E-ISSN: 2721-1145
DOI 10.xxxx/xxxx

Representasi Kompetensi Berbahasa Arab dalam Tes Objektif: Kajian Kritis Perspektif Positivisme Auguste Comte

Moh. Nazaruddin*
UIN Sunan Ampel Surabaya, Indonesia
Email: nadzirroby@gmail.com
**corresponding author*

Kamal Yusuf
UIN Sunan Ampel Surabaya, Indonesia
Email: kamalyusuf@uinsa.ac.id

Abstract:
Objective tests are widely used in Arabic language assessment because they offer consistent scoring, efficiency, and comparability. This article asks how far the scores of such tests represent Arabic language competence, using Auguste Comte's positivism as a critical lens rather than as a historical explanation of testing practice. The study is a library-based, critical-conceptual analysis of peer-reviewed journal articles on positivism, validity theory, communicative competence, Arabic diglossia, and the revised Bloom's Taxonomy. The analysis traces three positivist assumptions embedded in score-based assessment, namely observability, decomposability, and comparability, and examines each against validity theory. The findings show that objective tests provide strong evidence for discrete, form-focused, and largely receptive knowledge such as vocabulary, morphosyntax, and part of reading and listening comprehension. However, they risk construct underrepresentation when scores are read as evidence of speaking, writing, pragmatic appropriateness, and the ability to move between Modern Standard Arabic and spoken varieties. Item format does not determine cognitive level, and cognitive level does not exhaust communicative competence. The article proposes a competence–evidence mapping that pairs each Arabic language competence with a suitable assessment form, its representation risk, and the validity evidence to be sought. It concludes that objectivity of scoring is necessary but insufficient: the defensibility of a score interpretation depends on how well the tasks sample the intended construct.
Keywords: Arabic language assessment; objective tests; construct validity; positivism; Bloom's Taxonomy

Abstrak:
Tes objektif banyak digunakan dalam evaluasi pembelajaran bahasa Arab karena menawarkan konsistensi penskoran, efisiensi, dan keterbandingan hasil. Artikel ini

Copyright (c) 2026 by **Wissen:** Jurnal Pendiidkan Indonesia

menanyakan sejauh mana skor tes semacam itu merepresentasikan kompetensi berbahasa Arab, dengan menjadikan positivisme Auguste Comte sebagai lensa kritis, bukan sebagai penjelasan historis atas praktik pengujian. Penelitian ini merupakan kajian kepustakaan dengan analisis kritis-konseptual terhadap artikel jurnal bereputasi tentang positivisme, teori validitas, kompetensi komunikatif, diglosia bahasa Arab, dan Taksonomi Bloom revisi. Analisis menelusuri tiga asumsi positivistik dalam penilaian berbasis skor, yaitu keteramatan, keterpilahan, dan keterbandingan, lalu menguji masing-masing dengan teori validitas. Hasil kajian menunjukkan bahwa tes objektif memberikan bukti yang kuat untuk pengetahuan yang diskret, berfokus pada bentuk, dan bersifat reseptif, seperti kosakata, morfosintaksis, serta sebagian pemahaman bacaan dan simakan. Namun, tes objektif berisiko mengalami ketakterwakilan konstruk apabila skornya dibaca sebagai bukti kemampuan berbicara, menulis, ketepatan pragmatik, dan kemampuan berpindah antara bahasa Arab ragam standar dan ragam lisan. Format soal tidak menentukan tingkat kognitif, dan tingkat kognitif tidak mencakup seluruh kompetensi komunikatif. Artikel ini mengusulkan pemetaan kompetensi–bukti yang memasangkan setiap kompetensi bahasa Arab dengan bentuk asesmen yang sesuai, risiko representasinya, dan bukti validitas yang perlu dicari. Kajian ini menyimpulkan bahwa objektivitas penskoran diperlukan tetapi tidak memadai: tafsir skor hanya dapat dipertanggungjawabkan sejauh tugas-tugas dalam tes benar-benar mewakili konstruk yang dimaksud.
Kata Kunci: evaluasi bahasa Arab; tes objektif; validitas konstruk; positivisme; Taksonomi Bloom

PENDAHULUAN

Pembelajaran bahasa Arab diarahkan pada dua sasaran yang saling berkaitan: penguasaan unsur kebahasaan (*mufradat* dan *qawā'id*) dan penguasaan keterampilan berbahasa (*istimā'*, *kalām*, *qirā'ah*, dan *kitābah*). Karena sasarannya majemuk, evaluasi tidak berhenti pada pemberian nilai. Evaluasi menentukan apa yang dianggap sebagai keberhasilan belajar, bagaimana kemajuan peserta didik dibaca, dan bagian pembelajaran mana yang perlu diperbaiki (Zurqoni dkk. 2020; Muradi dkk. 2022; Nurzahira dkk. 2025). Dengan kata lain, bentuk evaluasi yang dipilih ikut membentuk pemahaman tentang apa yang dimaksud dengan "mampu berbahasa Arab".

Dalam praktik, tes objektif seperti pilihan ganda, benar-salah, menjodohkan, dan isian singkat menempati posisi yang kuat. Alasannya masuk akal. Kunci jawaban yang pasti menjadikan penskoran konsisten, pemeriksaan pada kelompok besar menjadi cepat, dan cakupan materi dapat diperluas (Putri dkk. 2022). Pedoman penulisan butirnya pun telah mapan (Haladyna, Downing, dan Rodriguez 2002). Keunggulan-keunggulan ini nyata dan tidak perlu dibantah oleh kajian yang bersikap kritis.

