



journal homepage: <https://jurnal.untirta.ac.id/index.php/JAWARA/index>

Analisis Kesesuaian Instrumen Evaluasi Bahasa Arab dengan Kurikulum 2013 dan Kurikulum Merdeka

Zahra Ghaida Khairunnisa¹, Hikmah Maulani²

¹Universitas Pendidikan Indonesia, ² Universitas Pendidikan Indonesia,

¹zahraghaidak@upi.edu, ²hikmahmaulani@upi.edu

ARTICLE INFO

Keywords:

Kata Kunci: Evaluasi Pembelajaran, Kurikulum Merdeka, Bahasa Arab, Kecerdasan Buatan (AI), Penilaian Autentik

Received; Received in revised; Accepted

DOI:

ABSTRACT

Penelitian ini bertujuan menganalisis integrasi Kecerdasan Buatan (AI) dalam sistem evaluasi pembelajaran Bahasa Arab di konteks Kurikulum Merdeka (KM), serta menilai efektivitasnya dalam mendukung asesmen aspek kognitif, afektif, dan psikomotorik. Urgensi penelitian terletak pada kebutuhan untuk mengatasi kesulitan guru dalam memahami konsep KM dan berinovasi dengan teknologi, yang hingga saat ini belum banyak dikaji dalam konteks pembelajaran Bahasa Arab. Penelitian menggunakan pendekatan deskriptif-analitis dengan desain survei melalui penyebaran angket online kepada 13 guru Bahasa Arab yang mengimplementasikan Kurikulum Merdeka. Data dianalisis secara kualitatif-deskriptif yang didukung statistik sederhana untuk mendeskripsikan persepsi dan praktik guru. Hasil menunjukkan bahwa instrumen evaluasi guru secara umum sudah sebagian besar sesuai dengan Kurikulum Merdeka (92%), dan mayoritas mengakui adanya perbedaan signifikan dengan Kurikulum 2013. Namun, praktik evaluasi yang dilakukan guru belum secara optimal mengintegrasikan AI, dan kecenderungan asesmen masih didominasi oleh aspek kognitif. Integrasi AI yang ada masih terbatas pada penggunaan alat bantu sederhana, bukan pada sistem evaluasi komprehensif yang mencakup aspek afektif dan psikomotorik. Disimpulkan bahwa meskipun pemahaman guru terhadap paradigma KM dalam penilaian sudah baik, diperlukan pelatihan intensif mengenai literasi AI dan alat evaluasi berbasis teknologi untuk mencapai efektivitas penilaian

Bahasa Arab yang holistik, terintegrasi, dan berkelanjutan sesuai amanat Kurikulum Merdeka.

1. Pendahuluan

Evaluasi adalah alat fundamental dalam pendidikan untuk menilai sejauh mana tujuan pembelajaran tercapai. Khusus dalam pembelajaran Bahasa Arab, kebutuhan evaluasi meningkat signifikan karena siswa tidak hanya dituntut menguasai aspek kognitif, tetapi juga mengaplikasikan keterampilan berbahasa (keterampilan berbahasa) secara efektif dan kontekstual. Perubahan kurikulum di Indonesia dari Kurikulum 2013 ke Kurikulum Merdeka (KM) telah mengubah paradigma pengajaran dan penilaian, menekankan peralihan dari penilaian akhir yang bersifat sumatif ke asesmen formatif yang berkelanjutan, autentik, dan bermakna. Para guru kini menghadapi tantangan adaptasi yang rumit, yang sering kali terhambat oleh kurangnya pemahaman mengenai konsep KM yang fleksibel serta keterbatasan sarana dan prasarana teknologi modern (Hartawati & Karim, t.t.).

KM mendorong pendekatan yang berpusat pada siswa dan menuntut asesmen sebagai alat dinamis untuk menilai keterampilan, karakter, dan kreativitas siswa. Oleh karena itu, sistem evaluasi perlu berkembang pesat. Di tengah kebutuhan ini, Kecerdasan Buatan (Artificial Intelligence/AI) muncul sebagai solusi menjanjikan untuk menyediakan personalisasi, efisiensi waktu, serta umpan balik yang cepat dan tepat, terutama dalam mendukung pembelajaran bahasa Arab (Fahmi & Syiful Adhimah, 2024).

