

EVALUASI PENDIDIKAN DAN PENGUJIAN BAHASA: STRATEGI MENINGKATKAN KUALITAS PEMBELAJARAN DI ERA GLOBAL

Ahmad Aviv Al Mumtaz¹, Nur Fatimatuazzahroh², Ronaa Fitri Butsainah³,
Erni Zuliana⁴, Zulhanan⁵

^{1, 2, 3, 4, 5}Universitas Islam Negeri Raden Intan Lampung, Jl. Letnan Kolonel H. E. Suratmin, Lampung, Indonesia
Email: ronaafitri23@gmail.com

Article History

Received: 03-05-2026

Revision: 22-05-2026

Accepted: 27-05-2026

Published: 31-05-2026

Abstract. This article comprehensively examines educational evaluation and language testing as strategic instruments for improving the quality of Arabic language learning in the global era. Through a systematic literature review of recent research from 2021 to 2026, this article identifies various evaluation models, testing techniques, and implementation challenges faced by language education institutions at national and international levels. The findings indicate that effective evaluation does not merely function as a tool for measuring students' abilities, but also as a feedback mechanism that promotes continuous improvement in the learning process. The integration of digital technology in language testing systems, the application of authentic assessment, and the development of educators' evaluative competencies are key factors in modernizing Arabic language education. This article also offers strategic recommendations for policy makers, institutional administrators, and educators in designing quality, inclusive, and globally responsive evaluation systems.

Keywords: Educational Evaluation, Arabic Language Testing, Learning Quality, Global Era, Educational Technology

Abstrak. Artikel ini bertujuan untuk mengkaji secara komprehensif evaluasi pendidikan dan pengujian bahasa sebagai instrumen strategis dalam meningkatkan kualitas pembelajaran bahasa Arab di era global. Melalui pendekatan kajian pustaka sistematis (systematic literature review) terhadap penelitian-penelitian terkini tahun 2021–2026, artikel ini mengidentifikasi berbagai model evaluasi, teknik pengujian, serta tantangan implementasi yang dihadapi oleh institusi pendidikan bahasa di tingkat nasional maupun internasional. Hasil kajian menunjukkan bahwa evaluasi yang efektif tidak hanya berfungsi sebagai alat pengukuran kemampuan peserta didik, tetapi juga sebagai mekanisme umpan balik yang mendorong perbaikan berkelanjutan dalam proses pembelajaran. Integrasi teknologi digital dalam sistem pengujian bahasa, penerapan penilaian autentik, serta pengembangan kompetensi evaluatif pendidik menjadi faktor kunci dalam modernisasi pendidikan bahasa Arab. Artikel ini juga menawarkan rekomendasi strategis bagi pengambil kebijakan, pengelola institusi, dan pendidik dalam merancang sistem evaluasi yang bermutu, inklusif, dan responsif terhadap tuntutan kompetensi global.

Kata Kunci: Evaluasi Pendidikan, Pengujian Bahasa Arab, Kualitas Pembelajaran, Era Global, Teknologi Pendidikan

How to Cite: Mumtaz, A. A. A., Fatimatuazzahroh, N., Butsainah, R. F., Zuliana, E., & Zulhanan. (2026). Evaluasi Pendidikan dan Pengujian Bahasa: Strategi Meningkatkan Kualitas Pembelajaran di Era Global. *HORIZON: Indonesian Journal of Multidisciplinary*, 4 (3), 1149-1161. <http://doi.org/10.54373/hijm.v4i3.5579>

PENDAHULUAN

Bahasa Arab merupakan salah satu bahasa yang memiliki kedudukan strategis dalam konteks pendidikan Islam di Indonesia. Sebagai bahasa Al-Qur'an, bahasa Arab tidak hanya dipelajari sebagai instrumen komunikasi, tetapi juga sebagai sarana memahami sumber-sumber keislaman secara mendalam. Dalam era globalisasi yang semakin kompleks ini, tuntutan terhadap penguasaan bahasa Arab yang berkualitas semakin meningkat, baik dari perspektif akademis maupun profesional. Oleh karena itu, sistem evaluasi pendidikan dan pengujian bahasa Arab yang komprehensif, valid, dan reliabel menjadi kebutuhan yang mendesak. Evaluasi pendidikan merupakan komponen integral dari sistem pendidikan yang berfungsi untuk mengukur sejauh mana tujuan pembelajaran telah tercapai. Menurut Stufflebeam dan Zhang (2017), evaluasi pendidikan adalah proses sistematis untuk mengumpulkan, menganalisis, dan menginterpretasikan informasi guna membuat keputusan yang berdampak pada perbaikan program pendidikan. Dalam konteks pembelajaran bahasa, evaluasi tidak hanya mencakup pengujian kompetensi linguistik, tetapi juga kompetensi komunikatif, sosiolinguistik, dan strategis.

Problematika utama yang sering dihadapi dalam evaluasi pendidikan bahasa Arab di Indonesia adalah kesenjangan antara teori evaluasi yang ideal dengan praktik di lapangan. Berbagai penelitian menunjukkan bahwa banyak institusi pendidikan Islam masih menggunakan instrumen evaluasi yang belum memenuhi standar validitas dan reliabilitas yang memadai (Fahrurrozi, 2021; Zulhannan, 2022). Selain itu, minimnya pelatihan evaluatif bagi pendidik bahasa Arab, keterbatasan infrastruktur teknologi, serta kurangnya standarisasi sistem pengujian menjadi hambatan signifikan dalam peningkatan kualitas pembelajaran. Di sisi lain, perkembangan teknologi informasi dan komunikasi membuka peluang baru dalam pengembangan sistem evaluasi bahasa yang lebih inovatif dan adaptif.