Persoalan muncul ketika keunggulan prosedural diperlakukan sebagai keunggulan epistemik, yakni ketika konsistensi penskoran dianggap sama dengan kebenaran tentang kompetensi. Skor tes sesungguhnya adalah hasil serangkaian representasi: konstruk (apa yang dimaksud dengan kompetensi) diterjemahkan menjadi butir atau tugas, respons diubah menjadi angka, dan angka itu ditafsirkan. Pada setiap tahap ada bagian kompetensi

yang mungkin tidak ikut terbawa. Messick (1995) menamai dua ancamannya: ketakterwakilan konstruk (*construct underrepresentation*), yaitu ketika tes terlalu sempit untuk mencakup konstruk yang dimaksud, dan varians yang tidak relevan dengan konstruk (*construct-irrelevant variance*), yaitu ketika skor dipengaruhi hal-hal di luar kompetensi yang ingin dinilai. Dalam konteks bahasa, kemampuan mengenali bentuk yang benar belum tentu disertai kemampuan menghasilkan ujaran, menangkap maksud mitra tutur, atau memilih ungkapan yang sesuai dengan situasi. Canale dan Swain (1980) menggambarkan kompetensi komunikatif sebagai gabungan kompetensi gramatikal, sosiolinguistik, dan strategis, sehingga tes yang hanya menyentuh satu komponen tidak dapat berbicara tentang keseluruhannya.

Bahasa Arab menambahkan satu lapisan lagi, yaitu diglosia: ragam standar dipakai untuk tulisan dan situasi formal, sedangkan ragam lisan sehari-hari berbeda menurut wilayah (Ferguson 1959). Tinjauan konseptual terbaru atas asesmen bahasa Arab sebagai bahasa asing mencatat bahwa tes kemahiran cenderung berfokus pada ragam standar sehingga kurang mewakili penggunaan bahasa dalam komunikasi nyata (Alharbi dan Sa'ad 2026). Dengan demikian, pertanyaan tentang representasi kompetensi dalam tes objektif bukan persoalan teknis semata, melainkan persoalan tentang apa yang dianggap sebagai bahasa Arab yang layak diukur.

Positivisme Auguste Comte dipakai dalam kajian ini sebagai lensa untuk menyingkap apa yang diandaikan ketika skor dibaca sebagai kompetensi. Dalam literatur, tahap positif Comte dicirikan oleh pengutamaan pengamatan dan penemuan keteraturan antargejala, dengan menyisihkan penjelasan spekulatif (Nugroho 2016; Jubaedah dan Hayati 2024). Studi kualitatif di Indonesia menunjukkan bahwa orientasi positivistik dalam asesmen cenderung menekankan capaian terukur dan penilaian terstandar (Nita dkk. 2025), sedangkan kajian tentang asesmen bahasa dari perspektif kritis menegaskan bahwa persoalan validitas dan etika dirumuskan secara berbeda ketika paradigma pengukuran itu sendiri dipertanyakan (Lynch 2001). Kajian ini tidak mengklaim bahwa tes objektif bersumber langsung dari pemikiran Comte. Klaim semacam itu memerlukan bukti historis tersendiri. Yang diajukan adalah kesejajaran analitis: asumsi tentang keterbandingan dan keterukuran yang melekat pada penilaian berbasis skor dapat diperiksa dengan bantuan kerangka epistemologis positivisme.

Penelitian terdahulu telah membahas konsep evaluasi pembelajaran bahasa Arab (Nurzahira dkk. 2025), instrumen penilaian kognitif pada tes objektif dan uraian (Putri dkk. 2022), serta evaluasi program pembelajaran bahasa Arab (Zurqoni dkk. 2020; Muradi dkk. 2022). Di sisi lain, kajian positivisme membahas asumsi epistemologisnya (Nugroho 2016; Jubaedah dan Hayati 2024) atau dampaknya pada asesmen guru (Nita dkk. 2025). Pada literatur yang ditelaah dalam kajian ini, kedua arus tersebut belum dipertemukan dalam satu pertanyaan: bagaimana kompetensi berbahasa Arab direpresentasikan oleh skor tes objektif, dan apa yang ikut hilang dalam representasi itu. Celah inilah yang hendak diisi oleh artikel ini, tanpa menyatakan bahwa penelitian terdahulu mengabaikan persoalan validitas.

Berdasarkan celah tersebut, penelitian ini bertujuan (1) mengidentifikasi asumsi positivistik yang bekerja dalam penilaian berbasis skor; (2) menganalisis sejauh mana tes

objektif merepresentasikan kompetensi berbahasa Arab, termasuk bagian yang kuat dan bagian yang berisiko tidak terwakili; dan (3) merumuskan pemilihan instrumen berdasarkan pembedaan antara tuntutan kognitif yang dipetakan Taksonomi Bloom dan kompetensi komunikatif. Sumbangan utama artikel ini adalah kerangka pemetaan kompetensi–bukti yang dapat dipakai pendidik untuk menggunakan hasil tes objektif secara proporsional.

