

Analisis Urgensi dan Implementasi Evaluasi Pembelajaran Bahasa Arab: Tinjauan Konseptual serta Praktik pada Empat Kemahiran Berbahasa

Marsyanti¹, Herdah²

^{1,2}Institut Agama Islam Negeri Parepare, Indonesia

E-mail: marsyantigradia123@gmail.com¹, Herdah@iainpare.ac.id²

Article History:

Received: 21 Juni 2026

Revised: 02 Juli 2026

Accepted: 07 Juli 2026

Keywords:

evaluasi pembelajaran; pembelajaran bahasa Arab; asesmen autentik; keterampilan berbahasa; 4C; library research.

Abstrak: Evaluasi merupakan komponen esensial dalam pembelajaran bahasa karena menentukan kualitas asesmen dan menjadi dasar pengambilan keputusan pedagogis. Namun, dalam pembelajaran bahasa Arab masih terdapat kerancuan konseptual antara pengukuran, penilaian, dan evaluasi yang berimplikasi pada kurang tepatnya perancangan sistem asesmen. Penelitian ini bertujuan menganalisis perbedaan konseptual ketiga istilah tersebut serta mengkaji implementasinya dalam evaluasi empat keterampilan berbahasa Arab, yaitu *istimā'*, *kalām*, *qirā'ah*, dan *kitābah*. Penelitian menggunakan pendekatan kualitatif dengan metode *library research* melalui analisis kritis terhadap literatur mengenai evaluasi pendidikan dan pembelajaran bahasa Arab. Hasil kajian menunjukkan bahwa pengukuran, penilaian, dan evaluasi merupakan konsep yang saling berkaitan, tetapi memiliki fungsi dan ruang lingkup yang berbeda dalam proses asesmen. Evaluasi yang efektif memerlukan penerapan prinsip objektivitas, validitas, reliabilitas, dan keadilan melalui asesmen autentik yang disesuaikan dengan karakteristik setiap keterampilan berbahasa. Selain itu, integrasi keterampilan abad ke-21 (*critical thinking*, *creativity*, *collaboration*, dan *communication*) memperkuat orientasi evaluasi yang tidak hanya mengukur capaian belajar, tetapi juga mengembangkan kompetensi berbahasa secara holistik. Temuan ini memberikan landasan konseptual bagi pengembangan sistem evaluasi pembelajaran bahasa Arab yang lebih komprehensif, kontekstual, dan berorientasi pada peningkatan kualitas pembelajaran.

PENDAHULUAN

Evaluasi pendidikan merupakan suatu proses yang dilakukan secara sistematis untuk menilai dan memahami sejauh manakah tujuan, efisiensi dan efektivitas, serta kualitas suatu sistem pendidikan dapat tercapai (Arifin, 2017; Stufflebeam & Shinkfield, 2007). Kegiatan evaluasi bertujuan untuk mengidentifikasi berbagai kelebihan, kelemahan, serta aspek-aspek yang masih memerlukan perbaikan dalam penyelenggaraan pendidikan, sehingga langkah-langkah perbaikan

dapat dirumuskan dalam pelaksanaan evaluasi, penting untuk memperhatikan teknik yang digunakan serta sasaran yang ingin dicapai agar hasil evaluasi selaras dengan tujuan pembelajaran dan mendukung keberhasilan proses pendidikan secara menyeluruh. Evaluasi pada dasarnya berlandaskan pada tujuan pembelajaran yang telah ditetapkan sebelumnya dalam tahap perencanaan, yang disusun secara sistematis oleh pendidik sebagai pedoman dalam proses pembelajaran peserta didik (Arifin, 2017; Tyler, 1949).

Evaluasi juga merupakan instrumen krusial dalam siklus pendidikan yang berfungsi sebagai kompas untuk menentukan arah dan kualitas pembelajaran (Brown & Abeywickrama, 2019). Dalam konteks Pembelajaran Bahasa Arab (PBA), peran evaluasi kini menghadapi tantangan ganda: tuntutan standar kompetensi global abad 21 dan realitas keberagaman sosiolinguistik peserta didik. Di era transformasi pendidikan saat ini, penguasaan bahasa tidak lagi sekadar kemampuan menghafal kaidah gramatikal (*qawa'id*) secara kaku, melainkan kemampuan untuk bernalar kritis (*critical thinking*), berkomunikasi secara efektif (*communication*), serta berekspresi secara kreatif (*creativity*) (OECD, 2019; Partnership for 21st Century Learning, 2019). Namun, realita di lapangan menunjukkan bahwa masih banyak pendidik yang terjebak dalam dikotomi pengukuran tradisional yang bersifat mekanistik, di mana nilai angka atau skor tes dianggap sebagai representasi tunggal dan final dari kecerdasan bahasa siswa (Brookhart, 2017; Brown & Abeywickrama, 2019).

