

**التقييم والتقدير في تدريس اللغة العربية:
مراجعة منهجية للأدبيات**

**Assessment and Evaluation in Arabic Language Teaching:
A Systematic Literature Review**

Prayogi Pangestu Hadi

Universitas Islam Negeri Sultan Syarif Kasim Riau

E-mail: hadiprayogipangestu@gmail.com

ABSTRAK	
<i>This systematic literature review (SLR) synthesizes assessment and evaluation practices in Arabic Language Teaching (ALT) based on Scopus-indexed publications from 2020 to 2025. This review appears to be among the earliest dedicated SLRs focusing specifically on assessment and evaluation in ALT, prior reviews in this domain have primarily addressed teaching methods, instructional media, or curriculum development, with assessment examined only tangentially. Guided by PRISMA 2020 and the PICO research question structure, a Boolean search string was applied to the Scopus TITLE-ABS-KEY field via Publish or Perish, yielding 122 initial records. After dual-reviewer independent screening (Cohen's $\kappa = 0.81$) and full-text eligibility assessment, 28 studies were included. Quality appraisal using an adapted MMAT/CASP checklist yielded predominantly high-quality studies ($n=22$, 78.6%) with a subset rated moderate quality ($n=6$, 21.4%), reflecting the heterogeneous methodological landscape of ALT assessment research. Narrative synthesis informed by Bachman and Palmer's (1996) validity framework, Black and Wiliam's (1998) formative assessment theory, and Kane's (2013) argument validity approach identified five thematic clusters: summative and achievement testing, technology-enhanced assessment, program evaluation, formative and diagnostic assessment, and AI-based evaluation. AI-mediated assessment emerged as a dominant trend from 2023 onward. Reading comprehension (35.7%) and oral proficiency (21.4%) dominate the skill focus. Six critical research gaps are identified, including the near-absence of formative assessment research, insufficient psychometric validation of digital instruments, and unaddressed ethical dimensions of AI-based assessment.</i>	
KEYWORD	ARTICLE INFO
<i>Arabic Language Teaching, Assessment, Evaluation, Formative Assessment, AI In Education</i>	Published: 30 July 2026
	COPYRIGHT
	 © Author(s) 2026 This work is licensed under a Creative Commons Attribution 4.0 International License.

المقدمة

يعد التقييم والتقدير عنصرين أساسيين في تعليم اللغة، حيث يعملان في آن واحد كأدوات لقياس النتائج، وتوجيه عملية التدريس، وتشكيل المسارات التعليمية للمتعلمين. وفي مجال تدريس اللغة العربية (ALT)، تكتسب هذه الوظائف أهمية خاصة. فالطابع الثنائي للغة العربية، أي التعايش بين اللغة العربية الفصحى الحديثة (MSA) ومجموعة متنوعة من اللهجات المنطوقة، يخلق تحديات مستمرة تتعلق بصحة البناء: ما هي المعرفة اللغوية التي يتم تقييمها، ولأي أغراض تواصلية؟ هذه ليست مجرد مخاوف منهجية مجردة. فهي تحدد بشكل مباشر ما إذا كانت قرارات التقييم المتعلقة بمتعلمي اللغة العربية ذات مغزى وعادلة ومناسبة.¹

وعلى الرغم من هذه الأهمية، فقد تطورت الأدبيات المتعلقة بتدريس اللغة العربية إلى حد كبير دون وجود قاعدة أدلة منهجية حول ممارسات التقييم. وقد ركزت المراجعات الأكاديمية في مجال تدريس اللغة العربية كلغة أجنبية (ALT) في الغالب على منهجيات التدريس، وأدوات التعلم الرقمية، وتصميم المناهج الدراسية. وعندما يظهر التقييم في هذه المراجعات، فإنه يظهر عادةً كمقياس للنتائج، أي درجة اختبار قبل وبعد التدخل تشير إلى ما إذا كان التدخل قد نجح أم لا، وليس باعتباره الموضوع الرئيسي للبحث. وهذا التباين له عواقب: فلا يمكن لأي مجال أن يحسن ممارسات التقييم الخاصة به دون أن يوثق أولاً وبشكل منهجي ماهية تلك الممارسات، وكيفية توزيعها عبر السياقات ومجالات المهارات، وأين تكون الأدلة ضعيفة.

وتزداد الحاجة الملحة إلى ذلك خلال الفترة 2020–2025. فقد أجبرت جائحة كوفيد-19 برامج اللغة العربية في جميع أنحاء العالم على الانتقال السريع إلى صيغ التقييم عبر الإنترنت والصيغ المختلطة، مما أدى إلى ظهور تحديات جديدة تتعلق بالنزاهة الأكاديمية، وإمكانية الوصول التقني، وتكافؤ القياس.² وفي الوقت نفسه، بدأت أدوات

¹ Lyle F. Bachman dan Adrian S. Palmer, *Language Testing in Practice* (Oxford University Press, 1996); Cristina Solimando, "Language Tests in the Arabic-speaking World: Between Ideology and Language Policy," *Lingue Culture Mediazioni - Languages Cultures Mediation (LCM Journal)* 8, no. 2 (2022), <https://doi.org/10.7358/lcm-2021-002-soli>.

² Ahmad Muradi dkk., "Evaluation of Arabic Learning Program at State Islamic Religious College During COVID-19," *Eurasian Journal of Educational Research*, no. 99 (2022), <https://doi.org/10.14689/ejer.2022.99.011>; Yusring Sanusi Baso dkk., "Reducing Cheating in Online Exams

الذكاء الاصطناعي التوليدية مثل النماذج اللغوية الضخمة، والتعرف الآلي على الكلام، والمحتوى التعليمي الذي يولده الذكاء الاصطناعي في إعادة تشكيل كيفية قياس إتقان اللغة العربية وتدريسها.³ هذه تحولات سريعة وذات آثار كبيرة لم يتم بعد تجميع آثارها على التقييم.

أُجريت هذه المراجعة المنهجية للأدبيات (SLR) وفقًا لبروتوكول مسجل مسبقًا (الإصدار 1.0، مايو 2026) بهدف تحديد وتوليف الأدبيات التي خضعت لمراجعة الأقران حول التقييم والتقدير في مجال تدريس اللغة العربية (ALT) المنشورة بين عامي 2020 و2025، والمفهرسة في قاعدة بيانات Scopus. وتوجه المراجعة أربعة أسئلة بحثية (RQ)، صيغت باستخدام إطار عمل PICo (Stern et al., 2014):⁴

RQ1: ما هي أساليب التقييم والتقدير الموثقة في أبحاث تعليم اللغات البديلة (ALT) المنشورة بين عامي 2020 و2025؟

RQ2: كيف تطورت اتجاهات البحث الزمنية والجغرافية في تقييم تدريس اللغة العربية كلغة ثانية (ALT) خلال هذه الفترة؟

RQ3: ما هي مهارات اللغة العربية التي غالبًا ما تكون محور أبحاث التقييم والتقدير؟

RQ4: ما هي الثغرات البحثية الموجودة، وما هي التوجهات الموصى بها للبحوث المستقبلية؟

تقدم هذه المراجعة توليفًا قائمًا على أسس نظرية ومصممًا ليكون قابلاً للتطبيق من قبل الباحثين والممارسين وصانعي السياسات في مجال تعليم اللغة العربية.

Through the Proctor Test Model,” *International Journal of Advanced Computer Science and Applications* 14, no. 1 (2023), <https://doi.org/10.14569/IJACSA.2023.0140139>.

³ Norah Alrashoudi dkk., “Improving Mispronunciation Detection and Diagnosis for Non- Native Learners of the Arabic Language,” *Discover Computing* 28, no. 1 (2025), <https://doi.org/10.1007/s10791-024-09489-8>; Ahmed Allaithy dan Mai Zaki, “Evaluation of AI-Generated Reading Comprehension Materials for Arabic Language Teaching,” *Computer Assisted Language Learning*, 6 Maret 2025, <https://doi.org/10.1080/09588221.2025.2474037>.

⁴ Cindy Stern dkk., “Developing the Review Question and Inclusion Criteria,” *AJN, American Journal of Nursing* 114, no. 4 (2014), <https://doi.org/10.1097/01.NAJ.0000445689.67800.86>.

منهج البحث

تصميم الدراسة وبروتوكولها

تستخدم هذه الدراسة منهجية مراجعة الأدبيات (SLR) وفقًا لإرشادات PRISMA 2020، مع تنظيم أسئلة البحث باستخدام إطار عمل PICO: السكان (متعلمو ومدرسو اللغة العربية كلغة أولى أو ثانية أو لغة أجنبية)، والموضوع (ممارسات التقييم والتقدير)، والسياق (بيئات تدريس اللغة العربية كلغة أجنبية الرسمية وغير الرسمية، 2020-2025).⁵ تم تسجيل البروتوكول مسبقًا قبل جمع البيانات (الإصدار 1.0، مايو 2026، متوافق مع PROSPERO).

استراتيجية البحث

تم البحث عن المراجع في قاعدة بيانات Scopus عبر برنامج Publish or Perish (PoP)، الذي تم تطبيقه على حقل TITLE-ABS-KEY. تم اختيار Scopus لتغطيتها الشاملة لمجالي التعليم واللغويات التطبيقية، واتساق البيانات الوصفية، وجودة الفهرسة.⁶ يُعتبر نهج قاعدة البيانات الواحدة أحد أوجه القصور (القسم 6)؛ وقد أُعطيت الأولوية لـ Scopus على البحث في قواعد البيانات المتعددة نظرًا لاتساق بياناتها الوصفية الفائقة بالنسبة للمجال والفترة الزمنية المستهدفة. تم تطوير سلسلة البحث بشكل تكراري واختبارها تجريبيًا على ثلاث مقالات معروفة ذات صلة:

TITLE-ABS-KEY(("تدريس اللغة العربية" أو "تدريس اللغة" أو "تعليم اللغة العربية" أو "تعلم اللغة العربية" أو "اللغة العربية كلغة أجنبية" أو "اللغة العربية كلغة ثانية") و ("تقييم" أو "تقدير" أو "اختبار" أو "تقييم اللغة" أو "التقييم التكويني" أو "التقييم التلخيصي" أو "التقييم التشخيصي" أو "التقييم الأصيل" أو "تقييم الملف" أو "التقييم الذاتي" أو "تقييم الأقران" أو "التقييم الإلكتروني" أو "التقييم الرقمي" أو rubric* أو الكفاءة) و 2020 > PUBYEAR و PUBYEAR > 2025 (LANGUAGE(English) و DOCTYPE(ar))

⁵ Matthew J. Page dkk., "The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews," Research Methods & Reporting, *BMJ* 372 (Maret 2021), <https://doi.org/10.1136/bmj.n71>.