METODE

Penelitian ini menggunakan desain kajian kepustakaan (*library research*) dengan pendekatan kualitatif-kritis. Kajian ini bersifat konseptual: sumber-sumber diperlakukan sebagai bahan analisis teoretis, bukan sebagai data empiris yang mewakili seluruh lembaga atau praktik evaluasi bahasa Arab. Pertanyaan yang memandu analisis berasal dari judul artikel, yaitu bagaimana kompetensi berbahasa Arab direpresentasikan dalam tes objektif dan apa yang dapat disingkap oleh perspektif positivisme Comte terhadap representasi tersebut.

Sumber utama berupa artikel jurnal berjejaring ilmiah yang dapat ditelusuri melalui DOI atau laman resmi jurnal. Penelusuran dilakukan pada Google Scholar, laman resmi penerbit jurnal, dan Garuda dengan kata kunci dalam bahasa Indonesia, Inggris, dan Arab yang berkaitan dengan positivisme Comte, evaluasi pembelajaran bahasa Arab, tes objektif, validitas konstruk, kompetensi komunikatif, diglosia, dan Taksonomi Bloom. Sumber dipilih berdasarkan tiga kriteria: (a) diterbitkan dalam jurnal dengan mekanisme penelaahan sejawat; (b) relevan langsung dengan salah satu dari empat tema analisis; dan (c) dapat diakses sehingga rujukannya dapat diperiksa kembali oleh pembaca. Karena pemikiran Comte dikaji melalui artikel jurnal sekunder, kesimpulan tentang Comte dibatasi pada ciri epistemologis yang dilaporkan secara konsisten dalam sumber-sumber tersebut. Artikel yang masih berstatus versi awal dalam proses penelaahan terbuka diperlakukan sebagai argumen konseptual, bukan sebagai bukti empiris.

Analisis dilakukan melalui lima langkah. Pertama, mengidentifikasi dan menyeleksi literatur berdasarkan kriteria di atas. Kedua, mengklasifikasikan isi literatur ke dalam empat tema: epistemologi positivisme, karakteristik tes objektif, representasi kompetensi bahasa (termasuk diglosia), dan tuntutan kognitif menurut Taksonomi Bloom. Ketiga, menurunkan tiga asumsi positivistik dari tema pertama, yaitu keteramatan, keterpilahan, dan keterbandingan, lalu menjadikannya pertanyaan analitis terhadap tema kedua sampai keempat. Keempat, membandingkan setiap asumsi dengan teori validitas, terutama gagasan ketakterwakilan konstruk dan varians tak relevan (Messick 1995) serta validasi berbasis argumen (Kane 2013). Kelima, menyintesiskan temuan menjadi pemetaan kompetensi–bukti.

Dua pembedaan dijaga sepanjang analisis. Pertama, kesejajaran analitis antara positivisme dan penilaian berbasis skor tidak dibaca sebagai hubungan sebab-akibat historis. Kedua, bentuk soal dibedakan dari tingkat kognitifnya: pilihan ganda dapat menuntut penalaran tinggi jika dirancang dengan tepat, sedangkan soal uraian tidak otomatis menuntutnya. Keterbatasan kajian ini terletak pada sifatnya yang berbasis

literatur. Hasilnya tidak dimaksudkan untuk mengukur seberapa sering tes objektif digunakan di lembaga tertentu.

TEMUAN

Analisis menghasilkan lima temuan pokok yang dirangkum pada uraian berikut ini.

Temuan 1: tiga asumsi positivistik melekat pada penilaian berbasis skor. Literatur tentang Comte menempatkan pengamatan dan penemuan keteraturan antargejala sebagai inti tahap positif (Nugroho 2016; Jubaedah dan Hayati 2024). Pada ranah penilaian, orientasi ini terbaca sebagai tiga asumsi: kompetensi dapat diketahui lewat respons yang teramati (keteramatan), kompetensi dapat dipecah menjadi butir yang dijumlahkan (keterpilahan), dan skor yang sama bermakna sama lintas peserta dan waktu (keterbandingan). Studi di Indonesia mencatat kecenderungan serupa pada asesmen yang menekankan capaian terukur dan penilaian terstandar (Nita dkk. 2025).

Temuan 2: tes objektif kuat pada konsistensi, efisiensi, dan cakupan, dan dapat dirancang untuk tuntutan yang lebih tinggi. Taksonomi 31 pedoman penulisan butir pilihan ganda telah tersedia untuk asesmen kelas (Haladyna dkk. 2002), dan kajian tentang penulisan butir menunjukkan bahwa butir pilihan ganda dapat dirancang untuk mengukur berpikir tingkat tinggi (Scully 2017). Kajian tentang instrumen penilaian kognitif juga membedakan fungsi tes objektif dan tes uraian (Putri dkk. 2022).

Temuan 3: representasi kompetensi menghadapi ancaman ketakterwakilan, terutama pada kompetensi produktif, pragmatik, dan kesadaran ragam. Teori validitas menegaskan bahwa yang divalidasi adalah tafsir skor, bukan tes (Messick 1995; Kane 2013). Model kompetensi komunikatif memuat komponen sosiolinguistik dan strategis di samping gramatikal (Canale dan Swain 1980), dan diglosia menuntut perhatian pada ragam bahasa yang dipakai dalam tugas tes (Ferguson 1959; Alharbi dan Sa'ad 2026).

