Al Matla : Journal of Arabic Language Education
e-ISSN: 3125-0505
https://doi.org/00.0000/ Al Matla.x0x0.000
Volume 1, Issue. 2, 2026, Page 68-86
journal.staimh.ac.id

TRANSFORMASI ASESMEN KETERAMPILAN BAHASA ARAB DI ERA DIGITAL DAN KURIKULUM MERDEKA: SEBUAH TINJAUAN LITERATUR SISTEMATIS

Nurhuda Abdillah[*1], Muslihin[*2], Isrun Abdurahman[*3]

[123] Sekolah Tinggi Agama Islam Minhajul Haq Purwakarta Indonesia
e-mail: nurhuda.abdillah@staimh.ac.id, muslihin@staimh.ac.id, isrun.abdurahman@staimh.ac.id

Submitted: 09-07-2026	Revised : 20-08-2026	Accepted: 27-09-2026	Published: 31-10-2026

ABSTRAK Asesmen dalam pembelajaran Bahasa Arab sebagai bahasa asing (*Arabic as a Foreign Language*) saat ini sedang mengalami perubahan paradigma yang cukup signifikan. Perubahan tersebut dipengaruhi oleh perkembangan teknologi digital, pemanfaatan Kecerdasan Buatan (AI), serta penerapan Kurikulum Merdeka di Indonesia yang menekankan fleksibilitas, diferensiasi, dan personalisasi pembelajaran. Artikel ini merupakan *Systematic Literature Review* (SLR) yang bertujuan mengidentifikasi, mengkaji, dan mensintesis perkembangan, inovasi, pendekatan metodologis, serta berbagai tantangan dalam asesmen empat keterampilan berbahasa Arab, yaitu menyimak, berbicara, membaca, dan menulis, selama periode 2020 hingga awal 2026. Kajian ini menggunakan protokol PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) untuk menelusuri dan menganalisis literatur yang relevan. Hasil kajian memperlihatkan adanya pergeseran dari sistem evaluasi konvensional berbasis tes diskrit menuju penerapan *Technology-Enhanced Assessment* (TEA) dan bentuk asesmen yang lebih autentik. Sintesis literatur menunjukkan beberapa temuan utama. Pertama, asesmen diagnostik, baik pada aspek kognitif maupun nonkognitif, memiliki posisi penting dalam perancangan pembelajaran berdiferensiasi pada kerangka Kurikulum Merdeka. Kedua, pengembangan instrumen penilaian *maharah* bahasa Arab semakin diarahkan pada pengukuran kemampuan berpikir tingkat tinggi (*Higher Order Thinking Skills*/HOTS). Ketiga, pemanfaatan teknologi *Generative AI*, termasuk ChatGPT, serta berbagai platform pembelajaran berbasis gamifikasi telah membuka pendekatan baru dalam penyediaan umpan balik (*feedback*) yang lebih cepat, adaptif, dan interaktif. Berdasarkan hasil sintesis, efektivitas asesmen modern tidak semata-mata ditentukan oleh tingkat kecanggihan teknologi yang digunakan, tetapi juga oleh kesesuaian teknologi dengan tujuan pedagogis, keterbukaan prosedur penilaian, serta kemampuannya dalam menggambarkan kompetensi komunikatif peserta didik secara menyeluruh dengan tetap menjaga prinsip integritas akademik.

KATA KUNCI: Asesmen Bahasa Arab; Kurikulum Merdeka; *Technology-Enhanced Assessment*; Asesmen Autentik; *Systematic Literature Review*; Kecerdasan Buatan; Keterampilan Berpikir Tingkat Tinggi.

ABSTRACT *Assessment in Arabic language learning as a foreign language (Arabic as a Foreign Language) is currently undergoing a significant paradigm shift. This transformation is influenced by the development of digital technology, the utilization of Artificial Intelligence (AI), and the implementation of the Merdeka Curriculum in Indonesia, which emphasizes flexibility, differentiation, and learning personalization. This article presents a Systematic Literature Review (SLR) aimed at identifying, examining, and synthesizing developments, innovations, methodological approaches, and various challenges in the assessment of the four Arabic language skills, namely listening, speaking, reading, and writing, during the period from 2020 to early 2026. The review employed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) protocol to identify and analyze relevant literature. The findings indicate a shift from conventional evaluation systems based on discrete tests toward the implementation of Technology-Enhanced Assessment (TEA) and more authentic forms of assessment. The literature synthesis reveals several key findings. First, diagnostic assessment, encompassing both*

© 2026 by the authors; This is an Open Access article distributed under the terms of the Creative Commons Attribution-ShareAlike 4.0 International License. (https://creativecommons.org/licenses/by-sa/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

cognitive and non-cognitive aspects, plays an important role in designing differentiated learning within the framework of the Merdeka Curriculum. Second, the development of assessment instruments for Arabic language maharah is increasingly oriented toward measuring higher-order thinking skills (Higher Order Thinking Skills/HOTS). Third, the utilization of Generative AI, including ChatGPT, as well as various gamification-based learning platforms, has introduced new approaches to providing faster, more adaptive, and interactive feedback. Based on the synthesis, the effectiveness of modern assessment is not determined solely by the sophistication of the technology employed, but also by the alignment between technology and pedagogical objectives, the transparency of assessment procedures, and its ability to comprehensively represent learners' communicative competence while maintaining the principles of academic integrity.

KEYWORDS: *Arabic Language Assessment; Merdeka Curriculum; Technology-Enhanced Assessment; Authentic Assessment; Systematic Literature Review; Artificial Intelligence; Higher-Order Thinking Skills.*

1. PENDAHULUAN

Pembelajaran Bahasa Arab di Indonesia, mulai dari jenjang pendidikan dasar dan menengah hingga perguruan tinggi, memiliki perjalanan pedagogis yang panjang serta terus mengalami perkembangan. Dalam praktik tradisional, pembelajaran dan evaluasi Bahasa Arab sering kali masih dipengaruhi oleh pendekatan *Grammar-Translation Method* (Thariqah Qawa'id wa Tarjamah), yang memberikan penekanan besar pada penguasaan kaidah tata bahasa, khususnya *nahwu* dan *sharaf*, serta penguasaan kosakata (*mufradat*) secara terpisah (Rahman et al., 2022). Orientasi tersebut turut memengaruhi karakteristik asesmen yang diterapkan. Instrumen evaluasi cenderung menggunakan tes tertulis konvensional (*paper-based testing*) dan lebih banyak mengukur kemampuan kognitif pada tingkat mengingat dan memahami. Meskipun model tes diskrit memiliki keunggulan dari sisi kemudahan pelaksanaan dan administrasi, pendekatan tersebut belum sepenuhnya mampu merepresentasikan kompetensi komunikatif peserta didik dalam situasi penggunaan bahasa yang autentik (Boeriswati et al., 2023).

Perkembangan pendidikan pada abad ke-21, ditambah dengan tuntutan globalisasi dan kemunculan paradigma *Society 5.0*, mendorong lembaga pendidikan untuk melakukan peninjauan kembali terhadap cara kompetensi berbahasa dikembangkan dan diukur (Ritonga et al., 2023). Dalam konteks Indonesia, perubahan tersebut semakin memperoleh landasan melalui penerapan Kurikulum Merdeka. Kurikulum ini tidak hanya membawa perubahan pada struktur dan pengorganisasian pembelajaran, tetapi juga memperkenalkan orientasi *Merdeka Belajar* yang menempatkan peserta didik sebagai pusat proses pendidikan. Konsekuensinya, fungsi asesmen tidak lagi terbatas sebagai instrumen untuk mengukur capaian belajar (*assessment of learning*) pada akhir proses pembelajaran, tetapi juga digunakan untuk mengarahkan dan memperbaiki proses belajar (*assessment for learning*) serta mendorong peserta didik melakukan refleksi terhadap proses belajarnya (*assessment as learning*) (Laudza'i et al., 2026). Dalam kondisi tersebut, pendidik perlu mengembangkan asesmen diagnostik, formatif, dan sumatif yang dapat merespons perbedaan kesiapan belajar, karakteristik peserta didik, serta kebutuhan pembelajaran yang beragam (Muktamar et al., 2023).

