

Implementasi Penilaian Non-Tes *Mahārah Kalām*: Studi Kasus Multisitrus di Perguruan Tinggi Yogyakarta

**Cahya Edi Setyawan^a, Zulaeha^b, Akhsan^c, Nada Ibrahim
Alribdi^d, Ahmadi Me^e**

^aSTAI Masjid Syuhada Yogyakarta, ^bIAIN Kendari, ^cUniversitas
Ibrahimi Situbondo, ^dPrinces Nourah Bint Abdulrahman University
Riyadh, ^eUniversitas Ibrahimi Situbondo
Corresponding author: cahya.edi24@gmail.com

Abstract

Non-test assessment is an evaluation method that focuses on students' processes and performance in real contexts without relying on written examinations. This study aims to analyze the implementation of non-test assessment in the mahārah kalām (speaking skills) course at three higher education institutions: STAI Masjid Syuhada (STAIMS), Universitas Muhammadiyah Yogyakarta (UMY), and UIN Sunan Kalijaga Yogyakarta (UIN Suka). Employing a multi-site case study with a descriptive qualitative approach, data were collected through in-depth interviews, participatory classroom observations, and analysis of assessment documents and student outputs. The findings reveal diverse strategies tailored to each institution's characteristics. STAIMS applies a project-based assessment by producing Arabic-language vlogs to enhance communication, creativity, and digital literacy. UMY adopts a process-based approach through simulations and role-play, enabling gradual skill development with direct feedback. UIN Suka relies on intensive muḥādathah and khiṭābah minbarīyah practices to refine fluency and formal rhetoric. All three models are grounded in authentic and performance-based assessment principles, evaluating competence contextually, applicatively, and holistically. This study underscores that diverse approaches can complement one another in developing Arabic speaking proficiency,

offering comparative insights into locally adapted pedagogical innovations aligned with learners' needs and learning contexts.

Abstrak

Penilaian non-tes merupakan metode evaluasi yang menekankan proses dan kinerja mahasiswa dalam konteks nyata tanpa bergantung pada tes tertulis. Penelitian ini bertujuan menganalisis implementasi penilaian non-tes pada mata kuliah *mahārah kalām* di tiga perguruan tinggi, yaitu STAI Masjid Syuhada (STAIMS), Universitas Muhammadiyah Yogyakarta (UMY), dan UIN Sunan Kalijaga Yogyakarta (UIN Suka). Menggunakan pendekatan studi kasus multi-situs dengan metode kualitatif deskriptif, data diperoleh melalui wawancara mendalam, observasi partisipatif, serta analisis dokumen penilaian dan karya mahasiswa. Hasil penelitian menunjukkan variasi strategi sesuai karakteristik institusi. STAIMS menerapkan *project-based assessment* melalui pembuatan vlog berbahasa Arab yang melatih komunikasi, kreativitas, dan keterampilan digital. UMY menggunakan *process-based assessment* lewat simulasi dan permainan peran, membantu perkembangan bertahap dengan umpan balik langsung. UIN Suka mengandalkan praktik *muḥādathah* dan *khiṭābah minbarīyah* intensif untuk mengasah kefasihan dan retorika formal. Ketiga model berlandaskan prinsip *authentic* dan *performance-based assessment*, menilai kompetensi secara kontekstual, aplikatif, dan holistik. Studi ini menegaskan bahwa keberagaman pendekatan dapat saling melengkapi dalam membentuk keterampilan berbicara bahasa Arab, sekaligus memberikan wawasan komparatif terhadap inovasi pedagogis yang adaptif terhadap kebutuhan dan situasi pembelajaran.

المخلص

التقويم غير الاختباري هو أسلوب تقييم يركز على عملية وأداء الطلاب في سياق واقعي دون الاعتماد على الاختبارات التحريرية. يهدف هذا البحث إلى تحليل تطبيق التقويم غير الاختباري في مقرر مهارة الكلام في ثلاث جامعات، وهي: جامعة مسجد الشهداء الإسلامية بجوكجاكرتا، والجامعة المحمدية بجوكجاكرتا، وجامعة سونان كاليجاكا الإسلامية الحكومية. استخدم البحث منهج دراسة الحالة متعددة المواقع بأسلوب وصفي نوعي، وجمعت البيانات من خلال المقابلات المتعمقة، والملاحظة بالمشاركة، وتحليل وثائق التقييم وأعمال الطلاب. أظهرت النتائج تنوع الاستراتيجيات

وفقا لخصائص كل مؤسسة. اعتمدت جامعة مسجد الشهداء الإسلامية على أسلوب التقييم القائم على المشروع من خلال إعداد مقاطع فيديو باللغة العربية، لتعزيز مهارات التواصل والإبداع والقدرات الرقمية. أما الجامعة المحمدية بجوكجاكرتا فاعتمدت التقييم القائم على العملية عبر المحاكاة ولعب الأدوار، مما أتاح التطور التدريجي مع التغذية الراجعة المباشرة. بينما اعتمدت جامعة سونان كاليجاكا الإسلامية الحكومية على الممارسة المكثفة للمحادثة والخطابة المنبرية لصقل الطلاقة والفصاحة الرسمية. تستند النماذج الثلاثة إلى مبادئ التقييم الأصيل والتقييم القائم على الأداء، لتقييم الكفاءة بشكل سياقي وتطبيقي وشمولي. يؤكد هذا البحث أن تنوع الأساليب يشكل قوة في تنمية مهارات التحدث بالعربية، ويقدم رؤية مقارنة للابتكارات البيداغوجية المحلية المتكيفة مع حاجات ومواقف التعلم.

Keywords: Arabic speaking skills; authentic assessment; maharah kalam; non-test assessment

Pendahuluan

Kemampuan berbicara merupakan keterampilan produktif dalam pembelajaran bahasa yang membutuhkan pendekatan penilaian yang autentik, kontekstual, dan berbasis performa.¹ Namun demikian, praktik evaluasi di berbagai perguruan tinggi, termasuk dalam konteks Pendidikan Bahasa Arab, masih didominasi oleh tes tertulis yang kurang mampu merepresentasikan kemampuan komunikasi lisan secara menyeluruh.² Evaluasi semacam ini gagal menangkap elemen penting seperti kefasihan, intonasi, kebermaknaan dalam konteks, dan respons spontan dalam interaksi verbal. Di Yogyakarta, sebagai pusat pendidikan dengan banyak perguruan tinggi Islam yang memiliki program studi Bahasa

¹ Barry O'Sullivan, "Assessing Speaking," in *The Concise Companion to Language Assessment*, ed. Antony John Kunnan (New Jersey: Wiley-Blackwell, 2024), 143–53.

² Marta Montenegro-Rueda et al., "Assessment in Higher Education during the COVID-19 Pandemic: A Systematic Review," *Sustainability* 13, no. 19 (2021): 10509, <https://doi.org/10.3390/su131910509>.

Arab, kebutuhan akan pendekatan penilaian alternatif semakin mendesak.³

Tuntutan kurikulum berbasis kompetensi dan orientasi pada pembelajaran aktif menuntut adanya inovasi dalam strategi evaluasi kemampuan berbicara.⁴ Penilaian non-tes muncul sebagai solusi untuk menjawab tantangan tersebut, karena mampu memberikan gambaran menyeluruh tentang performa komunikasi nyata mahasiswa.⁵ Penilaian non-tes merupakan evaluasi berbasis performa yang tidak bergantung pada tes tertulis, melainkan pada aktivitas autentik seperti diskusi, presentasi, atau simulasi komunikasi dalam konteks nyata.⁶ Sebagai contoh, dalam mata kuliah keterampilan

³ R. Umi Baroroh and Nafisatun Nisa, "Non-Test Assessment Innovation Performance for Maharah Kalam Through Youtube in the Pandemic Era," *Ijaz Arabi Journal of Arabic Learning* 5, no. 1 (2022): 222–33, <https://doi.org/10.18860/ijazarabi.v5i1.14789>; Abdul Munip, "Tantangan dan Prospek Studi Bahasa Arab di Indonesia," *al Mahāra: Jurnal Pendidikan Bahasa Arab* 5, no. 2 (2020): 303–18, <https://doi.org/10.14421/almahara.2019.052.08>.