Temuan 4: format soal tidak menentukan tingkat kognitif, dan tingkat kognitif tidak mencakup seluruh kompetensi bahasa. Taksonomi Bloom revisi memetakan proses kognitif (Krathwohl 2002; Gunawan dan Palupi 2016; Nafiati 2021) dan telah dibahas pula dalam kaitannya dengan peran asesmen (Airasian dan Miranda 2002). Kerangka ini merinci tuntutan berpikir tugas, bukan komponen kemampuan berbahasa seperti kelancaran, pengelolaan giliran bicara, atau ketepatan pragmatik.

Temuan 5: risiko subjektivitas pada penilaian performa dapat dikelola. Tinjauan atas 75 studi menunjukkan bahwa rubrik dapat meningkatkan reliabilitas penilaian apabila bersifat analitik, spesifik terhadap tugas, dan dilengkapi contoh serta pelatihan penilai, meskipun rubrik saja tidak menjamin validitas (Jonsson dan Svingby 2007).

Tabel 1
Ringkasan Temuan dan Implikasinya terhadap Representasi Kompetensi

Temuan	Dasar literatur	Implikasi bagi representasi kompetensi
Penilaian berbasis skor mengandaikan keteramatan, keterpilahan, dan keterbandingan	Nugroho (2016); Jubaedah dan Hayati (2024); Nita dkk. (2025)	Skor berisiko dibaca sebagai kompetensi itu sendiri tanpa jeda tafsir
Tes objektif kuat pada konsistensi, efisiensi, dan cakupan; dapat dirancang untuk tuntutan lebih tinggi	Haladyna dkk. (2002); Putri dkk. (2022); Scully (2017)	Tepat untuk pengetahuan bentuk dan sebagian pemahaman reseptif
Ancaman ketakterwakilan konstruk dan varians tak relevan; diglosia memperlebar persoalan	Messick (1995); Canale dan Swain (1980); Ferguson (1959); Alharbi dan Sa'ad (2026)	*Kalām*, *kitābah*, pragmatik, dan kesadaran ragam kurang terwakili
Format soal tidak menentukan tingkat kognitif; tingkat kognitif tidak mencakup seluruh kompetensi bahasa	Krathwohl (2002); Airasian dan Miranda (2002); Scully (2017)	Taksonomi Bloom perlu dipasangkan dengan model kompetensi bahasa
Risiko subjektivitas penilaian performa dapat dikelola	Jonsson dan Svingby (2007)	Rubrik analitik, contoh kinerja, dan pelatihan penilai memperkuat bukti

PEMBAHASAN

Tiga Asumsi Positivistik dalam Penilaian Berbasis Skor

Pembacaan melalui Comte tidak dimulai dari tes, melainkan dari sikap epistemik. Menurut literatur yang dikaji, tahap positif dalam gagasan Comte adalah tahap ketika pengetahuan disusun dari pengamatan dan penemuan hubungan keteraturan antargejala, bukan dari penjelasan spekulatif tentang sebab terakhir (Nugroho 2016; Jubaedah dan Hayati 2024). Sikap ini menghasilkan tiga asumsi yang dalam penilaian berbasis skor sering bekerja secara diam-diam. Ketiganya bukan kekeliruan, sebab justru ketiganya yang membuat pengukuran berguna. Persoalannya muncul ketika asumsi itu tidak lagi diperiksa.

Asumsi keteramatan menganggap kompetensi, yang tidak dapat diamati langsung, dapat diketahui melalui respons yang teramati, misalnya huruf yang dipilih pada lembar jawaban. Respons itu bukan kompetensi, melainkan bukti yang harus ditafsirkan. Karena itu Messick (1995) menegaskan bahwa yang divalidasi adalah inferensi dari respons dan kinerja, bukan tes sebagai benda. Asumsi keterpilahan menganggap kompetensi dapat dipecah menjadi butir diskret yang jumlahnya menggambarkan keseluruhan. Gagasan ini berhadapan dengan model kompetensi komunikatif yang memandang kemampuan berbahasa sebagai gabungan komponen yang saling berkaitan (Canale dan Swain 1980).

Menjumlahkan butir pengetahuan bentuk tidak serta-merta menghasilkan gambaran tentang kemampuan menggunakan bahasa. Asumsi keterbandingan menganggap skor yang sama bermakna sama. Dalam kerangka validasi berbasis argumen, kebermaknaan itu harus disusun sebagai rangkaian inferensi yang masing-masing didukung bukti (Kane 2013), bukan diasumsikan dari keseragaman prosedur.

Kesejajaran ini bersifat analitis, bukan genealogis. Penggunaan tes objektif juga dipengaruhi pertimbangan praktis seperti jumlah peserta, waktu, kebijakan, dan sumber daya, sehingga klaim bahwa tes objektif adalah "akibat" positivisme tidak dapat disimpulkan dari kemiripan prinsip. Kajian ini hanya mengklaim bahwa kerangka positivisme berguna untuk bertanya: bukti apa yang dianggap sah untuk menyatakan seseorang mampu berbahasa Arab, bagaimana bukti itu diperoleh, dan apa yang tersisa di luar bukti itu. Sikap serupa terlihat dalam pembacaan kritis atas asesmen bahasa, yang menunjukkan bahwa persoalan validitas dan etika tampak berbeda ketika paradigma pengukuran sendiri dipersoalkan (Lynch 2001).