Perubahan paradigma kurikulum tersebut berlangsung bersamaan dengan transformasi digital yang semakin cepat setelah pandemi COVID-19. Kondisi ini turut memperluas penggunaan *Technology-Enhanced Assessment* (TEA) dalam pembelajaran Bahasa Arab. Pada fase awal perkembangannya, TEA banyak diwujudkan melalui penggunaan platform berbasis gamifikasi, seperti Quizizz, Kahoot!, dan Wordwall. Berbagai platform tersebut memberikan alternatif terhadap pelaksanaan evaluasi konvensional dengan menghadirkan aktivitas penilaian yang lebih

interaktif dan menarik, terutama dalam konteks asesmen formatif (Ichsani et al., 2025; Syagif et al., 2021). Perkembangan berikutnya memperlihatkan meningkatnya perhatian terhadap pemanfaatan *Generative Artificial Intelligence* (GenAI). Teknologi seperti ChatGPT dan berbagai *Large Language Models* (LLMs) mulai dimanfaatkan untuk mendukung beragam aktivitas pembelajaran dan asesmen, mulai dari penyusunan teks bacaan hingga pengembangan sistem yang dapat membantu mengevaluasi performa berbicara (*speech recognition*) dan menganalisis struktur sintaksis atau *i'rab* (Xia et al., 2024; Wang et al., 2025; Nafis, 2026). Di sisi lain, pemanfaatan AI dalam asesmen juga menimbulkan sejumlah persoalan terkait validitas konstruk, keterbukaan mekanisme algoritmik, reliabilitas hasil penilaian, dan integritas akademik mahasiswa (Creely, 2025).

Selain perkembangan teknologi, kualitas instrumen asesmen Bahasa Arab kontemporer juga semakin berkaitan dengan kemampuannya dalam mengukur *Higher Order Thinking Skills* (HOTS). Penilaian kemampuan berbahasa tidak lagi cukup apabila hanya mengukur kemampuan peserta didik dalam menerjemahkan atau memahami informasi tekstual. Asesmen juga perlu memberikan ruang bagi peserta didik untuk menganalisis informasi, mengevaluasi argumen dan konteks budaya, serta menghasilkan wacana Bahasa Arab yang memiliki keterpaduan dan keterkaitan antargagasan (Delima et al., 2025; Rosyadi, 2021). Orientasi tersebut menunjukkan bahwa asesmen keterampilan berbahasa memiliki karakter yang semakin kompleks dan multidimensional. Oleh karena itu, pengembangan rubrik penilaian yang sistematis, komprehensif, dan objektif menjadi penting, terutama dalam mengukur keterampilan produktif seperti berbicara (*kalam*) dan menulis (*kitabah*).

Meskipun penelitian mengenai pembelajaran dan pengajaran Bahasa Arab telah berkembang secara luas, kajian yang secara sistematis mengintegrasikan perkembangan Kurikulum Merdeka, transformasi menuju asesmen autentik berbasis HOTS, serta perkembangan teknologi kecerdasan buatan dalam satu kerangka analisis masih relatif terbatas. Kesenjangan tersebut menjadi dasar bagi artikel ini untuk melakukan *Systematic Literature Review* secara komprehensif terhadap perkembangan asesmen Bahasa Arab. Kajian ini diarahkan pada tiga tujuan utama, yaitu: (1) memetakan perkembangan dan inovasi instrumen asesmen pada empat keterampilan berbahasa Arab; (2) mengkaji peran dan efektivitas asesmen diagnostik serta formatif dalam konteks penerapan Kurikulum Merdeka; dan (3) menganalisis peluang serta tantangan yang muncul dari pemanfaatan *Technology-Enhanced Assessment* dan kecerdasan buatan dalam asesmen Bahasa Arab. Sintesis hasil penelitian diharapkan dapat memperkaya landasan teoretis dalam bidang linguistik terapan sekaligus memberikan implikasi praktis bagi pendidik, pengembang kurikulum, dan pemangku kebijakan dalam merancang sistem asesmen Bahasa Arab yang relevan dengan perkembangan pendidikan dan teknologi.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan *Systematic Literature Review* (SLR), yaitu metode penelitian yang dilakukan melalui prosedur sistematis, eksplisit, dan dapat direproduksi untuk mengidentifikasi, mengevaluasi, serta mengintegrasikan berbagai hasil penelitian yang telah dilakukan oleh peneliti maupun praktisi (Booth, 2016). Untuk menjaga transparansi proses kajian sekaligus mengurangi potensi bias dalam pemilihan sumber, seluruh tahapan pelaksanaan SLR mengacu secara ketat pada pedoman PRISMA 2020 (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) (Page et al., 2021).

2.1 Tahapan Identifikasi dan Strategi Pencarian (*Search Strategy*)

Penelusuran literatur primer dilakukan secara sistematis dan komprehensif dengan memanfaatkan beberapa pangkalan data akademik utama, yaitu Google Scholar, *Directory of Open Access Journals* (DOAJ), Scopus, dan Garuda (Garba Rujukan Digital). Selain penelusuran berdasarkan basis data tersebut, pencarian juga diarahkan pada sejumlah jurnal yang memiliki reputasi dalam bidang pendidikan dan pembelajaran Bahasa Arab di Indonesia, di antaranya *Arabi*, *Arabiyat*, *Al-Bayan*, dan *Al-Mahara*.

Strategi pencarian dikembangkan dengan menggunakan kombinasi kata kunci dan operator Boolean (AND, OR). Kueri pencarian disusun berdasarkan tiga domain utama yang merepresentasikan fokus penelitian, yaitu domain evaluasi, domain subjek, dan domain konteks inovasi, sebagai berikut:

1. **Domain Evaluasi:** ("asesmen" OR "evaluasi" OR "penilaian" OR "assessment" OR "rubrik" OR "diagnostic assessment" OR "formative assessment").
2. **Domain Subjek:** ("bahasa Arab" OR "Arabic language" OR "maharah" OR "maharat al-lughah" OR "istima" OR "kalam" OR "qira'ah" OR "kitabah").
3. **Domain Konteks Inovasi:** ("Kurikulum Merdeka" OR "HOTS" OR "Technology-Enhanced Assessment" OR "Generative AI" OR "ChatGPT" OR "Authentic Assessment").

Kombinasi ketiga domain tersebut digunakan untuk memperoleh literatur yang relevan dengan fokus kajian, khususnya penelitian yang membahas asesmen Bahasa Arab dalam kaitannya dengan perkembangan kurikulum, keterampilan berpikir tingkat tinggi, teknologi digital, kecerdasan buatan, dan asesmen autentik.

2.2 Kriteria Inklusi dan Eksklusi (*Eligibility Criteria*)

Proses seleksi literatur dilakukan dengan menerapkan seperangkat kriteria inklusi dan eksklusi secara konsisten. Penerapan kriteria tersebut dimaksudkan untuk memastikan bahwa artikel yang dianalisis benar-benar memiliki keterkaitan dengan fokus penelitian, terutama mengenai instrumen, pendekatan, metodologi, dan inovasi dalam asesmen Bahasa Arab. Dengan demikian, literatur yang hanya membahas media pembelajaran atau model pembelajaran secara umum tanpa memiliki keterkaitan yang memadai dengan aspek asesmen tidak menjadi fokus utama dalam kajian ini.

Tabel 1. Kriteria inklusi dan eksklusi yang digunakan dalam proses penyaringan literatur.

Parameter	Kriteria Inklusi	Kriteria Eksklusi
Fokus Studi	Artikel yang secara spesifik membahas metodologi, pengembangan instrumen, rubrik, platform teknologi, atau evaluasi hasil asesmen dalam pembelajaran Bahasa Arab.	Studi umum tentang metode pengajaran, kurikulum, atau pengembangan media pembelajaran Bahasa Arab yang tidak membahas aspek evaluasi atau instrumen asesmen.
Konteks Tematik	Membahas minimal satu dari elemen berikut: asesmen empat keterampilan berbahasa (maharat), Kurikulum Merdeka, *Technology-Enhanced Assessment* (TEA), AI dalam pendidikan bahasa, atau evaluasi HOTS.	Studi pendidikan Islam umum, tafsir, atau kajian sastra Arab murni yang tidak memiliki elemen pedagogi linguistik dan asesmen.
Rentang Waktu	Literatur yang dipublikasikan dari tahun 2020 hingga tahun 2026.	Literatur yang dipublikasikan sebelum tahun 2020 (untuk menjaga relevansi dengan teknologi pasca-