Fenomena ini sering kali berakar pada kerancuan konseptual dalam memahami anatomi evaluasi. Masih terdapat tumpang tindih persepsi antara istilah pengukuran (*measurement*), penilaian (*assessment*), dan evaluasi (*evaluation*). Ketidakmampuan dalam membedakan ketiga konsep ini berdampak pada munculnya ketidakadilan instruksional (Arifin, 2017; Nitko & Brookhart, 2014). Sebagai contoh, siswa yang memiliki tingkat kreativitas tinggi dalam produksi bahasa (*Maharah Kalam* dan *Kitabah*) namun memiliki keterbatasan dalam ketepatan struktural sering kali mendapatkan label "gagal" atau tidak kompeten. Hal ini terjadi karena instrumen yang digunakan hanya bersifat pengukuran satu arah yang fokus pada kesalahan, bukan pada pencapaian kompetensi secara holistik. Padahal, dalam perspektif psikolinguistik, setiap individu memiliki ritme pemerolehan bahasa yang unik dan dipengaruhi oleh faktor motivasi serta sikap bahasa yang berbeda.

Tanpa adanya sistem penilaian yang sistematis, proses evaluasi hanya akan menjadi formalitas administratif yang kehilangan esensi edukatifnya. Oleh sebab itu, diperlukan redefinisi dan implementasi evaluasi yang komprehensif pada empat kemahiran bahasa—*istima'*, *kalam*, *qira'ah*, dan *kitabah* (Ainin, 2016; Hermawan, 2018). Evaluasi yang ideal harus mampu berperan sebagai sarana diagnostik untuk mengenali hambatan belajar sekaligus sebagai pendorong motivasi siswa (Black & Wiliam, 1998; Brown & Abeywickrama, 2019). Dengan memahami prinsip-prinsip evaluasi yang objektif dan adil, pendidik dapat menciptakan atmosfer pembelajaran yang inklusif (Brookhart, 2017; Nitko & Brookhart, 2014). Hal ini krusial untuk memastikan bahwa sistem penilaian yang diterapkan tidak hanya valid secara akademis, tetapi juga mampu menghargai keberagaman potensi siswa tanpa ada pihak yang merasa terdiskriminasi oleh standar penilaian yang sempit dan kaku.

METODE PENELITIAN

Penelitian ini menggunakan metode studi pustaka (*library research*) (Creswell & Creswell, 2023; Snyder, 2019). Metode ini dilakukan dengan cara mengkaji berbagai sumber literatur yang relevan dengan topik penelitian, seperti buku, artikel jurnal, serta dokumen ilmiah yang berkaitan dengan evaluasi pendidikan dan pembelajaran bahasa Arab (Xiao & Watson, 2019; Zed, 2018).

HASIL DAN PEMBAHASAN

Konsep Evaluasi Pembelajaran

Evaluasi pendidikan didefinisikan sebagai proses pengumpulan dan analisis data secara sistematis untuk menentukan sejauh mana tujuan pembelajaran telah tercapai, sebagaimana dirumuskan oleh Ralph W. Tyler sebagai penilaian apakah tujuan pendidikan dari program sekolah atau kurikulum terealisasi. Benjamin Bloom melengkapi definisi ini dengan menekankan evaluasi sebagai pengumpulan bukti empiris untuk mengukur perubahan perilaku siswa dalam ranah kognitif, afektif, dan psikomotorik. Daniel Stufflebeam melalui model CIPP memperluasnya menjadi evaluasi holistik yang mencakup konteks, input, proses, dan produk untuk pengambilan keputusan berkelanjutan (Arifin, 2017; Arikunto, 2021; Stufflebeam & Zhang, 2017; Tyler, 1949).

Tiga konsep ini berbeda secara mendasar: pengukuran (*measurement*) hanya menghasilkan data kuantitatif seperti skor tes, penilaian (*assessment*) menginterpretasikan data tersebut untuk umpan balik proses belajar, sedangkan evaluasi (*evaluation*) menghasilkan keputusan nilai secara komprehensif tentang efektivitas program. Dalam pembelajaran Bahasa Arab, pengukuran mungkin menghitung jumlah kosakata yang dikuasai siswa, *assessment* menganalisis pola kesalahan nahwu dalam kitabah, dan evaluasi menentukan apakah kurikulum muhadatsah telah mencapai tujuan komunikatif (Arifin, 2017; Nitko & Brookhart, 2014).

Tujuan evaluasi pembelajaran mencakup mengetahui pencapaian kompetensi siswa, memvalidasi efektivitas strategi pengajaran, dan menyediakan dasar perbaikan kurikulum, sebagaimana ditekankan Tyler dalam orientasi tujuan perilaku. Bloom melalui taksonomi pembelajaran menargetkan pengukuran hierarkis dari pengetahuan hingga penciptaan, sementara CIPP Stufflebeam memastikan evaluasi kontekstual yang adaptif terhadap kebutuhan madrasah (Arifin, 2017; Bloom & Reenen, 2013; Stufflebeam & Zhang, 2017; Tyler, 1949).