Representasi Kompetensi: dari Konstruk, Butir, Skor, hingga Tafsir

Istilah representasi dalam judul artikel ini dapat dirinci menjadi empat titik. Pertama, representasi konstruk: apa yang dimaksud dengan kompetensi berbahasa Arab pada jenjang dan tujuan tertentu. Kedua, representasi sampel: butir atau tugas apa yang dipilih untuk mewakili konstruk itu. Ketiga, representasi agregat: bagaimana respons diubah menjadi skor. Keempat, representasi tafsir: apa yang dikatakan tentang peserta berdasarkan skor itu. Ketakterwakilan konstruk dapat terjadi pada titik kedua, ketika sampel butir terlalu sempit, dan pada titik keempat, ketika skor yang sah untuk satu tafsir dipakai untuk tafsir lain yang lebih luas (Messick 1995).

Contoh dari kompetensi bahasa Arab memperjelas hal ini. Butir pilihan ganda yang meminta peserta memilih *harakat* akhir yang tepat memberikan bukti yang baik tentang pengetahuan *qawā'id*. Namun, butir itu tidak menunjukkan apakah peserta dapat menyusun kalimat bermakna dalam percakapan. Demikian pula, tes pemahaman bacaan tidak membuktikan kemampuan menulis teks atau merespons lawan bicara secara spontan. Selama skor tes dibaca sebagai bukti pengetahuan bentuk dan pemahaman teks, representasinya memadai. Ketika skor yang sama dibaca sebagai bukti "kemampuan berbahasa Arab", sebagian konstruk telah hilang.

Diglosia memperdalam persoalan ini. Jika tugas tes hampir seluruhnya memakai ragam standar tulis, skor tinggi dapat berdampingan dengan kesulitan menanggapi tuturan lisan sehari-hari. Argumen ini dikemukakan dalam tinjauan konseptual atas asesmen bahasa Arab sebagai bahasa asing, yang mengusulkan agar tugas tes memuat pula komprehensi dan produksi ragam lisan sesuai tuntutan komunikasi nyata (Alharbi dan Sa'ad 2026). Karena artikel tersebut berupa kerangka konseptual yang belum diuji secara

empiris, kajian ini memperlakukannya sebagai argumen yang masuk akal dan layak diuji, bukan sebagai bukti. Pada konteks pendidikan bahasa Arab di Indonesia, pertanyaan ragam mana yang menjadi sasaran, yakni bahasa Arab untuk memahami teks keagamaan, untuk akademik, atau untuk komunikasi lisan, perlu dijawab pada tahap perumusan konstruk, bukan setelah soal selesai ditulis.

Ancaman kedua menurut Messick, varians yang tidak relevan dengan konstruk, juga nyata pada tes objektif. Kualitas distraktor, kejelasan butir, dan kemungkinan menebak dapat memengaruhi skor tanpa berkaitan dengan kompetensi yang diukur. Pedoman penulisan butir (Haladyna dkk. 2002) pada dasarnya merupakan upaya sistematis mengurangi sumber varians tersebut. Dengan demikian, tes objektif tidak otomatis lebih "bersih" daripada penilaian performa. Keduanya hanya memiliki jenis ancaman yang berbeda.

Objektivitas Penskoran dan Validitas Interpretasi

Konsistensi penskoran adalah syarat yang diperlukan, tetapi bukan syarat yang memadai bagi tafsir yang sahih. Tes dengan kunci jawaban jelas masih dapat memiliki validitas terbatas bila butirnya hanya mewakili sebagian tujuan pembelajaran. Di sinilah pembedaan antara objektivitas prosedur dan validitas tafsir menjadi penting. Objektivitas menjawab pertanyaan "apakah dua pemeriksa memberi skor yang sama?", sedangkan validitas menjawab pertanyaan "apakah skor itu cukup untuk menyimpulkan apa yang hendak disimpulkan?" (Messick 1995; Kane 2013).

Sebaliknya, kekhawatiran bahwa penilaian performa terlalu subjektif tidak dapat dijadikan alasan untuk menghindarinya. Tinjauan Jonsson dan Svingby (2007) atas 75 studi menunjukkan bahwa rubrik analitik yang spesifik terhadap tugas, dilengkapi contoh kinerja dan pelatihan penilai, dapat meningkatkan reliabilitas penilaian, sekaligus mengingatkan bahwa rubrik saja tidak menjamin validitas. Dengan demikian, perbandingan yang tepat bukan antara tes objektif yang selalu baik dan penilaian performa yang selalu subjektif, melainkan antara bukti yang sesuai dan bukti yang tidak sesuai dengan konstruk yang ditargetkan. Pembedaan klasik "objektif" dan "subjektif" ternyata kurang membantu untuk merumuskan pertanyaan yang benar.