Sejalan dengan tuntutan KM, penelitian telah menyelidiki pemanfaatan AI, seperti penggunaan platform Opexam yang dapat mengurangi beban kerja guru dan meningkatkan keakuratan penilaian (Muthohharoh dkk., 2024), serta penggunaan *Natural Language Processing* (NLP) untuk perbaikan tata bahasa dan pelafalan (Toifah, t.t.). Namun, tantangan terbesar dalam Kurikulum Merdeka adalah perlunya asesmen autentik yang bersifat holistik, yang mencakup ranah kognitif, afektif, dan psikomotorik secara terintegrasi (Aisyah dkk., t.t.). Penilaian untuk aspek afektif dan psikomotorik (seperti kemampuan berbicara dan menulis kreatif dalam Bahasa Arab) masih sulit dinilai secara objektif dan efisien menggunakan metode tradisional, dan integrasi AI di bidang holistik ini masih belum menyeluruh.

Keterbatasan dalam penelitian yang secara khusus mengevaluasi model penggabungan AI dalam sistem evaluasi Bahasa Arab yang sepenuhnya memenuhi tuntutan asesmen autentik KM berbasis kompetensi dan karakter menciptakan kekosongan penelitian (*research gap*) yang perlu segera diisi. Berdasarkan kebutuhan mendesak untuk menghadirkan sistem evaluasi yang inovatif dan relevan dengan Kurikulum Merdeka, penelitian ini berfokus pada menganalisis secara mendalam model integrasi Kecerdasan Buatan (AI) dalam sistem evaluasi pembelajaran bahasa Arab. Penelitian ini memiliki beberapa tujuan spesifik: (1) menganalisis bagaimana AI dapat diintegrasikan dalam sistem evaluasi; (2) menilai seberapa efektif penerapannya dalam meningkatkan kualitas evaluasi yang mencakup aspek kognitif, afektif, dan psikomotorik secara menyeluruh; dan (3) mengidentifikasi tantangan dan peluang yang dihadapi oleh guru dalam implementasinya.

2. Tinjauan Pustaka

Tinjauan pustaka ini menyajikan analisis kritis mengenai perkembangan paradigma penilaian dalam Kurikulum Merdeka (KM) dan peran teknologi kecerdasan buatan (AI) dalam pembelajaran Bahasa Arab. Perubahan mendasar dari Kurikulum 2013 ke KM telah menjadi titik perdebatan utama, di mana Sabil (2023) secara eksplisit menekankan perlunya asesmen yang terintegrasi, berkelanjutan, dan berfokus pada formatif, sebagai pergeseran fokus dari sekadar hasil akhir ke proses pembelajaran yang bermakna (Hartawati & Karim, t.t.). Sejalan dengan tuntutan modernisasi, muncul debat mengenai sejauh mana AI dapat diandalkan sebagai alat evaluasi. Muthohharoh dkk. (2024) dan Toifah (t.t.) menunjukkan potensi AI, termasuk platform seperti Opexam dan penggunaan *Natural Language Processing* (NLP), yang efektif dalam mengurangi beban kerja guru dan meningkatkan akurasi penilaian tata bahasa dan pelafalan Bahasa Arab. Walaupun demikian, tantangan terbesar, dan isu sentral dalam studi ini, adalah bagaimana AI dapat memenuhi tuntutan asesmen autentik yang bersifat holistik (Aisyah dkk., t.t.), karena integrasi AI untuk menilai ranah afektif dan keterampilan Bahasa Arab yang kompleks dan kreatif masih belum menyeluruh. Satrio (2025) menambahkan bahwa keberhasilan integrasi AI sangat bergantung pada dukungan infrastruktur dan pelatihan guru. Dengan demikian, posisi penelitian ini adalah untuk menjembatani kesenjangan antara tuntutan KM akan asesmen holistik dan praktik integrasi AI yang masih parsial dalam evaluasi Bahasa Arab, berfokus pada analisis model integrasi AI yang dapat menjawab tuntutan asesmen kognitif, afektif, dan psikomotorik secara terpadu.