Fungsi diagnostik berperan mengidentifikasi kelemahan awal siswa, seperti kesulitan istima' akibat interferensi fonem L1, sebelum proses pembelajaran dimulai. Fungsi formatif memantau kemajuan secara berkala melalui observasi kalam dan portofolio qira'ah, memberikan umpan balik segera untuk remedial. Fungsi sumatif mengukur hasil akhir via ujian komprehensif empat kemahiran untuk sertifikasi kompetensi. Fungsi reflektif, sebagaimana dalam model Tyler, mendorong guru merefleksikan pencapaian tujuan secara keseluruhan untuk perencanaan siklus berikutnya (Arifin, 2017; Arikunto, 2021; Black & Wiliam, 1998).

Model Ralph Tyler (1949) menawarkan pendekatan berorientasi tujuan dengan tujuh langkah: menetapkan tujuan, mengklasifikasikan perilaku, merumuskan tes, memilih bukti, mengumpulkan data, menganalisis, dan menyimpulkan efektivitas. Benjamin Bloom (1956) mengembangkan taksonomi untuk mengevaluasi domain kognitif secara bertingkat, esensial untuk PBA yang kompleks. Model CIPP Stufflebeam (1960-an) paling komprehensif untuk PBA: konteks menilai kebutuhan siswa, input mengukur sumber daya guru/media, proses memantau pelaksanaan muhadatsah, dan produk mengevaluasi outcome kitabah (Stufflebeam & Zhang, 2017; Tyler, 1949).

Evaluasi pembelajaran pada hakikatnya tidak hanya berorientasi pada hasil akhir peserta didik, tetapi juga mencakup keseluruhan proses pembelajaran yang berlangsung di kelas. Dalam konteks pendidikan modern, evaluasi dipandang sebagai bagian integral dari sistem pendidikan yang berfungsi untuk mengetahui efektivitas tujuan, materi, metode, serta media pembelajaran yang digunakan guru. Evaluasi yang baik akan memberikan informasi yang akurat mengenai perkembangan peserta didik sehingga dapat dijadikan dasar pengambilan keputusan pendidikan secara objektif dan terukur. Menurut Arikunto, evaluasi merupakan kegiatan sistematis yang bertujuan menentukan kualitas suatu program pendidikan berdasarkan kriteria tertentu yang telah ditetapkan sebelumnya.

Evaluasi Dalam Pembelajaran Bahasa Arab

Evaluasi dalam pembelajaran bahasa memiliki karakteristik unik karena menargetkan kemampuan komunikatif holistik yang dinamis, melibatkan empat kemahiran terintegrasi (*listening, speaking, reading, writing*) serta ranah kognitif, afektif, dan psikomotorik siswa. Berbeda dengan mata pelajaran eksakta, evaluasi bahasa menekankan konteks autentik dan proses kreatif, bukan sekadar hafalan fakta. Karakteristik utamanya meliputi: (1) multidimensional (mengukur reseptif dan produktif secara simultan), (2) kontekstual (mengadopsi situasi kehidupan nyata), (3) berkelanjutan (formatif > sumatif), dan (4) subjektif-objektif hybrid yang membutuhkan rubrik jelas untuk mengurangi bias. H.D. Brown menegaskan evaluasi bahasa efektif jika merefleksikan penggunaan fungsional bahasa target, bukan drill gramatikal semata (Bachman & Palmer, 2010; Brown & Abeywickrama, 2019; O'Malley & Pierce, 1996)

Penilaian autentik, sebagaimana didefinisikan Brown, adalah pendekatan yang menilai kemampuan siswa dalam tugas-tugas mirip dunia nyata, seperti presentasi lisan (*maharah kalam*), debat Arab autentik, atau menulis esai respons Qur'ani untuk Bahasa Arab (Bachman & Palmer, 2010; Brown & Abeywickrama, 2019). Bachman & Palmer menambahkan bahwa autentik assessment mengukur "ability to use language in real communicative contexts" melalui performance tasks, portfolios, dan self-assessment, yang lebih valid daripada multiple-choice tradisional (Bachman & Palmer, 2010). Keunggulannya: meningkatkan motivasi intrinsik dan mengungkap gap antara tes formal dengan kompetensi komunikatif sebenarnya. Contoh aplikasi: role-play jual-beli Arab untuk kalam, atau analisis nash klasik untuk qira'ah-kitabah terintegrasi (Bachman & Palmer, 2010; Brown & Abeywickrama, 2019)

Prinsip validitas memastikan instrumen mengukur kemampuan berbahasa aktual (construct validity), bukan sekadar pengetahuan gramatika; Brown mengklasifikasikan menjadi content, criterion-related, dan consequential validity (Brown & Abeywickrama, 2019). Reliabilitas menjamin konsistensi skor antar-rater dan waktu, dicapai melalui inter-rater training dan rubrik analitik (e.g., skor kalam: *fluency* 30%, *accuracy* 40%, *pronunciation* 30%) (Bachman & Palmer, 2010). Objektivitas mengurangi subjektivitas guru via standardized criteria dan multiple assessors, krusial untuk penilaian lisan. Praktikalitas menekankan efisiensi waktu-biaya, seperti digital rubrics di Google Classroom untuk portofolio kitabah, memastikan skalabilitas di kelas besar (Brown & Abeywickrama, 2019).