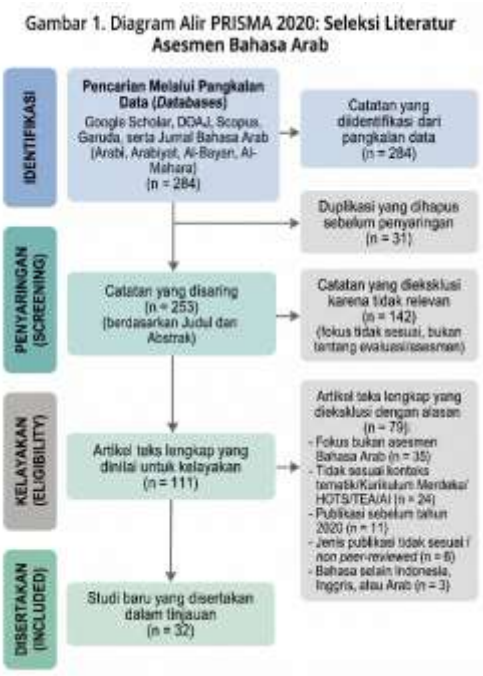
		pandemi dan Kurikulum Merdeka).
Jenis Publikasi	Artikel jurnal ilmiah *peer-reviewed*, prosiding konferensi akademik, serta tesis/disertasi yang dipublikasikan.	Artikel opini, blog, majalah populer, laporan tinjauan buku, atau dokumen tanpa proses *peer-review*.
Bahasa	Ditulis dalam Bahasa Indonesia, Bahasa Inggris, atau Bahasa Arab.	Ditulis dalam bahasa selain tiga bahasa tersebut.

2.3 Seleksi Literatur (PRISMA Flow) dan Analisis Data

Proses seleksi literatur dilakukan dengan mengikuti tahapan dalam alur PRISMA yang mencakup empat fase utama. Tahap pertama adalah Identifikasi, yaitu proses menghimpun records dari berbagai pangkalan data yang telah ditentukan. Tahap kedua adalah Screening, yang meliputi penghapusan artikel duplikat serta pemeriksaan judul dan abstrak untuk menilai relevansinya. Tahap ketiga adalah Eligibility, yakni penelaahan terhadap teks lengkap artikel guna memastikan kesesuaiannya dengan kriteria inklusi dan eksklusi. Tahap terakhir adalah Included, yaitu penetapan artikel yang memenuhi seluruh persyaratan untuk selanjutnya dianalisis dalam kajian sistematis (Laudza'i et al., 2026).

Artikel yang telah melewati seluruh proses seleksi kemudian dianalisis dan diekstraksi ke dalam suatu matriks data yang mencakup informasi konseptual dan temuan utama dari masing-masing penelitian. Selanjutnya, sintesis dilakukan dengan menerapkan analisis tematik secara induktif (Miles et al., 2014). Melalui proses pengodean yang dilakukan secara berulang, berbagai temuan dalam literatur dikelompokkan ke dalam lima tema utama yang menjadi dasar pembahasan artikel ini, yaitu: (a) rekonseptualisasi asesmen dalam kerangka Kurikulum Merdeka; (b) transformasi dari pengujian konvensional menuju asesmen autentik; (c) pengembangan kembali instrumen asesmen berbasis HOTS; (d) inovasi taksonomi dalam penilaian empat maharah; serta (e) transformasi yang ditimbulkan oleh teknologi digital dan Generative AI.

Gambar 1. Diagram Alir PRISMA 2020: Seleksi Literatur Asesmen Bahasa Arab



3. HASIL DAN DISKUSI

Hasil telaah sistematis terhadap berbagai publikasi akademik terkini menunjukkan adanya perkembangan yang dinamis dalam sistem asesmen Bahasa Arab. Transformasi tersebut tidak hanya terlihat pada perubahan bentuk dan instrumen penilaian, tetapi juga mencerminkan perubahan orientasi pedagogis yang mendasarinya. Berdasarkan hasil sintesis, pembahasan selanjutnya dikelompokkan ke dalam lima klaster utama yang menguraikan perkembangan prosedural sekaligus implikasi pedagogis dan konseptual dari perubahan dalam asesmen Bahasa Arab.

3.1 Restrukturisasi Asesmen Bahasa Arab dalam Arsitektur Kurikulum Merdeka

Kurikulum Merdeka membawa orientasi terhadap penyederhanaan dan pengurangan beban materi pembelajaran, dengan menempatkan penguasaan kompetensi esensial serta pembentukan karakter peserta didik melalui Profil Pelajar Pancasila sebagai salah satu fokus utama (Muktamar et al., 2023). Dalam pembelajaran Bahasa Arab, fleksibilitas tersebut tercermin melalui perubahan fungsi dan kedudukan asesmen. Sejumlah literatur menunjukkan bahwa asesmen tidak lagi diposisikan semata-mata sebagai sarana menghasilkan nilai pada rapor, tetapi berkembang menjadi instrumen yang terintegrasi dengan keseluruhan siklus pembelajaran (Khoiriyah, 2020; Asrori, 2021). Implementasinya dapat dipahami melalui tiga bentuk asesmen yang saling melengkapi, yaitu Asesmen Diagnostik, Asesmen Formatif, dan Asesmen Sumatif.

3.1.1 Urgensi Asesmen Diagnostik Kognitif dan Non-Kognitif

Salah satu perkembangan penting yang banyak ditemukan dalam literatur setelah tahun 2021 adalah semakin kuatnya penekanan terhadap pelaksanaan Asesmen Diagnostik pada awal pembelajaran atau ketika memasuki materi baru. Asesmen ini menjadi landasan penting dalam penerapan Pembelajaran Berdiferensiasi (*Differentiated Learning*) karena memberikan gambaran mengenai kondisi awal peserta didik sebelum proses pembelajaran dirancang (Budiono & Hatip, 2023; Fitri, 2022). Dalam pembelajaran bahasa asing, asesmen diagnostik dapat dikelompokkan ke dalam dua ranah utama sebagai berikut.

1. **Asesmen Diagnostik Kognitif:** Asesmen ini bertujuan mengidentifikasi pengetahuan dan kemampuan awal (*prior knowledge*) peserta didik yang berkaitan dengan materi Bahasa Arab. Instrumen yang digunakan dapat mengukur penguasaan kompetensi linguistik dasar, seperti kemampuan mengenali kata ganti (*isim dhamir*), memberikan respons terhadap ungkapan sapaan (*tahiyyat*), maupun membaca teks Bahasa Arab tanpa harakat (Aini & Alwi, 2024; Suprapti & Ridho, 2024). Informasi mengenai kompetensi awal tersebut dapat membantu guru menyesuaikan Capaian Pembelajaran (CP) dan Alur Tujuan Pembelajaran (ATP) dengan tingkat kesiapan peserta didik dalam suatu kelas.
2. **Asesmen Diagnostik Non-Kognitif:** Aspek ini menunjukkan perluasan perhatian asesmen terhadap kondisi peserta didik yang sebelumnya relatif kurang mendapatkan perhatian dalam pembelajaran Bahasa Arab. Sejumlah penelitian (Viani & Ripai, 2024; Hairida et al., 2024; Kizi & Shadjalilovna, 2022) menunjukkan bahwa asesmen diagnostik non-kognitif dapat digunakan untuk memperoleh gambaran mengenai karakteristik psikososial peserta didik. Pemanfaatan platform digital seperti Google Forms atau Jotform memungkinkan guru menghimpun

informasi mengenai berbagai faktor yang berpotensi memengaruhi proses belajar, termasuk minat terhadap pembelajaran, kecenderungan gaya belajar seperti visual, auditori, dan kinestetik, dukungan dari lingkungan keluarga, serta tingkat kecemasan dalam mempelajari bahasa asing (*foreign language anxiety*).

Berdasarkan kajian terhadap instrumen diagnostik non-kognitif yang diterapkan dalam praktik pembelajaran (Kiptiah et al., 2023), sejumlah pertanyaan dapat diarahkan pada aspek afektif peserta didik, misalnya mengenai respons mereka ketika menghadapi soal Bahasa Arab yang dianggap sulit atau perasaan yang muncul ketika mengikuti pembelajaran Bahasa Arab. Informasi semacam ini dapat memberikan perspektif yang lebih luas mengenai kondisi peserta didik. Secara konseptual, penggunaan instrumen tersebut menunjukkan adanya perubahan posisi guru Bahasa Arab, dari yang sebelumnya lebih berorientasi pada evaluasi aspek kebahasaan menuju peran yang lebih luas sebagai fasilitator pembelajaran yang memperhatikan dimensi akademik dan psikologis peserta didik.