ماتيو إي. فالاغاس وآخرون، «مقارنة بين PubMed و Scopus و Web of Science و Google Scholar: نقاط القوة والضعف»، *FASEB* 22، العدد 2 (2008)، <https://doi.org/10.1096/fj.07-9492LSF>.

معايير الأهلية

تم تحديد معايير الأهلية مسبقاً وتطبيقها بشكل متسق في مرحلتي الفرز⁷ يعرض الجدول 1 المعايير الكاملة.

الجدول 1. معايير الأهلية المطبقة في جميع مراحل الفرز

الرمز	البعد	شروط القبول	الاستبعاد
I1/E1	فترة النشر	2020–2025 (جمع البيانات: مايو 2026)	خارج الفترة
I2/E2	اللغة	الإنجليزية	غير الإنجليزية
I3/E3	نوع الوثيقة	مقالة في مجلة محكمة ومفهرسة في سكوبوس	ورقة مؤتمر، فصل كتاب، افتتاحية، رسالة
I4/E4	الموضوع الأساسي	التقييم باعتباره محوراً أساسياً أو جوهرياً في ALT	التقييم مذكور بشكل عابر فقط (≥1 فقرة)
I5/E5	اللغة المستهدفة	اللغة العربية كلغة مستهدفة (L1، L2، أو لغة أجنبية)	لغة أخرى كلغة مستهدفة أساسية
I6/E6	المنهجية	منهجية بحثية واضحة وقابلة للتتبع	مقال رأي، تعليق، لا توجد بيانات
I7/E7	إمكانية الوصول	النص الكامل متاح	النص الكامل غير متاح

اختيار الدراسة

تم الاختيار وفقاً لعملية PRISMA 2020 المكونة من أربع مراحل. أسفر البحث في قاعدة بيانات Scopus عن 122 سجلاً. بعد حذف 3 سجلات مكررة، خضع 119 سجلاً لفحص مستقل للعناوين والملخصات من قبل مراجعين اثنين. كان التوافق بين المراجعين

⁷ Julian P. T. Higgins dan Sally Green, *Cochrane Handbook for Systematic Reviews of Interventions*, Version 5.1.0, ed. oleh Julian P. T. Higgins dan Sally Green (The Cochrane Collaboration, 2011), <https://handbook-5-1.cochrane.org/>.

كبيرًا (معامل كوهين $K = 0.81$)⁸. تم استبعاد 62 سجلًا (E2: n=8؛ E3: n=14؛ E4: n=38)؛ E1: n=2). تم استرجاع وتقييم 57 نصًا كاملاً، واستبعد منها 29 (E4: n=11؛ E3: n=12)؛ E6: n=6)، مما أدى إلى اختيار 28 دراسة للتوليف. يوضح الشكل 1 مخطط تدفق PRISMA.

الشكل 1. مخطط تدفق PRISMA 2020

المرحلة	الوصف	n
التحديد	السجلات من Scopus عبر Publish or Perish	122
التعريف	تم حذف التكرارات	3
الفرز	السجلات التي خضعت للفحص (العنوان/الملخص)؛ مراجعة ثنائية، $K=0.81$	119
الفرز	تم استبعادها (E1: n=2؛ E2: n=8؛ E3: n=14؛ E4: n=38)	62
الأهلية	النصوص الكاملة التي تم تقييمها	57
الأهلية	المستبعدة (E6: n=6؛ E4: n=11؛ E3: n=12)	29
تم تضمينها	التوليف النهائي	28

استخراج البيانات

وقد تم استخدام نموذج استخراج موحد، تم اختباره تجريبياً على ثلاث مقالات تم اختيارها عشوائياً، لتسجيل: المؤلف (المؤلفون)، والسنة، والمجلة، والبلد، والمستوى التعليمي، وتصميم البحث، والمشاركين، والمهارة اللغوية التي ركز عليها البحث، ونوع التقييم، والأدوات المستخدمة، واستخدام التكنولوجيا، والنتائج الرئيسية، والقيود المذكورة. وقد أجرى المراجع الأول عملية الاستخراج الأولية، وقام المراجع الثاني بالتحقق من 30% من الدراسات المشمولة. وترد بيانات الاستخراج الكاملة في الملحق أ.

⁸ J. Richard Landis dan Gary G. Koch, "The Measurement of Observer Agreement for Categorical Data," *Biometrics* 33, no. 1 (1977): 159, <https://doi.org/10.2307/2529310>.

تقييم الجودة

تم تقييم الجودة باستخدام قائمة مراجعة مقبسة من MMAT (Hong et al, 2018) و (2022) CASP عبر خمسة أبعاد: وضوح الهدف (K1)، وتوافق التصميم مع السؤال (K2)، وأدلة صحة الأدوات (K3)، وجودة التحليل (K4)، وتوافق الاستنتاجات مع النتائج (K5). تم تقييم كل بُعد بدرجة 0 (غير مستوفى)، أو 0.5 (مستوفى جزئياً)، أو 1 (مستوفى بالكامل)؛ والحد الأقصى للدرجة هو 5.0. تُصنف الدراسات التي حصلت على درجة ≤ 3.5 على أنها ذات جودة عالية؛ و $2.5-3.0$ على أنها متوسطة؛ و ≥ 2.0 على أنها منخفضة (مبرر للاستبعاد). لم يتم استبعاد أي دراسة لأسباب تتعلق بالجودة. وترد النتائج في الجدول 2.

الجدول 2. نتائج تقييم الجودة لجميع الدراسات الـ 28 المشمولة

المؤلفون والسنة	التصميم	K1	K2	K3	K4	K5	النتيجة	الفئة
القطونة وآخرون (2020)	شبه تجريبي	1	1	1	1	1	5.0	مرتفع
ميدوا والحواري (2020)	شبه تجريبي	1	1	1	1	1	5.0	مرتفع
نصيري وآخرون (2022)	الحوسبة/معالجة اللغة الطبيعية	1	1	1	1	1	5.0	مرتفع

⁹ Q. N. Hong dkk., *Mixed Methods Appraisal Tool (MMAT), Version 2018* (Canadian Intellectual Property Office, 2018), <http://mixedmethodsappraisaltoolpublic.pbworks.com/>; Critical Appraisal Skills Programme, *CASP Checklists* (Critical Appraisal Skills Programme, 2022), <https://casp-uk.net/casp-tools-checklists/>.

مرتفع	5.0	1	1	1	1	1	كهي مقارن	لانغستن وآخرون (2022)
مرتفع	5.0	1	1	1	1	1	شبه تجريبي	روسلي وآخرون (2024)
مرتفع	5.0	1	1	1	1	1	الحوسبة/معالجة اللغة الطبيعية	Alrashoudi وآخرون (2025)
مرتفع	5.0	1	1	1	1	1	البحث والتطوير (ADDIE)	أحسان الدين وآخرون (2025)
مرتفع	5.0	1	1	1	1	1	شبه تجريبي	هيجريا وآخرون (2025)
مرتفع	4.5	1	1	0.5	1	1	نوعي	زورقوني وآخرون (2020)
مرتفع	4.5	1	1	0.5	1	1	دراسة استقصائية وصفية	الأولي واللاهاني (2020)
مرتفع	4.5	1	1	0.5	1	1	مراجعة نقدية	سوليمانندو (2021)

مرتفع	4.5	1	1	0.5	1	1	تقييمي نوعي	ريتونغا وآخرون (2021)
مرتفع	4.5	1	1	0.5	1	1	تقييم CIPP	باندو وآخرون (2021)
مرتفع	4.5	1	1	0.5	1	1	استبيان	معمون وآخرون (2021)
مرتفع	4.5	1	1	0.5	1	1	تقييم البرامج	مرادی وآخرون (2022)
مرتفع	4.5	1	1	0.5	1	1	وصفي-تحليلي	مؤذن وجوما (2022)
مرتفع	4.5	1	1	0.5	1	1	الأساليب المختلطة	باسو وآخرون (2023)
مرتفع	4.5	1	1	0.5	1	1	PAR/ABCD	مكرم وباروكا (2023)
مرتفع	4.5	1	1	0.5	1	1	الأساليب المختلطة	ريتونغا وآخرون (2024)

مرتفع	4.5	1	1	0.5	1	1	الأساليب المختلطة	& Allaithy Zaki (2025)
مرتفع	4.5	1	1	0.5	1	1	نوعي	الكعبي والمامي (2025)
مرتفع	4.5	1	1	0.5	1	1	المنهجيات المختلطة	ماوجود وشاهارودين (2025)
معتدل	3.0	0.5	0.5	0.5	0.5	1	الأساليب المختلطة	مينغازوفا وآخرون (2020)
معتدل	3.0	0.5	0.5	0.5	0.5	1	تحليل الوثائق	إلهادي (2021)
مرتفع	4.0	1	0.5	0.5	1	1	شبه تجريبي	Gu وآخرون (2025)
مرتفع	4.5	1	1	0.5	1	1	الارتباط الكمي	بهارودين ورحمان (2025)
مرتفع	3.5	1	0.5	0.5	0.5	1	وصفي-تقيني	سويب وباهارودين (2025)
عالية	3.5	1	0.5	0.5	1	0.5	طرق مختلطة	زولهانان وآخرون (2025)

تشير نتائج تقييم الجودة إلى أن 22 دراسة (78.6٪) حصلت على تصنيف جودة عالية (درجات ≤ 3.5)، وست دراسات حصلت على تصنيف جودة متوسطة (درجات 2.5–3.5)، مما يعكس التنوع المنهجي في أبحاث تقييم ALT. تشترك الدراستان ذواتا الجودة المتوسطة (Mingazova et al., 2020; Elhadky, 2021) في قيود مشتركة: توثيق محدود لصلاحية الأدوات وشفافية جزئية في تحليل «¹⁰». وقد أدرجت نتائج هاتين الدراستين في التوليف، لكن تم تفسيرها بحذر مناسب.