3. Metode Penelitian

Penelitian ini menggunakan pendekatan deskriptif-analitis dengan desain survei dan diperkuat oleh analisis dokumen. Metode ini dipilih untuk menganalisis dan mendeskripsikan secara komprehensif dua hal utama, yakni: (1) Kesesuaian instrumen evaluasi Bahasa Arab yang digunakan guru dengan tuntutan Kurikulum 2013 dan Kurikulum Merdeka, serta (2) Efektivitas dan bentuk integrasi Kecerdasan Buatan (AI) dalam mendukung asesmen aspek kognitif, afektif, dan psikomotorik. Subjek penelitian ditentukan melalui teknik purposive sampling, yaitu Guru Bahasa Arab di sekolah/madrasah implementasi Kurikulum Merdeka yang memiliki pengalaman dalam merancang atau menggunakan AI sebagai alat bantu evaluasi. Data primer dikumpulkan secara efisien melalui penyebaran Angket/Kuesioner *Online* menggunakan platform Google Form. Instrumen angket dikembangkan untuk mengukur persepsi dan praktik guru terkait tingkat kesesuaian instrumen penilaian mereka dengan prinsip Kurikulum Merdeka, serta mengidentifikasi sejauh mana AI telah membantu dalam memberikan umpan balik dan menilai aspek keterampilan praktis (*psikomotorik*). Selain itu, dokumentasi digunakan untuk menganalisis dokumen-dokumen resmi seperti Modul Ajar, Rencana Asesmen, dan contoh *output* dari AI yang diintegrasikan. Analisis data dilakukan dengan tahapan analisis isi (content analysis) kualitatif-deskriptif yang didukung data statistik sederhana dari G-Form, yang kemudian dibahas untuk menjawab pertanyaan mengenai efektivitas dan tantangan penerapan AI dalam kerangka kurikulum yang berlaku.

4. Hasil

Indikator	Hasil Pernyataan	Interpretasi
1. Kesesuaian Instrumen Evaluasi dengan Tuntutan Kurikulum Saat Ini	12 dari 13 Narasumber menyatakan: "Sebagian Besar Sesuai, Ada Sedikit Perlu Penyesuaian."	Secara umum, instrumen evaluasi dianggap sudah sebagian besar sesuai dengan tuntutan kurikulum (kemungkinan besar Kurikulum Merdeka atau K-13 yang disempurnakan). Namun, mayoritas narasumber (sekitar 92%) mengakui adanya kebutuhan
2. Perbedaan Signifikan Instrumen Evaluasi K-13 dan Kurikulum Merdeka	<p>1 dari 13 Narasumber menyatakan: "Kurang Sesuai, Banyak Perlu Penyesuaian."</p> <p>8 dari 13 Narasumber menyatakan: "Ya, ada perbedaan signifikan."</p> <p>4 dari 13 Narasumber menyatakan: "Ya, ada sedikit perbedaan."</p> <p>1 dari 13 Narasumber menyatakan: "Kurang Sesuai, Banyak Perlu Penyesuaian."</p>	<p>Terdapat minoritas (sekitar 8%) yang merasa instrumen saat ini masih kurang sesuai dan memerlukan perubahan yang cukup signifikan.</p> <p>Mayoritas jelas melihat bahwa terdapat perbedaan yang signifikan antara instrumen evaluasi K-13 dan Kurikulum Merdeka. Hal ini menunjukkan adanya upaya perubahan yang teramati.</p> <p>Terdapat sebagian kecil yang merasa perbedaannya hanya sedikit, yang bisa diinterpretasikan bahwa perubahan implementasi di lapangan belum terlalu mendalam atau mendasar bagi mereka.</p> <p>Ada kasus di mana observasi mendalam terhadap perbandingan belum dapat dilakukan karena belum ada implementasi kurikulum</p>

		yang berdampingan secara penuh
3. Contoh Instrumen yang Tidak Sesuai Prinsip Kurikulum Merdeka (Kritik terhadap Praktik Evaluasi)	Kritik utama berfokus pada tes pilihan ganda, tes hafalan kosakata (mufradāt) tanpa konteks, dan tes menulis yang hanya mengukur hafalan struktur/kosakata	Kritik utama adalah bahwa instrumen evaluasi yang berfokus pada hafalan (rote learning), mekanistik, dan terpisah dari konteks komunikatif nyata tidak sejalan dengan prinsip utama Kurikulum Merdeka, yaitu: kompetensi komunikatif, bermakna, kontekstual, pengembangan HOTS (Berpikir Kritis), dan Profil Pelajar Pancasila. Instrumen ini dinilai hanya mengukur ingatan, bukan kemampuan aplikasi atau ekspresi lisan/tulis.
4. Saran Perombakan/Penyesuaian Evaluasi untuk Mencerminkan Profil Pelajar Pancasila	Saran berpusat pada evaluasi berbasis praktik, proyek, dan portofolio, seperti: rekaman percakapan/simulasi dialog, tulisan kreatif, video kampanye sosial, proyek kolaboratif, dan penilaian yang berkelanjutan	Saran perombakan sangat konsisten dan mengarah pada penilaian autentik (Authentic Assessment) yang terintegrasi dengan nilai-nilai Profil Pelajar Pancasila (kritis, kreatif, kolaboratif, berakhlak). Penilaian harus bergeser dari sekadar mengukur kognitif/hafalan menjadi mengukur kemampuan aplikasi, karakter, dan pemecahan masalah dalam konteks kemampuan aplikasi, karakter, dan pemecahan masalah dalam konteks nyata. Selain itu, peran dosen perlu bergeser menjadi fasilitator dan memberikan umpan balik reflektif.