Dimensi lain muncul ketika evaluasi dipandang dari tujuan pendidikan. Biesta (2009) mengingatkan bahwa ketika perbincangan pendidikan hampir seluruhnya terfokus pada pengukuran dan perbandingan hasil, pertanyaan tentang tujuan pendidikan cenderung tersisih. Peringatan ini relevan bagi pembelajaran bahasa Arab yang tujuannya bukan hanya skor yang tinggi, melainkan kemampuan memahami teks, berkomunikasi, dan dalam banyak konteks juga memahami sumber keagamaan. Studi Nita dkk. (2025) pada sekolah dan madrasah atau pesantren, meskipun berfokus pada penilaian kinerja

guru, melaporkan adanya penilaian terstandar yang menekankan capaian terukur berdampingan dengan penilaian lisan dan pengamatan langsung pada mata pelajaran keagamaan. Temuan itu mengisyaratkan bahwa praktik di lapangan sudah memadukan berbagai jenis bukti, hanya saja perpaduan itu jarang dirumuskan secara eksplisit sebagai desain asesmen.

Taksonomi Bloom: Tuntutan Kognitif Tidak Sama dengan Kompetensi Komunikatif

Taksonomi Bloom yang direvisi memetakan proses kognitif ke dalam enam kategori, yakni mengingat, memahami, menerapkan, menganalisis, mengevaluasi, dan mencipta (Krathwohl 2002; Gunawan dan Palupi 2016; Nafiati 2021). Menurut Krathwohl (2002), kerangka asalnya dimaksudkan untuk memudahkan pertukaran butir soal antardosen sehingga bank butir yang mengukur tujuan pendidikan yang sama dapat dibangun. Fakta ini menarik untuk pembacaan kritis: kerangka yang kini dipakai untuk memperluas tujuan evaluasi sendiri lahir dari kepedulian pada keterbandingan pengukuran. Itu tidak mengurangi manfaatnya, tetapi mengingatkan bahwa Taksonomi Bloom adalah alat untuk menamai tuntutan kognitif, bukan teori tentang kemampuan berbahasa.

Pemetaan pada bahasa Arab dapat dilakukan tanpa mengaitkan tingkat tertentu dengan format tertentu. Pada tingkat mengingat, peserta mengenali arti kosakata. Pada tingkat memahami, peserta menyimpulkan isi teks. Pada tingkat menerapkan, peserta memakai kaidah untuk melengkapi ungkapan dalam situasi tertentu. Pada tingkat menganalisis, peserta membandingkan pilihan ungkapan atau menjelaskan mengapa suatu bentuk tidak cocok dengan konteks. Pada tingkat mengevaluasi, peserta menilai ketepatan terjemahan berdasarkan kriteria. Pada tingkat mencipta, peserta menghasilkan dialog atau paragraf untuk tujuan komunikatif tertentu. Pemetaan ini menunjukkan bahwa butir pilihan ganda yang dirancang dengan teks atau data baru dapat menuntut analisis dan evaluasi (Haladyna dkk. 2002; Scully 2017), sedangkan tugas uraian yang hanya meminta penyalinan definisi tetap berada pada tingkat mengingat. Tingkat kognitif ditentukan oleh proses mental yang benar-benar dituntut tugas, bukan oleh format responsnya (Airasian dan Miranda 2002).

Namun, bahkan tugas pada tingkat mencipta belum tentu menilai seluruh kompetensi komunikatif. Kelancaran berbicara, pengelolaan giliran bicara, pelafalan, dan kesesuaian pragmatik memerlukan kriteria khusus yang diturunkan dari model kompetensi bahasa, bukan dari taksonomi proses berpikir (Canale dan Swain 1980). Maka dua kerangka itu bekerja pada sumbu yang berbeda. Taksonomi Bloom menjawab pertanyaan seberapa berat tuntutan berpikir suatu tugas. Model kompetensi bahasa menjawab pertanyaan kemampuan berbahasa apa yang sedang dibuktikan. Perancang

asesmen perlu memeriksa keduanya agar tidak menyamakan kompleksitas kognitif dengan kecakapan komunikatif.

Pemetaan Kompetensi–Bukti sebagai Rumusan Implikatif

Berdasarkan analisis di atas, artikel ini mengusulkan pemetaan yang menghubungkan kompetensi bahasa Arab dengan bentuk asesmen, risiko representasi jika hanya mengandalkan tes objektif, dan bukti validitas yang perlu dicari (Tabel 2). Pemetaan ini merupakan sintesis konseptual penulis, bukan hasil pengujian empiris, dan tidak bersifat preskriptif: bentuk asesmen dapat dipadukan sesuai tujuan, jenjang, waktu, jumlah peserta, dan sumber daya.

Tabel 2
Pemetaan Kompetensi Bahasa Arab, Bentuk Asesmen, Risiko Representasi, dan Bukti Validitas

Kompetensi	Bentuk asesmen yang sesuai	Risiko representasi jika hanya tes objektif	Bukti validitas yang dicari
Mufradat	Tes objektif (pilihan ganda, menjodohkan)	Rendah untuk pengenalan makna; tinggi jika yang dimaksud adalah penggunaan kata dalam produksi	Kesesuaian butir dengan daftar kosakata target; kualitas distraktor
Qawā'id (nahwu–sharaf)	Tes objektif berbasis kalimat atau teks	Menunjukkan pengetahuan memilih bentuk, bukan kemampuan menyusun kalimat sendiri	Cakupan kaidah; daya pembeda butir; kejelasan konteks
Istimā' dan *qirā'ah*	Tes objektif berbasis teks, dilengkapi respons singkat	Sedang: bergantung pada ragam bahasa dan jenis teks yang dipakai	Kesesuaian teks dengan jenjang; tingkat kognitif yang dituntut
Kalām	Tugas performa: dialog, wawancara, simulasi situasi	Tinggi: kelancaran, interaksi, dan pragmatik tidak tertangkap	Rubrik analitik; pelatihan penilai; moderasi skor
Kitābah	Produk tulisan dan portofolio	Tinggi: produksi teks utuh tidak tertangkap	Rubrik (keterpaduan, ketepatan bahasa, tujuan komunikatif); contoh kinerja
Kesadaran ragam dan kesesuaian pragmatik	Tugas berbasis situasi dengan ragam standar dan lisan sesuai tujuan; observasi	Tinggi: konstruk sering tidak pernah dirumuskan sehingga tidak muncul dalam skor	Rumusan konstruk dalam kisi-kisi; penilaian ahli terhadap tugas