⁴ Liesbeth K.J. Baartman et al., "Evaluating Assessment Quality in Competence-Based Education: A Qualitative Comparison of Two Frameworks," *Educational Research Review* 2, no. 2 (2007): 114–29, <https://doi.org/10.1016/j.edurev.2007.06.001>; Sin Wang Chong and Hayo Reinders, "Introduction: Learning-Oriented Language Assessment—Insights for Evidence-Based Practices," in *Innovation in Learning-Oriented Language Assessment*, ed. Sin Wang Chong and Hayo Reinders (Cham: Palgrave Macmillan, 2023), 1–11, https://doi.org/10.1007/978-3-031-18950-0_1.

⁵ Sinta Purnama Sari, Muhammad Sayuti, and Tasha Ayu Azzahra, "Inovasi Tes Maharah Kalam Berbasis Penilaian Kurikulum Merdeka Pada Buku Durus Al-Lughah Al-Arabiyah Jilid 1," *Jurnal Pelita Manajemen Pendidikan* 1, no. 1 (2024): 70–79, <https://jurnalpelitanegribelantaraya.com/index.php/JMPMP/article/view/26>.

⁶ Mabruuri Mabruuri, Siti Rauhillah, and Hani Nurlaeli Wijayanti, "Implementasi Penilaian Non-Tes Dalam Perkuliahan Di Prodi Pendidikan Bahasa Arab Se-Yogyakarta," *Jurnal Ihtimam* 7, no. 2 (2024): 182–98, <https://doi.org/10.36668/jih.v7i02.1185>; Ikhrom Ikhrom, "The Impact of Learning Assessment Emergency on the Education Quality," *Lentera Pendidikan : Jurnal Ilmu Tarbiyah Dan Keguruan* 26, no. 1 (2023): 31–44, <https://doi.org/10.24252/lp.2023v26n1i4>.

berbicara bahasa Arab (*mahārah kalām*), mahasiswa dapat dinilai melalui simulasi percakapan spontan atau proyek video berbahasa Arab,⁷ yang memungkinkan penilaian terhadap kefasihan dan kemampuan berkomunikasi dengan bahasa Arab. Oleh karena itu, kajian mendalam mengenai implementasi penilaian non-tes pada mata kuliah *mahārah kalām* menjadi sangat penting dan relevan dengan perkembangan pembelajaran abad ke-21.

Dalam Program Studi Pendidikan Bahasa Arab, penilaian non-tes berperan penting dalam menumbuhkan keterampilan *mahārah kalām* melalui praktik langsung dan kegiatan berbasis performa.⁸ Mahasiswa dilibatkan dalam aktivitas seperti *muḥādathah*, *khiṭābah*, hingga pembuatan konten digital, yang memungkinkan mahasiswa mengaplikasikan bahasa dalam situasi nyata.⁹ Penilaian ini juga memberikan ruang bagi dosen untuk memberikan umpan balik formatif secara berkelanjutan, sehingga proses pembelajaran menjadi lebih reflektif dan bermakna.¹⁰ Penilaian non-tes tidak hanya meningkatkan

⁷ Junjun Muhamad Ramdani and R. Rahmat, "Promoting Speaking Spontaneity in Large Classes: An Action Research Study in an Indonesian EFLuniversity Setting," *Indonesian Journal of Applied Linguistics* 8, no. 2 (2018): 388–401, <https://doi.org/10.17509/ijal.v8i2.13304>; Azman Che Mat et al., "An Authentic Learning Environment Based on Video Project among Arabic Learners," *International Journal of Applied Linguistics and English Literature* 6, no. 4 (2017): 143–48, <https://doi.org/10.7575/aiac.ijalel.v.6n.4p.143>.

⁸ Katrien Vanpee and Dan Soneson, "Arabic Proficiency Improvement Through a Culture of Assessment," in *Foreign Language Proficiency in Higher Education*, ed. Paula Winke and Susan M. Gass, vol. 37, Educational Linguistics (Cham: Springer, 2019), 197–216, https://doi.org/10.1007/978-3-030-01006-5_11.

⁹ Muhammad Nidom Hamami Abicandra, "Application of Authentic Assessment in Arabic Language Teaching in the Program of Studying the Arabic Language at the Idrisi Islamic Boarding School for Boys, Jember," *Journal of Arabic Language Teaching* 2, no. 1 (2022): 1–16, <https://doi.org/10.35719/arkhas.v2i1.1267>.

¹⁰ Risna Rianti Sari and Alfatus Syarofah, "Measuring Students' Creativity in Arabic Speaking Class Based on Project Based Learning

kualitas pembelajaran *mahārah kalām*, tetapi juga berkontribusi terhadap pengembangan *soft skills* mahasiswa.¹¹ Oleh karena itu, pendekatan ini sangat potensial untuk diimplementasikan secara sistematis dalam rangka memperkuat kualitas pendidikan Bahasa Arab di perguruan tinggi.¹²

Berbagai studi tentang penilaian non-tes dalam pembelajaran bahasa telah dilakukan. Penelitian yang dilakukan oleh Fathi, dkk. Menunjukkan bahwa latihan interaksi komunikatif dengan bantuan AI dapat meningkatkan keterampilan berbicara dan *willingness to communicate* mahasiswa. Studi ini menegaskan pentingnya pembelajaran keterampilan berbicara melalui praktik dialogis bermakna, bukan hanya sekadar *drilling* bentuk-bentuk pola.¹³

Sementara itu, penelitian oleh Widodo dan Qosim berfokus pada penerapan penilaian kinerja (*performance assessment*) dalam kursus bahasa Arab tingkat dasar, yang mengutamakan aspek pelafalan, intonasi, dan kefasihan dalam evaluasi langsung melalui praktik. Kedua penelitian tersebut memberikan kontribusi berharga dalam menguatkan pendekatan penilaian berbasis performa untuk keterampilan berbicara. Namun demikian, keduanya belum secara eksplisit menelusuri bagaimana proses perencanaan dan pelaksanaan penilaian non-tes dilakukan secara sistematis dalam konteks

Model," *Abjadia : International Journal of Education* 8, no. 1 (2023): 9–16, <https://doi.org/10.18860/abj.v8i1.21058>.

¹¹ Elen Nokalia Angelina, *The Effect of Project-Based Learning in Improving Students' Speaking Ability (Undergraduate Thesis)* (Banda Aceh: Ar-Raniry State Islamic University, 2020), <https://repository.ar-raniry.ac.id/id/eprint/13865/>.

¹² Maimun, "Evaluasi dalam pembelajaran bahasa Arab di STAIN Pamekasan," *NUANSA: Jurnal Penelitian Ilmu Sosial dan Keagamaan Islam* 8, no. 2 (2011): 285–96, <https://doi.org/10.19105/nuansa.v8i2.17>.

¹³ Jalil Fathi, Masoud Rahimi, and Ali Derakhshan, "Improving EFL Learners' Speaking Skills and Willingness to Communicate via Artificial Intelligence-Mediated Interactions," *System* 121 (2024): 103254, <https://doi.org/10.1016/j.system.2024.103254>.

institusi pendidikan tinggi Islam di Indonesia, khususnya dalam mata kuliah *mahārah kalām* yang memiliki karakteristik pedagogis dan kurikulum yang lebih kompleks.¹⁴

Penelitian ini hadir untuk mengisi kekosongan tersebut dengan mengkaji secara holistik dan mendalam bagaimana dosen di perguruan tinggi Islam Yogyakarta merancang, melaksanakan, dan merefleksikan penilaian non-tes pada mata kuliah *mahārah kalām*, mulai dari perencanaan instrumen, strategi pelaksanaan di kelas, hingga teknik umpan balik. Berbeda dari penelitian Fathi, dkk. yang lebih menekankan faktor praktik dialogis bermakna, serta studi Widodo dan Qosim yang fokus pada evaluasi akhir, penelitian ini mengambil perspektif proses dan kelembagaan yang belum banyak diangkat dalam literatur.