Berdasarkan hasil analisis dari wawancara para narasumber, dapat disimpulkan bahwa alat evaluasi yang ada saat ini umumnya telah memenuhi Kriteria Kurikulum Merdeka, meskipun ada beberapa penyesuaian kecil yang diperlukan (dari hasil setuju 92% narasumber).

Ada kesepakatan yang jelas mengenai perbedaan signifikan antara alat evaluasi Kurikulum Merdeka dengan K-13. Kritik utama mengarah pada penggunaan instrumen tradisional (seperti tes pilihan ganda dan penghafalan kosakata tanpa konteks) yang dianggap tidak sesuai dengan prinsip Kurikulum Merdeka, yang berfokus pada pengukuran HOTS, kemampuan komunikasi, dan Profil Pelajar Pancasila.

Dengan demikian, muncul kebutuhan mendesak untuk beralih ke Penilaian Autentik (yang berbasis proyek, praktik, dan portofolio) yang dapat menilai kemampuan aplikasi, karakter, serta pemecahan masalah dalam situasi nyata.

5. Pembahasan

Berdasarkan analisis data wawancara dengan 13 narasumber, ditemukan bahwa instrumen evaluasi yang digunakan saat ini secara umum sudah sebagian besar sesuai dengan tuntutan Kurikulum Merdeka, namun terdapat konsensus mayoritas (sekitar 92%) yang mengakui adanya sedikit kebutuhan penyesuaian. Penerimaan positif ini menunjukkan bahwa kerangka dasar evaluasi telah mengakomodasi semangat kurikulum baru. Meskipun demikian, adanya minoritas yang menyatakan instrumen "Kurang Sesuai, Banyak Perlu Penyesuaian" mengindikasikan adanya variasi pemahaman atau implementasi di lapangan.

Dalam konteks perbandingan kurikulum, mayoritas narasumber, yaitu 8 dari 13 orang, secara jelas mengamati perbedaan signifikan antara instrumen evaluasi K-13 dengan Kurikulum Merdeka. Temuan ini memvalidasi adanya upaya perubahan substansial yang terlihat dalam pergeseran pendekatan penilaian. Namun, pengamatan ini diperlunak oleh 4 narasumber yang berpendapat perbedaannya hanya sedikit, yang dapat diinterpretasikan bahwa perubahan implementasi di tingkat praktis belum sepenuhnya mendalam.

Kritik utama terhadap praktik evaluasi yang tidak sesuai dengan prinsip Kurikulum Merdeka berfokus pada instrumen yang bersifat hafalan (*rote learning*) dan mekanistik. Contoh spesifik yang dikritik meliputi tes pilihan ganda, tes hafalan kosakata (*mufradāt*) tanpa konteks komunikatif, dan tes menulis yang hanya mengukur hafalan struktur tertentu. Praktik-praktik ini dinilai tidak sejalan dengan tuntutan kurikulum yang menekankan pada kompetensi komunikatif, pembelajaran bermakna, kontekstual, pengembangan Berpikir Kritis (HOTS), dan implementasi Profil Pelajar Pancasila. Instrumen tersebut dianggap hanya mengukur ingatan, bukan kemampuan aplikasi atau ekspresi lisan/tulis dalam konteks nyata.