توليف البيانات

اتبع التوليف السردى إرشادات (Popay et al. 2006)، وتم تنظيمه حول الأسئلة البحثية الأربعة. وطبق التحليل الموضوعي إطار العمل المكون من ست مراحل الذي وضعه Braun و Clarke (2006)¹¹ وتُعرض النتائج والمناقشة التفسيرية في أقسام منفصلة للحفاظ على الفصل بين عرض الأدلة (القسم 4) والمناقشة النظرية والتحليلية (القسم 5). وأجري تحليل الفجوات من خلال القراءة المتقاطعة المنهجية للقيود المذكورة وتوصيات البحوث المستقبلية عبر جميع المقالات الـ 28.

نتائج البحث

نظرة عامة على الدراسات المشمولة

تمتد الدراسات الـ 28 المشمولة في الدراسة من عام 2020 إلى عام 2025، مع زيادة كبيرة في وتيرة النشر اعتبارًا من عام 2023. يلخص الجدول 3 خصائص الدراسات.

¹⁰ Islam Elhadky, "Arap Olmayanlara Arapça Öğretimine Yönelik Araştırmacı ve Uzman Yetiştirilmede Dil Yeterliliği Dersine Yönelik Bir Müfredat Önerisi," *darülfünun ilahiyat* 32, no. 2 (2021), <https://doi.org/10.26650/di.2021.32.2.995627>; N. G. Mingazova dkk., "Usage of Digital Educational Sources in Arabic Teaching," *Utopia y Praxis Latinoamericana* 25, no. Extra 12 (2020), <https://doi.org/10.5281/zenodo.4280165>.

¹¹ Jennie Popay dkk., *Guidance on the Conduct of Narrative Synthesis in Systematic Reviews* (ESRC Methods Programme, 2006), <https://www.lancaster.ac.uk/media/lancaster-university/content-assets/documents/fhm/dhr/chir/NSsynthesisguidanceVersion1-April2006.pdf>; Virginia Braun dan Victoria Clarke, "Using Thematic Analysis in Psychology," *Qualitative Research in Psychology* 3, no. 2 (2006), <https://doi.org/10.1191/1478088706qp063oa>.

الجدول 3. خصائص الدراسات المشمولة (N=28)

السمة	الفئة	n	%
السنة	2020	5	17.9%
	2021	5	17.9%
	2022	4	14.3%
	2023	2	7.1%
	2024	2	7.1%
	2025 (جمع البيانات: مايو 2026)	10	35.7%
الدولة	إندونيسيا	11	39.3%
	ماليزيا	5	17.9%
	الدول ذات الأغلبية العربية (المملكة العربية السعودية، الأردن، مصر، الإمارات العربية المتحدة، عمان)	6	21.4%
	متعددة البلدان	3	10.7%
	أخرى (الصين، روسيا، نيوزيلندا)	3	10.7%
	التعليم العالي	18	64.3%
المستوى التعليمي	التعليم الثانوي / المدارس الدينية	6	21.4%
	المرحلة الابتدائية	2	7.1%
	مختلط / سياسي	2	7.1%
	كمية / شبه تجريبية	10	35.7%
المنهجية	نوعي	6	21.4%
	الأساليب المختلطة	7	25.0%
	البحث والتطوير / الحسابي	4	14.3%
	تقييم البرامج / تحليل الوثائق	1	3.6%
	اختبار تجميعي / اختبار الإنجاز	8	28.6%
نوع التقييم			

28.6	8	مدعوم بالتكنولوجيا / رقمي
21.4	6	تقييم البرامج / المناهج الدراسية
14.3	4	التقييم التكويني / التشخيصي
7.1	2	التقييم القائم على الذكاء الاصطناعي

تساهم إندونيسيا بأكبر حصة من الدراسات (39.3%)، مما يعكس نظام المدارس الدينية الواسع النطاق في البلاد حيث تُعد اللغة العربية مادة إلزامية. وتُعد مجموعة عام 2025 أكبر مجموعة في عام واحد (35.7%)؛ وقد اكتمل جمع البيانات في مايو 2026. ويُعد التعليم العالي هو البيئة السائدة (64.3%).

طرق وممارسات التقييم (السؤال البحثي 1)

انبثقت خمس مجموعات مواضيعية من التحليل. يعرض الجدول 4 مصفوفة الترميز الخاصة بالربط بين المقالات والمواضيع.

الجدول 4. مصفوفة الترميز المواضيعي: ربط المقالات بالمواضيع

الموضوع	المقالات (n)	النتائج الأساسية
الاختبارات التلخيصية واختبارات التحصيل (n=8)	زرقوني وآخرون (2020)؛ القطونة وآخرون (2020)؛ ميدوة والحواري (2020)؛ لأنغستن وآخرون (2022)؛ موزين وجوما (2022)؛ الحادي (2021)؛ روسل وآخرون (2024)؛ بهارودين ورحسن (2025)	تسود الاختبارات قبل وبعد التدريب؛ وتغيب إلى حد كبير الوثائق المتعلقة بصحة البناء؛ ونادراً ما تُستخدم أدوات قياس موحدة
التقييم المعزز بالتكنولوجيا (n=8)	مينغازوفا وآخرون (2020)؛ باسو وآخرون	المنصات الإلكترونية تعزز المشاركة؛ التحقق من

الصحة النفسية محدود؛ مخاوف بشأن النزاهة في مرحلة ما بعد الجائحة	(2023)؛ معروف وباروكه (2023)؛ ريتونغا وآخرون (2024)؛ أحسان الدين وآخرون (2025)؛ الهجرية وآخرون (2025)؛ غو وآخرون (2025)؛ الكعبي والمامي (2025)	
تم تحديد ثغرات في الرصد المنهجي؛ نموذج CIPP هو الأكثر استخدامًا؛ أطر ضمان الجودة غير مطورة بشكل كافٍ	ريتونغا وآخرون (2021)؛ باندو وآخرون (2021)؛ مراد وآخرون (2022)؛ الأوالي ولاهياني (2020)؛ زولحنان وآخرون (2025)؛ موجود وشاهارودين (2025)	تقييم البرامج والمناهج الدراسية (n=6)
المجموعة الأكثر ابتكارًا من الناحية المنهجية؛ أدوات تشخيصية آلية (قابلية القراءة، MDD) جديدة بالملاحظة؛ قدرة المعلمين التكوينية محدودة	معمون وآخرون (2021)؛ ناصر وآخرون (2022)؛ سويب وباهارودين (2025)؛ الرشودي وآخرون (2025)	التقييم التكويني والتشخيصي (n=4)
نمو سريع منذ عام 2023؛ مخاوف بشأن الصلاحية الثقافية؛ أبعاد أيدولوجية ذات صلة بأنظمة الذكاء الاصطناعي؛	Allaithy & Zaki (2025) Alrashoudi et al (2025)؛ Alkaabi & (2025)؛ Almaamari Solimando (2021)	التقييم القائم على الذكاء الاصطناعي ناشئ (n=4)

استعداد المعلمين يمثل عائقاً		
---------------------------------	--	--

الاختبارات التلخيصية واختبارات التحصيل

يسود التقييم التلخيصي في المجموعة، ويتميز باختبارات الإنجاز قبل وبعد التدخلات التعليمية. قام زرقوني وآخرون (2020) بتقييم تعلم اللغة العربية في تسع مدارس ثانوية إندونيسية (Madrasah Aliyah) من خلال أدوات نوعية مجموعات النقاش المركزة، والاستبيانات، والمقابلات وكشفوا عن نقص مستمر في المفردات وقيود بيئية حددتها الاختبارات ولكنها لم تتمكن من معالجتها.¹² أظهر القطاونة وآخرون (2020) مكاسب ذات دلالة إحصائية في التعلم المدمج في جامعة عجمان باستخدام اختبارات قياسية قبل وبعد التدخل، حيث أبدى الطلاب أيضاً مواقف إيجابية تجاه هذه الطريقة.¹³ قدم ميدوا والحواري (2020) الدراسة القياسية الأكثر وضوحاً في المجموعة، حيث بحثا في دور علامات التشكيل في القراءة لدى 54 متعلماً للغة الإنجليزية كلغة ثانية من الناطقين باللغة العربية على ثلاثة مستويات من الكفاءة ووجدوا تأثيرات ميسرة في المستويين المبتدئ والمتوسط، ولكن ليس في المستوى المتقدم.¹⁴ ويتميز تصميم دراستهما بتفعيل المتغيرات ذات الصلة بالمفهوم بشكل صريح.

استخدم لانغستن وآخرون (2022) بيانات التقييم الوطني لمهارات القراءة والكتابة لمقارنة المهارات اللغوية العربية بين المدارس المجتمعية والحكومية في المناطق الريفية بصعيد مصر، موثقين تباينات في المساواة نادراً ما يتم بحثها في أدبيات تقييم تعليم اللغة العربية كلغة ثانية.¹⁵ وطبق موزين وجوما (2022) تحليلاً مرجعياً للمعايير على

¹² Zurqoni Zurqoni dkk., "Has Arabic Language Learning Been Successfully Implemented?," *International Journal of Instruction* 13, no. 4 (2020), <https://doi.org/10.29333/iji.2020.13444a>.