Prinsip	Deskripsi (Brown/Bachman)	Aplikasi Bahasa Arab
Validitas	Mengukur construct komunikatif	Tes muhadatsah fungsional [Brown, 2004]
Reliabilitas	Konsistensi skor	Rubrik ganda untuk istima' [Bachman & Palmer, 1996]
Objektivitas	Kurangi bias penilai	Audio recording kalam dinilai blind
Praktikalitas	Efisien pelaksanaan	Apps untuk self-assessment qira'ah

Implementasi prinsip ini transformasi evaluasi bahasa dari mekanistik menjadi pembelajaran bermakna

Empat Kemahiran Berbahasa Arab

Empat kemampuan inti berbahasa Arab, yaitu *maharah al-istima'* (keterampilan menyimak), *maharah al-kalam* (keterampilan berbicara), *maharah al-qira'ah* (keterampilan membaca), dan *maharah al-kitabah* (keterampilan menulis), membentuk kerangka utama penguasaan *lughah al-'arabiyyah* dalam pendidikan formal, khususnya di madrasah dan perguruan tinggi Islam. Kemahiran ini tidak berdiri sendiri, melainkan saling terkait membentuk kompetensi komunikatif menyeluruh sesuai skala CEFR-Arabic, dengan aplikasi kontekstual mulai dari dialog

sehari-hari hingga analisis teks keislaman klasik.

1. *Maharah Istima'* (Katrampilan Menyimak)

Maharah al-istima' mengembangkan kemampuan menyerap dan memahami informasi lisan dalam Bahasa Arab, meliputi pengenalan makna global, rincian spesifik, serta penarikan kesimpulan dari materi audio. Aktivitas pembelajaran mencakup mendengarkan rekaman tafsir ayat, wawancara tokoh Arab, atau pidato Jumu'at dengan latihan seperti note-taking dan pertanyaan inferensial. Ciri khas: bersifat pasif-reseptif namun menuntut fokus intensif pada variasi dialek dan kecepatan irtijal. Kendala umum berupa interferensi fonetik bahasa ibu diatasi melalui latihan bertahap dari audio lambat menuju native speaker (Ainin, 2016; Hermawan, 2018; Nuha, 2016; Thu'aimah, 1989).

2. *Maharah Kalam* (Ketrampilan Berbicara)

Kemampuan *kalam* menitik beratkan produksi ujaran Arab yang fasih dan spontan, dinilai berdasarkan kelancaran (*tholaqah*), ketepatan gramatikal (*shohhah*), serta ekspresi nonverbal (*isyarah*). Bentuk latihan autentik terdiri dari simulasi peran (pengenalan diri, negosiasi pasar), monolog bertema sejarah Islam, dan diskusi kelompok isu kontemporer. Penilaian dilakukan melalui rekaman suara untuk analisis mandiri dan koreksi rekan, mengatasi hambatan seperti rasa malu berbicara atau inhibisi bilingual melalui sesi *muhadatsah* rutin (Ainin, 2016; Hermawan, 2018; Nuha, 2016; Thu'aimah, 1989).

3. *Maharah Qiro'ah* (Ketrampilan Membaca)

Qira'ah membagi menjadi dua aspek: pembacaan lantang (*qira'ah jahriyyah*) dengan pengamalan tajwid dan pemahaman bacaan diam (*qira'ah khafi*) untuk analisis semantik-struktural. Progresi latihan meliputi pembacaan intensif (surah pendek dengan terjemah), baca luas (buku nahwu dasar), hingga baca kritis (interpretasi balaghah Al-Qur'an). Tes meliputi cloze procedure, soal pemahaman pilihan ganda, dan ringkasan teks. Masalah utama seperti keterbatasan kosa kata diatasi dengan kartu glosarium tematik dan teknik *chunking* (Nurfauzan et al., 2026).

4. *Maharah Kitabah* (Ketrampilan Menulis)

Keterampilan *kitabah* melatih penyusunan wacana tulis kohesif dari tingkat kalimat dasar hingga karangan ekspositori atau argumentatif, dengan kriteria i'rab lengkap, alur logis, dan keberagaman stilistik. Tahapan pengajaran: penyalinan model → pengisian celah → komposisi terkontrol → esai bebas (contoh: pidato ta'dhim dalam Arab). Skoring analitik menimbang konten (40%), struktur (30%), dan bahasa (30%) melalui revisi berulang dan tinjauan teman. Transfer kesalahan dari bahasa pertama dan mental block diatasi dengan latihan penggabungan kalimat dan jurnal harian Arab (Ainin, 2016; Hermawan, 2018; Nuha, 2016; Thu'aimah, 1989).

