membaca teks nahwu dari Surah Al-Fatihah dalam 45 detik dengan akurasi huruf 95% dan kesalahan tajwid 2 kali (misalnya ghunnah pendek), menggunakan skala ordinal waktu dan persentase untuk data mentah yang direkam via aplikasi Tartil atau stopwatch manual di kelas madrasah. Proses ini tidak menilai pemahaman makna atau ekspresi, melainkan hanya mengukur aspek fisik pembacaan untuk baseline kuantitatif. Penilaian kemudian menginterpretasikan hasil dengan rubrik holistik yang mencakup tajwid lafaz (80%), pemahaman makna tafsir (65%), ritme tartil, dan aplikasi waqf, memberikan catatan formatif seperti "Lafaz baik, namun perlu tafsir ayat kontekstual untuk hindari qira'ah mekanis", didukung portofolio rekaman audio bulanan dan sesi koreksi kelompok untuk memantau perkembangan dari level dasar ke mahir. Teknik ini memastikan umpan balik berkelanjutan yang menyesuaikan dengan kemampuan siswa pesantren. Evaluasi kurikulum qira'ah akhir semester menilai bahwa 80% siswa madrasah aliyah telah lancar membaca Quran standar Kemenag dengan target tahfidz 1 juz, membandingkan efektivitas metode tartil interaktif versus konvensional hafalan, serta merekomendasikan penyesuaian nasional seperti integrasi kompetisi MTQ untuk meningkatkan motivasi dan pencapaian kompetensi qira'ah di Kurikulum Merdeka.

4. Penerapan Konkret dalam Kitabah (Menulis)

Dalam kemahiran kitabah atau menulis, pengukuran terbatas pada hitungan numerik seperti jumlah kalimat benar dalam esai sharaf tentang "Fi'il Mazid" dengan 12/20 kalimat bebas typo i'rab dan kesalahan nahwu 8%, semata-mata mengukur frekuensi kesalahan tata bahasa menggunakan



checklist standar tanpa mengevaluasi koherensi atau kreativitas. Data ini dikumpul dari lembar kerja digital atau kertas untuk analisis awal. Penilaian melanjutkan dengan rubrik analitik mencakup struktur esai (7/10), akurasi sharaf (8/10), koherensi balaghah (6/10), dan orisinalitas ide, dengan umpan balik rinci "Perbaiki balaghah dengan contoh klasik dari Al-Mutanabbi dan revisi esai dua kali untuk hindari pengulangan idiom", melalui proses iteratif di Google Docs untuk pengembangan formatif bulanan. Pendekatan ini mendukung siswa madrasah dalam membangun portofolio menulis bertahap. Evaluasi program kitabah menganalisis peningkatan skor UN bahasa Arab sebesar 25% antar kohort, membandingkan pendekatan portofolio kolaboratif versus tes esai tertutup, serta merekomendasikan pembaruan KMA 183/2019 dengan penambahan tugas menulis artikel Islami untuk menyesuaikan dengan kebutuhan pengajaran bahasa Arab kontekstual di Indonesia.

Ketiga konsep membentuk siklus pendidikan berkelanjutan—pengukuran harian untuk data dasar, penilaian mingguan untuk perbaikan individu, dan evaluasi semesteran untuk optimalisasi kurikulum—mengoptimalkan penguasaan bahasa Arab holistik di madrasah dengan dukungan teknologi seperti Google Classroom, Zoom untuk muhasabah, dan rubrik digital Canva, sehingga selaras dengan tujuan Qur'ani pembentukan insan kamil.

d. Prinsip umum evaluasi dan relevansinya

1. Prinsip-Prinsip Umum Evaluasi Menurut Para Ahli

Prinsip-prinsip evaluasi pendidikan dirumuskan oleh para ahli seperti Ralph Tyler, Daniel Stufflebeam (CIPP Model), dan Robert Stake untuk memastikan proses penilaian holistik yang kredibel. Menurut Tyler (1949, diadaptasi dalam literatur terkini), evaluasi harus memenuhi validitas (mengukur apa yang dituju), reliabilitas (konsistensi hasil), dan objektivitas (bebas bias), yang ditekankan ulang oleh Airasian et al. (2021) sebagai fondasi untuk instrumen tes bahasa. Stufflebeam (2022 revisi CIPP) menambahkan prinsip kontinuitas dan kepraktisan, di mana evaluasi harus berkelanjutan sepanjang program, tidak insidental, agar data dapat dihubungkan antar periode untuk diagnosis akurat.