¹³ Sami Al-Qatawneh dkk., "The Effect of Blended Learning on the Achievement of HDE Students in the Methods of Teaching Arabic Language Course and Their Attitudes towards Its Use at Ajman University: A Case Study," *Education and Information Technologies* 25, no. 3 (2020), <https://doi.org/10.1007/s10639-019-10046-w>.

¹⁴ Ali Al Midhwah dan Mohammad T. Alhawary, "Arabic Diacritics and Their Role in Facilitating Reading Speed, Accuracy, and Comprehension by English L2 Learners of Arabic," *The Modern Language Journal* 104, no. 2 (2020): 418–38, <https://doi.org/10.1111/modl.12642>.

¹⁵ Ray Langsten dkk., "Arabic Language Skills: A Comparative Study of Community and Government Schools in Rural Upper-Egypt," *Compare: A Journal of Comparative and International Education* 52, no. 6 (2022): 879–94, <https://doi.org/10.1080/03057925.2020.1843003>.

الكفاءة المعجمية بين طلاب الجامعات الأتراك المتعلمين للغة العربية، ووجدوا أداءً متدنياً منهجياً مقارنةً بمعايير الدورة الدراسية.¹⁶ قام (Elhadky (2021)، الذي صُنفت دراسته على أنها ذات جودة متوسطة، بفحص أطر الكفاءة لإعداد معلمي اللغة العربية من خلال تحليل الوثائق مما وفر خريطة مفاهيمية مفيدة على الرغم من محدودية التحقق التجريبي.¹⁷ أثبت Baharudin و (Rahman (2025) وجود علاقات ترابطية بين الوعي المعرفي الفوقي وكفاءة الاستماع.¹⁸ قارن Rusli وآخرون (2024) نتائج فهم النصوص السردية مقابل النصوص المعلوماتية، ووجدوا أداءً أعلى بشكل ملحوظ مع النصوص السردية.¹⁹

التقييم المعزز بالتكنولوجيا والتقييم الرقمي

يمتد التقييم المعزز بالتكنولوجيا من التقييم المبكر للمصادر الرقمية (Mingazova et al., 2020) وصولاً إلى التطوير المتطور القائم على الألعاب.²⁰ استخدم Makruf و (Barokah (2023) منهجية PAR و ABCD لتحسين التقييم القائم على تكنولوجيا المعلومات والاتصالات في المدارس الدينية الإندونيسية، حيث أظهر Wordwall تحسينات من خلال دورات التعلم المهني التكرارية وهو نهج موجه نحو المجتمع غير شائع في مجال تدريس اللغة الأجنبية (ALT).²¹ وثَّق باسو وآخرون حلولاً عملية لمكافحة الغش من خلال نموذج اختبار تحت إشراف المراقبين تم تطبيقه على امتحانات اللغة العربية عبر الإنترنت.²² ووجد ريتونغا وآخرون (2024) أن منصة الحرم

¹⁶ Ahmet Müezzini dan Mai Goma, "Ana Dili Arapça Olmayan Öğrenciler İçin Sözcüksel Yeterlilikler ve Öğrencilerin Beklenen Performansına Uygunluğu 'Uygulamalı Bir Çalışma,'" *darülfunun ilahiyat* 33, no. 1 (2022), <https://doi.org/10.26650/di.2022.33.1.1087872>.

¹⁷ Elhadky, "Arap Olmayanlara Arapça Öğretimine Yönelik Araştırmacı ve Uzman Yetiştirmede Dil Yeterliliği Dersine Yönelik Bir Müfredat Önerisi."

¹⁸ Harun Baharudin dan Nur Farahhuda Abdul Rahman, "The Impact of Metacognitive Awareness on Arabic Listening Proficiency among Religious Secondary School Students," *The Open Psychology Journal* 18, no. 1 (2025): e18743501422272, <https://doi.org/10.2174/0118743501422272251006211402>.

¹⁹ Radif Khotamir Rusli dkk., "The Impact of Informative and Narrative Texts on Arabic Language Comprehension in Islamic Education," *International Journal of Language Education* 8, no. 4 (2025), <https://doi.org/10.26858/ijole.v8i4.70009>.

²⁰ Mingazova dkk., "Usage of Digital Educational Sources in Arabic Teaching."

²¹ Imam Makruf dan Anisatul Barokah, "Improving the Quality of ICT-Based Arabic Learning Assessment With Online Applications," *Journal of Higher Education Theory and Practice* 23, no. 11 (2023), <https://doi.org/10.33423/jhetp.v23i11.6216>.

²² Baso dkk., "Reducing Cheating in Online Exams Through the Proctor Test Model."

الجامعي الإلكتروني عززت شفافية التقييم والمراقبة في التعليم العالي الماليزي.²³ وقام أحسان الدين وآخرون بتطوير وتقييم صحة تقييم قائم على الألعاب باستخدام Construct 2 ونهج ADDIE، وأبلغوا عن صحة المحتوى من خلال مراجعة الخبراء وقبول الطلاب العالي، على الرغم من أن الصحة المرتبطة بالمعايير لا تزال جزئية.²⁴ قامت Hijriyah وآخرون (2025) بتقييم SUNO.AI لتعزيز مهارات الاستماع للغة العربية، وأفادوا بتحسين كبير في الأداء.²⁵ وجد Gu وآخرون (2025) أن دمج محو الأمية الرقمية أدى إلى تحسين الأداء الشفوي في اللغة العربية في برامج إعداد معلمي اللغة الصينية.²⁶ قدمت Mingazova وآخرون (2020)، التي صُنفت بجودة متوسطة بسبب محدودية الوثائق التي تثبت صحة النتائج، أدلة مبكرة عبر الثقافات على اعتماد الأدوات الرقمية في تدريس اللغة العربية.²⁷

تقييم البرامج والمناهج الدراسية

توفر دراسات تقييم البرامج أدلة ذات صلة بالسياسات. وقد وثق ريتونغا وآخرون (2021) الفجوات في التنفيذ بين تصميم المناهج الدراسية والواقع الفعلي داخل الفصول الدراسية في التعليم العالي الإندونيسي.²⁸ طبق باندو وآخرون (2021) نموذج CIPP الذي وضعه ستافليم (1983) في المعهد الإسلامي العالي في بالو (IAIN Palu)، مما وفر إطاراً تقييمياً قابلاً للتطبيق في سياقات أخرى.^{29,30} حدد مراد وآخرون (2022) إخفاقات التوحيد

²³ Mahyudin Ritonga dkk., "Arabic language learning reconstruction as a response to strengthen Al-Islam studies at higher education," *International Journal of Evaluation and Research in Education (IJERE)* 10, no. 1 (2021), <https://doi.org/10.11591/ijere.v10i1.20747>.

²⁴ Mohammad Ahsanuddin dkk., "Development and Validation of a Construct 2-Based Gamified Assessment for Arabic Language Learning," *Asia Pacific Journal of Educators and Education* 40, no. 3 (2026), <https://doi.org/10.21315/apjee2025.40.3.26>.

²⁵ Umi Hijriyah, "How Effective Is SUNO.AI in Enhancing Arabic Listening Skills? An Evaluation of AI-Based Personalized Learning," *International Journal of Information and Education Technology* 15, no. 2 (2025), <https://doi.org/10.18178/ijiet.2025.15.2.2251>.

²⁶ Bingxin Gu dkk., "The Effectiveness of a Digital Literacy-Integrated Syllabus for Arabic-speaking Courses in Teacher Education Universities in China," *Journal of Curriculum and Teaching* 14, no. 1 (2025): 202, <https://doi.org/10.5430/jct.v14n1p202>.

²⁷ Mingazova dkk., "Usage of Digital Educational Sources in Arabic Teaching."

²⁸ Mahyudin Ritonga dkk., "Arabic language learning reconstruction as a response to strengthen Al-Islam studies at higher education," *International Journal of Evaluation and Research in Education (IJERE)* 10, no. 1 (2021): 355, <https://doi.org/10.11591/ijere.v10i1.20747>.

²⁹ Darwis Jauhari Bandu dkk., "Context Evaluation on Implementation of English for Islamic Studies Program in IAIN Palu, Indonesia," *Kasetsart Journal of Social Sciences* 42, no. 2 (2021), <https://so04.tci-thaijo.org/index.php/kjss/article/view/250927>.

³⁰ Daniel L. Stufflebeam, "The CIPP Model for Program Evaluation," dalam *Evaluation Models*, ed. oleh George F. Madaus dkk. (Kluwer-Nijhoff Publishing, 1983).

القياسي وغياب الرقابة في برنامج اللغة العربية في مؤسسة التعليم العالي الدينية الإسلامية (PTKIN) خلال جائحة كوفيد-19. قام Alwaely و Lahiani (2020) بتقييم كفاءة الكتابة الحجاجية لدى مدرسي اللغة العربية في سياقات التعلم عبر الإنترنت في الإمارات العربية المتحدة.³¹ قام Zulhannan وآخرون بتقييم كتاب «العربية بين يديك» من حيث دمج الكفاءة التواصلية.³² قام Maujud و Syaharuddin (2025) بتطوير إطار لضمان الجودة لتقييم أداء المعلمين بعد الحصول على الشهادة وهو استجابة مباشرة لغياب البنية التحتية للرصد المنهجي في المدارس الدينية الإندونيسية.³³

التقييم التكويني والتشخيصي

الدراسات التكوينية والتشخيصية قليلة (n=4) لكنها متميزة من الناحية التقنية. طور ناصري وآخرون (2022) نظامًا آليًا لقابلية قراءة النصوص العربية كلغة ثانية يصنف النصوص إلى خمسة مستويات من الصعوبة بدقة تبلغ 86.15% مما يتيح مطابقة المواد التشخيصية على نطاق واسع.³⁴ أجرى معمون وآخرون (2021) مسحًا حول نموذج TPACK بين معلمي اللغة العربية الإندونيسيين، ووجدوا أوجه قصور محددة في المجال الفرعي «التكنولوجيا في التقييم»، مما يشير إلى أنه حتى في حالة وجود الأدوات، فإن قدرة المعلمين على استخدامها بشكل تكويني محدودة.³⁵ أنتج الرشودي وآخرون (2025) الأداة الأكثر تطورًا من الناحية التقنية في المجموعة: نظام «الكشف عن الأخطاء النطقية وتشخيصها» (MDD) القائم على محول (transformer)، والذي حقق دقة

³¹ Suad Abdelkareem Alwaely dan Hanene Lahiani, "Assessing Argumentative Writing Skills in Online Environments among Arabic Language Teachers of Grades 5 to 7 in UAE Schools," *International Journal of Emerging Technologies in Learning (IJET)* 15, no. 18 (2020), <https://doi.org/10.3991/ijet.v15i18.13581>.