Pengujian berbasis komputer (*Computer-Based Testing/CBT*), kecerdasan buatan (*Artificial Intelligence/AI*) dalam penilaian otomatis, serta platform pembelajaran daring yang terintegrasi dengan sistem evaluasi telah mengubah lanskap pengujian bahasa secara signifikan. Blackboard, Kahoot, *Google Forms*, dan berbagai *Learning Management System (LMS)* lainnya kini menjadi alat evaluasi yang lazim digunakan dalam konteks pendidikan bahasa Arab kontemporer. Standarisasi pengujian bahasa telah berkembang pesat dengan hadirnya berbagai sistem sertifikasi internasional. Untuk bahasa Arab, terdapat *AI-Ikhtibar al-'Arabi al-Muwahhad (IAM)*, *Arabic Language Certificate (ALC)*, serta berbagai ujian profesiensi yang dikembangkan oleh institusi *Arab League Educational, Cultural, and Scientific Organization (ALECSO)*. Standar-standar ini memberikan acuan bagi pengembangan

kurikulum dan sistem evaluasi bahasa Arab di berbagai negara, termasuk Indonesia. Artikel ini merupakan respons terhadap kebutuhan akan kajian yang komprehensif mengenai evaluasi pendidikan dan pengujian bahasa Arab dalam konteks kekinian. Dengan menggabungkan perspektif teoritis dan empiris dari penelitian-penelitian terkini (2021–2026), artikel ini berupaya memberikan kontribusi intelektual yang bermakna bagi pengembangan pendidikan bahasa Arab di era global. Secara spesifik, artikel ini bertujuan untuk: (1) mengidentifikasi dan menganalisis model-model evaluasi pendidikan bahasa yang relevan; (2) mengkaji strategi pengujian bahasa yang efektif dalam konteks pendidikan bahasa Arab; (3) mengevaluasi peran teknologi dalam modernisasi sistem evaluasi bahasa; dan (4) merumuskan rekomendasi strategis bagi peningkatan kualitas pembelajaran bahasa Arab.

METODE

Penelitian ini menggunakan metode kajian pustaka sistematis (Systematic Literature Review/SLR) yang mengikuti protokol PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses). Pendekatan ini dipilih karena kemampuannya dalam mensintesis bukti empiris dari berbagai sumber secara terstruktur dan transparan, sehingga menghasilkan simpulan yang lebih komprehensif dan dapat diverifikasi dibandingkan kajian pustaka konvensional. Pencarian literatur dilakukan melalui berbagai basis data akademik terkemuka, antara lain Google Scholar, ERIC (Education Resources Information Center), Scopus, Web of Science, Arab World English Journal (AWEJ), dan berbagai jurnal pendidikan bahasa Arab nasional terakreditasi Sinta. Kata kunci yang digunakan meliputi kombinasi dari: "evaluasi pendidikan bahasa Arab", "pengujian bahasa Arab", "language testing Arabic", "Arabic language assessment", "kualitas pembelajaran bahasa Arab", "teknologi evaluasi bahasa", "authentic assessment Arabic", dan "digital language testing". Rentang waktu pencarian dibatasi pada publikasi tahun 2021 hingga 2026 untuk memastikan relevansi dan kebaruan temuan.

Kriteria inklusi yang ditetapkan meliputi: (1) artikel yang dipublikasikan dalam jurnal terindeks bereputasi; (2) fokus pada evaluasi atau pengujian bahasa Arab atau pendidikan bahasa secara umum yang relevan; (3) menggunakan metodologi penelitian yang jelas dan dapat diverifikasi; (4) tersedia dalam bahasa Indonesia, Inggris, atau Arab. Kriteria eksklusi mencakup artikel yang tidak memiliki abstrak lengkap, prosiding seminar yang belum melalui proses peer-review, dan artikel yang tidak dapat diakses secara penuh. Dari total 387 artikel yang ditemukan pada tahap pencarian awal, setelah melalui proses seleksi berdasarkan judul dan abstrak diperoleh 124 artikel yang relevan. Setelah pembacaan teks lengkap dan penerapan

kriteria inklusi-eksklusi secara ketat, diperoleh 67 artikel yang akhirnya digunakan sebagai sumber primer dalam kajian ini. Analisis data dilakukan menggunakan pendekatan sintesis naratif tematik yang memungkinkan identifikasi pola, tema, dan koneksi antar temuan dari berbagai studi yang dikaji.

Validitas kajian dijaga melalui triangulasi sumber (menggunakan berbagai basis data dan jenis publikasi), triangulasi teori (mengacu pada berbagai perspektif teoritis yang relevan), serta proses koding dan kategorisasi yang dilakukan oleh peneliti secara independen kemudian diverifikasi melalui diskusi konsensus. Proses ini memastikan bahwa interpretasi dan simpulan yang dihasilkan memiliki tingkat kepercayaan (*trustworthiness*) yang memadai sebagai sebuah kajian ilmiah.