Apabila hasil asesmen menunjukkan adanya tingkat kecemasan yang tinggi dalam suatu kelompok belajar, misalnya karena pengalaman belajar tata bahasa (*nahwu*) yang kurang menyenangkan, guru dapat mempertimbangkan penyesuaian strategi pembelajaran. Pendekatan seperti *Total Physical Response* (TPR) maupun permainan bahasa interaktif dapat digunakan sebagai alternatif untuk menciptakan suasana belajar yang lebih nyaman sebelum peserta didik diarahkan pada materi gramatika yang lebih kompleks.

3.1.2 Revitalisasi Asesmen Formatif dan Reposisi Asesmen Sumatif

Dalam implementasi Kurikulum Merdeka, Asesmen Formatif memperoleh posisi yang semakin penting karena dilaksanakan secara berkelanjutan selama proses pembelajaran. Literatur menunjukkan bahwa asesmen formatif Bahasa Arab dapat dilakukan melalui berbagai teknik, seperti observasi, pencatatan perkembangan peserta didik (*anecdotal records*), serta pemberian umpan balik secara langsung (Shifany et al., 2025; Paulsen & Valdivia, 2022). Pendekatan tersebut memungkinkan guru memperoleh informasi secara berkala mengenai perkembangan pemahaman peserta didik dan menggunakannya untuk menyesuaikan strategi pembelajaran.

Dalam praktiknya, asesmen formatif dapat berfungsi sebagai mekanisme pemantauan yang membantu guru mengidentifikasi kesulitan belajar secara lebih cepat. Sebagai contoh, apabila sebagian besar peserta didik belum memahami konsep sintaksis (*tarkib*) tertentu, guru dapat segera mengubah metode, memberikan penjelasan tambahan, atau menyediakan aktivitas penguatan. Dengan demikian, asesmen tidak hanya menghasilkan informasi mengenai capaian belajar, tetapi juga menjadi bagian integral dari proses perbaikan pembelajaran.

Sementara itu, Asesmen Sumatif tetap memiliki fungsi dalam mengukur pencapaian kompetensi pada akhir periode pembelajaran. Namun, bentuk pelaksanaannya tidak lagi harus terbatas pada ujian tulis standar. Literatur menunjukkan adanya kecenderungan untuk mengembangkan bentuk asesmen yang lebih kontekstual dan beragam, seperti gelar karya (*exhibition*), portofolio, maupun pementasan proyek kebahasaan (Khasawneh, 2024). Perubahan ini menunjukkan bahwa asesmen sumatif dapat dirancang untuk memberikan gambaran capaian peserta didik melalui produk atau performa yang lebih representatif

terhadap kompetensi yang telah dikembangkan.

3.2 Transformasi Epistemologis: Dari Tes Konvensional Menuju Asesmen Autentik

Dalam sejarah pembelajaran Bahasa Arab, sistem evaluasi sering kali menghadapi persoalan berupa pemisahan keterampilan linguistik ke dalam komponen-komponen yang terpisah. Tes objektif berbentuk pilihan ganda (*multiple-choice*) dan isian singkat (*fill-in-the-blanks*) banyak digunakan karena relatif mudah dalam pelaksanaan, penskoran, dan pengelolaannya (Rahman et al., 2022). Meskipun memiliki keunggulan dalam aspek efisiensi dan objektivitas penskoran, instrumen tersebut memiliki keterbatasan dalam menggambarkan kemampuan peserta didik ketika menggunakan bahasa dalam konteks komunikasi yang sebenarnya.

Sebagai ilustrasi, tes yang hanya meminta peserta didik mengingat bentuk *fi'il madhi* dapat menunjukkan penguasaan terhadap aspek gramatikal tertentu, tetapi belum tentu dapat menggambarkan kemampuan mereka menggunakan bentuk tersebut secara tepat dalam situasi komunikasi nyata, seperti melakukan percakapan atau negosiasi ketika bertransaksi di pasar.

Kondisi tersebut mendorong berkembangnya perhatian terhadap penerapan Asesmen Autentik (*Authentic Assessment*). Sejumlah penelitian menunjukkan kecenderungan yang kuat terhadap penggunaan asesmen autentik dalam pembelajaran bahasa (Ardianto & Fakhirah, 2025; Marden & Herrington, 2025; Cahyani et al., 2024). Pendekatan ini menempatkan peserta didik pada tugas atau aktivitas yang memiliki keterkaitan dengan penggunaan bahasa dalam kehidupan nyata, sehingga kompetensi yang dinilai tidak hanya berupa pengetahuan mengenai bahasa, tetapi juga kemampuan menerapkannya dalam konteks komunikasi yang relevan (Fahruddin, 2025).

Dengan demikian, asesmen autentik memberikan ruang bagi peserta didik untuk memperlihatkan kemampuan berbahasa melalui tugas-tugas yang menyerupai situasi penggunaan bahasa secara nyata (*real-world tasks*). Pendekatan tersebut memungkinkan penilaian mencakup aspek pengetahuan, keterampilan, penerapan, dan kemampuan berkomunikasi secara lebih terpadu.

Tabel 2. Perbedaan karakteristik antara pendekatan asesmen konvensional dan asesmen autentik dapat disajikan secara lebih sistematis melalui.

Parameter Evaluasi	Penilaian Konvensional (Diskrit)	Penilaian Autentik (Kinerja/Proyek)
Fokus Kompetensi	Menilai elemen linguistik (tata bahasa, ejaan, kosakata) secara terpisah.	Menilai kemampuan komunikasi secara terintegrasi (kompetensi sosiolinguistik, wacana, pragmatik).
Sifat Instrumen	Reaktif; peserta didik memilih jawaban yang telah disediakan (Pilihan Ganda, Benar/Salah).	Produktif & Konstruktif; peserta didik membangun respons, menciptakan karya (Presentasi, Proyek, Simulasi).
Konteks Kognitif	Dekontekstual; kalimat yang diuji sering kali lepas dari konteks makna (hapalan kaidah).	Kontekstual; bahasa digunakan untuk menyelesaikan masalah nyata (misal: menyusun surat resmi bisnis).
Model Umpan Balik	Kuatitatif absolut (skor biner benar/salah), berorientasi pada produk akhir.	Kualitatif dan berkelanjutan (menggunakan rubrik bertingkat), berorientasi pada proses.
Validitas Ekologis	Rendah; tidak mencerminkan penggunaan bahasa di luar ruang kelas.	Tinggi; selaras dengan tuntutan dunia kerja dan interaksi lintas budaya.

Lebih lanjut, pergeseran menuju asesmen autentik juga membawa perubahan pada pola keterlibatan peserta didik dalam proses penilaian. Peserta didik tidak lagi hanya ditempatkan sebagai objek yang dinilai, tetapi mulai dilibatkan secara aktif melalui kegiatan penilaian diri (*self-assessment*) dan penilaian sejawat (*peer-assessment*). Hasil kajian menunjukkan bahwa keterlibatan peserta didik dalam memahami dan menggunakan rubrik penilaian dapat memperkuat kemampuan metakognitif, mengurangi ketergantungan pada penilaian dari satu pihak, serta meningkatkan rasa memiliki terhadap proses pembelajaran Bahasa Arab (Alshakhi, 2021).

3.3 Integrasi Keterampilan Berpikir Tingkat Tinggi (HOTS)

Karakteristik pembelajaran abad ke-21 menuntut adanya perubahan dari pola pembelajaran yang berorientasi pada hafalan (*rote learning*) menuju pengembangan kemampuan berpikir yang lebih kompleks. Dalam konteks tersebut, asesmen Bahasa Arab perlu diarahkan pada pengukuran *Higher Order Thinking Skills* (HOTS), terutama kemampuan kognitif tingkat tinggi berdasarkan Taksonomi Bloom yang telah direvisi, yaitu menganalisis (C4), mengevaluasi (C5), dan mencipta (C6) (Anderson & Krathwohl, 2001; Delima et al., 2025; Rosyadi, 2021).

Hasil penelusuran literatur memperlihatkan bahwa penerapan instrumen berbasis HOTS dalam pembelajaran Bahasa Arab mulai memperoleh ruang dalam buku teks maupun instrumen Penilaian Akhir Semester (PAS) (Fauziah et al., 2020; Elmi et al., 2024; Widodo, 2021). Pada asesmen yang masih berorientasi *Lower Order Thinking Skills* (LOTS), pertanyaan umumnya hanya mengharuskan peserta didik menerjemahkan suatu kalimat atau mengidentifikasi arti sebuah kosakata. Sebaliknya, asesmen berbasis HOTS menuntut peserta didik untuk menghubungkan informasi, menyusun argumentasi, melakukan evaluasi, dan menghasilkan pemikiran berdasarkan informasi yang tersedia.