³² Zulhannan dkk., "Enhancing Communicative Arabic Teaching: Evaluating the AI-Arabiyah Baina Yadaik Model," *Journal of Educational and Social Research* 15, no. 3 (2025), <https://doi.org/10.36941/jesr-2025-0100>.

³³ Fathul Maujud dan Syaharuddin Syaharuddin, "A Quality Assurance Framework for Strengthening Post-Certification Arabic Language Teacher Performance in Madrasah," *Jurnal Pendidikan Islam* 11, no. 2 (2025): 343–60, <https://doi.org/10.15575/jpi.v11i2.49517>.

³⁴ Naoual Nassiri dkk., "Arabic L2 Readability Assessment: Dimensionality Reduction Study," *Journal of King Saud University - Computer and Information Sciences* 34, no. 6 (2022), <https://doi.org/10.1016/j.jksuci.2020.12.021>.

³⁵ Titin Nurhayati Ma'mun dkk., "Foreign language teachers' technological and pedagogical content knowledge: A study with AFL teachers in Indonesia," *Journal of Language and Linguistic Studies* 17, no. 4 (2021), <https://doi.org/10.52462/jlls.145>.

كشف بلغت 91.3% ودقة تشخيص بلغت 80.8%³⁶ قام سويب وباهارودين (2025) بتقييم قدرة المعلمين على تقييم العناصر الإيقاعية في اللغة العربية، ووجدوا ثغرات منهجية في التدريب على التقييم الخاص بالإيقاع.³⁷

التقييم القائم على الذكاء الاصطناعي: موضوع ناشئ

تمثل الدراسات التي تركز على الذكاء الاصطناعي (n=4 كمساهمين رئيسيين) المجموعة الأسرع نموًا. قام Allaithy و Zaki (2025) بتقييم مخرجات ChatGPT و Microsoft Copilot و Gemini لتقييم فهم القراءة باللغة العربية، ووجدوا ملاءمة نحوية ولكن مع قيود ثقافية وسياقية منهجية.³⁸ وثق الكعبي والمامي (2025) أن المعلمين يقدرون إمكانات الذكاء الاصطناعي في مجال التقييم، لكنهم يفتقرون إلى المهارات النقدية اللازمة لتقييم مخرجات الذكاء الاصطناعي بمسؤولية وهي نتيجة لها آثار مباشرة على التطوير المهني.³⁹ نظام MDD الذي طوره الرشودي وآخرون (2025)، على الرغم من قوته التقنية، تم التحقق من صلاحيته على مجموعات محددة من المتعلمين، مما ترك قابلية تعميمه عبر المجموعات السكانية دون اختبار. يقدم سوليمانندو (2021) الأساس النظري لهذه المجموعة، حيث يجادل بأن الاختبارات العربية عملت تاريخيًا كأدوات أيديولوجية وهو إطار قابل للتطبيق مباشرة على أنظمة التقييم الخوارزمية التي تشفر خيارات تصميمها افتراضات اجتماعية لغوية.⁴⁰

الاتجاهات الزمنية والجغرافية (السؤال البحثي 2)

³⁶ Alrashoudi dkk., "Improving Mispronunciation Detection and Diagnosis for Non- Native Learners of the Arabic Language."

³⁷ Nur Syukriyah Mohd Suib dan Harun Baharudin, "Effective Strategies for Teaching Arabic Prosody in Oral Communication Skills," *Qubahan Academic Journal* 5, no. 1 (2025), <https://doi.org/10.48161/qaj.v5n1a855>.

³⁸ Allaithy dan Zaki, "Evaluation of AI-Generated Reading Comprehension Materials for Arabic Language Teaching."

³⁹ Mozah H. Alkaabi dan Asma Saeed Almaamari, "Generative AI Implementation and Assessment in Arabic Language Teaching," *International Journal of Online Pedagogy and Course Design* 15, no. 1 (2025), <https://doi.org/10.4018/IJOPCD.368037>.

⁴⁰ Solimando, "Language Tests in the Arabic-speaking World."

الجدول 5. التوزيع الزمني والمواضيع السائدة حسب الفترة

الفترة	n (%)	المواضيع السائدة	دراسات تمثيلية
2020-2021	10 (35.7%)	الاختبارات التلخيصية؛ تقييم البرامج؛ تأثير التعلم المدمج؛ أطر الكفاءة	القطونة وآخرون (2020)؛ الزرقوني وآخرون (2020)؛ ريتونغا وآخرون (2021)؛ باندو وآخرون (2021)
2022-2023	6 (21.4%)	قابلية القراءة الآلية؛ تقييم برامج COVID-19؛ جودة تقييم تكنولوجيا المعلومات والاتصالات؛ نزاهة الامتحانات عبر الإنترنت	نصيري وآخرون (2022)؛ مراد وآخرون (2022)؛ معروف وباروكا (2023)؛ باسو وآخرون (2023)
2024-2025	12 (42.9%)	تقييم محتوى الذكاء الاصطناعي؛ تطوير التقييم القائم على الألعاب؛ MDD القائم على المحولات؛ أطر ضمان الجودة للمعلمين؛ الاستماع التلوي	علائي وزكي (2025)؛ أحسان الدين وآخرون (2025)؛ الرشودي وآخرون (2025)؛ موجود وشاهار الدين (2025)

تشكل الأبحاث الإندونيسية 39.3% من المجموعة، مدفوعةً بالعدد الكبير للسكان المسلمين في البلاد، وتدرّس اللغة العربية الإلزامي في المدارس الدينية، وزيادة معدلات النشر في سكوبوس. وتعكس الدراسات الماليزية (17.9%) استثمارات سياسية مماثلة. وتقدم مساهمات البلدان ذات الأغلبية العربية (21.4%) وجهات نظر مؤسسية من متحدثي اللغة العربية كلغة أم. ويلاحظ غياب سياقات شمال إفريقيا والمشرق العربي

والشتات العربي وهي فجوات تعكس على الأرجح تفاوتات في الفهرسة في سكوبوس أكثر من كونها انعدامًا للأبحاث.
التركيز على المهارات اللغوية (السؤال البحثي 3)

الجدول 6. توزيع التركيز على المهارات اللغوية (N=28)

المهارة	n	%	دراسات تمثيلية	نهج التقييم في المرحلة الابتدائية
فهم القراءة	10	35.7	Midhwah & Alhawary (2020) Nassiri et al (2022) Rusli et al (2022) & Allaithy (2024) Zaki (2025)	مهام الأداء؛ قابلية القراءة الآلية؛ اختبارات الفهم قبل وبعد
الشفوية / النطق	6	21.4	الراشودي وآخرون (2025) سويب وبهارودين (2025) الحادي (2021)	Transformer MDD؛ معايير النغمة؛ ربط إطار الكفاءة
عام / متعدد	6	21.4%	زرقوني وآخرون (2020) مراد وآخرون (2022) معروف وباروكا (2023)	تقييم البرامج؛ تقييم المهارات المتعددة باستخدام تكنولوجيا المعلومات والاتصالات
الكتابة	3	10.7	الأويلي واللاهاني (2020) لانغستن وآخرون (2022) موزين وجوما (2022)	معايير تقييم الكتابة؛ البيانات الوطنية؛ التحليل المرجعي

تقييم منصة الذكاء الاصطناعي؛ دراسة ارتباطية باستخدام نموذج MALQ	الهجرية وآخرون (2025)؛ باهارودين ورحسن (2025)	7.1%	2	الاستماع
التقييم المعجمي المرجعي	مؤذن وجوما (2022)	3.6%	1	المفردات

يهيمن فهم القراءة (35.7%)، بما يتوافق مع أهميته المركزية في السياقات التعليمية الإسلامية حيث يُعد الإمام الديني القائم على النصوص هدفاً أساسياً. ويعكس تقييم المحادثة والنطق (21.4%) الاهتمام المتزايد بالكفاءة التواصلية وظهور أدوات الكلام الآلية. أما تقييم الكتابة (10.7%) فهو غير ممثل بشكل كافٍ نظراً للتحديات الفريدة التي يفرضها الحرف العربي على المتعلمين غير الناطقين باللغة؛ ولم يرق سوى Lahiani و Alwaely بتقييم الكتابة بشكل مباشر باعتبارها مكوناً أساسياً.⁴¹ ويُعتبر الاستماع (7.1%) والمفردات (3.6%) هامشين كمجالات تركيز قائمة بذاتها.

ثغرات البحث (RQ4)

الفجوة 1: عدم وجود تصميمات طويلة

جميع الدراسات الـ 28 هي دراسات مستعرضة. لا توجد دراسة تتابع ممارسات التقييم أو تطور المتعلم بشكل طولي. ولا يزال نظام المدارس الدينية الإندونيسية الذي يتلقى فيه الطلاب تعليم اللغة العربية عبر مستويات تعليمية متعددة غير متابع تماماً من حيث التأثير التراكمي للتقييم.⁴²

الفجوة 2: غياب التقييم التكويني وتقييم الملفات و أقران التقييم

على الرغم من قاعدة الأدلة القوية التي قدمها بلاك وويليام (1998)، فإن مناهج التقييم التكويني وتقييم الملفات وأقران التقييم غائبة عملياً عن مجموعة الدراسات.⁴³

⁴¹ Alwaely dan Lahiani, "Assessing Argumentative Writing Skills in Online Environments among Arabic Language Teachers of Grades 5 to 7 in UAE Schools."