Meskipun sejumlah penelitian terdahulu telah mengkaji implementasi penilaian non-tes dalam pembelajaran bahasa Arab, fokus kajiannya berbeda dan belum secara spesifik menyoroti pelaksanaan penilaian non-tes pada keterampilan *mahārah kalām* di level perguruan tinggi. Selain itu, belum banyak studi yang membandingkan secara mendalam variasi pendekatan penilaian non-tes yang diterapkan di berbagai institusi pendidikan tinggi dalam konteks lokal tertentu, seperti di perguruan tinggi Islam Yogyakarta. Padahal, perbedaan institusi dapat memengaruhi strategi pengajaran dan evaluasi yang digunakan. Oleh karena itu, penelitian ini menempati posisi penting dengan mengkaji secara khusus bagaimana penilaian non-tes *mahārah kalām* diimplementasikan pada tiga perguruan tinggi di Yogyakarta, serta mengidentifikasi tantangan dan keunggulan masing-masing pendekatan dalam

¹⁴ Joko Widodo and Muhammad Nanang Qosim, "Penilaian Kinerja pada Pembelajaran Maharah Al-Kalam Level 1 di Kursus Bahasa Arab Al-Arobiya Surakarta," *Uktub: Journal of Arabic Studies* 1, no. 2 (2021): 84–101, <https://doi.org/10.32678/uktub.v1i2.5814>.

membentuk kompetensi komunikatif mahasiswa secara kontekstual dan autentik.

Implementasi penilaian non-tes *mahārah kalām* di beberapa perguruan tinggi di Yogyakarta, seperti STAI Masjid Syuhada (STAIMS), Universitas Muhammadiyah Yogyakarta (UMY), dan UIN Sunan Kalijaga Yogyakarta (UIN Suka), menunjukkan pendekatan yang beragam sesuai dengan karakter masing-masing institusi. STAIMS menonjol dengan proyek digital kreatif,¹⁵ UMY menerapkan simulasi berbasis skenario interaktif,¹⁶ sedangkan UIN Suka mempertahankan model tradisional berbasis *muhādathah* dan *khiṭābah*.¹⁷ Ketiga model ini memiliki ciri khas tersendiri dalam menyasar kompetensi berbicara bahasa Arab mahasiswa. Namun demikian, variasi pendekatan ini belum banyak dikaji secara komparatif dalam kerangka teori penilaian autentik dan kompetensi komunikatif secara utuh. Keunikan masing-masing strategi tersebut memunculkan pertanyaan mendalam tentang efektivitas, relevansi, serta keterpaduan antara instrumen, proses, dan hasil penilaian. Oleh karena itu, penting dilakukan studi lebih lanjut yang tidak hanya mendeskripsikan perbedaan, tetapi juga mengeksplorasi potensi integratif dari masing-masing pendekatan dalam membentuk model evaluasi non-tes *mahārah kalām* yang lebih kontekstual dan transformatif.

Berdasarkan dinamika yang terjadi di STAIMS, UMY, dan UIN Suka, maka sangat relevan untuk melakukan penelitian lebih mendalam mengenai implementasi penilaian non-tes pada mata kuliah *mahārah kalām*. Setiap institusi memiliki pendekatan, keunggulan, dan tantangan tersendiri yang mencerminkan keberagaman praktik di lapangan, sehingga

¹⁵ Wawancara dengan Kaprodi Pendidikan Bahasa Arab pada Senin, 02 Juli 2024.

¹⁶ Wawancara dengan TN, dosen Pendidikan Bahasa Arab, Rabu, 04 Juli 2024.

¹⁷ Wawancara dengan Sekprodi Pendidikan Bahasa Arab, Kamis, 05 Juli 2024.

tidak cukup hanya dilihat dari satu perspektif umum. Untuk itu, pendekatan studi kasus multi-situs dipilih oleh peneliti karena memungkinkan peneliti mengeksplorasi secara kontekstual bagaimana setiap kampus mengembangkan dan menerapkan model penilaian non-tes, tahapan implementasi penilaiannya serta pengembangan rubriknya. Arah temuan penelitian ini terletak pada integrasi pedagogis berbasis performa dengan prinsip asesmen autentik dalam penilaian non-tes *mahārah kalām* di pendidikan tinggi Islam Indonesia melalui pendekatan multi-situs yang menghasilkan model konseptual evaluasi kontekstual, aplikatif, dan responsif terhadap dinamika pembelajaran abad ke-21.

Hasil Penelitian

Perencanaan dan Pelaksanaan Penilaian Non-Tes Mahārah Kalām di STAIMS, UMY, dan UIN Suka

Berdasarkan hasil penelitian studi kasus multi-situs, implementasi penilaian non-tes *mahārah kalām* di tiga perguruan tinggi menunjukkan tahapan yang relatif serupa namun dengan kekhasan masing-masing institusi. Tahapan pertama dimulai dari perencanaan penilaian, yang melibatkan penyusunan indikator keterampilan berbicara sesuai capaian pembelajaran. STAIMS dalam perencanaan penilaian non-tes difokuskan pada pengembangan proyek berbasis vlog berbahasa Arab yang dikaitkan dengan tema-tema lokal, seperti pariwisata. UMY merancang skenario permainan peran dan simulasi diskusi tematik. Sementara itu, di UIN Suka, perencanaan diarahkan pada penguatan metode tradisional seperti *muḥadathah* dan *khiṭābah minbarīyah*, yang dilaksanakan secara bertahap dan terstruktur. Perbedaan ini menunjukkan bahwa tahapan awal implementasi sangat dipengaruhi oleh pendekatan pedagogis masing-masing institusi.

Tabel 1

Matriks penerapan penilaian non-tes mahārah kalām

Aspek	STAIMS	UMY	UIN Suka
Tujuan Penilaian	Mengukur kelancaran berbicara dan ketepatan struktur kalimat dalam konteks keagamaan.	Menilai kefasihan, keterlibatan, dan daya nalar dalam presentasi dan diskusi tematik.	Mengevaluasi kemampuan komunikasi fungsional dalam konteks akademik dan sosial.
Teknik Non-Tes	Observasi langsung saat praktik <i>muḥāḍarah</i> ; wawancara.	Portofolio video presentasi; observasi kelas dan <i>peer feedback</i> .	Observasi partisipasi kelas; asesmen berbasis proyek (tugas pidato, debat).
Waktu Pelaksanaan	Setiap akhir sesi mingguan (dalam praktik intensif).	Tengah dan akhir semester (terintegrasi dengan tugas akhir).	Secara berkala tiap modul/topik selesai (berbasis rubrik).
Kriteria Keberhasilan	Kesesuaian isi, kelancaran bicara, penggunaan <i>mufradāt</i> dan <i>naḥw</i> .	Kecakapan menyampaikan gagasan, argumentasi logis, dan penggunaan kosakata tematik.	Keterpahaman pesan, koherensi logika, dan keberanian berkomunikasi di depan publik.
Pelaksana/ Penilai	Dosen pengampu dan sesekali kolaborasi dengan teman sebaya.	Dosen, dengan penilaian mandiri dan penilaian teman sejawat.	Tim pengampu mata kuliah dengan asesmen berlapis (<i>multi-rater</i>).
Hambatan/ Tantangan	Keterbatasan waktu latihan dan minimnya literatur latihan lisan.	Variasi kemampuan mahasiswa dan kurangnya keberanian tampil.	Kelas besar, sulit observasi mendalam, serta minimnya pelatihan non-verbal.