Sebagai contoh, dalam asesmen keterampilan *Qira'ah* berbasis HOTS, peserta didik dapat diberikan sebuah teks opini berbahasa Arab yang membahas dampak perkembangan ekonomi digital di kawasan Timur Tengah. Pertanyaan tidak berhenti pada pemahaman isi teks, tetapi dapat meminta peserta didik mengevaluasi kesimpulan penulis mengenai pengaruh teknologi terhadap budaya lokal serta menyampaikan pandangan mereka dengan menggunakan struktur sintaksis Bahasa Arab yang logis. Model penilaian semacam ini memungkinkan kemampuan membaca sekaligus kemampuan berpikir kritis dan kemampuan produksi bahasa dinilai secara terpadu.

Pengembangan instrumen HOTS menunjukkan bahwa penguasaan tata bahasa (*nahwu/sharaf*) tidak lagi ditempatkan sebagai tujuan akhir pembelajaran, tetapi berfungsi sebagai perangkat linguistik yang mendukung peserta didik dalam mengungkapkan gagasan dan melakukan penalaran. Dengan demikian, hasil asesmen HOTS dapat memberikan gambaran yang lebih luas mengenai kesiapan kognitif peserta didik dalam menghadapi dan berpartisipasi dalam wacana yang kompleks (Muradi et al., 2020).

3.4 Inovasi Asesmen pada Taksonomi Empat Keterampilan Berbahasa (*Maharat al-Lughah*)

Ketepatan suatu instrumen asesmen sangat dipengaruhi oleh jenis keterampilan bahasa yang hendak diukur. Dalam pembelajaran Bahasa Arab, *maharat al-lughah* mencakup keterampilan reseptif, yaitu menyimak dan membaca, serta keterampilan produktif, yaitu berbicara dan menulis. Setiap keterampilan memiliki karakteristik yang berbeda sehingga membutuhkan

pendekatan dan teknik asesmen yang disesuaikan dengan kompetensi yang hendak diukur (Ma'arif & Bakhiet, 2026).

3.4.1 Asesmen Maharah Istima' (Menyimak)

Keterampilan *istima'* dalam sistem evaluasi konvensional sering memperoleh perhatian yang relatif terbatas. Kondisi tersebut antara lain berkaitan dengan keterbatasan fasilitas pendukung dan belum optimalnya ketersediaan instrumen asesmen yang inovatif. Dalam praktik tradisional, asesmen menyimak bahkan sering hanya dilakukan dengan cara guru membacakan teks atau melaksanakan kegiatan *imla'*. Namun, kajian mutakhir menunjukkan adanya perkembangan penggunaan media audio-visual yang menghadirkan materi dari penutur asli (*native speaker*) (Chabib et al., 2025; Sa'adah, 2025).

Salah satu inovasi yang berkembang adalah pemanfaatan kode QR pada Lembar Kerja Siswa (LKS) yang terhubung dengan file audio autentik. Model tersebut memberikan kesempatan kepada peserta didik untuk mengakses materi audio secara lebih mandiri dan mengontrol proses asesmen secara lebih fleksibel. Selain itu, aplikasi pengembangan animasi seperti Plotagon dapat digunakan untuk membuat skenario percakapan Bahasa Arab yang kontekstual. Penggunaan animasi tersebut memberikan stimulus pendengaran yang lebih variatif sekaligus menghadirkan dukungan visual dalam proses pemahaman bahasa (Rini et al., 2023).

3.4.2 Asesmen Maharah Kalam (Berbicara)

Asesmen keterampilan *kalam* atau berbicara merupakan salah satu bentuk penilaian yang memiliki tantangan cukup kompleks, terutama berkaitan dengan validitas dan reliabilitas. Penilaian performa berbicara memiliki kemungkinan yang lebih besar untuk dipengaruhi subjektivitas guru sebagai *rater*. Oleh sebab itu, salah satu inovasi penting dalam bidang ini adalah pengembangan serta standardisasi **Rubrik Penilaian Kinerja** yang disusun berdasarkan indikator yang jelas dan terukur (Setyawan et al., 2025).

Pada jenjang pendidikan menengah maupun perguruan tinggi, asesmen *kalam* dapat dilakukan melalui berbagai bentuk performa, seperti pembuatan *vlog* berbahasa Arab, debat, dan simulasi percakapan melalui kegiatan *role-play*. Bentuk-bentuk tersebut memungkinkan peserta didik menunjukkan kemampuan menggunakan Bahasa Arab dalam situasi komunikasi yang lebih kontekstual.

Rubrik penilaian untuk tugas performatif tersebut dapat mencakup beberapa dimensi utama sebagai berikut:

1. **Akurasi (Ketepatan):** Kemampuan menggunakan kaidah tata bahasa secara tepat, menghasilkan artikulasi fonologis yang sesuai dengan *makharijul huruf*, serta memilih kosakata (*mufradat*) yang relevan dengan konteks komunikasi.
2. **Fluensi (Kelancaran):** Kemampuan mempertahankan kesinambungan tuturan dengan kecepatan yang wajar serta meminimalkan jeda yang dapat mengganggu kelancaran komunikasi.
3. **Intonasi:** Kemampuan mengatur tinggi-rendah nada suara (*pitch*) sesuai dengan fungsi dan konteks pragmatis tuturan, misalnya dalam membedakan kalimat interogatif dan deklaratif.

4. **Ekspresi dan Komunikasi Non-Verbal:** Kemampuan menggunakan ekspresi wajah, kontak mata, dan gestur tubuh untuk mendukung penyampaian pesan secara efektif (Widodo & Qosim, 2021).

   Penerapan rubrik yang terstruktur dan konsisten dapat membantu meningkatkan keadilan dalam proses penilaian sekaligus mengurangi variasi subjektivitas antarpeserta didik. Selain itu, hasil penilaian dapat dimanfaatkan sebagai umpan balik formatif yang lebih spesifik sehingga peserta didik mengetahui aspek kemampuan berbicara yang masih perlu dikembangkan (Fahruddin, 2025).

3.4.3 Asesmen Maharah Qira'ah (Membaca)

Asesmen *qira'ah* mengalami perkembangan dari model yang semata-mata berorientasi pada penerjemahan teks secara literal menuju aktivitas literasi yang menuntut kemampuan membaca secara analitis dan kritis. Perkembangan teknologi digital turut mendukung perubahan tersebut melalui pemanfaatan *hypertext* yang memungkinkan peserta didik menjelajahi teks sekaligus memperoleh dukungan multimedia ketika menemukan kosakata atau informasi yang sulit dipahami (Ma'arif & Bakhiet, 2026).

Dalam perkembangannya, asesmen *qira'ah* semakin diarahkan pada kemampuan literasi yang lebih kompleks. Peserta didik tidak hanya dituntut memahami informasi eksplisit, tetapi juga mampu melakukan inferensi, menemukan gagasan utama yang tersirat, serta menganalisis sudut pandang atau bias penulis dalam teks kontemporer. Perkembangan teknologi AI juga mulai dimanfaatkan untuk menghasilkan bahan bacaan yang dapat disesuaikan dengan tingkat kemampuan peserta didik, termasuk berdasarkan tingkat kompleksitas kosakata atau *Lexile* (Allaithy & Zaki, 2025).

3.4.4 Asesmen Maharah Kitabah (Menulis)

Sebagai salah satu keterampilan produktif yang membutuhkan kemampuan kognitif kompleks, asesmen *kitabah* mengalami perubahan dari model yang berfokus pada dikte, kaligrafi, dan modifikasi struktur kalimat menuju kemampuan menghasilkan wacana akademik (*insya'*) (Munawarah & Zulkiflih, 2021). Perubahan tersebut menunjukkan adanya pergeseran cara pandang terhadap aktivitas menulis, dari sekadar menghasilkan produk akhir menjadi suatu proses yang berlangsung secara bertahap dan berulang.

Konsep menulis sebagai proses iteratif dapat diwujudkan melalui asesmen berbasis portofolio. Dalam pendekatan ini, tulisan peserta didik tidak hanya dinilai pada produk akhirnya, tetapi juga melalui proses penyusunan draf, pemberian umpan balik, perbaikan, dan revisi yang dilakukan secara berulang sepanjang periode pembelajaran (Purnomo et al., 2025).