⁴² Ritonga dkk., "Arabic language learning reconstruction as a response to strengthen Al-Islam studies at higher education," 2021; Zurqoni dkk., "Has Arabic Language Learning Been Successfully Implemented?"; Muradi dkk., "Evaluation of Arabic Learning Program at State Islamic Religious College During COVID-19."

⁴³ Paul Black dan Dylan Wiliam, "Assessment and Classroom Learning," *Assessment in Education: Principles, Policy & Practice* 5, no. 1 (1998), <https://doi.org/10.1080/0969595980050102>.

يوثق معمون وآخرون (2021) أوجه القصور في قدرات المعلمين؛ ويدعو مراد وآخرون (2022) ومكروف وباروكا (2023) إلى المراقبة المستمرة لكن لا توجد دراسة أساسية تختبر استراتيجيات التقييم التكويني في إطار تدريس اللغة العربية كلغة ثانية (ALT).⁴⁴

الفجوة 3: عدم كفاية التحقق من الصحة السيكومترية

يتم تطوير الأدوات الرقمية والأدوات القائمة على الألعاب دون الحجج المتعلقة بالصحة التي يتطلبها إطار عمل كين (2013).⁴⁵ تعد دراسة أحسان الدين وآخرون هي الدراسة الأكثر توجهاً نحو التحقق من الصحة، ومع ذلك لا تزال الصحة المرتبطة بالمعايير وقابلية تعميم الموثوقية غير موثقة.⁴⁶ غالباً ما تكون آليات التقييم في أنظمة الذكاء الاصطناعي غير شفافة.⁴⁷

الفجوة 4: عدم إجراء أبحاث حول معرفة المعلمين بالتقييم

يتم الإشارة باستمرار إلى فجوات المعرفة التقييمية عبر دراسات متعددة.⁴⁸ لكن لم يتم التحقيق فيها مباشرة كمتغير أساسي. لا تطبق أي دراسة إطار عمل راسخاً للمعرفة التقييمية للمعلمين (على سبيل المثال، Fulcher، 2012) على سياقات التقييم المدعوم بالذكاء الاصطناعي (ALT).⁴⁹

الفجوة 5: الأبعاد الأخلاقية للتقييم باستخدام الذكاء الاصطناعي

⁴⁴ Makruf dan Barokah, "Improving the Quality of ICT-Based Arabic Learning Assessment With Online Applications"; Nurhayati Ma'mun dkk., "Foreign language teachers' technological and pedagogical content knowledge"; Muradi dkk., "Evaluation of Arabic Learning Program at State Islamic Religious College During COVID-19."

⁴⁵ Michael T. Kane, "Validating the Interpretations and Uses of Test Scores," *Journal of Educational Measurement* 50, no. 1 (2013), <https://doi.org/10.1111/jedm.12000>.

⁴⁶ Ahsanuddin dkk., "Development and Validation of a Construct 2-Based Gamified Assessment for Arabic Language Learning."

⁴⁷ Alrashoudi dkk., "Improving Mispronunciation Detection and Diagnosis for Non- Native Learners of the Arabic Language"; Allaithy dan Zaki, "Evaluation of AI-Generated Reading Comprehension Materials for Arabic Language Teaching."

⁴⁸ Nurhayati Ma'mun dkk., "Foreign language teachers' technological and pedagogical content knowledge"; Maujud dan Syaharuddin, "A Quality Assurance Framework for Strengthening Post-Certification Arabic Language Teacher Performance in Madrasah"; Muradi dkk., "Evaluation of Arabic Learning Program at State Islamic Religious College During COVID-19."

⁴⁹ Glenn Fulcher, "Assessment Literacy for the Language Classroom," *Language Assessment Quarterly* 9, no. 2 (2012), <https://doi.org/10.1080/15434303.2011.642041>.

يوثق Allaithy و (2025) Zaki قيود الصلاحية الثقافية؛ وي طرح Solimando (2021) نظرية حول الوظائف الأيديولوجية لاختبارات اللغة العربية.⁵⁰ ولا توجد أي أبحاث تجريبية تدرس عدالة الخوارزميات، وتحيز بيانات التدريب، وخصوصية البيانات في التقييم باللغة العربية القائم على الذكاء الاصطناعي.

الفجوة 6: الفجوات الجغرافية والمؤسسية

لا يمثل البحث في مجال التقييم في المرحلة الابتدائية سوى دراستين فقط.⁵² ولم يتم تمثيل شمال إفريقيا والشرق الأدنى ومجتمعات الشتات وسياقات تعلم اللغة العربية غير الرسمية.

المناقشة

توثق النتائج مجاًلاً يمر بمرحلة انتقال هيكلية، حيث ينتقل من الممارسات التلخيصية التي تركز على الاختبارات إلى نهج تعتمد على التكنولوجيا وتدمج الذكاء الاصطناعي بشكل متزايد، بينما تكشف في الوقت نفسه عن أوجه قصور نظرية ومنهجية مستمرة تحد من قدرة هذا المجال على توليد أدلة تقييم صالحة ومفيدة. ويُعد النقص في الصحة هو الأكثر جوهرية. يشترط باخمان وبالمر (1996) وجود توافق صريح بين المفاهيم والمهام ومتطلبات اللغة في العالم الواقعي.⁵³ ويشترط كين (2013) وجود سلسلة استدلالية واضحة من أداء الاختبار إلى ادعاءات الكفاءة.⁵⁴ وعلى مستوى المجموعة الكاملة من الدراسات، لم يتم استيفاء أي من هذين الشرطين بشكل متسق. فغالبية الدراسات شبه التجريبية⁵⁵ تعامل الصلاحية على أنها أمر مفروغ منه فهي تجري الاختبارات، وتبلغ عن الدرجات، وتستخلص الاستنتاجات دون تحديد البناء الذي

⁵⁰ Allaithy dan Zaki, "Evaluation of AI-Generated Reading Comprehension Materials for Arabic Language Teaching."

⁵¹ Solimando, "Language Tests in the Arabic-speaking World."

⁵² Langsten dkk., "Arabic Language Skills"; Midhwah dan Alhawary, "Arabic Diacritics and Their Role in Facilitating Reading Speed, Accuracy, and Comprehension by English L2 Learners of Arabic."

⁵³ Bachman dan Palmer, *Language Testing in Practice*.

⁵⁴ Kane, "Validating the Interpretations and Uses of Test Scores."

⁵⁵ Baharudin dan Rahman, "The Impact of Metacognitive Awareness on Arabic Listening Proficiency among Religious Secondary School Students"; Al-Qatawneh dkk., "The Effect of Blended Learning on the Achievement of HDE Students in the Methods of Teaching Arabic Language Course and Their Attitudes towards Its Use at Ajman University"; Rusli dkk., "The Impact of Informative and Narrative Texts on Arabic Language Comprehension in Islamic Education."

يتم قياسه، أو ما تعنيه الدرجات بالنسبة لاستخدام اللغة العربية في العالم الواقعي، أو ما هي الأدلة التي تدعم تلك الاستدلالات. يمثل ميدوا والحواري (2020)، في دراستهما المنشورة في مجلة اللغات الحديثة (Modern Language Journal)، الاستثناء: فقد حددوا المفهوم (تسهيل القراءة من خلال السمات الإملائية)، وقاموا بتطبيقه من خلال مهام تستند إلى النظرية، وقصروا استنتاجاتهم وفقاً لذلك.⁵⁶ وهذا هو المعيار الذي ينبغي أن يتجه إليه هذا المجال.

تترتب على فجوة التقييم التكويني عواقب مباشرة على التعلم. وقد أثبت التوليف الذي أجراه بلاك وويليام (1998) لأكثر من 250 دراسة أن التقييم التكويني يُعد أحد التدخلات التعليمية الأكثر تأثيراً المتاحة.⁵⁷ يشير التوجه شبه الكامل لأدبيات تعليم اللغة العربية كلغة ثانية (ALT) نحو النتائج التلخيصية إلى أن قاعدة الأدلة هذه لم تتغلغل في الممارسة أو البحث في هذا المجال. والآثار المترتبة على ذلك جسيمة: يتم تقييم متعلمي اللغة العربية في السياقات الموثقة هنا المدارس الدينية الإندونيسية، والجامعات الماليزية، والمؤسسات الأردنية والسعودية في المقام الأول للتصديق على النتائج بدلاً من تحسين التعلم. تشير النتائج التي توصل إليها معمون وآخرون (2021) إلى أن معلمي اللغة العربية الإندونيسيين يفتقرون على وجه التحديد إلى «التصورات والممارسات والمعرفة والمهارات» (TPACK) في مجالات التقييم، مما يشير إلى أن الحلقة المفقودة ليست دافع المعلم بل قدراته.⁵⁸ أما إطار ضمان الجودة الذي وضعه موجود وشاهارودين (2025)، فعلى الرغم من أنه يمثل مساهمة هيكلية مهمة، إلا أنه يتناول المسألة المؤسسية بدلاً من الممارسات التكوينية داخل الفصل الدراسي.⁵⁹

⁵⁶ Midwhah dan Alhawary, "Arabic Diacritics and Their Role in Facilitating Reading Speed, Accuracy, and Comprehension by English L2 Learners of Arabic."

⁵⁷ Black dan Wiliam, "Assessment and Classroom Learning."

⁵⁸ Nurhayati Ma'mun dkk., "Foreign language teachers' technological and pedagogical content knowledge."

⁵⁹ Maujud dan Syaharuddin, "A Quality Assurance Framework for Strengthening Post-Certification Arabic Language Teacher Performance in Madrasah."