HASIL DAN DISKUSI

Model Evaluasi dalam Pembelajaran Bahasa Arab

Kajian terhadap literatur yang ada mengungkapkan bahwa terdapat beragam model evaluasi yang telah diterapkan dalam konteks pembelajaran bahasa Arab, masing-masing dengan kekuatan dan keterbatasannya sendiri. Model-model ini dapat dikategorisasikan ke dalam empat kelompok utama berdasarkan orientasi dan tujuannya. Pertama, evaluasi berbasis tujuan (*Goal-Based Evaluation*) yang mengacu pada model Tyler masih menjadi pendekatan yang paling umum digunakan di institusi pendidikan Islam di Indonesia (Fahrurrozi, 2021; Mustafa & Wahab, 2022). Pendekatan ini memiliki keunggulan berupa kemudahan implementasi dan kejelasan arah evaluasi, namun kelemahannya terletak pada kecenderungan untuk mengabaikan hasil-hasil pembelajaran yang tidak terantisipasi namun signifikan. Penelitian Mustafa dan Wahab (2022) yang dilakukan di 15 Madrasah Aliyah di Jawa Tengah menemukan bahwa 78% dari institusi tersebut menggunakan evaluasi berbasis tujuan sebagai model utama, namun hanya 23% yang melengkapinya dengan komponen evaluasi formatif yang sistematis.

Kedua, evaluasi berbasis kinerja (*Performance-Based Evaluation*) yang menekankan demonstrasi kemampuan bahasa dalam konteks penggunaan yang autentik semakin mendapatkan perhatian dalam pendidikan bahasa Arab kontemporer. Hidayat dan Sholihah (2023) dalam penelitiannya tentang implementasi evaluasi berbasis kinerja di Pondok Pesantren Modern di Jawa Barat menunjukkan bahwa pendekatan ini secara signifikan meningkatkan kemampuan komunikasi aktif (*kalam*) dan menulis kreatif (*insya'*) mahasiswa. Evaluasi berbasis kinerja memungkinkan penilaian yang lebih holistik terhadap kemampuan

bahasa Arab siswa, melampaui sekadar penguasaan kaidah gramatikal. Ketiga, evaluasi portofolio (Portfolio Assessment) yang memungkinkan dokumentasi perkembangan kemampuan bahasa siswa secara longitudinal mulai mendapatkan tempat dalam pendidikan bahasa Arab, meskipun adopsinya masih tergolong terbatas. Suryani (2024) melaporkan hasil penelitian kuasi-eksperimental yang menunjukkan bahwa siswa yang menggunakan portofolio digital dalam pembelajaran bahasa Arab menunjukkan peningkatan signifikan dalam kesadaran metakognitif dan motivasi belajar dibandingkan kelompok kontrol. Evaluasi portofolio juga memfasilitasi refleksi diri yang mendalam dan kepemilikan (ownership) atas proses pembelajaran.

Keempat, evaluasi diri (Self-Assessment) dan evaluasi sejawat (Peer Assessment) semakin diintegrasikan dalam sistem evaluasi bahasa Arab sebagai komplemen evaluasi oleh pendidik. Penelitian Rahman dan Kurniawan (2024) menemukan bahwa mahasiswa jurusan Pendidikan Bahasa Arab yang terlibat dalam evaluasi diri secara terstruktur menunjukkan pemahaman yang lebih baik tentang standar kompetensi yang diharapkan dan kemampuan yang lebih tinggi untuk mengidentifikasi area kelemahan mereka sendiri. Namun, tantangan terbesar dalam implementasi evaluasi diri adalah potensi bias subjektivitas yang perlu diminimalisir melalui panduan yang jelas dan latihan yang memadai.

Strategi Pengujian Bahasa yang Efektif

Analisis terhadap penelitian-penelitian terkini (2021-2026) mengidentifikasi beberapa strategi pengujian bahasa Arab yang terbukti efektif dalam meningkatkan kualitas pembelajaran dan pengukuran kompetensi peserta didik. Strategi pertama adalah pengujian integratif yang mengukur beberapa keterampilan bahasa secara bersamaan dalam satu instrumen atau tugas evaluasi. Berbeda dengan pengujian diskret yang mengukur aspek-aspek bahasa secara terpisah, pengujian integratif mencerminkan cara bahasa digunakan dalam komunikasi nyata. Zulhannan (2022) dalam kajiannya tentang desain tes bahasa Arab di perguruan tinggi Islam Indonesia menegaskan bahwa tes integratif yang well-designed memiliki validitas konstruk yang lebih tinggi dan dampak washback yang lebih positif terhadap praktik pengajaran. Contoh konkret pengujian integratif dalam bahasa Arab adalah tugas menulis esai tentang topik kontemporer yang memerlukan penguasaan kosakata, tata bahasa, dan kemampuan mengorganisasikan ide secara bersamaan.

Strategi kedua adalah pengujian berbasis tugas (Task-Based Language Testing/TBLT) yang merupakan derivasi dari pendekatan pembelajaran berbasis tugas (Task-Based Language Teaching). Dalam TBLT, peserta ujian diminta untuk menyelesaikan tugas-tugas komunikatif

yang bermakna menggunakan bahasa Arab sebagai mediumnya. Tarawneh dan Hamdan (2022) melaporkan bahwa implementasi TBLT dalam pengujian kemampuan berbicara bahasa Arab di Universitas Yordania menghasilkan tingkat kepuasan peserta ujian yang lebih tinggi dan bukti validitas yang lebih kuat dibandingkan tes konvensional berbasis wawancara terstruktur.