Perkembangan lainnya terlihat pada upaya pengembangan rubrik penilaian *kitabah* yang diselaraskan dengan parameter Kerangka Kualifikasi Nasional Indonesia (KKNI). Penyelarasan tersebut diharapkan dapat menghasilkan standar kompetensi yang lebih terukur sehingga kompetensi lulusan program studi Bahasa Arab di Indonesia memiliki keterbandingan dengan standar kualifikasi yang berlaku secara lebih luas (Boeriswati et al., 2023).

3.5 Disrupsi *Technology-Enhanced Assessment* (TEA) dan Kecerdasan Buatan (AI)

Periode 2021 hingga 2026 menunjukkan percepatan transformasi asesmen dari pendekatan analog menuju *Technology-Enhanced Assessment* (TEA) (Laudza'i et al., 2026). Perkembangan ini dapat dipahami melalui dua gelombang utama, yaitu pemanfaatan gamifikasi digital dan munculnya teknologi *Generative Artificial Intelligence* (GenAI).

3.5.1 Ekosistem Gamifikasi dan Asesmen Digital Interaktif

Penggunaan platform asesmen berbasis permainan, seperti Kahoot!, Quizizz, Wordwall, dan iSpring Suite, berkembang sebagai salah satu alternatif dalam pelaksanaan asesmen formatif pada berbagai jenjang pendidikan (Chabib et al., 2025; Ichsani et al., 2025). Platform tersebut mengintegrasikan elemen permainan, termasuk *Points, Badges, and Leaderboards* (PBL), ke dalam kegiatan penilaian. Pendekatan tersebut dapat menciptakan suasana evaluasi yang lebih interaktif, menghadirkan unsur kompetisi, dan membantu mengurangi ketegangan peserta didik ketika mengikuti proses penilaian (Aftina & Muasomah, 2025; Syagif et al., 2021).

Hasil tinjauan meta-analisis yang dilakukan oleh Li et al. (2024) menunjukkan bahwa gamifikasi dalam asesmen dapat memberikan kontribusi terhadap motivasi intrinsik dan persepsi otonomi peserta didik. Namun, kajian tersebut juga menunjukkan adanya keterbatasan apabila gamifikasi hanya diterapkan pada bentuk kuis objektif yang mengutamakan kecepatan menjawab. Pengaruhnya terhadap pengembangan kompetensi linguistik yang lebih kompleks, seperti kemampuan menyusun argumentasi secara lisan, dapat menjadi terbatas. Oleh karena itu, penggunaan kuis digital perlu diseimbangkan dengan asesmen autentik yang berbasis performa agar pengukuran kompetensi bahasa tidak berhenti pada kemampuan menjawab soal secara cepat.

3.5.2 *Generative AI*: Peluang, Otomatisasi, dan Tantangan Etis

Kemunculan ChatGPT pada akhir 2022 menjadi salah satu titik penting dalam perkembangan teknologi asesmen pembelajaran Bahasa Arab. Kehadiran GenAI membawa perubahan signifikan terhadap cara aktivitas evaluasi dirancang, dilaksanakan, dan diberikan umpan baliknya (Xia et al., 2024). Dari sisi teknologi, pemrosesan bahasa alami (*Natural Language Processing*/NLP) untuk Bahasa Arab memiliki sejumlah tantangan karena karakteristik morfologi yang kompleks, sistem penulisan dari kanan ke kiri (*right-to-left*/RTL), serta penggunaan diakritik atau *harakat*. Perkembangan AI menunjukkan kemampuan yang semakin baik dalam menangani karakteristik tersebut.

Sejumlah penelitian (Dhouib et al., 2022; Khasawneh et al., 2025) menunjukkan beberapa bentuk pemanfaatan AI dalam asesmen Bahasa Arab, antara lain:

1. **Analisis Sintaksis (*I'rab*) Otomatis:** Teknologi AI, termasuk ChatGPT, dapat digunakan untuk membantu menganalisis fungsi sintaksis kata dalam struktur kalimat yang kompleks, seperti *Mubtada, Khabar, Fa'il,* dan *Maf'ul bihi*. Kemampuan tersebut berpotensi dimanfaatkan sebagai sarana pendukung *self-assessment* bagi peserta didik.
2. **AI-Mediated Language Assessment:** Pemanfaatan *Automatic Speech Recognition* (ASR) untuk Bahasa Arab membuka peluang bagi evaluasi pelafalan dan aspek fonologis tuturan secara komputasional. Teknologi tersebut dapat membantu

mengurangi beban waktu yang diperlukan untuk melakukan evaluasi lisan secara manual oleh guru (Dhouib et al., 2022).

3. **Automasi Pembangkitan Butir Soal (*Item Generation*):** AI dapat dimanfaatkan untuk memproses teks berbahasa Arab, memperkirakan tingkat kompleksitas bacaan, serta menghasilkan pertanyaan yang mengarah pada kemampuan HOTS dalam waktu yang relatif singkat (Allaithy & Zaki, 2025).

Meskipun memberikan berbagai peluang dalam efisiensi proses asesmen, integrasi AI juga menghadirkan persoalan mendasar mengenai validitas konstruk (Jurāne-Brēmane, 2023; Ma'arif & Bakhiet, 2026). Ketika peserta didik menggunakan ChatGPT dalam menyelesaikan tugas *insya'* Bahasa Arab, muncul pertanyaan mengenai kompetensi yang sebenarnya sedang diukur. Apakah asesmen tersebut merepresentasikan kemampuan menulis Bahasa Arab peserta didik atau justru kemampuan mereka dalam memberikan instruksi dan memanfaatkan AI melalui *prompt engineering*?

Persoalan tersebut menunjukkan pentingnya merumuskan batas penggunaan AI dalam proses asesmen. Berbagai penelitian menekankan perlunya transparansi prosedural, akuntabilitas, dan literasi etis dalam pemanfaatan teknologi AI agar penggunaan teknologi tidak mengurangi validitas penilaian maupun integritas akademik (Nafis, 2026; Creely, 2025).

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Berdasarkan hasil sintesis *Systematic Literature Review* terhadap publikasi pada periode 2020 hingga 2026, dapat dilihat bahwa asesmen keterampilan berbahasa Arab sedang mengalami transformasi yang signifikan. Sistem evaluasi yang sebelumnya banyak berorientasi pada pengujian diskrit dan penguasaan aspek gramatikal mulai berkembang menuju model asesmen yang lebih holistik, autentik, dan berorientasi pada performa. Dalam konteks Indonesia, penerapan Kurikulum Merdeka turut memperkuat perubahan tersebut dengan memberikan posisi penting kepada Asesmen Diagnostik, baik pada aspek kognitif maupun non-kognitif, sebagai dasar dalam merancang pembelajaran yang sesuai dengan keberagaman kebutuhan peserta didik. Pada saat yang sama, instrumen asesmen semakin diarahkan pada pengukuran *Higher Order Thinking Skills* (HOTS), sehingga pembelajaran Bahasa Arab tidak hanya berfungsi untuk menguasai struktur kebahasaan, tetapi juga menjadi sarana untuk mengembangkan kemampuan bernalar, menganalisis, mengevaluasi, dan menghasilkan gagasan.

Dari sisi perkembangan teknologi, pemanfaatan *Technology-Enhanced Assessment* (TEA) dan *Generative Artificial Intelligence* (GenAI) telah membawa perubahan terhadap mekanisme pelaksanaan asesmen serta pemberian umpan balik. Gamifikasi memberikan alternatif dalam menciptakan pengalaman asesmen yang lebih interaktif dan dapat membantu mengurangi kecemasan dalam pembelajaran bahasa asing. Sementara itu, teknologi AI membuka peluang bagi pengembangan asesmen yang lebih adaptif, personal, dan otomatis, termasuk dalam analisis bahasa, pengenalan tuturan, serta pembangkitan butir soal.

Namun demikian, efektivitas digitalisasi asesmen tidak semata-mata ditentukan oleh tingkat kecanggihan teknologi yang digunakan. Keberhasilannya juga bergantung pada kesesuaian teknologi dengan tujuan dan prinsip pedagogis, kualitas rubrik yang digunakan, keautentikan proses asesmen, serta kemampuan lembaga pendidikan dalam mengantisipasi risiko yang

berkaitan dengan integritas akademik. Dengan demikian, teknologi sebaiknya ditempatkan sebagai instrumen pendukung asesmen, bukan sebagai pengganti pertimbangan pedagogis dan profesional pendidik.

4.2 Saran

Hasil kajian sistematis ini menghasilkan sejumlah rekomendasi yang dapat dipertimbangkan oleh berbagai pemangku kepentingan dalam pengembangan asesmen Bahasa Arab.