تثير مجموعة الذكاء الاصطناعي (AI) أسئلة تتجاوز الأداء التقني. يحقق نظام MDD الذي طوره Alrashoudi وآخرون (2025) دقة اكتشاف مثيرة للإعجاب.⁶⁰ لكن النتائج التي توصل إليها Allaithy و Zaki (2025) بأن أدوات الذكاء الاصطناعي تنتج محتوى تقييم غير ملائم ثقافياً، وتوثيق Alkaabi و Almaamari (2025) لعدم استعداد المعلمين لتقييم مخرجات الذكاء الاصطناعي تقييماً نقدياً، تشير معاً إلى أن القدرة التقنية تتقدم على البنية التحتية التربوية والأخلاقية.⁶¹ إن حجة سوليمانندو (2021) التي تفيد بأن اختبارات اللغة العربية تنطوي على افتراضات أيديولوجية حول الاستخدام المشروع للغة تنطبق بشكل مباشر هنا: فقد تؤدي أنظمة الذكاء الاصطناعي التي تم تدريبها بشكل أساسي على مجموعات نصوص اللغة العربية الفصحى إلى أداء أقل من المتوقع بشكل منهجي بالنسبة لمتعلمي اللهجات العامية، أو المتعلمين من خلفيات لغوية أولية معينة، أو المتعلمين في سياقات تعليمية غير سائدة ولن تكون هذه الفوارق في الأداء مرئية دون إجراء اختبارات مخصصة لضمان العدالة.⁶²

تتفق عدة نتائج على وجود تحدٍ هيكلي لا يمكن حله من خلال المساهمات الفردية في الدراسات: يفتقر هذا المجال إلى بنية تحتية مشتركة لضمان جودة التقييم. لا يوجد إطار متفق عليه لمستويات إتقان اللغة العربية يماثل الإطار المرجعي الأوروبي العام للغات (CEFR) في السياقات الأوروبية. ولا توجد أدوات موحدة تم التحقق من صحتها على نطاق واسع لسياقات تدريس اللغة العربية كلغة ثانية (ALT) الشائعة. لا توجد مجتمع مهني مستدام ينظم تطوير معرفة المعلمين في مجال التقييم. ويُعد تحليل إطار الكفاءة الذي أجراه إلهادي (2021) وإطار ضمان الجودة للمعلمين الذي وضعه موجود وشاهارودين (2025) جهوداً منفردة تشير إلى الشكل الذي قد تبدو عليه البنية التحتية الأكثر منهجية لكنها تظل مساهمات معزولة مؤسسياً.⁶³

⁶⁰ Alrashoudi dkk., "Improving Mispronunciation Detection and Diagnosis for Non- Native Learners of the Arabic Language."

⁶¹ Allaithy dan Zaki, "Evaluation of AI-Generated Reading Comprehension Materials for Arabic Language Teaching."

⁶² Solimando, "Language Tests in the Arabic-speaking World."

⁶³ Elhadky, "Arap Olmayanlara Arapça Öğretimine Yönelik Araştırmacı ve Uzman Yetiştirilmede Dil Yeterliliği Dersine Yönelik Bir Müfredat Önerisi"; Maujud dan Syaharuddin, "A Quality Assurance Framework for Strengthening Post-Certification Arabic Language Teacher Performance in Madrasah."

هناك نتيجة تستحق الاعتراف الصريح بها باعتبارها أحد قيود تمثيلية هذه المراجعة نفسها: فالتركيز الجغرافي في إندونيسيا وماليزيا يعني أن قاعدة الأدلة المجمعة تميل نحو نماذج مؤسسية معينة (التعليم العالي الإسلامي، والمناهج الدراسية الوطنية للمدارس الدينية)، وملامح معينة للمتعلمين (متعلمو اللغة العربية كلغة ثانية (L2) الذين تكون لغتهم الأم (L1) هي الإندونيسية أو الماليزية)، وسياقات سياسية معينة (اللغة العربية كلغة دينية). قد تبدو النتائج والثغرات مختلفة بشكل جوهري في أنظمة التعليم ثنائي اللغة في شمال إفريقيا، أو برامج لغة التراث في الشتات، أو أقسام اللغة العربية في الجامعات العلمانية. وينبغي أن تراعي المراجعات المستقبلية هذا التنوع الجغرافي بشكل صريح.

الخلاصة

جمعت هذه المراجعة المنهجية للأدبيات 28 دراسة مُفهرسة في قاعدة بيانات سكوبوس (Scopus) حول التقييم والتقدير في تدريس اللغة العربية، نُشرت بين عامي 2020 و2025. تم استرجاع الدراسات عبر موقع «Publish or Perish» وفحصها من خلال عملية مستقلة بمراجعين اثنين ($K = 0.81$)، وتم تقييم جودة الدراسات المدرجة باستخدام معايير MMAT/CASP المعدلة: حصلت 22 دراسة على تصنيف «جودة عالية» (78.6٪) وست دراسات على تصنيف «جودة متوسطة» (21.4٪). يبدو أن هذه المراجعة تُعد من أوائل الجهود المخصصة للمراجعة المنهجية للأدبيات (SLR) التي تركز بشكل خاص على التقييم والتقدير كمجال أساسي في تدريس اللغة العربية.

تتميز مشهد التقييم بخمس مجموعات مواضيعية: اختبارات الإنجاز التلخيصية، والتقييم المعزز بالتكنولوجيا، وتقييم البرامج، والتقييم التكويني والتشخيصي، والتقييم القائم على الذكاء الاصطناعي (،)، حيث برز هذا الأخير كاتجاه سائد اعتبارًا من عام 2023 فصاعدًا. ويهيمن فهم القراءة والكفاءة الشفوية على المهارات التي ركزت عليها الدراسات (57.1% مجتمعة). وتعد إندونيسيا (39.3٪) والتعليم العالي (64.3٪) السياق والبيئة

الأكثر تمثيلاً على التوالي. وقد تم تحديد ست فجوات بحثية، حيث برز غياب أبحاث التقييم التكويني وعدم كفاية التحقق من الصحة السيكمومترية باعتبارهما الأكثر أهمية. تنبثق ثلاث توصيات عن هذا التوليف. بالنسبة للباحثين: تتمثل الثغرات ذات الأولوية القصوى في التصاميم الطولية، ودراسات التدخل في مجال التقييم التكويني، ودراسات حجج الصحة الخاصة بأدوات التقييم الرقمية والقائمة على الذكاء الاصطناعي، ويمكن معالجة كل منها بشكل مباشر من خلال الأطر المنهجية القائمة. بالنسبة للممارسين: يوفر هذا التوليف أساساً تجريبياً للاستثمار في دمج التقييم التكويني وتنمية معرفة المعلمين بالتقييم باعتبارهما أنشطة ذات تأثير أكبر مقارنةً بالانتشار المستمر للأدوات الرقمية التلخيصية. بالنسبة لصانعي السياسات: يقدم إطار ضمان الجودة الذي وضعه ماوجود وشاهارودين (2025) نموذجاً قابلاً للتكرار لإضفاء الطابع المؤسسي على التقييم المنهجي لأداء المعلمين؛ ومن شأن توسيع نطاق هذا النموذج ليشمل السياقات الوطنية أن يعالج الفجوة في البنية التحتية الهيكلية التي حددها هذا التوليف. تشمل القيود الاعتماد على قاعدة بيانات واحدة (Scopus)، وقصر البحث على اللغة الإنجليزية، والتركيز الجغرافي على مؤسسة واحدة، وكون التوليف سردياً وليس تحليلياً. وينبغي أن تتوسع المراجعات المستقبلية لتشمل قواعد بيانات Web of Science وقواعد البيانات باللغة العربية؛ وأن تستخدم مراجعين مزدوجين مسجلين مسبقاً؛ وأن تشمل الأدبيات الرمادية لتغطية النطاق الكامل لممارسات تقييم تقنيات التعلم البديلة (ALT) على الصعيد العالمي.

المراجع

- Ahsanuddin, Mohammad, Zawawi Ismail, Muhammad Ahsan Thoriq, dkk. "Development and Validation of a Construct 2-Based Gamified Assessment for Arabic Language Learning." *Asia Pacific Journal of Educators and Education* 40, no. 3 (2026). <https://doi.org/10.21315/apjee2025.40.3.26>.
- Alkaabi, Mozah H., dan Asma Saeed Almaamari. "Generative AI Implementation and Assessment in Arabic Language Teaching." *International Journal of Online Pedagogy and Course Design* 15, no. 1 (2025). <https://doi.org/10.4018/IJOPCD.368037>.
- Allaithy, Ahmed, dan Mai Zaki. "Evaluation of AI-Generated Reading Comprehension Materials for Arabic Language Teaching." *Computer Assisted Language Learning*, 6 Maret 2025. <https://doi.org/10.1080/09588221.2025.2474037>.