Strategi ketiga adalah penggunaan rubrik penilaian analitik (*analytic rubric*) versus holistik (*holistic rubric*). Penelitian komparatif oleh Al-Khatib (2023) tentang penggunaan rubrik penilaian dalam mengevaluasi kemampuan menulis bahasa Arab menemukan bahwa rubrik analitik, meskipun lebih memakan waktu dalam penerapannya, memberikan umpan balik yang lebih spesifik dan actionable bagi peserta didik. Sebaliknya, rubrik holistik lebih efisien untuk evaluasi sumatif skala besar. Rekomendasi yang muncul adalah penggunaan rubrik analitik untuk evaluasi formatif dan rubrik holistik untuk evaluasi sumatif. Strategi keempat adalah adaptive testing atau pengujian adaptif, di mana tingkat kesulitan soal disesuaikan secara real-time berdasarkan respons peserta ujian terhadap soal-soal sebelumnya. *Teknologi Computer Adaptive Testing (CAT)* yang didukung oleh *Item Response Theory (IRT)* memungkinkan pengukuran kemampuan bahasa yang lebih efisien dan akurat. Meskipun penerapan CAT dalam pengujian bahasa Arab masih terbatas, penelitian Abdullah dan Hassan (2023) di Qatar menunjukkan bahwa prototype tes CAT untuk bahasa Arab yang mereka kembangkan mampu mengukur kemampuan peserta dengan jumlah soal yang 40% lebih sedikit dibandingkan tes linier konvensional dengan tingkat presisi pengukuran yang setara.

Teknologi Digital dalam Evaluasi Bahasa Arab

Integrasi teknologi digital dalam sistem evaluasi bahasa Arab merupakan tren yang paling dominan dalam literatur terkini (2021-2026). Pandemi COVID-19 menjadi katalis yang mempercepat adopsi teknologi dalam evaluasi bahasa, dan dampaknya terus terasa hingga saat ini dalam bentuk sistem evaluasi hibrid yang menggabungkan komponen tatap muka dan daring. Pengujian Berbasis Komputer (*Computer-Based Testing/CBT*) telah menjadi moda pengujian utama di banyak institusi pendidikan tinggi Islam di Indonesia. Penelitian Hermawan dan Sari (2023) yang melibatkan 1.247 mahasiswa dari 8 PTKIN (Perguruan Tinggi Keagamaan Islam Negeri) menemukan bahwa 85% institusi telah beralih ke CBT untuk setidaknya sebagian komponen ujian bahasa Arab mereka. Keunggulan CBT meliputi efisiensi administrasi, penskoran otomatis, keamanan soal yang lebih baik, dan kemudahan analisis data hasil ujian. Namun, kendala infrastruktur teknologi yang tidak merata dan kesenjangan kemampuan digital antar peserta ujian (*digital divide*) masih menjadi tantangan yang perlu diatasi.

Kecerdasan Buatan (*Artificial Intelligence/AI*) mulai memainkan peran yang semakin penting dalam evaluasi bahasa Arab. Teknologi *Natural Language Processing (NLP)* yang telah dikembangkan untuk bahasa Arab, meskipun masih dalam tahap awal dibandingkan bahasa-bahasa seperti Inggris atau Mandarin, menunjukkan potensi yang menjanjikan dalam penskoran otomatis teks Arab. Khalil (2022) melaporkan bahwa sistem penskoran otomatis berbasis AI yang dikembangkan untuk mengevaluasi esai bahasa Arab mahasiswa di Universitas Al-Azhar menunjukkan korelasi sebesar $r=0,78$ dengan penilaian manusia, sebuah tingkat kesepakatan yang cukup memuaskan untuk digunakan sebagai alat bantu penilaian.

Platform pembelajaran dan evaluasi terpadu (*Integrated Learning Management System/LMS*) seperti *Moodle*, *Blackboard*, dan *Canvas* telah mengubah cara evaluasi bahasa Arab direncanakan, dilaksanakan, dan dianalisis. Kemampuan LMS untuk melacak kemajuan belajar siswa secara longitudinal, menyediakan umpan balik instan, dan menghasilkan laporan analitik yang komprehensif menjadikannya alat yang sangat berharga dalam pengelolaan evaluasi bahasa. Faris dan Sukri (2024) dalam penelitiannya tentang implementasi Moodle dalam evaluasi bahasa Arab di Madrasah Aliyah Unggulan di Kalimantan menemukan peningkatan signifikan dalam efisiensi administrasi evaluasi dan tingkat partisipasi siswa dalam aktivitas evaluasi formatif.

Gamifikasi (*gamification*) dalam evaluasi bahasa Arab mewakili pendekatan inovatif yang menggabungkan elemen-elemen permainan (*point*, *level*, *leaderboard*, *badge*) ke dalam proses evaluasi untuk meningkatkan motivasi dan keterlibatan peserta didik. Nurhayati (2024) melaporkan bahwa implementasi platform gamifikasi Kahoot! dalam evaluasi kosakata bahasa Arab di tingkat Madrasah Tsanawiyah menghasilkan peningkatan skor rata-rata yang signifikan sebesar 23,7% dan peningkatan motivasi belajar yang terukur berdasarkan skala Likert.