Keseimbangan pengembangan keempat *maharah* sangat menentukan kefasihan komprehensif, di mana kelebihan satu kemampuan tanpa penguatan lainnya menghalangi kemampuan berbahasa fungsional holistik.

Urgensi Evaluasi Dalam Pembelajaran Bahasa Arab

Evaluasi memegang peran strategis dalam Pembelajaran Bahasa Arab (PBA) sebagai parameter utama yang mengukur efektivitas proses pengajaran dan pencapaian kompetensi siswa pada empat kemahiran berbahasa. Di tengah tantangan pendidikan Islam modern, urgensi evaluasi tidak hanya teknis, tetapi juga esensial untuk memastikan penguasaan Bahasa Arab sebagai pintu gerbang pemahaman sumber utama agama (Ainin, 2016; Arifin, 2017; Brown & Abeywickrama, 2019). Adapun urgensi evaluasi dalam pembelajaran bahasa Arab sebagai berikut:

1. *Evaluasi sebagai indikator keberhasilan pembelajaran.*

Evaluasi berfungsi sebagai tolok ukur objektif keberhasilan program PBA, menunjukkan

apakah tujuan kurikulum seperti penguasaan nahwu dasar atau muhadatsah lancar telah terealisasi melalui data kuantitatif (skor tes) dan kualitatif (observasi). Tanpa indikator ini, pengajaran berisiko stagnan, gagal memenuhi standar nasional seperti Kurikulum Merdeka yang menuntut outcome komunikatif (Nurfauzan et al., 2026)

2. Evaluasi sebagai alat diagnosis kesulitan belajar

Sebagai instrumen diagnostik, evaluasi mengungkap hambatan spesifik siswa, seperti kelemahan istima' akibat interferensi fonetik atau kesulitan kitabah karena i'rab tidak presisi, memungkinkan intervensi remedial dini. Pre-test dan analisis error pattern membantu guru menyesuaikan strategi individual, mencegah kegagalan berkelanjutan (Black & Wiliam, 1998; Nitko & Brookhart, 2014).

3. Evaluasi sebagai dasar perbaikan metode pembelajaran

Hasil evaluasi menyediakan bukti empiris untuk merevisi metode, misalnya beralih dari hafalan ke task-based learning jika kalam lemah, atau integrasi media digital untuk qira'ah interaktif. Refleksi pasca-evaluasi mendorong inovasi guru, meningkatkan kualitas pengajaran secara iteratif (Arifin, 2017; Black & Wiliam, 1998)..

4. Evaluasi sebagai sarana peningkatan kompetensi komunikasi

Evaluasi autentik memperkuat kemampuan komunikatif holistik, di mana umpan balik formatif mendorong siswa dari receptive skills (*istima'-qira'ah*) ke productive (*kalam-kitabah*), membentuk kefasihan fungsional untuk konteks nyata (Ainin, 2016; Arifin, 2017; Brown & Abeywickrama, 2019).

5. Urgensi Khusus dalam pendidikan Islam

Dalam ranah pendidikan Islam, evaluasi PBA krusial karena Bahasa Arab sebagai *lisān al-Qur'ān*, memfasilitasi akses langsung ke:

- Al-Qur'an: Pemahaman tafsir dan balaghah ayat memerlukan qira'ah-istima' fasih untuk tadabbur yang mendalam
- Hadis: Analisis sanad-matanat menuntut kalam presisi untuk diskusi fiqh dan kitabah risalah
- Literatur Keislaman Klasik: Teks seperti *Alfiyyah Ibn Malik* atau *Tafsir Jalalain* bergantung pada kompetensi terintegrasi untuk ijtihad kontemporer

Evaluasi yang komprehensif menjamin PBA tidak hanya akademis, tetapi juga transformatif bagi pemahaman keislaman autentik (Patriani et al., 2022).

Implementasi Evaluasi pada Maharah Istima'

Evaluasi dalam pembelajaran Bahasa Arab bertujuan untuk mengukur sejauh mana peserta didik mampu menguasai kompetensi reseptif dan produktif. Implementasi evaluasi harus disesuaikan dengan karakteristik masing-masing kemahiran guna memastikan validitas hasil penilaian.

1. Implementasi Evaluasi pada Maharah Istimā' (Keterampilan Menyimak)

Valuasi pada kemahiran ini difokuskan pada kemampuan peserta didik dalam mengonstruksi makna dari input lisan. Strategi yang dapat diterapkan meliputi:

- Asesmen Pemahaman Auditori: Mengukur kemampuan diskriminasi bunyi dan pemahaman pesan melalui rekaman penutur asli (*native speaker*).
- Respon Komprehensif: Peserta didik memberikan jawaban atas pertanyaan terstruktur (pilihan ganda atau esai singkat) berdasarkan teks lisan yang didengar.
- Teknik Cloze-Listen: Melengkapi rumpang pada naskah teks berdasarkan informasi spesifik yang tertangkap melalui indra pendengaran.