Tahapan pelaksanaan penilaian non-tes *mahārah kalām* di tiga kampus menunjukkan variasi strategi. STAI Masjid Syuhada menilai *vlog* sebagai tugas akhir berdasarkan kefasihan, ketepatan, dan kreativitas. UMY menerapkan observasi berkelanjutan dalam simulasi dan diskusi untuk umpan balik formatif. UIN Suka menilai pidato dan dialog spontan secara langsung di kelas. Evaluasi akhir di STAIMS dan UMY menggunakan portofolio dan dokumentasi video, sementara UIN Suka mengandalkan penilaian lisan berbasis observasi.

Tabel 2
Analisis perbandingan implementasi antar kampus

Aspek	STAIMS	UMY	UIN Suka
Jenis Penilaian	Non-tes berbasis proyek (<i>Vlog</i>)	Non-tes berbasis proses (bertahap)	Non-tes berbasis praktik (<i>Muhādathah</i> & <i>Khiṭābah</i>)
Tahapan Implementasi	Perencanaan tema, penyusunan naskah, rekaman, unggah YouTube	Pengenalan konteks dialog terpandu → <i>role-play</i> → refleksi	→ <i>Muhādathah</i> kelompok → <i>khiṭābah</i> individu
Teknologi & Media	Media digital (YouTube)	Latihan lisan langsung dalam kelas	Praktik langsung tanpa media digital
Pendekatan Didaktik	Kreativitas dan produksi mandiri	Pendekatan konten komunikatif-kolaboratif	Retorika klasik dan formal
Frekuensi Penilaian	Satu kali tugas akhir (sumatif)	Berulang sepanjang semester (formatif)	Dua momen utama (formatif dan sumatif)
Fokus Kompetensi	<i>Fluency, grammar, Situational awareness, Structured speech, vocabulary, spontaneity, confidence, presentation creativity</i>		<i>formal Arabic, rhetoric</i>
Peran Dosen & Teman	Dosen sebagai fasilitator dan evaluator akhir	Dosen dan teman sebagai sejawat pemberi umpan balik	Dosen sebagai penilai formal, teman sebagai pendukung latihan
Refleksi & Umpan Balik	Minim (terjadi di akhir)	Sangat kuat (refleksi langsung berkelanjutan)	Moderat (lebih pada evaluasi, bukan refleksi)

Penilaian non-tes *mahārah kalām* di ketiga kampus (Tabel 3) menunjukkan bahwa penguatan keterampilan berbicara dalam bahasa Arab memerlukan variasi pendekatan: berbasis proyek, proses, maupun praktik formal. Masing-masing mencerminkan filosofi pembelajaran yang berbeda namun bersatu dalam tujuan utama: membentuk mahasiswa yang fasih, percaya diri, dan mampu berbicara bahasa Arab dalam konteks autentik.

Implementasi penilaian non-tes *mahārah kalām* tidak bisa lepas dari instrumen penilaian yang di susun oleh dosen. Instrumen yang disusun berbentuk rubrik. Berikut rubrik dari masing-masing kampus. Masing-masing rubrik dikembangkan berdasarkan karakteristik metode penilaian yang diterapkan di tiap institusi.

Tabel 3
Analisis tematik penilaian non-tes antar kampus

Tema Tematik	Penjabaran
Fokus pada Kemampuan Komunikatif	Semua kampus mengevaluasi kemampuan berbicara fungsional dalam bahasa Arab, meskipun dengan metode berbeda (<i>vlog</i> , <i>role-play</i> , pidato).
Penguatan Kompetensi Lisan dalam Konteks Autentik	Ketiganya menggunakan konteks nyata/simulatif: ekspresi digital (STAIMS), peran sosial (UMY), dan dakwah formal (UIN Suka).
Pengembangan <i>Soft Skill & Public Speaking</i>	Ketiga metode mengembangkan keberanian berbicara di ruang publik melalui media digital, diskusi kelompok, atau mimbar terbuka.
Variasi Media & Metode Sesuai Kultur Institusi	Pendekatan berbeda mencerminkan kultur: STAIMS (kreatif-digital), UMY (kontekstual-modern), UIN Suka (klasik-formal).
Kesadaran akan Proses, Bukan Sekadar Hasil	Penilaian tidak hanya hasil akhir. UMY eksplisit pada penilaian proses; STAIMS & UIN Suka juga mengakomodasi tahapan latihan.

***Rubrik Penilaian Non-Tes Mahārah Kalām di STAIMS
(Penilaian Proyek Pembuatan Vlog Bahasa Arab)***

Rubrik penilaian non-tes *mahārah kalām* di STAIMS merefleksikan pendekatan *project based assessment* berbasis *authentic assessment* ala *Wiggins*. Evaluasi dilakukan melalui produk nyata berupa *vlog* berbahasa Arab yang menuntut penerapan kompetensi komunikasi secara otentik. Penilaian mencakup aspek proses dan produk, termasuk isi, relevansi tema, media digital, dan kejelasan penyampaian. Berdasarkan teori kompetensi komunikatif *Canale & Swain*, rubrik ini mencakup empat dimensi utama: gramatikal, sociolinguistik, strategis, dan wacana. Hal ini terlihat dari indikator seperti kefasihan, kesesuaian konteks, pemanfaatan media, serta koherensi pesan. Dengan demikian, rubrik ini memadukan secara efektif prinsip penilaian autentik dengan kerangka teori kompetensi komunikatif dalam evaluasi *mahārah kalām*.

Tabel 4
Rubrik penilaian non-tes mahārah kalām di STAIMS

Kriteria	Skor 4	Skor 3	Skor 2	Skor 1
Kefasihan Berbicara	Lancar tanpa kesalahan berarti	Lancar dengan sedikit kesalahan	Agak terbata, banyak kesalahan	Tidak lancar dan sulit dipahami
Kosakata dan Tata Bahasa	Variatif dan tepat	Umumnya tepat	Beberapa kesalahan	Banyak kesalahan dasar
Penggunaan Media (YouTube)	Kreatif dan profesional	Cukup menarik	Kurang menarik	Minim kreativitas
Konten dan Relevansi	Informatif, sesuai tema	Cukup informatif	Kurang relevan	Tidak sesuai tema
Kejelasan Penyampaian	Jelas dan mudah diikuti	Umumnya jelas	Kurang jelas	Tidak bisa dipahami

Keterangan: Skor 4: Sangat baik; Skor 3: Baik; Skor 2: Cukup; Skor 1: kurang

***Rubrik Penilaian Non-Tes Mahārah Kalām di UMY
(Penilaian Proses Simulasi dan Role Play Bahasa Arab)***

Rubrik non-tes *mahārah kalām* di UMY (Tabel 5) menggunakan pendekatan simulasi dan *role-play* yang menekankan praktik komunikasi autentik dalam suasana menyerupai dunia nyata. Penilaian dilakukan secara berkelanjutan dengan fokus pada aspek partisipasi aktif, keberanian, intonasi, dan responsivitas dialog mahasiswa. Pendekatan ini mencerminkan penerapan prinsip *authentic assessment* menurut *Wiggins* yang menekankan evaluasi performatif dan berorientasi proses. Dari teori kompetensi komunikatif *Canale & Swain*, rubrik ini mencakup dimensi gramatikal, sociolinguistik, strategis, dan wacana. Aspek-aspek seperti adaptasi konteks, ekspresi non-verbal, dan kelancaran bicara turut menjadi elemen penting dalam penilaiannya. Secara keseluruhan, rubrik ini mengintegrasikan teori penilaian autentik dengan kerangka kompetensi komunikatif secara aplikatif dan kontekstual.