- Al-Qatawneh, Sami, Mohd. Elmagzoub Eltahir, dan Najeh Rajeh Alsalhi. "The Effect of Blended Learning on the Achievement of HDE Students in the Methods of Teaching Arabic Language Course and Their Attitudes towards Its Use at Ajman University: A Case Study." *Education and Information Technologies* 25, no. 3 (2020). <https://doi.org/10.1007/s10639-019-10046-w>.
- Alrashoudi, Norah, Hend Al-Khalifa, dan Yousef Alotaibi. "Improving Mispronunciation Detection and Diagnosis for Non- Native Learners of the Arabic Language." *Discover Computing* 28, no. 1 (2025). <https://doi.org/10.1007/s10791-024-09489-8>.
- Alwaely, Suad Abdelkareem, dan Hanene Lahiani. "Assessing Argumentative Writing Skills in Online Environments among Arabic Language Teachers of Grades 5 to 7 in UAE Schools." *International Journal of Emerging Technologies in Learning (iJET)* 15, no. 18 (2020). <https://doi.org/10.3991/ijet.v15i18.13581>.
- Bachman, Lyle F., dan Adrian S. Palmer. *Language Testing in Practice*. Oxford University Press, 1996.
- Baharudin, Harun, dan Nur Farahhuda Abdul Rahman. "The Impact of Metacognitive Awareness on Arabic Listening Proficiency among Religious Secondary School Students." *The Open Psychology Journal* 18, no. 1 (2025). <https://doi.org/10.2174/0118743501422272251006211402>.
- Bandu, Darwis Jauhari, Ishak Abdulhak, Dinn Wahyudin, Rusman, dan Rohmani Nur Indah. "Context Evaluation on Implementation of English for Islamic Studies Program in IAIN Palu, Indonesia." *Kasetsart Journal of Social Sciences* 42, no. 2 (2021). <https://so04.tci-thaijo.org/index.php/kjss/article/view/250927>.
- Baso, Yusring Sanusi, Nurul Murtadho, Syihabuddin -, dkk. "Reducing Cheating in Online Exams Through the Proctor Test Model." *International Journal of Advanced Computer Science and Applications* 14, no. 1 (2023). <https://doi.org/10.14569/IJACSA.2023.0140139>.
- Black, Paul, dan Dylan Wiliam. "Assessment and Classroom Learning." *Assessment in Education: Principles, Policy & Practice* 5, no. 1 (1998). <https://doi.org/10.1080/0969595980050102>.
- Braun, Virginia, dan Victoria Clarke. "Using Thematic Analysis in Psychology." *Qualitative Research in Psychology* 3, no. 2 (2006). <https://doi.org/10.1191/1478088706qp063oa>.
- Critical Appraisal Skills Programme. *CASP Checklists*. Critical Appraisal Skills Programme, 2022. <https://casp-uk.net/casp-tools-checklists/>.
- Elhadky, Islam. "Arap Olmayanlara Arapça Öğretimine Yönelik Araştırmacı ve Uzman Yetiştirmede Dil Yeterliliği Dersine Yönelik Bir Müfredat Önerisi." *darulfunun ilahiyat* 32, no. 2 (2021). <https://doi.org/10.26650/di.2021.32.2.995627>.
- Falagas, Matthew E., Eleni I. Pitsouni, George A. Malietzis, dan Georgios Pappas. "Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and Weaknesses." *The FASEB Journal* 22, no. 2 (2008). <https://doi.org/10.1096/fj.07-9492LSF>.
- Fulcher, Glenn. "Assessment Literacy for the Language Classroom." *Language Assessment Quarterly* 9, no. 2 (2012). <https://doi.org/10.1080/15434303.2011.642041>.
- Gu, Bingxin, Saedah Binti Siraj, dan Zawawi Bin Ismail. "The Effectiveness of a Digital Literacy-Integrated Syllabus for Arabic-speaking Courses in Teacher Education Universities in China." *Journal of Curriculum and Teaching* 14, no. 1 (2025). <https://doi.org/10.5430/jct.v14n1p202>.

- Higgins, Julian P. T., dan Sally Green. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0. Disunting oleh Julian P. T. Higgins dan Sally Green. The Cochrane Collaboration, 2011. <https://handbook-5-1.cochrane.org/>.
- Hijriyah, Umi. "How Effective Is SUNO.AI in Enhancing Arabic Listening Skills? An Evaluation of AI-Based Personalized Learning." *International Journal of Information and Education Technology* 15, no. 2 (2025). <https://doi.org/10.18178/ijiet.2025.15.2.2251>.
- Hong, Q. N., Pierre Pluye, Sergi Fàbregues, dkk. *Mixed Methods Appraisal Tool (MMAT), Version 2018*. Canadian Intellectual Property Office, 2018. <http://mixedmethodsappraisaltoolpublic.pbworks.com/>.
- Kane, Michael T. "Validating the Interpretations and Uses of Test Scores." *Journal of Educational Measurement* 50, no. 1 (2013). <https://doi.org/10.1111/jedm.12000>.
- Landis, J. Richard, dan Gary G. Koch. "The Measurement of Observer Agreement for Categorical Data." *Biometrics* 33, no. 1 (1977). <https://doi.org/10.2307/2529310>.
- Langsten, Ray, Fatma Abdelkhalek, dan Tahra Hassan. "Arabic Language Skills: A Comparative Study of Community and Government Schools in Rural Upper-Egypt." *Compare: A Journal of Comparative and International Education* 52, no. 6 (2022). <https://doi.org/10.1080/03057925.2020.1843003>.
- Makruf, Imam, dan Anisatul Barokah. "Improving the Quality of ICT-Based Arabic Learning Assessment With Online Applications." *Journal of Higher Education Theory and Practice* 23, no. 11 (2023). <https://doi.org/10.33423/jhetp.v23i11.6216>.
- Maujud, Fathul, dan Syaharuddin Syaharuddin. "A Quality Assurance Framework for Strengthening Post-Certification Arabic Language Teacher Performance in Madrasah." *Jurnal Pendidikan Islam* 11, no. 2 (2025). <https://doi.org/10.15575/jpi.v11i2.49517>.
- Midhwah, Ali Al, dan Mohammad T. Alhawary. "Arabic Diacritics and Their Role in Facilitating Reading Speed, Accuracy, and Comprehension by English L2 Learners of Arabic." *The Modern Language Journal* 104, no. 2 (2020). <https://doi.org/10.1111/modl.12642>.
- Mingazova, N. G., V. G. Subich, dan A. I. Gazizova. "Usage of Digital Educational Sources in Arabic Teaching." *Utopia y Praxis Latinoamericana* 25, no. Extra 12 (2020). <https://doi.org/10.5281/zenodo.4280165>.
- Müezzini, Ahmet, dan Mai Gomaa. "Ana Dili Arapça Olmayan Öğrenciler İçin Sözcüksel Yeterlilikler ve Öğrencilerin Beklenen Performansına Uygunluğu 'Uygulamalı Bir Çalışma.'" *darulfunun ilahiyat* 33, no. 1 (2022). <https://doi.org/10.26650/di.2022.33.1.1087872>.
- Muradi, Ahmad, La Ode Amaluddin, Danial Hilmi, dkk. "Evaluation of Arabic Learning Program at State Islamic Religious College During COVID-19." *Eurasian Journal of Educational Research*, no. 99 (2022). <https://doi.org/10.14689/ejer.2022.99.011>.
- Nassiri, Naoual, Abdelhak Lakhouaja, dan Violetta Cavalli-Sforza. "Arabic L2 Readability Assessment: Dimensionality Reduction Study." *Journal of King Saud University - Computer and Information Sciences* 34, no. 6 (2022). <https://doi.org/10.1016/j.jksuci.2020.12.021>.
- Nurhayati Ma'mun, Titin, Ade Kosasih, Yani Rohmayani, dan Eka Kurnia Fimansyah. "Foreign language teachers' technological and pedagogical content knowledge: A study with AFL teachers in Indonesia." *Journal of Language and Linguistic Studies* 17, no. 4 (2021): 1998–2021. <https://doi.org/10.52462/jlls.145>.

- Page, Matthew J., Joanne E. McKenzie, Patrick M. Bossuyt, dkk. "The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews." *Research Methods & Reporting. BMJ* 372 (Maret 2021): n71. <https://doi.org/10.1136/bmj.n71>.
- Popay, Jennie, Helen Roberts, Amanda Sowden, dkk. *Guidance on the Conduct of Narrative Synthesis in Systematic Reviews*. ESRC Methods Programme, 2006. <https://www.lancaster.ac.uk/media/lancaster-university/content-assets/documents/fhm/dhr/chir/NSsynthesisguidanceVersion1-April2006.pdf>.
- Ritonga, Mahyudin, Hendro Widodo, Munirah Munirah, dan Talqis Nurdianto. "Arabic language learning reconstruction as a response to strengthen Al-Islam studies at higher education." *International Journal of Evaluation and Research in Education (IJERE)* 10, no. 1 (2021). <https://doi.org/10.11591/ijere.v10i1.20747>.
- Ritonga, Mahyudin, Hendro Widodo, Munirah Munirah, dan Talqis Nurdianto. "Arabic language learning reconstruction as a response to strengthen Al-Islam studies at higher education." *International Journal of Evaluation and Research in Education (IJERE)* 10, no. 1 (2021): 355. <https://doi.org/10.11591/ijere.v10i1.20747>.
- Rusli, Radif Khotamir, Abraham Yazdi Martin, Zahra Khusnul Lathifah, Nova Monaya, dan Mukarama Yakşi. "The Impact of Informative and Narrative Texts on Arabic Language Comprehension in Islamic Education." *International Journal of Language Education* 8, no. 4 (2025). <https://doi.org/10.26858/ijole.v8i4.70009>.
- Solimando, Cristina. "Language Tests in the Arabic-speaking World: Between Ideology and Language Policy." *Lingue Culture Mediazioni - Languages Cultures Mediation (LCM Journal)* 8, no. 2 (2022). <https://doi.org/10.7358/lcm-2021-002-soli>.
- Stern, Cindy, Zoe Jordan, dan Alexa McArthur. "Developing the Review Question and Inclusion Criteria." *AJN, American Journal of Nursing* 114, no. 4 (2014). <https://doi.org/10.1097/01.NAJ.0000445689.67800.86>.
- Stufflebeam, Daniel L. "The CIPP Model for Program Evaluation." Dalam *Evaluation Models*, disunting oleh George F. Madaus, Michael Scriven, dan Daniel L. Stufflebeam. Kluwer-Nijhoff Publishing, 1983.
- Syukriyah Mohd Suib, Nur, dan Harun Baharudin. "Effective Strategies for Teaching Arabic Prosody in Oral Communication Skills." *Qubahan Academic Journal* 5, no. 1 (2025). <https://doi.org/10.48161/qaj.v5n1a855>.
- Zulhannan, Guntur Cahaya Kesuma, Umi Musyarrofah, dan Yuli Andriansyah. "Enhancing Communicative Arabic Teaching: Evaluating the Al-Arabiyyah Baina Yadaik Model." *Journal of Educational and Social Research* 15, no. 3 (2025). <https://doi.org/10.36941/jesr-2025-0100>.
- Zurqoni, Zurqoni, Heri Retnawati, Syarifatur Rahmatullah, Hasan Djidu, dan Ezi Apino. "Has Arabic Language Learning Been Successfully Implemented?" *International Journal of Instruction* 13, no. 4 (2020). <https://doi.org/10.29333/iji.2020.13444a>.