Tabel 2 menunjukkan bahwa tes objektif paling sesuai untuk kompetensi yang bukti validitasnya berkaitan dengan cakupan butir dan konsistensi skor, yakni *mufradat*, *qawā'id*, dan sebagian *istimā'* serta *qirā'ah*. Pada *kalām* dan *kitābah*, risiko representasinya tinggi bila tes objektif menjadi satu-satunya sumber bukti, sehingga bukti validitas bergeser pada

mutu rubrik, pelatihan penilai, dan moderasi skor. Baris terakhir menegaskan bahwa kesadaran ragam dan kesesuaian pragmatik tidak akan muncul dalam skor jika konstruknya tidak pernah dirumuskan sejak awal.

Sedikitnya empat prinsip dapat ditarik untuk perancangan asesmen. Pertama, penentuan instrumen dimulai dari tujuan pembelajaran dan rumusan konstruk, bukan dari preferensi pada satu format. Kedua, setiap tafsir skor perlu dilengkapi pernyataan tentang batasnya, misalnya "skor ini menunjukkan penguasaan kosakata target, bukan kemampuan berbicara". Ketiga, bukti dari berbagai instrumen dipadukan sesuai pertanyaan asesmen, dengan proporsi yang disesuaikan pada waktu dan sumber daya. Keempat, ragam bahasa yang menjadi sasaran, yakni standar, lisan, atau keduanya, dinyatakan secara eksplisit dalam kisi-kisi. Dalam bahasa epistemologis, langkah-langkah ini tidak menolak pengukuran dan objektivitas, tetapi menempatkannya secara proporsional: objektivitas prosedur tetap diperlukan, sementara validitas tafsir juga harus dijaga.

Keterbatasan dan Agenda Penelitian

Karena kajian ini berbasis literatur, ia tidak membuktikan seberapa sering tes objektif digunakan dalam pembelajaran bahasa Arab atau apakah tes tersebut dominan pada semua jenjang. Rujukan tentang Comte juga diperoleh melalui artikel sekunder, sedangkan sebagian argumen tentang diglosia bertumpu pada kerangka konseptual yang belum diuji. Penelitian lanjutan dapat menguji Tabel 2 melalui analisis dokumen soal ujian bahasa Arab untuk memetakan komposisi konstruk yang terwakili, penilaian ahli terhadap kisi-kisi, wawancara atau *think-aloud* dengan peserta didik untuk memeriksa proses yang sebenarnya dilakukan saat menjawab, serta studi korelasi antara skor tes objektif dan skor tugas performa pada peserta yang sama.

KESIMPULAN

Kajian ini menunjukkan bahwa positivisme Auguste Comte dapat digunakan sebagai lensa kritis untuk menyingkap tiga asumsi yang bekerja dalam penilaian berbasis skor, yaitu keteramatan, keterpilahan, dan keterbandingan. Ketiganya membuat pengukuran bermanfaat, tetapi juga membuka kemungkinan skor dibaca sebagai kompetensi itu sendiri. Kesejajaran ini bersifat analitis dan tidak menunjukkan bahwa praktik tes objektif secara langsung bersumber dari pemikiran Comte.

Tes objektif merepresentasikan dengan baik pengetahuan kebahasaan yang diskret dan sebagian pemahaman reseptif, dengan manfaat berupa konsistensi penskoran, efisiensi, dan standardisasi. Representasinya berisiko tidak lengkap ketika skornya dipakai sebagai bukti kemampuan berbicara, menulis, ketepatan pragmatik, dan kemampuan berpindah ragam bahasa Arab. Taksonomi Bloom membantu membedakan tuntutan

kognitif tugas, tetapi tingkat kognitif tidak ditentukan oleh format soal dan tidak mencakup seluruh kompetensi komunikatif.

Implikasinya, pemilihan instrumen perlu didasarkan pada keselarasan antara tujuan pembelajaran, konstruk, tugas, dan bukti validitas, sebagaimana dirumuskan dalam pemetaan kompetensi–bukti. Tes objektif dianjurkan digunakan secara proporsional dan dipadukan dengan tugas performa yang dinilai melalui rubrik ketika kompetensi yang dinilai menuntut bukti tambahan. Penelitian empiris lanjutan diperlukan untuk menguji pemetaan ini pada dokumen ujian, kisi-kisi, dan data skor bahasa Arab di konteks pendidikan tertentu.