Tabel 5
Rubrik penilaian non-tes mahārah kalām di UMY

Kriteria	Skor 4	Skor 3	Skor 2	Skor 1
Partisipasi dalam Simulasi	Aktif, berinisiatif	Cukup aktif	Pasif sebagian waktu	Tidak aktif sama sekali
Ketepatan Kosakata	Sangat tepat dan variatif	Umumnya tepat	Beberapa kesalahan	Banyak kesalahan
Responsivitas Dialog	Tanggap dan kontekstual	Umumnya tanggap	Kadang bingung	Tidak responsif
Keberanian & Intonasi	Percaya diri dan ekspresif	Cukup percaya diri	Terlihat ragu-ragu	Gugup dan tidak jelas
Perkembangan Progresif	Meningkat signifikan	Ada peningkatan	Sedikit peningkatan	Tidak berkembang

Keterangan: Skor 4: Sangat baik; Skor 3: Baik; Skor 2: Cukup; Skor 1: kurang

***Rubrik Penilaian Non-Tes Mahārah Kalām di UIN Suka
(Penilaian muḥādathah dan khiṭābah minbarīyah)***

Instrumen penilaian non-tes di UIN Sunan Kalijaga (Tabel 6) menilai keterampilan *muḥādathah* dan *khiṭābah minbarīyah* melalui aktivitas lisan yang formal dan kontekstual, mencerminkan praktik komunikasi di ruang akademik dan keagamaan. Fokus utama rubrik adalah performativitas mahasiswa, seperti struktur ujaran, artikulasi, dan retorika, bukan sekadar penguasaan linguistik. Berdasarkan teori *Canale & Swain*, penilaian ini telah mencakup dimensi gramatikal, sosiolinguistik, strategis, dan wacana, menandakan pendekatan yang holistik. Penggunaan indikator seperti gestur dan intonasi memperkuat aspek strategis dalam penyampaian pesan lisan. Dengan demikian, pendekatan ini mendorong penilaian berorientasi praktik nyata yang relevan dengan tuntutan komunikasi publik.

Tabel 6

Rubrik penilaian non-tes mahārah kalām di UIN Suka

Kriteria	Skor 4	Skor 3	Skor 2	Skor 1
Struktur Teks Lisan	Terstruktur jelas dan logis	Cukup terstruktur	Kurang teratur	Tidak terstruktur
Kejelasan Artikulasi	Sangat jelas dan tegas	Umumnya jelas	Kurang jelas	Tidak terdengar jelas
Retorika & Gaya Bicara	Sangat menarik dan meyakinkan	Cukup menarik	Biasa saja	Monoton dan membosankan
Tata Bahasa Arab	Sangat tepat dan kompleks	Umumnya tepat	Cukup tepat	Banyak kesalahan
Kesesuaian Tema	Sangat relevan dan aktual	Relevan	Kurang sesuai	Tidak sesuai tema

Keterangan: Skor 4: Sangat baik; Skor 3: Baik; Skor 2: Cukup; Skor 1: kurang

Dalam konteks pembelajaran Bahasa Arab di perguruan tinggi keislaman, evaluasi terhadap keterampilan *mahārah kalām* memerlukan pendekatan yang lebih kontekstual dan

otentik, terutama melalui metode penilaian non-tes. Berdasarkan temuan di tiga institusi pendidikan tinggi Islam yang menjadi objek studi, terungkap bahwa implementasi penilaian non-tes dilakukan secara sistematis melalui dua tahapan utama, yaitu tahap perencanaan dan pelaksanaan. Tahap perencanaan mencakup penyusunan rubrik evaluasi dengan indikator yang mencerminkan kompetensi komunikatif mahasiswa secara utuh, sementara tahap pelaksanaan diwujudkan dalam bentuk kegiatan evaluatif seperti simulasi dialog, pidato tematik, presentasi berbasis proyek, serta interaksi lisan dalam konteks dunia nyata.

Matriks pada tabel 7 menyajikan hasil analisis implementasi tersebut yang dikaitkan dengan dimensi-dimensi teori asesmen autentik, seperti yang dikembangkan oleh *Wiggins*¹⁸ dan diintegrasikan dengan kerangka kompetensi komunikatif *Canale & Swain*.¹⁹ Penyajian ini bertujuan untuk memberikan gambaran konseptual dan aplikatif atas praktik evaluasi non-tes yang diterapkan, sekaligus mengidentifikasi kontribusi teoretis dan celah pengembangan lebih lanjut di masa depan.

Hasil penelitian ini menunjukkan bahwa implementasi penilaian non-tes *mahārah kalām* di tiga perguruan tinggi Islam telah bergerak menuju pendekatan autentik dan transformatif yang tidak hanya menilai aspek linguistik, tetapi juga mengakomodasi konteks sosial, performa komunikatif, dan keragaman karakter mahasiswa. Perencanaan yang matang melalui rubrik berbasis kompetensi komunikatif serta pelaksanaan evaluasi dalam bentuk simulasi, proyek tematik, dan interaksi nyata mencerminkan respons institusi terhadap

¹⁸ Grant Wiggins, "The Case for Authentic Assessment," *Practical Assessment, Research, and Evaluation* 2, no. 2 (1990): 1-3, <https://doi.org/10.7275/ffb1-mm19>.

¹⁹ Michael Canale and Merrill Swain, "Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing," *Applied Linguistics* I, no. 1 (1980): 1-47, <https://doi.org/10.1093/applin/I.1.1>.

tuntutan dunia nyata dan kebutuhan pembelajaran abad ke-21. Hal ini menegaskan bahwa keberhasilan evaluasi non-tes terletak pada kemampuan institusi dan dosen dalam memadukan dimensi pedagogis, psikologis, dan pragmatis secara adaptif. Dengan demikian, praktik penilaian ini tidak hanya menjadi alat ukur capaian belajar, melainkan juga sebagai wahana pembentukan kompetensi komunikatif mahasiswa secara holistik dan berkelanjutan.

Tabel 7
Matriks implementasi non-tes mahārah kalām

Aspek	UIN Suka	UMY	STAIMS	Analisis Teoretis
Perencanaan: Penyusunan Rubrik	Rubrik berbasis performa, mencakup aspek retorika, ketepatan makna, dan keberanian berbicara	Rubrik kolaboratif dengan masukan dari mahasiswa dan dosen, mengintegrasikan konteks dan gaya komunikasi	Rubrik sederhana namun fokus pada keberanian, kosakata, dan spontanitas dalam berbicara	Merujuk pada teori <i>Communicative Competence</i> (Canale & Swain), rubrik sudah mengakomodasi dimensi sosiolinguistik dan strategi
Pelaksanaan: Bentuk Evaluasi	Simulasi wawancara, ceramah keagamaan, presentasi kelompok	Debat terbuka, forum diskusi, proyek video tematik	Monolog naratif, permainan peran, <i>story telling</i>	Mengadopsi prinsip-prinsip <i>Authentic Assessment</i> (Wiggins, 1998) dengan keterlibatan konteks nyata
Pelaksanaan: Tahapan Evaluasi	1) <i>Briefing</i> ; 2) Penugasan individu/kelompok; 3) Umpan balik langsung	1) Penyusunan topik; 2) Diskusi; 3) Presentasi performatif; 4) Refleksi pasca tampil	1) Pengenalan model; 2) Latihan terbimbing; 3) Evaluasi performa dalam kelas	Tahapan mengikuti kerangka <i>Experiential Learning</i> (Kolb), memfasilitasi refleksi dan pembelajaran berkelanjutan