- d. Imlā' (Dikte): Menguji ketepatan antara persepsi bunyi dengan representasi grafis (penulisan) kata atau kalimat.
2. **Implementasi Evaluasi pada *Maharah Kalām* (Keterampilan Berbicara)**
Evaluasi kemahiran berbicara bersifat performatif, yang menuntut peserta didik memproduksi bahasa secara lisan dalam konteks komunikatif.
 - a. Instrumen Evaluasi:
 - 1) Wawancara Terstruktur: Interaksi tanya jawab langsung untuk mengukur respon spontan.
 - 2) Dialog Partisipatif: Percakapan berpasangan yang menekankan pada fungsi sosial bahasa.
 - 3) Produksi Lisan (Presentasi & Storytelling): Mengukur kemampuan menyampaikan gagasan secara kronologis dan sistematis.
 - 4) Simulasi Sosial (Role Play): Bermain peran dalam situasi kontekstual (misal: di bandara atau pasar).
 - b. Aspek Penilaian: Mencakup kelancaran (*fluency*), akurasi gramatikal (*accuracy*), kekayaan diksi (*vocabulary*), serta ketepatan artikulasi (*pronunciation*).
3. **Implementasi Evaluasi pada *Maharah Qirā'ah* (Keterampilan Membaca)**
Evaluasi ini bertujuan mengukur kemampuan literasi dan analisis teks tertulis.
 - a. Pemahaman Tekstual: Menjawab pertanyaan literal maupun inferensial dari teks bacaan.
 - b. Analisis Gagasan: Kemampuan mengidentifikasi ide pokok (*al-fikrah al-raisiyyah*) dan informasi pendukung.
 - c. Analisis Struktural: Membedah genre teks, kohesi, dan organisasi wacana dalam bacaan.
 - d. Membaca Kritis (*Critical Reading*): Mengevaluasi argumen penulis dan membedakan antara fakta serta opini dalam teks Arab.
4. **Implementasi Evaluasi pada *Maharah Kitābah* (Keterampilan Menulis)**
Keterampilan menulis merupakan kemahiran produktif yang paling kompleks karena melibatkan aspek mekanis dan kognitif.
 - a. Bentuk Evaluasi:
 - 1) Produksi Paragraf: Menyusun kalimat menjadi satu kesatuan ide yang padu.
 - 2) Keterampilan Meringkas: Mengambil intisari dari sebuah teks tanpa menghilangkan makna orisinal.
 - 3) Menulis Akademik (Esai Pendek): Mengembangkan argumen dalam bentuk tulisan formal.
 - 4) Penulisan Kreatif (Naratif/Deskriptif): Menggambarkan objek atau peristiwa secara detail dengan kaidah bahasa yang tepat.
 - b. Aspek Penilaian: Menitikberatkan pada ortografi (ejaan), struktur sintaksis (*nahwu*), kohesi dan koherensi antar-kalimat, serta ketepatan penggunaan mufradāt.

Implikasi Pengembangan Evaluasi Bahasa Arab

Pengembangan evaluasi dalam pembelajaran Bahasa Arab tidak hanya berfungsi sebagai alat ukur keberhasilan, tetapi juga sebagai instrumen untuk meningkatkan kualitas instruksional secara keseluruhan. Berikut adalah empat implikasi strategis dalam pengembangan evaluasi modern:

1. Penggunaan penilaian autentik (*Authentic Assessment*)

Penilaian autentik menuntut peserta didik untuk mendemonstrasikan kemampuan kebahasaan dalam konteks dunia nyata, bukan sekadar menjawab pertanyaan teoritis. Arifin

menegaskan bahwa penilaian harus mencakup ranah kognitif, afektif, dan psikomotorik secara seimbang. Dalam Bahasa Arab, hal ini berarti menilai kemampuan siswa dalam menggunakan bahasa untuk fungsi sosial, seperti berbelanja di pasar simulasi atau menulis surat resmi, bukan sekadar menghafal kaidah *Nahwu* dan *Sharaf* (Arifin, 2017; Brown & Abeywickrama, 2019; O'Malley & Pierce, 1996).

2. Integrasi evaluasi dengan aktifitas komunikatif

Evaluasi yang efektif harus terintegrasi langsung dengan aktivitas komunikatif di kelas. Pendekatan ini menggeser fokus dari "mengetahui bahasa" menjadi "menggunakan bahasa". Menurut (Bachman & Palmer, 2010; Brown & Abeywickrama, 2019), evaluasi kemahiran produktif seperti *Kalām* (berbicara) harus dilakukan melalui interaksi spontan yang menyerupai komunikasi asli. Dengan mengintegrasikan evaluasi ke dalam diskusi kelompok atau debat, kecemasan bahasa (*language anxiety*) siswa dapat diminimalisir karena mereka merasa sedang berkomunikasi, bukan sedang diuji (Holmes et al., 2022; Redecker, 2017).