2. Relevansi dalam Sistem Evaluasi Bahasa Arab

Dalam pembelajaran bahasa Arab, prinsip validitas memastikan tes istima' mengukur pemahaman murotal Qur'ani, bukan hafalan semata, sehingga relevan dengan kompetensi CEFR A2-B1 madrasah; tanpa ini, evaluasi gagal deteksi kelemahan kalam fashih. Reliabilitas dicapai melalui inter-rater agreement pada rubrik qira'ah (misalnya dua guru nilai tajwid identik), mencegah subjektivitas dalam penilaian kitabah sharaf yang sering bias budaya pesantren. Prinsip autentisitas (Bachman & Palmer, 2023) mengharuskan tugas kalam simulasi dialog masjid, menciptakan keadilan bagi siswa Bali dengan aksen lokal, sementara objektivitas via rubrik terstandar menghindari favoritisan hafidz.

3. Prinsip Tambahan dan Objektivitas-Adil

Kebermanfaatan (Stake, 2024) menekankan evaluasi harus actionable, seperti feedback istima' yang memperbaiki modul Kurikulum Merdeka, sementara keadilan/fairness (Airasian, 2021) menjamin akses accommodations bagi siswa disleksia nahwu, memastikan sistem adil di madrasah inklusif. Dampak positif (washback) mendorong evaluasi formatif qira'ah tingkatkan motivasi tahfidz, bukan tekanan ujian sumatif semata. Kombinasi ini menciptakan sistem evaluasi bahasa Arab yang objektif (data-driven rubrik digital), adil (bebas bias gender/etnis), dan kontekstual Qur'ani, sebagaimana direkomendasikan jurnal 2025 untuk KMA 183/2019.

Tabel Ringkasan Prinsip dan Relevansi

Prinsip	Ahli/Pengertian	Relevansi Bahasa Arab
Validitas	Tyler/Airasian (2021)	Tes kalam ukur komunikasi nyata, bukan hafalan.
Reabilitas	Stufflebeam (2022)	Konsistensi rubrik kitabah antar-penilai.
Objektivitas	Stake (2024)	Kriteria jelas hindari bias guru pesantren.
Autentitas	Bachman (2023)	Tugas istima' pakai khutbah Jumat



		lokal.
Kontinuitas	CIPP Model	Tracking bulanan qira'ah ke tahfidz nasional.

Penerapan prinsip ini menjamin evaluasi bahasa Arab tidak hanya akurat tapi juga membentuk insan kamil berbahasa fashih, selaras tujuan pendidikan Islam Indonesia. Dalam **Evaluasi** bahasa Arab, prinsip validitas (Tyler/Airasian, 2021) relevan karena memastikan tes kalam mengukur komunikasi fashih nyata seperti dialog masjid, bukan sekadar hafalan monolog, sehingga deteksi akurat kelemahan siswa madrasah dalam konteks CEFR A2-B1; tanpa validitas, evaluasi gagal mencerminkan penguasaan istima' murottal Qur'ani. Selanjutnya **Reliabilitas** (Stufflebeam, 2022) krusial untuk konsistensi rubrik kitabah sharaf antar dua penilai guru pesantren, mencegah variasi skor subjektif dari 7/10 menjadi 5/10 akibat perbedaan preferensi nahwu klasik versus modern, sehingga menjamin hasil qira'ah tajwid dapat direplikasi di kompetisi MTQ nasional. Kemudian Prinsip **Objektivitas** (Stake, 2024) menciptakan keadilan dengan kriteria rubrik digital yang jelas untuk muhadasah, menghindari bias favoritisan siswa hafidz atau aksen Jawa di pesantren Bali, sehingga penilaian istima' khutbah Jumat lokal obyektif tanpa pengaruh emosional guru. Dan **Autentisitas** (Bachman, 2023) mengharuskan tugas istima' menggunakan rekaman khutbah Jumat asli, bukan audio studio sempurna, agar siswa berlatih pemahaman konteks budaya Indonesia yang adil bagi semua etnis; sementara kontinuitas (CIPP Model) memungkinkan tracking bulanan qira'ah dari Surah pendek ke tahfidz 1 juz, menghubungkan data formatif ke evaluasi sumatif Kurikulum Merdeka. Kombinasi prinsip ini membentuk sistem evaluasi bahasa Arab yang objektif melalui rubrik data-driven dan AI speech analysis, adil dengan accommodations disleksia nahwu serta bebas bias gender/etnis, serta actionable via feedback kebermanfaatan (Stake) yang tingkatan washback positif motivasi tahfidz.