Keterangan: Skor 4: Sangat baik; Skor 3: Baik; Skor 2: Cukup; Skor 1: kurang

Diskusi

Hasil penelitian ini menunjukkan bahwa model penilaian non-tes *mahārah kalām* di tiga perguruan tinggi cenderung mengadopsi pendekatan berbasis performa dan proyek (*performance-based and project-based assessment*), yang sejalan dengan prinsip *authentic assessment*. Pendekatan ini tidak hanya menilai hasil akhir keterampilan berbicara, tetapi juga menekankan proses konstruksi makna dalam konteks komunikasi nyata. Melalui simulasi percakapan, pidato formal, dan produksi *vlog* berbahasa Arab, penilaian mencerminkan prinsip Wiggins bahwa asesmen harus merepresentasikan kompleksitas tugas dunia nyata, bukan sekadar penguasaan struktur bahasa secara terisolasi.²⁰

Temuan ini konsisten dengan hasil Menggo & Gunas dan Loomis dkk. yang menyatakan bahwa *performance-based assessment* lebih akurat dalam mengukur kompetensi komunikatif dibandingkan tes tertulis tradisional.²¹ Kajian lintas bahasa oleh Murphy dkk. menunjukkan bahwa lingkungan belajar berbasis interaksi otentik secara signifikan meningkatkan keterampilan lisan, termasuk bahasa Arab.²²

²⁰ Salwa Mohamed, "The Development of an Arabic Curriculum Framework Based on a Compilation of Salient Features from CEFR Level Descriptors," *The Language Learning Journal* 51, no. 1 (2023): 33–47, <https://doi.org/10.1080/09571736.2021.1923781>; John Dylan Burton, "How Scripted Is This Going to Be? Raters' Views of Authenticity in Speaking-Performance Tests," *Language Assessment Quarterly* 17, no. 3 (2020): 244–61, <https://doi.org/10.1080/15434303.2020.1754829>.

²¹ Sebastianus Menggo and Tobias Gunas, "College Students' Perceptions on Performance-Based Assessment Use in Boosting Speaking Ability," *International Journal of Language Education* 6, no. 4 (2022): 423–36, <https://doi.org/10.26858/ijole.v6i4.22910>; Summer Loomis, "Using Speaking Test Data to Define the Advanced Proficiency Level for L2 Arabic Speakers," *Foreign Language Annals* 48, no. 4 (December 2015): 604–17, <https://doi.org/10.1111/flan.12167>.

²² Dianna Murphy and Sonya K. Sedivy, "The Speaking Proficiency Outcomes of Face-to-face and Online Intensive Summer LCTL Programs," *Foreign Language Annals* 57, no. 4 (2024): 872–99, <https://doi.org/10.1111/flan.12758>.

Penelitian Abd Satar dan Yusoff lebih jauh menegaskan bahwa *classroom-based assessment* holistik mampu memfasilitasi perkembangan kefasihan, ketepatan, dan kepercayaan diri siswa dalam berbicara.²³ Hal ini selaras dengan studi Mahbubi yang menunjukkan bahwa meskipun metode tradisional efektif dalam membangun fondasi tata bahasa, pendekatan tersebut cenderung kurang menekankan keterampilan komunikatif praktis yang justru menjadi fokus utama penilaian non-tes dalam penelitian ini.²⁴ Dengan demikian, temuan penelitian ini tidak hanya konsisten dengan studi terdahulu, tetapi juga memberikan bukti empiris tentang pentingnya integrasi konteks sosial dan tujuan komunikatif dalam desain tugas untuk mengoptimalkan perkembangan *mahārah kalām*.

Temuan per-situs menunjukkan konsekuensi pedagogis yang berbeda. Model sumatif-produk di STAIMS memperkuat kreativitas, kohesi wacana, dan penguasaan media digital, tetapi membatasi kesempatan refleksi iteratif karena penilaian dilakukan pada akhir proyek. Model formatif-proses di UMY meningkatkan *responsiveness*, keberanian, dan perkembangan progresif mahasiswa melalui umpan balik berkelanjutan dari dosen dan teman sejawat. Model praktik formal di UIN Suka mengoptimalkan keterampilan retorika dan struktur ujaran formal, namun menghadapi kendala observasi mendalam pada kelas besar yang memengaruhi detail penilaian. Variasi ini menjelaskan capaian yang berbeda pada dimensi gramatikal, sosiolinguistik, strategis, dan wacana (Tabel 1–3).

²³ Hafilah Abd Satar and Nik Mohd Rahimi Nik Yusoff, "Improvement of Lower Secondary Arabic Language Teaching and Learning through the Implementation of Classroom-Based Assessments (PBD)," *Creative Education* 10, no. 12 (2019): 2555–63, <https://doi.org/10.4236/ce.2019.1012183>.

²⁴ Abdillah Mahbubi, "Conventional and Contemporary Arabic Language Teaching Methods: A Comparative Analysis," *Alsina : Journal of Arabic Studies* 6, no. 2 (2024): 205–28, <https://doi.org/10.21580/alsina.6.2.23307>.

Analisis rubrik memperlihatkan pemetaan yang tidak seragam terhadap dimensi kompetensi komunikatif Canale & Swain. Indikator *responsivitas dialog* dan *perkembangan progresif* di UMY menonjolkan dimensi strategis dan wacana. Indikator *retorika & gaya bicara* di UIN mengekskansi dimensi sosiolinguistik/retoris, sedangkan indikator *media & konten* di STAIMS menggabungkan performa lisan dengan kompetensi multimodal—sebuah perluasan di luar formulasi awal kompetensi komunikatif. Hal ini menunjukkan bahwa setiap model memiliki kekuatan spesifik sekaligus potensi untuk saling melengkapi demi mencapai profil kompetensi yang lebih utuh.

Namun demikian, berbagai studi juga menyoroti keterbatasan pendekatan ini. Isbell dkk. menggarisbawahi *format effects* pada asesmen berbasis wawancara yang dapat mengurangi validitas konstruk.²⁵ Dalam konteks pemanfaatan teknologi, Al-Ghezi dkk. dan Handley dkk. mengingatkan bahwa asesmen berbantuan AI, meskipun efisien, sering mengabaikan aspek strategis komunikasi seperti jeda retorik, intonasi persuasif, dan adaptasi terhadap audiens—unsur esensial dalam keterampilan berbicara.²⁶ Hal ini sejalan dengan Dua dan Daniel yang menekankan perlunya mekanisme kontrol untuk menjaga keautentikan interaksi manusia.²⁷

²⁵ Daniel R. Isbell, Paula Winke, and Susan M. Gass, "Using the ACTFL OPIc to Assess Proficiency and Monitor Progress in a Tertiary Foreign Languages Program," *Language Testing* 36, no. 3 (2019): 439–65, <https://doi.org/10.1177/0265532218798139>.

²⁶ Ragheb Al-Ghezi et al., "Automatic Speaking Assessment of Spontaneous L2 Finnish and Swedish," *Language Assessment Quarterly* 20, no. 4–5 (2023): 421–44, <https://doi.org/10.1080/15434303.2023.2292265>; Zoe L. Handley and Haiping Wang, "What Do the Measures of Utterance Fluency Employed in Automatic Speech Evaluation (ASE) Tell Us About Oral Proficiency?," *Language Assessment Quarterly* 21, no. 1 (2024): 3–32, <https://doi.org/10.1080/15434303.2023.2283839>.

²⁷ Jinming Du and Ben Kei Daniel, "Transforming Language Education: A Systematic Review of AI-Powered Chatbots for English as a Foreign

Hambatan yang teridentifikasi di lapangan, keterbatasan waktu latihan dan minimnya literatur latihan lisan (STAIMS), variasi kemampuan dan keberanian tampil (UMY), serta kelas besar dengan minim pelatihan non-verbal (UIN), menuntut penyesuaian strategi. Pada model proyek, *mini-tasks* formatif dapat ditambahkan untuk memberi ruang refleksi dan revisi. Pada model proses, kalibrasi *peer-assessment* dengan *anchor samples* dapat meminimalkan bias. Pada model praktik formal, *multi-rater sampling* atau penggunaan *checklist* mikro dapat meningkatkan reliabilitas di kelas besar.