3. Pengembangan rubrik penilaian ketrampilan bahasa

Untuk menjamin objektivitas, pengembangan rubrik analitik sangat diperlukan, terutama pada kemahiran *Kalām* dan *Kitābah*. Rubrik yang baik harus memecah kompetensi menjadi kriteria yang terukur seperti kelancaran, akurasi gramatikal, dan ketepatan diksi. Brookhart (2017), Brown & Abeywickrama (2019) menyatakan bahwa kriteria yang jelas dalam bahan ajar dan alat evaluasi membantu peserta didik memahami standar capaian yang harus mereka raih, sekaligus memudahkan guru dalam memberikan umpan balik yang konstruktif.

4. Pemanfaatan teknologi dalam evaluasi bahasa

Transformasi digital menuntut penggunaan *Computer-Based Testing* (CBT) dan platform interaktif dalam evaluasi. Penggunaan aplikasi seperti *Quizizz*, *Kahoot*, atau sistem manajemen pembelajaran (LMS) memungkinkan guru melakukan evaluasi secara *real-time* dan memberikan hasil instan kepada siswa. Selain meningkatkan efisiensi, teknologi juga mendukung evaluasi kemahiran *Istimā'* melalui integrasi audio-visual yang lebih jernih dan variatif (Holmes et al., 2022; Redecker, 2017).

Analisis Konseptual Evaluasi Pembelajaran Bahasa Arab

Dalam pembelajaran bahasa Arab, evaluasi bukan sekedar instrumen pengukur capaian akhir, melainkan komponen sistemik yang krusial untuk mengoptimalkan kualitas pendidikan (Arifin, 2017; Stufflebeam & Zhang, 2017; Wijayati & Widhiyoga, 2023). Evaluasi berfungsi mendeteksi diskrepansi antara target kurikulum dan realitas perolehan belajar peserta didik, yang dilakukan secara kontinu dan adaptif terhadap dinamika kebutuhan pembelajaran.

Esensi evaluasi yang holistik terletak pada integrasi ranah kognitif, efektif dan psikomotorik. Artinya, asesmen tidak hanya terpaku pada penguasaan teoritis kaidah kebahasaan (gramatika), tetapi juga mencakup kemahiran komunikatif serta internalisasi sikap positif siswa. Pendekatan komprehensif ini bertujuan untuk memotret kompetensi mahasiswa secara utuh dan autentik (Arifin, 2017).

Tantangan Evaluasi dalam Pembelajaran Bahasa Arab

Implementasi evaluasi dalam pendidikan Bahasa Arab dihadapkan pada probelamtika teknis dan pedagogis yang berlapis. Persoalan krusial muncul pada upaya standarisasi penilaian kompetensi produktif, seperti kemahiran berbiacara (*maharah istima'*) dan menulis (*maharah kitabah*), yang rentan terhadap subjektivitas pengujian (Brown & Abeywickrama, 2019). Oleh sebab itu, penyusunan instrumen yang baku menjadi urgensi demi menjamin validitas serta reliabilitas

data evaluatif.

Di sisi lain, kendala operasional berupa keterbatasan durasi dan rasio guru-siswa yang tidak proporsional menghambat pelaksanaan asesmen individual yang mendalam (Bachman & Palmer, 2010). Penggunaan model evaluasi tradisional yang masih dominan juga dinilai kurang mampu menstimulasi kecakapan komunikatif mahasiswa. Sebagai solusinya, pembaruan teknik evaluasi yang inovatif sangat diperlukan untuk memitigasi hambatan-hambatan tersebut (O'Malley & Pierce, 1996).

KESIMPULAN

Evaluasi merupakan instrumen fundamental dalam ekosistem pendidikan Bahasa Arab yang tidak hanya berfungsi sebagai alat ukur capaian hasil belajar, tetapi juga sebagai mekanisme kendali mutu untuk menyempurnakan proses instruksional. Melalui proses evaluasi yang sistematis, pendidik dapat memperoleh data komprehensif mengenai efektivitas metode pengajaran dan kesesuaian materi ajar dengan kebutuhan peserta didik. Dengan demikian, kedudukan evaluasi dalam pembelajaran bukan sekadar kegiatan administratif di akhir semester, melainkan jantung dari perbaikan berkelanjutan demi tercapainya tujuan pendidikan yang berkualitas.

Selain itu, implementasi evaluasi dalam pembelajaran Bahasa Arab menuntut keseimbangan yang proporsional antara kemahiran reseptif dan produktif, yang meliputi *istimā'*, kalam, *qirā'ah*, dan *kitābah*. Penilaian yang timpang hanya pada salah satu aspek kemahiran akan mengakibatkan distorsi terhadap potret kompetensi bahasa siswa yang sebenarnya. Oleh karena itu, kurikulum dan instrumen penilaian harus dirancang sedemikian rupa agar dapat memotret kemampuan linguistik serta keterampilan komunikasi peserta didik secara holistik dan integratif.