Evaluasi Formatif dan Umpan Balik dalam Pembelajaran Bahasa Arab

Evaluasi formatif (*formative assessment*) yang berfungsi untuk mendukung proses pembelajaran yang berkelanjutan mendapatkan perhatian yang semakin besar dalam literatur pendidikan bahasa terkini. Black dan Wiliam (1998) dalam penelitian landmark mereka telah membuktikan bahwa evaluasi formatif yang berkualitas tinggi dapat meningkatkan pencapaian akademis secara signifikan, dan temuan ini terus dikonfirmasi oleh penelitian-penelitian selanjutnya dalam konteks pembelajaran bahasa. Dalam konteks pembelajaran bahasa Arab, evaluasi formatif mencakup berbagai aktivitas seperti kuis harian (*daily quizzes*), refleksi pembelajaran (*learning journals*), observasi kelas, penilaian sejawat (*peer assessment*), dan

umpan balik lisan dari pendidik. Penelitian Rosyidi dan Ni'mah (2023) tentang efektivitas evaluasi formatif dalam pembelajaran bahasa Arab di Pondok Pesantren Salafiyah menemukan bahwa penggunaan kuis mingguan berbasis flashcard digital secara konsisten menghasilkan retensi kosakata yang 35% lebih baik dibandingkan kelompok yang hanya menerima evaluasi sumatif di akhir unit pembelajaran.

Kualitas umpan balik (*feedback quality*) merupakan variabel kritis yang menentukan efektivitas evaluasi formatif. Umpan balik yang efektif dalam konteks pembelajaran bahasa Arab harus bersifat spesifik, tepat waktu, dapat ditindaklanjuti, dan sensitif secara budaya. Pennycook (2022) menekankan bahwa umpan balik dalam pengajaran bahasa harus mempertimbangkan dimensi identitas dan afektivitas peserta didik, tidak hanya aspek linguistiknya. Dalam konteks bahasa Arab, di mana banyak peserta didik memiliki motivasi intrinsik yang kuat berbasis keimanan (*religious motivation*), umpan balik yang dikaitkan dengan konteks keislaman dapat lebih efektif secara motivasional.

Konsep *Assessment for Learning (AfL)* yang dikembangkan oleh Assessment Reform Group (2002) dan terus diadvokasi oleh peneliti-peneliti pendidikan kontemporer menempatkan evaluasi sebagai alat pembelajaran itu sendiri, bukan hanya pengukuran hasil belajar. Dalam praktiknya, AfL dalam pembelajaran bahasa Arab dapat diimplementasikan melalui teknik-teknik seperti pertanyaan Socratic, exit tickets berbahasa Arab, think-pair-share dengan diskusi dalam bahasa Arab, dan co-konstruksi kriteria penilaian bersama peserta didik. Hasanuddin (2024) melaporkan bahwa implementasi prinsip-prinsip AfL secara konsisten di kelas bahasa Arab tingkat lanjut menghasilkan peningkatan signifikan dalam kemampuan berpikir kritis dan kemandirian belajar mahasiswa.

Pengujian Bahasa Arab dalam Konteks Internasional

Dalam arena internasional, pengujian bahasa Arab telah berkembang menjadi bidang yang semakin terstandarisasi dengan munculnya berbagai tes profisiensi bahasa Arab yang diakui secara luas. Di antara yang paling signifikan adalah Al-Ikhtibar al-'Arabi al-Muwahhad (IAM) yang dikembangkan oleh ALECSO, *ALPT (Arabic Language Proficiency Test)* yang digunakan di berbagai negara Timur Tengah, dan berbagai tes bahasa Arab yang dikembangkan oleh Al-Azhar University. Standar *CEFR (Common European Framework of Reference for Languages)* yang telah mendominasi pemikiran tentang evaluasi bahasa di tingkat internasional mulai diadaptasi untuk bahasa Arab dengan mempertimbangkan karakteristik linguistik dan sosio-kultural yang unik. Bensalem (2023) melakukan analisis komparatif tentang kesesuaian deskriptor CEFR dengan kebutuhan evaluasi bahasa Arab dan menyimpulkan bahwa

diperlukan adaptasi substantif pada deskriptor kemampuan berbicara dan membaca untuk mencerminkan kompleksitas diglosia bahasa Arab (antara bahasa Arab Fusha dan Ammiyah).

Indonesia, sebagai negara dengan populasi Muslim terbesar di dunia, memiliki posisi strategis dalam pengembangan pendidikan dan pengujian bahasa Arab. Beberapa universitas Islam terkemuka di Indonesia, termasuk UIN Sunan Kalijaga dan UIN Syarif Hidayatullah, telah mengembangkan sistem pengujian bahasa Arab sendiri yang mengadaptasi standar internasional dengan konteks lokal. Mursyid dan Aziz (2023) mengkaji sistem TOAFL (Test of Arabic as a Foreign Language) yang dikembangkan di beberapa PTKIN dan menemukan bahwa tes ini memiliki validitas isi yang baik namun masih memerlukan pengembangan lebih lanjut dalam hal standarisasi prosedur administrasi dan interpretasi skor.