Aspek teknis lain yang krusial adalah reliabilitas dan keadilan penilaian. Bijani dkk. serta Sureeyatanapas dkk. mengemukakan bahwa *rater training* dan kalibrasi antarpemilai merupakan prasyarat penting untuk meminimalkan bias subjektif.²⁸ Dalam pembelajaran bahasa Arab, Mohamed dkk. menegaskan bahwa keyakinan guru terhadap asesmen otentik dan ketersediaan infrastruktur digital yang memadai menjadi faktor penentu keberhasilan implementasi.²⁹

Dengan mempertimbangkan temuan positif dan tantangan tersebut, penelitian ini menegaskan bahwa keberhasilan model penilaian non-tes *mahārah kalām* sangat bergantung pada kesesuaian bentuk asesmen dengan kultur institusi dan tujuan

Language Speaking Practice," *Computers and Education: Artificial Intelligence* 6 (2024): 100230, <https://doi.org/10.1016/j.caeai.2024.100230>.

²⁸ Housman Bijani et al., "Investigating the Effect of Classroom-Based Feedback on Speaking Assessment: A Multifaceted Rasch Analysis," *Language Testing in Asia* 12, no. 1 (2022): 26, <https://doi.org/10.1186/s40468-022-00176-3>; Pornphan Sureeyatanapas et al., "The Analysis of Marking Reliability through the Approach of Gauge Repeatability and Reproducibility (GR&R) Study: A Case of English-Speaking Test," *Language Testing in Asia* 14, no. 1 (2024): 1-28, <https://doi.org/10.1186/s40468-023-00271-z>.

²⁹ Ahmed Mohamed et al., "Arabic Language Teachers' Beliefs and Practices of Using Evidence-Based Practices for Students with Learning Disabilities in the UAE: A Mixed-Methods Study," *Cogent Education* 11, no. 1 (2024), <https://doi.org/10.1080/2331186X.2024.2418807>.

pedagogis. Model digital-kreatif (STAIMS), komunikatif-kolaboratif (UMY), dan retorik-formal (UIN) masing-masing memberikan kontribusi unik terhadap pengembangan kompetensi berbicara bahasa Arab. Sebagai studi kasus multi-situs, penelitian ini berkontribusi menunjukkan *fit-praktik* antara konteks lokal dan bentuk *authentic assessment*, sekaligus mengidentifikasi elemen yang dapat ditransfer ke konteks lain dengan karakteristik serupa.

Secara teoretis, temuan ini memperluas konsep *authentic assessment* Wiggins dengan menegaskan pentingnya siklus formatif, multimodalitas, dan dimensi retorik dalam penilaian berbicara. Secara praktis, hasil penelitian merekomendasikan integrasi tugas otentik yang kontekstual, pemanfaatan teknologi dengan tetap menjaga keautentikan interaksi, serta penguatan kapasitas penilai melalui pelatihan dan kalibrasi yang terstruktur. Meskipun pendekatan ini telah terbukti efektif di berbagai konteks, penelitian lanjutan perlu mengeksplorasi pengaruh perbedaan latar belakang budaya, variasi dialek, dan motivasi belajar terhadap efektivitas asesmen otentik berbasis performa. Dengan demikian, kontribusi penelitian ini tidak hanya memperkuat dasar teoretis penilaian bahasa Arab, tetapi juga memberikan arah strategis bagi pengembangan praktik asesmen yang berkelanjutan dan inklusif di masa depan.

Simpulan

Berdasarkan hasil penelitian dapat disimpulkan bahwa ketiga perguruan tinggi tersebut mengimplementasikan penilaian non-tes *Mahārah Kalām* dengan pendekatan yang beragam namun saling melengkapi, mencerminkan pemahaman bersama akan pentingnya evaluasi berbasis kinerja dalam pembelajaran bahasa Arab. STAIMS menonjol dengan pendekatan proyek digital melalui pembuatan *vlog* berbahasa Arab, UMY menekankan simulasi dan permainan peran yang berorientasi pada proses, sedangkan UIN Sunan Kalijaga mengandalkan metode tradisional namun efektif seperti

muhādatsah dan *khiṭābah minbariyah*. Ketiga model ini berakar pada prinsip *authentic assessment* yang menilai kemampuan mahasiswa secara kontekstual, fungsional, dan holistik. Sintesis temuan ini menegaskan bahwa keberagaman strategi penilaian non-tes justru menjadi kekuatan dalam menjawab kebutuhan pembelajaran *Mahārah Kalām* yang bersifat dinamis, serta menuntut integrasi antara teori dan praktik berbahasa dalam situasi nyata.

Berdasarkan temuan tersebut, pemangku kebijakan institusi perlu memberikan pelatihan intensif dan berkelanjutan kepada dosen dalam perancangan serta penerapan penilaian non-tes berbasis autentik. Dukungan infrastruktur dan kebijakan akademik juga perlu diperkuat untuk memfasilitasi praktik pembelajaran berbasis kinerja, termasuk pengintegrasian teknologi digital sebagai media penilaian yang relevan dengan kebutuhan generasi pembelajar saat ini. Kolaborasi antar institusi dalam berbagi praktik baik perlu difasilitasi agar inovasi dalam penilaian *mahārah kalām* dapat berkembang secara lebih luas, sistematis, dan berkelanjutan.

Daftar Pustaka

- Abicandra, Muhammad Nidom Hamami. "Application of Authentic Assessment in Arabic Language Teaching in the Program of Studying the Arabic Language at the Idrisi Islamic Boarding School for Boys, Jember." *Journal of Arabic Language Teaching* 2, no. 1 (2022): 1–16. <https://doi.org/10.35719/arkhas.v2i1.1267>.
- Al-Ghezi, Ragheb et al. "Automatic Speaking Assessment of Spontaneous L2 Finnish and Swedish." *Language Assessment Quarterly* 20, no. 4–5 (2023): 421–44. <https://doi.org/10.1080/15434303.2023.2292265>.
- Angelina, Elen Nokalia. *The Effect of Project-Based Learning in Improving Students' Speaking Ability (Undergraduate Thesis)*. Banda Aceh: Ar-Raniry State Islamic University, 2020. <https://repository.ar-raniry.ac.id/id/eprint/13865/>.

- Baartman, Liesbeth K.J. et al. "Evaluating Assessment Quality in Competence-Based Education: A Qualitative Comparison of Two Frameworks." *Educational Research Review* 2, no. 2 (2007): 114–29. <https://doi.org/10.1016/j.edurev.2007.06.001>.
- Baroroh, R. Umi, and Nafisatun Nisa. "Non-Test Assessment Innovation Performance for Maharah Kalam Through Youtube in the Pandemic Era." *Ijaz Arabi Journal of Arabic Learning* 5, no. 1 (2022): 222–33. <https://doi.org/10.18860/ijazarabi.v5i1.14789>.
- Bijani, Houman et al. "Investigating the Effect of Classroom-Based Feedback on Speaking Assessment: A Multifaceted Rasch Analysis." *Language Testing in Asia* 12, no. 1 (2022): 26. <https://doi.org/10.1186/s40468-022-00176-3>.
- Burton, John Dylan. "'How Scripted Is This Going to Be?' Raters' Views of Authenticity in Speaking-Performance Tests." *Language Assessment Quarterly* 17, no. 3 (2020): 244–61. <https://doi.org/10.1080/15434303.2020.1754829>.
- Canale, Michael, and Merrill Swain. "Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing." *Applied Linguistics* 1, no. 1 (1980): 1–47. <https://doi.org/10.1093/applin/I.1.1>.
- Che Mat, Azman et al. "An Authentic Learning Environment Based on Video Project among Arabic Learners." *International Journal of Applied Linguistics and English Literature* 6, no. 4 (2017): 143–48. <https://doi.org/10.7575/aiac.ijalel.v.6n.4p.143>.
- Chong, Sin Wang, and Hayo Reinders. "Introduction: Learning-Oriented Language Assessment—Insights for Evidence-Based Practices." In *Innovation in Learning-Oriented Language Assessment*, edited by Sin Wang Chong and Hayo Reinders, 1–11. Cham: Palgrave Macmillan, 2023. https://doi.org/10.1007/978-3-031-18950-0_1.
- Du, Jinming, and Ben Kei Daniel. "Transforming Language Education: A Systematic Review of AI-Powered Chatbots for English as a Foreign Language Speaking Practice." *Computers and Education: Artificial Intelligence* 6 (2024): 100230. <https://doi.org/10.1016/j.caeai.2024.100230>.
- Fathi, Jalil et al. "Improving EFL Learners' Speaking Skills and