Terakhir, efektivitas evaluasi sangat bergantung pada penggunaan instrumen yang bervariasi dan bersifat autentik guna meningkatkan kompetensi komunikatif peserta didik secara riil. Penggunaan berbagai teknik penilaian, mulai dari rubrik analitik hingga pemanfaatan teknologi digital, memungkinkan guru untuk menangkap keberagaman potensi siswa secara lebih adil dan objektif. Dengan mengadopsi pendekatan evaluasi yang mendekati konteks kehidupan nyata, peserta didik diharapkan tidak hanya menguasai kaidah kebahasaan secara teoretis, tetapi juga mampu mengaplikasikannya sebagai alat komunikasi yang aktif dan fungsional di era global.

REFERENSI

- Ainin, M. (2016). *Evaluasi dalam Pembelajaran Bahasa Arab*. Misykat.
- Arifin, Z. (2017). *Evaluasi Pembelajaran: Prinsip, Teknik, dan Prosedur*. PT Remaja Rosdakarya.
- Arikunto, S. (2021). *Dasar-Dasar Evaluasi Pendidikan* (Edisi Revisi). Bumi Aksara.
- Bachman, L. F., & Palmer, A. S. (2010). *Language Assessment in Practice*. Oxford University Press.
- Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy and Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Bloom, N., & Reenen, J. Van. (2013). Bab Ii Kajian Teoretik. *NBER Working Papers*, 1, 89.
- Brookhart, S. M. (2017). *How to Give Effective Feedback to Your Students* (2nd ed.). ASCD.
- Brown, H. D., & Abeywickrama, P. (2019). *Language Assessment: Principles and Classroom Practices* (3rd ed.). Pearson.
- Creswell, J. W., & Creswell, J. D. (2023). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (6th ed.). SAGE Publications.
- Hermawan, A. (2018). *Metodologi Pembelajaran Bahasa Arab*. PT Remaja Rosdakarya.
- Holmes, W., Bialik, M., & Fadel, C. (2022). *Artificial Intelligence in Education*. Center for

Curriculum Redesign.

- Nitko, A. J., & Brookhart, S. M. (2014). *Educational Assessment of Students* (7th ed.). Pearson.
- Nuha, U. (2016). *Ragam Metodologi \& Media Pembelajaran Bahasa Arab*. Diva Press.
- Nurfauzan, M. F., Tri, M., Iqram, A., Ridlo, U., Islam, U., Syarif, N., & Jakarta, H. (2026). *Evaluasi Pembelajaran Bahasa Arab sebagai Bahasa Asing* : 2(1), 564–573.
- O'Malley, J. M., & Pierce, L. V. (1996). *Authentic Assessment for English Language Learners*. Addison-Wesley.
- OECD. (2019). *OECD Learning Compass 2030: A Series of Concept Notes*. OECD Publishing.
- Partnership for 21st Century Learning. (2019). *Framework for 21st Century Learning*. Battelle for Kids.
- Patriani, Y., Suskha, F., & Warsah, I. (2022). Urgensi Evaluasi Pembelajaran Bahasa Arab Program Studi Pendidikan Agama Islam UIN Fatmawati Sukarno Bengkulu. *At-Ta'lim : Media Informasi Pendidikan Islam*, 21(1), 53. <https://doi.org/10.29300/attalim.v21i1.7036>
- Redecker, C. (2017). *European Framework for the Digital Competence of Educators (DigCompEdu)*.
- Snyder, H. (2019). Literature review as a research methodology. *Journal of Business Research*. <https://doi.org/10.1016/j.jbusres.2019.07.039>
- Stufflebeam, D. L., & Shinkfield, A. J. (2007). *Evaluation Theory, Models, and Applications*. Jossey-Bass.
- Stufflebeam, D. L., & Zhang, G. (2017). *The CIPP Evaluation Model: How to Evaluate for Improvement and Accountability*. Guilford Press.
- Thu'aimah, R. A. (1989). *Ta'lim al-'Arabiyyah li Ghair al-Nathiqin Biha*. ISESCO.
- Tyler, R. W. (1949). *Basic Principles of Curriculum and Instruction*. University of Chicago Press.
- Wijayati, H., & Widhiyoga, G. (2023). *PENINGKATAN KUALITAS PENDIDIKAN INDONESIA MELALUI*. 4(2), 3598–3605.
- Xiao, Y., & Watson, M. (2019). Guidance on Conducting a Systematic Literature Review. *Journal of Planning Education and Research*, 39(1), 93–112. <https://doi.org/10.1177/0739456X17723971>
- Zed, M. (2018). *Metode Penelitian Kepustakaan* (3rd ed.). Yayasan Obor Indonesia.