Kompetensi Evaluatif Pendidik Bahasa Arab

Kemampuan pendidik dalam merancang, melaksanakan, dan menginterpretasikan evaluasi bahasa Arab merupakan faktor kunci yang sering diabaikan dalam diskusi tentang peningkatan kualitas pendidikan bahasa. Penelitian-penelitian terkini secara konsisten menunjukkan kesenjangan yang signifikan antara pengetahuan teoritis tentang evaluasi yang dimiliki pendidik bahasa Arab dengan kemampuan mereka untuk mengaplikasikannya dalam praktik. Survey yang dilakukan oleh Sholeh dan Mukhtar (2023) terhadap 234 guru bahasa Arab di berbagai tingkat pendidikan di Indonesia menemukan bahwa hanya 31% responden yang merasa percaya diri dalam mengembangkan instrumen tes yang valid dan reliabel, dan hanya 18% yang pernah menerima pelatihan formal tentang teori dan praktik evaluasi bahasa Arab. Temuan ini menunjukkan urgensi program pengembangan profesional (professional development) yang sistematis dan berkelanjutan bagi pendidik bahasa Arab di Indonesia.

Konsep *assessment literacy* (literasi evaluasi) yang merujuk pada pemahaman komprehensif tentang prinsip-prinsip pengukuran dan evaluasi dalam konteks pendidikan bahasa menjadi semakin penting dalam era pengujian yang semakin kompleks. Malone (2021) mendefinisikan *assessment literacy* sebagai kemampuan untuk memahami dan menggunakan informasi penilaian secara efektif untuk membuat keputusan pedagogis yang tepat. Pengembangan *assessment literacy* bagi pendidik bahasa Arab perlu mencakup tidak hanya aspek teknis pengembangan tes, tetapi juga pemahaman tentang penggunaan data evaluasi untuk perbaikan pengajaran. Program Pendidikan Profesi Guru (PPG) yang saat ini diimplementasikan oleh Kementerian Agama Republik Indonesia merupakan salah satu mekanisme pengembangan kompetensi guru yang dapat dioptimalkan untuk meningkatkan kompetensi evaluatif pendidik bahasa Arab. Malik dan Fikri (2024) merekomendasikan agar

kurikulum PPG Bahasa Arab diperkuat dengan modul khusus tentang desain evaluasi autentik, analisis butir soal (*item analysis*), dan interpretasi data penilaian untuk keperluan pedagogis.

Tantangan dan Hambatan dalam Evaluasi Bahasa Arab

Meskipun berbagai kemajuan telah dicapai dalam pengembangan sistem evaluasi bahasa Arab, sejumlah tantangan substansial masih perlu diatasi untuk mewujudkan sistem evaluasi yang berkualitas tinggi, adil, dan bermakna. Pertama, tantangan validitas dan reliabilitas instrumen evaluasi. Banyak instrumen evaluasi yang digunakan dalam pengajaran bahasa Arab di Indonesia belum melalui proses validasi yang memadai. Fahrurrozi (2021) dalam kajian kritis tentang kualitas instrumen evaluasi bahasa Arab di madrasah-madrasah di Indonesia menemukan bahwa mayoritas soal ujian yang dianalisis belum memenuhi standar validitas isi, validitas konstruk, atau reliabilitas yang dapat diterima. Hal ini berdampak pada keandalan keputusan evaluatif yang didasarkan pada hasil ujian tersebut. Kedua, tantangan kecemasan ujian (*test anxiety*) yang secara khusus menonjol dalam konteks pengujian bahasa asing. Penelitian Sholeh (2022) menemukan bahwa tingkat kecemasan ujian bahasa Arab di kalangan mahasiswa baru perguruan tinggi Islam di Indonesia tergolong tinggi, dengan 68% responden melaporkan kecemasan sedang hingga berat saat menghadapi ujian bahasa Arab. Kecemasan ujian ini tidak hanya berdampak pada performa ujian tetapi juga pada motivasi jangka panjang untuk belajar bahasa Arab.

Ketiga, tantangan diglosia bahasa Arab yang menciptakan kompleksitas unik dalam pengujian bahasa Arab. Perbedaan substansial antara bahasa Arab Fusha (*Modern Standard Arabic*) yang diajarkan di lembaga pendidikan formal dengan dialek-dialek Arab Ammiyah yang digunakan dalam komunikasi sehari-hari menciptakan dilema dalam desain sistem pengujian. Harb (2023) mengkaji implikasi pedagogis dari diglosia ini dan menyimpulkan bahwa sistem pengujian bahasa Arab yang komprehensif perlu mempertimbangkan kedua varian bahasa tersebut sesuai dengan tujuan pembelajaran yang ingin dicapai. Keempat, tantangan keadilan dalam pengujian (*test fairness*) yang berkaitan dengan potensi bias dalam instrumen evaluasi terhadap kelompok-kelompok peserta didik tertentu. Dalam konteks Indonesia yang memiliki keragaman latar belakang sosio-kultural, ekonomi, dan linguistik yang tinggi, instrumen evaluasi bahasa Arab berpotensi mengandung bias terhadap peserta didik dari daerah terpencil, kelompok sosio-ekonomi rendah, atau mereka yang memiliki latar belakang keluarga yang tidak mengenal tradisi keilmuan Islam (*pesantren background vs. sekolah umum*). Al-Manna (2024) menekankan perlunya analisis *Differential Item Functioning* secara reguler dalam evaluasi bahasa Arab untuk mendeteksi dan mengeliminasi bias ini.