1. **Bagi Pendidik dan Institusi Pendidikan:** Dalam menghadapi perkembangan pendidikan pada era *Society 5.0*, pendidik Bahasa Arab perlu meningkatkan literasi asesmen digital dan pemahaman mengenai pemanfaatan AI secara berkelanjutan. Institusi pendidikan juga perlu menyediakan pelatihan yang berkaitan dengan pengembangan rubrik untuk asesmen performa dan portofolio guna meningkatkan konsistensi penilaian, khususnya pada keterampilan produktif seperti *kalam* dan *kitabah*. Selain itu, asesmen formatif yang memperhatikan aspek diagnostik non-kognitif perlu diperkuat untuk membantu menciptakan lingkungan pembelajaran yang mendukung kondisi emosional peserta didik.
2. **Bagi Pengembang Kurikulum dan Pembuat Kebijakan:** Pengembangan buku ajar, modul pembelajaran, serta standar asesmen perlu mempertimbangkan keseimbangan antara tuntutan HOTS dan pemanfaatan teknologi secara etis. Pedoman yang lebih jelas mengenai penggunaan AI generatif dalam penyelesaian tugas Bahasa Arab juga diperlukan agar pemanfaatan teknologi tetap berada dalam batas yang mendukung proses pembelajaran serta menjaga validitas kompetensi lulusan.
3. **Bagi Peneliti Lanjutan:** Penelitian selanjutnya dapat diarahkan pada studi eksperimental dan *longitudinal* untuk mengkaji lebih mendalam mekanisme pengurangan bias serta permasalahan *halusinasi* data dalam sistem *AI-mediated assessment*, khususnya pada konteks bahasa Semit seperti Bahasa Arab. Penelitian mendatang juga perlu mengembangkan kerangka kebijakan yang berkaitan dengan keamanan, transparansi, dan integritas akademik dalam pelaksanaan asesmen Bahasa Arab di tengah perkembangan ekosistem *Generative AI*.

DAFTAR PUSTAKA

Afrianingsih, B. I., Azizah, N. L., & Sanah, S. (2025). *Evaluasi Keterampilan Berbahasa Arab*. Pustaka Edukasi.

Aini, N., & Alwi, I. M. (2024). Implementasi Asesmen Diagnostik Mata Pelajaran Bahasa Arab Kelas X MAN 1 Cilacap Tahun Pelajaran 2023/2024. *Jurnal Inovasi Pendidikan Madrasah*, 3(1), 45-56.

Allaithy, A., & Zaki, M. (2025). Evaluation of AI-Generated Reading Comprehension Materials for Arabic Language Teaching. *Journal of Arabic Language Learning*, 1-33. https://doi.org/10.1080/09588221.2025.24740377

Alshakhi, A. (2021). The Practice of Cross-Grading in Assessing Writing: The Case of EFL Teachers and Students in a Saudi Arabian Context. *Education Sciences*, 11(10), 1116. https://doi.org/10.3390/educsci1410111624

Abdurahman, I., Salim, A., Wangsadanureja, M., Alkhotiri, I., & Alpiansyah, R. (2026). Modernization of Human Resource Management in Islamic Boarding Schools: An Analysis of the Outlines of Personnel and Work Systems (GBKSK) Minhajul Haq. *Labina : Journal of Islamic Education Management*, *1*(1), 29-39. https://doi.org/10.66891/8wxyjc16

Alpiansyah, R., Affandy, A., & Dzulfikri. (2026). Branding Strategies For Islamic Educational Institutions: The Use Of Digital Marketing And Influencer Marketing To Attract Gen Z Prospective Students To Islamic-Based Schools. *Labina : Journal of Islamic Education Management*, *1*(2), 01-14. https://doi.org/10.66891/5brb6v48

Alpiansyah, R., Abdurahman, I., Alkhotiri, I., Wangsadanureja, M., & Dzulfikri. (2026). MANAJEMEN KURIKULUM DAN EKSTRAKURIKULER BERBASIS KARAKTER: INTEGRASI PENDIDIKAN JASMANI, PEMBIASAAN SPIRITUAL, DAN NEUROSAINS DALAM PEMBENTUKAN ADAB ISLAMI. *Labina : Journal of Islamic Education Management*, *1*(1), 62-90. https://doi.org/10.66891/v4dpda25

Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Longman.

Ardianto, D., & Fakhirah, N. (2025). Menyusun Asesmen Autentik. *Jurnal Pendidikan Bahasa dan Sastra*, 8(1), 12-25. https://doi.org/10.51278/aj.v8i1.21541

Asrori, M. (2021). Kurikulum Merdeka: Konsep, implementasi, dan tantangannya dalam meningkatkan kualitas pendidikan. *Jurnal Ilmu Pendidikan Islam*, 5(2), 146-153.

Boeriswati, E., Mayuni, I., Fahmi, A. K., & Surong, K. (2023). The Development of Assessment Rubric of Arabic Language Writing Based on National Qualifications Framework of Indonesia. *Jurnal Al Bayan: Jurnal Jurusan Pendidikan Bahasa Arab*, 15(1), 261–277. https://doi.org/10.24042/albayan.v15i1.163021

Booth, A. (2016). Searching for Qualitative Research for Inclusion in Systematic Reviews: A Structured Methodological Review. *Systematic Reviews*, 5(1), 1–23.

https://doi.org/10.1186/s13643-016-0249-x1

Budiono, A. N., & Hatip, M. (2023). Asesmen Pembelajaran Pada Kurikulum Merdeka. *Jurnal Axioma: Jurnal Matematika Dan Pembelajaran*, 8(1), 109-123.

Cahyani, V. P., Ahmad, F., Zulfakar, M., & Nawi, M. (2024). Authentic Assessment as a Solution to Enhance Deep Learning and Maintain Academic Integrity in Higher Education. *Higher Education Evaluation*, 3(3), 531–544. https://doi.org/10.1177/004723952211387911

Chabib, M., Baroroh, U., Umasugi, M. K., & Saifuddin, T. (2025). Inovasi Penilaian Pembelajaran Bahasa Arab berbasis HOTS dengan Media Audio Visual untuk Meningkatkan Keterampilan Menyimak. *Jurnal Pendidikan dan Evaluasi*, 5(2), 194–203.

Creely, E. (2025). Exploring the Role of Generative AI in Enhancing Language Learning: Opportunities and Challenges. *International Journal of Changes in Education*, 1(1), 45-60. https://doi.org/10.24059/olj.v28i2.44267

Delima, D. A., Hikmah, & Masrun. (2025). Analisis Higher Order Thinking Skills (HOTS) Pada Soal Ujian Asesmen Madrasah Aliyah Tahun 2024. *Maharaat Lughawiyyat: Jurnal Pendidikan Bahasa Arab*, 4(2), 118-125.

Dhouib, A., Othman, A., el Ghoul, O., Khribi, M. K., & al Sinani, A. (2022). Arabic Automatic Speech Recognition: A Systematic Literature Review. *Applied Sciences*, 12(17), 8898. https://doi.org/10.3390/app121788987

Elmi, B. A. O., Baroroh, R. U., Labibah, F. S., & Azzahra, L. Q. (2024). Innovation of HOTS-Based Test Instruments for Receptive Skills in Arabic Textbooks based on Kurikulum Merdeka. *Jurnal Al Bayan: Jurnal Jurusan Pendidikan Bahasa Arab*, 16(2), 577. https://doi.org/10.24042/albayan.v16i2.2469724

Fahruddin, F. (2025). Pengembangan Instrumen Asesmen Autentik dalam Pembelajaran untuk Meningkatkan Keterampilan Berpikir Kritis Siswa. *Jurnal Evaluasi Pendidikan*, 11(1), 34-45.

Fauziah, I. R. N., Syihabudin, S., & Sopian, A. (2020). Analisis Kualitas Tes Bahasa Arab Berbasis Higher Order Thinking Skill (HOTS). *LISANUNA: Jurnal Ilmu Bahasa Arab Dan Pembelajarannya*, 10(1), 45-54. https://doi.org/10.22373/ls.v10i1.780527

Fitri, D. K. (2022). Pembelajaran Berdiferensiasi dalam Perspektif Progresivisme pada Mata Pelajaran. *Jurnal Filsafat Indonesia*, 5(3), 250–258. https://doi.org/10.23887/jfi.v5i3.4124916

Hairida, H., Hartoyo, A., Pranata, R., & Munawar. (2024). Non-Cognitive Diagnostic Assessment Instrument of Learning Interest Based on Website for Elementary School Students. *Journal of Educational Evaluation*, 12(2), 216-225.