DAFTAR PUSTAKA

Airasian, Peter W., dan Helen Miranda. 2002. "The Role of Assessment in the Revised Taxonomy." *Theory Into Practice* 41(4):249–54. doi:10.1207/s15430421tip4104_8.

Alharbi, Bader Hudayban, dan Auwal Adam Sa'ad. 2026. "Designing a Unified Model for Measuring Arabic Language Competence among Non-Native Learners." *F1000Research* 15:325. doi:10.12688/f1000research.173332.1.

Biesta, Gert. 2009. "Good Education in an Age of Measurement: On the Need to Reconnect with the Question of Purpose in Education." *Educational Assessment, Evaluation and Accountability* 21(1):33–46. doi:10.1007/s11092-008-9064-9.

Canale, Michael, dan Merrill Swain. 1980. "Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing." *Applied Linguistics* 1(1):1–47. doi:10.1093/applin/1.1.1.

Ferguson, Charles A. 1959. "Diglossia." *Word* 15(2):325–40. doi:10.1080/00437956.1959.11659702.

Gunawan, Imam, dan Anggarini Retno Palupi. 2016. "Taksonomi Bloom – Revisi Ranah Kognitif: Kerangka Landasan untuk Pembelajaran, Pengajaran, dan Penilaian." *Premiere Educandum: Jurnal Pendidikan Dasar dan Pembelajaran* 2(2):98–117. doi:10.25273/pe.v2i02.50.

Haladyna, Thomas M., Steven M. Downing, dan Michael C. Rodriguez. 2002. "A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment." *Applied Measurement in Education* 15(3):309–33. doi:10.1207/S15324818AME1503_5.

Jonsson, Anders, dan Gunilla Svingby. 2007. "The Use of Scoring Rubrics: Reliability, Validity and Educational Consequences." *Educational Research Review* 2(2):130–44. doi:10.1016/j.edurev.2007.05.002.

Jubaedah, Siti, dan Ela Hikmah Hayati. 2024. "Epistemologi Positivisme Auguste Comte." *Indonesian Journal of Islamic Theology and Philosophy* 6(2):121–38. doi:10.24042/ijitp.v6i2.24108.

Kane, Michael T. 2013. "Validating the Interpretations and Uses of Test Scores." *Journal of Educational Measurement* 50(1):1–73. doi:10.1111/jedm.12000.

Krathwohl, David R. 2002. "A Revision of Bloom's Taxonomy: An Overview." *Theory Into Practice* 41(4):212–18. doi:10.1207/s15430421tip4104_2.

Lynch, Brian K. 2001. "Rethinking Assessment from a Critical Perspective." *Language Testing* 18(4):351–72. doi:10.1177/026553220101800403.

Messick, Samuel. 1995. "Validity of Psychological Assessment: Validation of Inferences from Persons' Responses and Performances as Scientific Inquiry into Score Meaning." *American Psychologist* 50(9):741–49. doi:10.1037/0003-066X.50.9.741.

Muradi, Ahmad, La Ode Amaluddin, Danial Hilmi, Budiawan Hery, Fatwiah Noor, Hasni Noor, dan M. Syahran Jailani. 2022. "Evaluation of Arabic Learning Program at State Islamic Religious College During COVID-19." *Eurasian Journal of Educational Research* 2022(99):189–200. doi:10.14689/ejer.2022.99.011.

Nafiati, Dewi Amaliah. 2021. "Revisi Taksonomi Bloom: Kognitif, Afektif, dan Psikomotorik." *Humanika* 21(2):151–72. doi:10.21831/hum.v21i2.29252.

Nita, Kamalasia Rio, Mahdum, Jimmy Copiriady, dan Isjoni. 2025. "Reconciling the Empirical and the Divine: Islamic Epistemology as a Critical Lens for Positivist Teacher Assessment in Indonesia." *Jurnal Ushuluddin* 33(2). doi:10.24014/Jush.v33i2.38838.

Nugroho, Irham. 2016. "Positivisme Auguste Comte: Analisa Epistemologis dan Nilai Etisnya terhadap Sains." *Cakrawala: Jurnal Studi Islam* 11(2):167–77.

Nurzahira, Fadhilah, M. Irwan Jayadi, dan Ubaid Ridlo. 2025. "Konsep Evaluasi Pembelajaran Bahasa Arab." *Ihya Al-Arabiyah: Jurnal Pendidikan Bahasa dan Sastra Arab* 11(3). doi:10.30821/ihya.v11i3.26379.

Putri, Hellin, Desty Susiani, Nabilla Setya Wandani, dan Fia Alifah Putri. 2022. "Instrumen Penilaian Hasil Pembelajaran Kognitif pada Tes Uraian dan Tes Objektif." *Jurnal Papeda: Jurnal Publikasi Pendidikan Dasar* 4(2):139–48. doi:10.36232/jurnalpendidikandasar.v4i2.2649.

Scully, Darina. 2017. "Constructing Multiple-Choice Items to Measure Higher-Order Thinking." *Practical Assessment, Research, and Evaluation* 22(4). doi:10.7275/swgt-rj52.

Zurqoni, Zurqoni, Heri Retnawati, Syarifatur Rahmatullah, Hasan Djidu, dan Ezi Apino. 2020. "Has Arabic Language Learning Been Successfully Implemented?" *International Journal of Instruction* 13(4):715–30. doi:10.29333/iji.2020.13444a.