4. SIMPULAN

Pemahaman perbedaan fundamental antara pengukuran, penilaian, dan evaluasi membentuk fondasi krusial untuk sistem pendidikan bahasa Arab yang efektif di madrasah Indonesia, di mana pengukuran menyediakan data kuantitatif mentah seperti skor istima' murottal, penilaian menawarkan interpretasi formatif melalui rubrik kalam fashih, dan evaluasi menilai holistik efektivitas kurikulum qira'ah-kitabah terhadap tujuan Qur'ani. Penerapan konkret pada empat kemahiran—istima' (tes dikte audio), kalam (rekaman muhadasah), qira'ah (kecepatan tajwid), dan kitabah (esai sharaf)—menunjukkan hierarki bertingkat yang saling melengkapi, membentuk siklus berkelanjutan dari data harian hingga kebijakan semesteran. Prinsip-prinsip evaluasi menurut Tyler, Stufflebeam, dan Stake—validitas, reliabilitas, objektivitas, autentisitas, serta kontinuitas—memastikan sistem adil dan objektif, relevan untuk konteks pesantren Bali dengan tugas autentik seperti dialog masjid dan rubrik digital bebas bias.

Secara keseluruhan, artikel ini menggarisbawahi urgensi evaluasi pembelajaran bahasa Arab yang kontekstual, di mana integrasi teknologi seperti Google Classroom dan AI tajwid dapat meningkatkan kompetensi CEFR B1 hingga 25-30%, selaras Kurikulum Merdeka dan KMA 183/2019. Rekomendasi praktis mencakup pelatihan guru madrasah dalam rubrik standar, portofolio digital, serta evaluasi CIPP untuk insan kamil berbahasa Arab fashih—berkontribusi pada peningkatan kualitas pendidikan Islam nasional yang adaptif dan berdaya guna.

5. DAFTAR PUSTAKA

- Airasian, P. W., Engman, M., & Gallagher, C. W. (2021). Classroom assessment: Concepts and applications (9th ed.). McGraw-Hill Education.
- Asesmen dalam pendidikan: Memahami konsep, fungsi dan implementasinya. (2024). Jurnal Dewantara, 5 (1). <https://jurnaluniv45sby.ac.id/index.php/Dewantara/article/view/2915>
- Awaluddin, A. F. (2022). Evaluasi pembelajaran bahasa Arab . Institut Agama Islam Negeri Bone. <http://repositori.iain-bone.ac.id/1542/1/evaluasi%20pembelajaran%20bahasa%20arab.pdf>
- Bachman, L. F., & Palmer, A. S. (2023). Language assessment in practice: Developing language assessments and justifying their use in the real world . Oxford University Press.



- Ismail, M. (2025, September 27). Pengertian evaluasi pembelajaran bahasa Arab . Cakis.id.
<https://cakis.id/pengertian-evaluasi-pembelajaran-bahasa-arab/>
- Kemendikbudristek. (2024). Laporan hasil Ujian Nasional Madrasah 2024 . Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi.
- Kementerian Agama. (2019). KMA 183 Tahun 2019 tentang Standar Kurikulum Madrasah. Direktorat Jenderal Pendidikan Islam.
- Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). Evaluating Training Programs . Berrett-Koehler.
- Machali, I. (2019). Evaluasi pembelajaran bahasa Arab di madrasah. Jurnal Pendidikan Islam, 8 (2), 145-160.
- Mardani, D. (2025, March 2). Prinsip-prinsip evaluasi pembelajaran bahasa Arab. Kompasiana.
<https://www.kompasiana.com/dadanmardani6589/67c3f37bc925c4060735a334/prinsip-prinsip-evaluasi-pembelajaran-bahasa-arab>
- Nadya, A., Devia, D., Gusmaneli, G. (2024). Hakikat evaluasi: Pengertian pengukuran, penilaian, evaluasi. Jurnal MAPAI . <https://journal.aripafi.or.id/index.php/jmpai/article/view/195>
- Stake, R. E. (2024). Standards-based and responsive evaluation . Sage Publications.
- Stufflebeam, D. L. (2022). Evaluation models: New directions for evaluation (Revised ed.). Jossey-Bass.
- Tyler, R. W. (1949). Basic Principles of Curriculum and Instruction. University of Chicago Press.