- Willingness to Communicate via Artificial Intelligence-Mediated Interactions." *System* 121 (2024): 103254. <https://doi.org/10.1016/j.system.2024.103254>.
- Handley, Zoe L., and Haiping Wang. "What Do the Measures of Utterance Fluency Employed in Automatic Speech Evaluation (ASE) Tell Us About Oral Proficiency?" *Language Assessment Quarterly* 21, no. 1 (2024): 3–32. <https://doi.org/10.1080/15434303.2023.2283839>.
- Ikhrom, Ikhrom. "The Impact of Learning Assessment Emergency on the Education Quality." *Lentera Pendidikan : Jurnal Ilmu Tarbiyah Dan Keguruan* 26, no. 1 (2023): 31–44. <https://doi.org/10.24252/lp.2023v26n1i4>.
- Isbell, Daniel R. et al. "Using the ACTFL OPIc to Assess Proficiency and Monitor Progress in a Tertiary Foreign Languages Program." *Language Testing* 36, no. 3 (2019): 439–65. <https://doi.org/10.1177/0265532218798139>.
- Loomis, Summer. "Using Speaking Test Data to Define the Advanced Proficiency Level for L2 Arabic Speakers." *Foreign Language Annals* 48, no. 4 (December 2015): 604–17. <https://doi.org/10.1111/flan.12167>.
- Mabruri, Mabruri et al. "Implementasi Penilaian Non-Tes Dalam Perkuliahan Di Prodi Pendidikan Bahasa Arab Se-Yogyakarta." *Jurnal Ihtimam* 7, no. 2 (2024): 182–98. <https://doi.org/10.36668/jih.v7i02.1185>.
- Mahbubi, Abdillah. "Conventional and Contemporary Arabic Language Teaching Methods: A Comparative Analysis." *Alsina : Journal of Arabic Studies* 6, no. 2 (2024): 205–28. <https://doi.org/10.21580/alsina.6.2.23307>.
- Maimun. "Evaluasi dalam pembelajaran bahasa Arab di STAIN Pamekasan." *NUANSA: Jurnal Penelitian Ilmu Sosial dan Keagamaan Islam* 8, no. 2 (2011): 285–96. <https://doi.org/10.19105/nuansa.v8i2.17>.
- Menggo, Sebastianus, and Tobias Gunas. "College Students' Perceptions on Performance-Based Assessment Use in Boosting Speaking Ability." *International Journal of Language Education* 6, no. 4 (2022): 423–36. <https://doi.org/10.26858/ijole.v6i4.22910>.
- Mohamed, Ahmed et al. "Arabic Language Teachers' Beliefs

- and Practices of Using Evidence-Based Practices for Students with Learning Disabilities in the UAE: A Mixed-Methods Study." *Cogent Education* 11, no. 1 (2024). <https://doi.org/10.1080/2331186X.2024.2418807>.
- Mohamed, Salwa. "The Development of an Arabic Curriculum Framework Based on a Compilation of Salient Features from CEFR Level Descriptors." *The Language Learning Journal* 51, no. 1 (2023): 33–47. <https://doi.org/10.1080/09571736.2021.1923781>.
- Montenegro-Rueda, Marta et al. "Assessment in Higher Education during the COVID-19 Pandemic: A Systematic Review." *Sustainability* 13, no. 19 (2021): 10509. <https://doi.org/10.3390/su131910509>.
- Munip, Abdul. "Tantangan dan Prospek Studi Bahasa Arab di Indonesia." *al Mahāra: Jurnal Pendidikan Bahasa Arab* 5, no. 2 (2020): 303–18. <https://doi.org/10.14421/almahara.2019.052.08>.
- Murphy, Dianna, and Sonya K. Sedivy. "The Speaking Proficiency Outcomes of Face-to-face and Online Intensive Summer LCTL Programs." *Foreign Language Annals* 57, no. 4 (2024): 872–99. <https://doi.org/10.1111/flan.12758>.
- O'Sullivan, Barry. "Assessing Speaking." In *The Concise Companion to Language Assessment*, edited by Antony John Kunnan, 143–53. New Jersey: Wiley-Blackwell, 2024.
- Ramdani, Junjun Muhamad, and R. Rahmat. "Promoting Speaking Spontaneity in Large Classes: An Action Research Study in an Indonesian EFL University Setting." *Indonesian Journal of Applied Linguistics* 8, no. 2 (2018): 388–401. <https://doi.org/10.17509/ijal.v8i2.13304>.
- Sari, Risna Rianti, and Alfiatus Syarofah. "Measuring Students' Creativity in Arabic Speaking Class Based on Project Based Learning Model." *Abjadia : International Journal of Education* 8, no. 1 (2023): 9–16. <https://doi.org/10.18860/abj.v8i1.21058>.
- Sari, Sinta Purnama et al. "Inovasi Tes Maharah Kalam Berbasis Penilaian Kurikulum Merdeka Pada Buku Durus Al-Lughah Al-Arabiyah Jilid 1." *Jurnal Pelita Manajemen Pendidikan* 1, no. 1 (2024): 70–79.

- <https://jurnalpelitanegribelantaraya.com/index.php/JPM/article/view/26>.
- Satar, Hafilah Abd, and Nik Mohd Rahimi Nik Yusoff. "Improvement of Lower Secondary Arabic Language Teaching and Learning through the Implementation of Classroom-Based Assessments (PBD)." *Creative Education* 10, no. 12 (2019): 2555–63. <https://doi.org/10.4236/ce.2019.1012183>.
- Sureeyatanapas, Pornphan et al. "The Analysis of Marking Reliability through the Approach of Gauge Repeatability and Reproducibility (GR&R) Study: A Case of English-Speaking Test." *Language Testing in Asia* 14, no. 1 (2024): 1–28. <https://doi.org/10.1186/s40468-023-00271-z>.
- Vanpee, Katrien, and Dan Soneson. "Arabic Proficiency Improvement Through a Culture of Assessment." In *Foreign Language Proficiency in Higher Education*, edited by Paula Winke and Susan M. Gass, 37:197–216. Educational Linguistics. Cham: Springer, 2019. https://doi.org/10.1007/978-3-030-01006-5_11.
- Widodo, Joko, and Muhammad Nanang Qosim. "Penilaian Kinerja pada Pembelajaran Maharah Al-Kalam Level 1 di Kursus Bahasa Arab Al-Arobiya Surakarta." *Uktub: Journal of Arabic Studies* 1, no. 2 (2021): 84–101. <https://doi.org/10.32678/uktub.v1i2.5814>.
- Wiggins, Grant. "The Case for Authentic Assessment." *Practical Assessment, Research, and Evaluation* 2, no. 2 (1990): 1–3. <https://doi.org/10.7275/ffb1-mm19>.

Setyawan, dkk.