Ichsani, H. D., Hanifah, U., & Hayati, A. L. (2025). Pemanfaatan Platform Quizizz sebagai Instrumen Penilaian dalam Pembelajaran Bahasa Arab di Madrasah Tsanawiyah. *Jurnal Pendidikan Teknologi*, 4(1), 45-58.

Jurāne-Brēmane, A. (2023). Digital assessment in technology-enriched education: Thematic review. *Education Sciences*, 13(5), 522. https://doi.org/10.3390/educsci1305052213

Khasawneh, M. A. S. (2024). Practical approaches to technology-mediated formative assessment for middle school educators. *Educational Technology Research*, 15(2), 526. https://doi.org/10.24042/albayan.v15i2.1908923

Khasawneh, M. A. S., Aladini, A., Assi, S. A., & Ajanil, B. (2025). Portfolio Assessment in AI-Enhanced Learning Environments: A Pathway to Emotion Regulation, Mindfulness, and Language Learning Attitudes. *Language Testing in Asia*, 15(1), 5. https://doi.org/10.1186/s40468-025-00345-07

Khoiriyah, S. (2020). Pembelajaran Bahasa Arab Berbasis Kurikulum Merdeka: Konsep dan Implementasinya. *Jurnal Pendidikan Bahasa Arab*, 5(2), 120-135.

Kiptiah, Priyono, & Purnomo. (2023). Asesmen Diagnostik, Perlukah Untuk Sekolah?. *Resonasi Pemikiran*, 25(1), 12-25.

Kizi, S., & Shadjalilovna, A. (2022). Non-cognitive diagnostic assessment for language learners. *International Journal of Language Education*, 4(2), 150-162.

Laudza'i, I. H., Bakar, M. Y. A., & Arifin, Z. (2026). The Transformation of Arabic Language Assessment: A Systematic Literature Review of Technology-Enhanced Assessment (2021–2026). *Aphorisme: Journal of Arabic Language, Literature, and Education*, 7(1), 697-711. https://doi.org/10.37680/aphorisme.v7i1.96692

Li, L., Foon, K., & Jiahui, H. (2024). Gamification Enhances Student Intrinsic Motivation, Perceptions of Autonomy and Relatedness, but Minimal Impact on Competency: A Meta-Analysis and Systematic Review. *Educational Technology Research and Development*, 72(2). https://doi.org/10.1007/s11423-023-10337-71

Ma'arif, A. S., & Bakhiet, B. M. (2026). Implementing Digital Assessment in Arabic Language Learning: A Case Study of Listening, Speaking, Reading, and Writing Skills. *Funun Arabiyyah Journal of Arabic*, 2(2), 147-175. https://doi.org/10.64131/fununarabiyyah.v2i2.14613

Marden, P., & Herrington, J. (2025). Designing authentic assessments for foreign language learning environments. *Journal of Education, Innovation and Communication*, 7(2), 4-15. https://doi.org/10.34097/jeicom-7-2-423

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative Data Analysis: A Methods Sourcebook* (3rd ed.). SAGE Publications.

Muktamar, A., Jalil, A., Tang, M., & Miharja, J. (2023). Kurikulum Merdeka Berbasis Pembelajaran Berdiferensiasi Perspektif Pendidikan Islam. *Journal of International Multidisciplinary Research*, 1(2), 246-251.

Munawarah, M., & Zulkiflih, Z. (2021). Pembelajaran Keterampilan Menulis (Maharah al-Kitabah) dalam Bahasa Arab. *Jurnal Al-Ittijah*, 16(1), 90-105. https://doi.org/10.32678/alittijah.v16i1.905124

Muradi, A., Mubarak, F., Darmawaty, R., & Hakim, A. R. (2020). Higher Order Thinking Skills dalam Kompetensi Dasar Bahasa Arab. *Arabi: Journal of Arabic Studies*, 5(2), 105-115.

Nafis, D. (2026). Procedural Transparency and Legal Accountability to Sustain AI-Mediated Language Assessment in Saudi Arabia. *SAGE Open*, 1-13.

https://doi.org/10.1177/215824402513961138

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372, n71. https://doi.org/10.1136/bmj.n711

Paulsen, R., & Valdivia, A. (2022). Formative and diagnostic assessments in foreign language classrooms. *Language Testing*, 39(1), 112-130.

Purnomo, H. H., Baroroh, R., & Arju, S. (2025). Inovasi Penerapan Penilaian Portofolio dalam Mahārah Qirā'ah Mata Pelajaran Bahasa Arab. *Jurnal Alfazuna: Jurnal Pembelajaran Bahasa Arab Dan Kebahasaaraban*, 9(2), 45-65. https://doi.org/10.15642/alfazuna.v9i2.456523

Rahman, N. F., Masdar, M. I., & Mustika, A. (2022). Al-Ta'rib Transformation of Arabic Assessment in Indonesia: Conventional Assessment Toward Digital Assessment. *Al-Ta'rib: Jurnal Ilmiah Program Studi Pendidikan Bahasa Arab*, 10(2), 199–212.

Rini, R., Mustofa, M., & Kurnia, K. (2023). Using the Plotagon Application On Arabic Language Learning Media Design. *Arabiyatuna: Jurnal Bahasa Arab*, 7(2), 637–654. https://doi.org/10.29240/jba.v7i2.61013

Ritonga, A. M., Suib, M. S., & Zaky, A. (2023). Tantangan Dan Hambatan: Analisis Problematika Pembelajaran Bahasa Arab Era Society 5.0 Di Perguruan Tinggi Agama Islam Di Indonesia. *Syntax Literate Jurnal Ilmiah Indonesia*, 8(5), 3272–3280. https://doi.org/10.36418/syntax-literate.v8i5.115403

Rosyadi, F. I. (2021). Muatan Higher Order Thinking Skills (HOTS) Pada Buku Teks Bahasa Arab Madrasah Aliyah Terbitan Kementerian Agama Tahun 2020. *Journal of Business Theory and Practice*, 10(2), 1-15.

Sa'adah, N. (2025). Inovasi Penilaian Mufrodat Menggunakan Media Audio Visual Berbasis Teknologi QR Code pada LKS Bahasa Arab. *Mahira: Journal of Arabic Studies*, 9, 272–291. https://doi.org/10.55380/mahira.v2i2.386.11

Setyawan, C. E., et al. (2025). Rubrik Penilaian Non-Tes Mahārah Kalām di Perguruan Tinggi Islam. *Alsina: Journal of Arabic Studies*, 7(1), 1-20.

Shifany, M. H., Ridlo, U., & Raswan. (2025). Implementasi Asesmen Autentik Kurikulum Merdeka Pada Mata Pelajaran Bahasa Arab SMP Islam Cikal Harapan 1 BSD. *Al-Muyassar: Journal of Arabic Education*, 4(1), 172-183.

Suprapti, S., & Ridho, A. (2024). Penilaian diagnostik kognitif dan non-kognitif dalam pembelajaran berdiferensiasi. *Jurnal Pendidikan Dasar*, 10(1), 15-28.

Syagif, A., et al. (2021). Penggunaan Media Google Form Dalam Evaluasi Pembelajaran Bahasa Arab. *Jurnal Naskhi: Jurnal Kajian Pendidikan Dan Bahasa Arab*, 3(1), 30–40. https://doi.org/10.47435/naskhi.v3i1.5535

Viani, Y. O., & Ripai, A. (2024). Implementasi Asesmen Diagnostik Non-Kognitif Gaya Belajar dengan Media Google Form Di SMK. *Jurnal Riset Pendidikan Kejuruan*, 7(1), 45-56.

Wang, Y., Zhang, T., Yao, L., & Seedhouse, P. (2025). A Scoping Review of Empirical Studies on Generative Artificial Intelligence in Language Education. *Innovation in Language*

Learning and Teaching, 1–28. https://doi.org/10.1080/17501229.2025.1234567

Widodo, J., & Qosim, M. N. (2021). Penilaian Kinerja Pada Pembelajaran Maharah Al-Kalam Level 1 Di Kursus Bahasa Arab Al-Arobiya Surakarta. *Uktub: Journal of Arabic Studies*, 1(2), 84-101. https://doi.org/10.32678/uktub.v1i2.581415

Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. F. (2024). A Scoping Review on How Generative Artificial Intelligence Transforms Assessment in Higher Education. *Assessment in Education: Principles, Policy & Practice*, 32(4), 386–400. https://doi.org/10.1080/0969594X.2025.25653861