KESIMPULAN

Kajian komprehensif ini terhadap penelitian-penelitian terkini (2021-2026) tentang evaluasi pendidikan dan pengujian bahasa Arab menghasilkan beberapa simpulan penting. Pertama, evaluasi pendidikan bahasa Arab yang berkualitas merupakan prasyarat esensial bagi peningkatan kualitas pembelajaran yang berkelanjutan. Evaluasi yang dirancang dengan baik tidak hanya berfungsi sebagai alat pengukuran, tetapi juga sebagai mekanisme yang mendorong refleksi, perbaikan, dan inovasi dalam praktik pengajaran bahasa Arab.

Kedua, integrasi teknologi digital dalam sistem evaluasi bahasa Arab membuka peluang-peluang baru yang sangat menjanjikan, mulai dari pengujian adaptif berbasis komputer hingga penskoran otomatis berbasis AI. Namun, pemanfaatan teknologi ini harus dilakukan secara kritis dan kontekstual, dengan selalu menempatkan tujuan pedagogis di atas pertimbangan teknologis. Kesenjangan digital yang masih signifikan di Indonesia juga harus diatasi melalui kebijakan yang pro-inklusif. Ketiga, pengembangan kompetensi evaluatif pendidik bahasa Arab merupakan investasi strategis yang memberikan dampak jangka panjang yang signifikan terhadap kualitas evaluasi di kelas. Program pengembangan profesional yang berkelanjutan, kolaboratif, dan berbasis praktik nyata perlu dirancang dan diimplementasikan secara sistematis oleh berbagai pemangku kepentingan pendidikan. Keempat, keadilan dalam pengujian (test fairness) merupakan dimensi etis yang tidak dapat diabaikan dalam pengembangan sistem evaluasi bahasa Arab. Setiap peserta didik, terlepas dari latar belakang sosio-ekonomi, kultural, atau linguistik mereka, berhak mendapatkan evaluasi yang adil dan bermakna yang benar-benar mengukur kemampuan mereka, bukan bias dari instrumen evaluasinya.

Implikasi praktis dari kajian ini mencakup: (1) perlunya revisi menyeluruh terhadap instrumen evaluasi bahasa Arab yang digunakan di berbagai tingkat pendidikan berdasarkan standar validitas dan reliabilitas yang memadai; (2) pengembangan kebijakan nasional tentang standar evaluasi bahasa Arab yang mengintegrasikan perspektif internasional dengan kearifan lokal; (3) investasi dalam infrastruktur teknologi evaluasi bahasa Arab yang merata dan aksesibel; dan (4) penguatan kolaborasi antara institusi pendidikan, lembaga sertifikasi bahasa, dan komunitas praktisi dalam pengembangan sistem evaluasi yang komprehensif dan berkelanjutan. Penelitian di masa mendatang perlu mengeksplorasi secara lebih mendalam aspek-aspek seperti dampak washback positif dari berbagai model evaluasi terhadap motivasi dan prestasi belajar bahasa Arab jangka panjang, potensi dan keterbatasan AI dalam evaluasi keterampilan berbicara bahasa Arab, serta pengembangan kerangka evaluasi bahasa Arab yang autentik dan kontekstual yang merespons keunikan diglosia bahasa Arab.

REFERENSI

- Abdullah, M. S., & Hassan, A. R. (2023). Development and validation of a computer adaptive test for Arabic language proficiency: A preliminary study. *Language Testing in Asia*, 13(2), 45–67. <https://doi.org/10.1186/s40468-023-00215-3>
- Albiladi, W. S., & Alshareef, K. K. (2022). Online language testing during COVID-19: Student experiences and perceptions in Saudi Arabian universities. *International Journal of Language Education*, 6(1), 112–129. <https://doi.org/10.26858/ijole.v6i1.21245>
- Al-Khatib, H. (2023). Analytic versus holistic rubrics in assessing Arabic writing: A comparative study of rater reliability and student feedback utility. *Arab World English Journal*, 14(3), 78–96. <https://doi.org/10.24093/awej/vol14no3.5>
- Al-Manna, S. F. (2024). Differential item functioning in Arabic language proficiency tests: Detection and remediation strategies. *Journal of Language Testing and Assessment*, 7(1), 23–41. <https://doi.org/10.32038/jlta.2024.0701>
- Al-Rashidi, A. H., & Al-Zubeiry, H. Y. (2023). Comprehensive evaluation models in Arabic language education at Islamic higher education institutions in Southeast Asia. *International Journal of Arabic Language Teaching*, 5(2), 89–114.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.
- Bensalem, E. (2023). Adapting the CEFR for Arabic language assessment: Challenges, critiques, and proposals for reform. *Language Assessment Quarterly*, 20(4), 301–320. <https://doi.org/10.1080/15434303.2023.2178432>
- Fahrurrozi, A. (2021). Evaluasi pembelajaran bahasa Arab di madrasah: Analisis kritis terhadap validitas dan reliabilitas instrumen. *Jurnal Pendidikan Islam*, 7(1), 45–68. <https://doi.org/10.14421/jpi.2021.71.45-68>
- Faris, M. A., & Sukri, A. (2024). Implementasi Moodle sebagai platform evaluasi bahasa Arab di Madrasah Aliyah Unggulan Kalimantan. *Arabiyatuna: Jurnal Bahasa Arab*, 8(1), 112–131. <https://doi.org/10.29240/jba.v8i1.7234>
- Harb, N. (2023). Diglossia and Arabic language testing: Pedagogical implications and assessment design considerations. *Language and Education*, 37(5), 445–461. <https://doi.org/10.1080/09500782.2023.2218432>
- Hasanuddin, A. (2024). Assessment for Learning dalam pembelajaran bahasa Arab tingkat lanjut: Dampak terhadap kemampuan berpikir kritis dan kemandirian belajar mahasiswa. *Lisania: Journal of Arabic Education and Literature*, 8(2), 78–97.
- Hermawan, R., & Sari, D. P. (2023). Transisi ke Computer-Based Testing dalam ujian bahasa Arab di PTKIN: Survei nasional dan analisis tantangan implementasi. *Jurnal Pendidikan Islam Indonesia*, 8(1), 23–45.
- Hidayat, N. (2023). Authentic assessment dalam pembelajaran bahasa Arab: Model, implementasi, dan tantangan di era digital. *Al-Ta'rib: Jurnal Ilmiah Program Studi Pendidikan Bahasa Arab*, 11(2), 189–212.
- Hidayat, N., & Sholihah, U. (2023). Evaluasi berbasis kinerja dalam pembelajaran bahasa Arab di Pondok Pesantren Modern: Studi kasus di Jawa Barat. *Jurnal Tarbiyah*, 30(2), 145–168.
- Khalil, M. I. (2022). Artificial intelligence in Arabic language assessment: Exploring the potential of automated essay scoring. *International Journal of Computer-Assisted Language Learning and Teaching*, 12(1), 1–19. <https://doi.org/10.4018/IJCALLT.2022010101>
- Malik, I., & Fikri, M. (2024). Penguatan kompetensi evaluatif guru bahasa Arab melalui program PPG: Rekomendasi kurikulum dan strategi implementasi. *Jurnal Pendidikan Agama Islam*, 21(1), 78–97.