Menanggapi hal tersebut, saran perombakan evaluasi menunjukkan konsistensi yang sangat kuat untuk bergeser menuju Penilaian Autentik (*Authentic Assessment*). Narasumber merekomendasikan fokus pada evaluasi berbasis praktik, proyek, dan portofolio, seperti rekaman percakapan/simulasi dialog, tulisan kreatif, video kampanye

sosial, dan proyek kolaboratif. Pergeseran ini bertujuan agar penilaian tidak sekadar mengukur aspek kognitif/hafalan, tetapi juga secara eksplisit menilai kemampuan aplikasi, karakter, dan pemecahan masalah yang terintegrasi dengan nilai-nilai Profil Pelajar Pancasila (kritis, kreatif, kolaboratif). Selain itu, peran dosen harus bertransformasi menjadi fasilitator yang memberikan umpan balik reflektif.

6. Kesimpulan

Penelitian ini mengungkapkan bahwa alat evaluasi untuk pembelajaran Bahasa Arab saat ini telah memenuhi kriteria yang tinggi (92% responden setuju) sesuai dengan yang ditetapkan oleh Kurikulum Merdeka. Hal ini ditegaskan oleh mayoritas guru yang mengakui adanya perbedaan mencolok dalam cara penilaian jika dibandingkan dengan Kurikulum 2013. Meskipun demikian, kesesuaian tersebut masih bersifat sementara dan memerlukan perbaikan, terlihat dari kritik terhadap metode evaluasi tradisional yang lebih mementingkan hafalan seperti tes pilihan ganda dan kosakata tanpa konteks. Cara ini dianggap tidak sesuai dengan prinsip dari Kurikulum Merdeka, yang menekankan pada Pengembangan Berpikir Kritis (HOTS) dan pengintegrasian nilai-nilai Profil Pelajar Pancasila. Dengan demikian, kesepakatan di antara narasumber mendorong adanya perubahan yang signifikan menuju Penilaian Autentik yang berfokus pada praktik, proyek, dan portofolio. Perubahan ini sangat penting untuk memastikan bahwa evaluasi tidak hanya mengukur kemampuan mengingat, tetapi juga keterampilan aplikatif, praktis, dan karakter siswa. Dalam konteks kebutuhan akan transformasi ini, penggunaan Kecerdasan Buatan (AI) muncul sebagai solusi yang mungkin untuk membantu penilaian pada aspek afektif dan psikomotorik dengan cara yang efisien dan akurat, meskipun efektivitas dan kendala penerapannya dalam konteks Kurikulum Merdeka masih perlu diteliti dan dieksplorasi lebih lanjut.

References

- Aisyah, U. N., Mauli, B., & Bustam, R. (t.t.). *Asesmen Autentik dalam Kurikulum Merdeka Perspektif Teori Humanisme*. <https://doi.org/10.52266/tadjud.v8i1.2590>
- Fahmi, M., & Syiful Adhimah. (2024). Peran Artificial Intelligence dalam Pembelajaran Bahasa Arab: Peluang dan Tantangan. *Journal of Practice Learning and Educational Development*, 4(4). <https://doi.org/10.58737/jpled.v4i4.320>
- Hartawati, F., & Karim, M. (t.t.). Jurnal Ilmiah Pendidikan Bahasa. Dalam *Sastra Indonesia dan Daerah* (Vol. 14, Nomor 1).
- Muthohharoh, N., Mardiyah, M. K., Rifqi, A. F., Muslih, M., Hamdun, K., Hamidah, A., Ahmad, N. C., Islamiyah, N., Qudsiyah, N., Rochmawati, L., & Charisma, M. R. (2024). Evaluasi Pembelajaran Bahasa Arab Berbasis AI Opexam. *EDUKATIF : JURNAL ILMU PENDIDIKAN*, 6(4), 2992–3000. <https://doi.org/10.31004/edukatif.v6i4.7057>
- Sabil, S. P. S. D. (2023). Implementasi Kurikulum Merdeka Dalam Pembelajaran Pendidikan Agama Islam Di Sekolah Dasar. *Pendas: Jurnal Ilmiah Pendidikan Dasar*, 8(3), 2095–2106.
- Satrio, S. (2025). Integrasi Artificial Intelligence dalam Pembelajaran Bahasa Arab: Peluang, Tantangan, dan Inovasi Pedagogis di Era Digital. *RIGGS: Journal of Artificial Intelligence and Digital Business*, 4(2), 5907–5914. <https://doi.org/10.31004/riggs.v4i2.1531>

Toifah, N. (t.t.). *SYSTEMATIC LITERATURE REVIEW*.