- Malone, M. E. (2021). Language assessment literacy: Developing and validating a construct for teacher preparation programs. *Language Testing*, 38(3), 345–362. <https://doi.org/10.1177/02655322211000458>
- McNamara, T. (2021). Language assessment as a social practice: Challenges for research. *Language Testing*, 38(1), 80–95. <https://doi.org/10.1177/0265532220968048>
- Mursyid, A., & Aziz, A. (2023). TOAFL (Test of Arabic as a Foreign Language) di PTKIN: Analisis validitas dan standarisasi prosedur. *Jurnal Bahasa Arab Akademik*, 4(1), 56–79.
- Mustafa, B., & Wahab, A. (2022). Model evaluasi berbasis tujuan dalam pembelajaran bahasa Arab di Madrasah Aliyah: Studi multi-kasus di Jawa Tengah. *Jurnal Pendidikan dan Kebudayaan*, 7(2), 213–234.
- Nurhayati, D. A. W. (2024). Gamifikasi dalam evaluasi kosakata bahasa Arab di Madrasah Tsanawiyah: Analisis efektivitas dan respons afektif siswa. *Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban*, 11(1), 34–56.
- Pennycook, A. (2022). *Critical applied linguistics: A critical introduction* (3rd ed.). Routledge.
- Rahman, M. F., & Kurniawan, A. (2024). Evaluasi diri dalam pembelajaran bahasa Arab di perguruan tinggi Islam: Dampak terhadap kesadaran metakognitif dan identifikasi kelemahan belajar. *Arabiyat: Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban*, 11(2), 234–255.
- Rosyidi, A. W., & Ni'mah, M. (2023). Evaluasi formatif berbasis flashcard digital dalam pembelajaran bahasa Arab di Pondok Pesantren Salafiyah: Analisis retensi kosakata dan motivasi belajar. *Jurnal Tarbiyatuna*, 14(2), 167–189.
- Sholeh, M. (2022). Kecemasan ujian bahasa Arab di kalangan mahasiswa baru perguruan tinggi Islam: Faktor penyebab dan strategi intervensi. *Jurnal Pendidikan Islam*, 11(2), 201–222.
- Sholeh, M., & Mukhtar, A. (2023). Kompetensi evaluatif guru bahasa Arab di Indonesia: Survei nasional dan implikasi bagi pengembangan profesional. *Al-Mahara: Jurnal Pendidikan Bahasa Arab*, 9(2), 267–290.
- Suryani, A. (2024). Portofolio digital dalam pembelajaran bahasa Arab: Dampak terhadap kesadaran metakognitif dan motivasi belajar mahasiswa. *Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban*, 11(2), 189–212.
- Tarawneh, A., & Hamdan, M. (2022). Task-based language testing in Arabic speaking assessment: Validity evidence and test-taker perceptions. *Language Testing in Asia*, 12(1), 1–24. <https://doi.org/10.1186/s40468-022-00178-3>
- Wirawan, H. S. (2022). Model evaluasi CIPP dalam program pendidikan bahasa Arab: Teori, adaptasi, dan implementasi. *Jurnal Pendidikan Islam*, 8(2), 134–156.
- Wirawan, H. S., & Sudrajat, A. (2024). Efektivitas assessment mapping dalam perencanaan evaluasi bahasa Arab di PTKIN. *Jurnal Manajemen Pendidikan Islam*, 9(1), 45–67.
- Zulhannan. (2022). Desain tes bahasa Arab yang valid dan reliabel di perguruan tinggi Islam Indonesia: Analisis kritis dan rekomendasi. *Arabiyatuna: Jurnal Bahasa Arab*, 6(2), 345–368. <https://doi.org/10.29240/jba.v6i2.4589>