ARKHAS: Journal of Arabic Language Teaching
Vol. 5 No. 2 (2025) : 337-354
Available online at https://jurnalpasca.uinkhas.ac.id/index.php/ARKHAS/

تحليل بنود الأسئلة مهارات اللغة العربية في المدرسة الابتدائية الإسلامية طريق الهدى مالانج: دراسة حالة

Analysis of Arabic Language Skills Test Items at Al-Huda Islamic Elementary School, Malang: A Case Study

Putri Damara Chaniago[1], Ziana Walida S[2], Yessita Amanda Putri[3], Nur Qomari [4]
[1,2,3,4]Universitas Islam Negeri Maulana Malik Ibrahim Malang
Email: putridamaraa@gmail.com[1], ziyanawalida97@gmail.com[2], yessitaamanda410@gmail.com[3], qomari@uin-malang.ac.id[4]

Article Information:

Received June 30, 2025
Revised December 23, 2025
Accepted December 25, 2025
Published December 27, 2025

Keywords: *Learning evaluation, test items, Arabic language skills, test validity, case study*

Abstract:
Evaluation is the process of giving a value to an object based on specific criteria. This study aims to analyze the quality of Arabic language skills test items at SD Islam MI Thoriqul Huda Malang and assess their suitability with the principles of compiling language skills test items. This study uses a qualitative approach with a case study design. Primary data sources include Arabic language skills test documents, while secondary data are in the form of interviews with teachers and other supporting documents. Data collection was carried out through documentation and interviews, and analyzed using the Miles and Huberman model (data reduction, data presentation, and conclusion). The study's results showed that the test items covered four main language skills: listening, speaking, reading, and writing. The test items were presented in various forms, including multiple-choice questions, essay questions, and performance tasks. The questions were compiled by the teacher with reference to the learning objectives and student abilities, and using a question bank as a reference. Daily assessments were conducted separately for each skill, whereas summative assessments typically combined all skills into a single test that focused more on reading and writing. Judging from the assessment principles, the test items were generally valid and relevant; however, there were shortcomings in evaluating speaking and listening skills due to limited resources available. This study recommends refining evaluation tools, increasing the clarity of instructions, and aligning test formats to measure all language skills optimally.

Correspondent Author: putridamaraa@gmail.com (Putri Damara Chaniago)
How to cite: Chaniago, P. D., S, Z. W., Putri, Y. A., & Qomari, N. (2025). تحليل بنود الأسئلة مهارات اللغة العربية في المدرسة الابتدائية الإسلامية طريق الهدى مالانج: دراسة حالة. Journal of Arabic Language Teaching, 5(2), 337–354. https://doi.org/10.35719/arkhas.v5i2.2366
Publisher: Arabic Language Education Department, Postgraduate of UIN Kiai Haji Achmad Siddiq Jember

المقدمة

تعد عملية تقييم المهارات اللغوية في المرحلة الابتدائية من الركائز الأساسية لقياس فاعلية المنظومة التعليمية، خاصةً في تعليم اللغات الأجنبية مثل اللغة العربية التي تُدرَّس في العديد من المؤسسات

الإسلامية (Baroroh & Hamani, 2022). ومع ذلك، تشير العديد من التقارير الحديثة إلى وجود خلل واضح في توزيع الاختبارات بين المهارات الأربع (الاستماع، والتحدث، والقراءة، والكتابة)، حيث تُهمَّل مهار الاستماع والتحدث غالباً لصالح التقييمات الكتابية التقليدية (Li & Zhang, 2021). ويتجلّى هذا التحوّل غير المتوازن في اتجاهات التقييم الحالية، التي لا تزال تركز على القياس المعرفي على حساب الكفاءات التواصلية العملية، مما يُضعف من قدرة الطلاب على استخدام اللغة في سياقات حقيقية (Fulcher, 2012). على سبيل المثال، أظهرت دراسة أجريت في مدرسة إسلامية ندونيسيا أن ٨٠٪ من أسئلة الاختبارات النهائية اقتصرت على مهارتي القراءة والكتابة، بينما ندرت الأسئلة المخصصة لتقييم المحادثة والاستيعاب السمعي (Al-Rawafi et al., 2021). وتُعدُّ هذه الظاهرة جزءاً من تحوّل أوسع في سياسات التقييم التربوي، التي تتجه نحو توحيد الأدوات لسهولة التصحيح، لكنها تتناقض مع المبادئ الحديثة التي تؤكد على التقييم الشامل والمتوازن (Wong et al., 2020). كما أن محدودية الموارد والبنية التحتية في كثير من المدارس، مثل نقص المعامل الصوتية وغرف المحادثة، تُفاقم من هذه الإشكالية وتحدّ من قدرة المعلمين على تطبيق تقييمات أصيلة (Dobrinić et al., 2025). لذلك، تبرز الحاجة الملحة لإجراء دراسات تحليلية عميقة لواقع بنود الاختبارات، كالدراسة الحالية، لاكتشاف مواطن القوة والضعف واقتراح حلول عملية قابلة للتطبيق، كدمج الوسائط المتعددة وتدريب المعلمين على بناء أدوات تقييمية شاملة (Kaya & Adiguzel, 2021). وعليه، يسهم هذا البحث في سدّ فجوة تطبيقية من خلال تشخيص واقع تقييم المهارات العربية في مدرسة "طريق الهدى"، سعياً نحو تحويل اتجاه التقييم إلى نموذج أكثر تكاملاً ومواءمةً لطبيعة اكتساب اللغة.

تستند هذه الدراسة إلى إطار نظري موسع في مجال تقييم تعلم اللغة العربية، حيث تؤكد الأدبيات الحديثة على ضرورة تصميم أدوات تقييم شاملة تغطي المهارات اللغوية الأربع: الاستماع، والتحدث، والقراءة، والكتابة (Rakhlin et al., 2021). وقد كشفت العديد من الدراسات السابقة عن وجود خلل في توزيع هذه المهارات داخل الاختبارات، حيث تميل إلى التركيز على القراءة والكتابة على حساب المهارات الشفهية (Soliman & Khalil, 2024). على سبيل المثال، وجدت دراسة أجراها عبد (٢٠٢٢) أن ٨٠٪ من أسئلة الاختبارات النهائية في بعض المدارس الابتدائية في إندونيسيا ركزت على الفهم القرائي والنحوي، بينما تم تهميش تقييم الكلام والاستماع بشكل فعلي. كما أشارت دراسة حديثة إلى أن معايير الصدق والموثوقية في بناء بنود الأسئلة لا تزال غير مطبقة بشكل كافٍ في السياقات التعليمية غير الناطقة لعربية (Almelhes, 2024). ومع ذلك، فإن غالبية الأبحاث السابقة ركزت على

التحليل الكمي للبنود ستخدام البرامج الإحصائية، مما أغفل الجانب النوعي المتعلق بملاءمة السياق التعليمي والقيود العملية التي تواجه المعلمين (Aloudah, 2022). لذلك، تهدف هذه الدراسة إلى سد هذه الفجوة من خلال توفير تحليل نوعي معمق لبنود أسئلة المهارات في بيئة مدرسية محددة، مع الأخذ في الاعتبار العوامل المؤسسية واللوجستية التي تؤثر على جودة التقييم، وهو جانب لم تحظَ به الأبحاث السابقة إلا بقدر محدود.

تحليل الفجوة: تقدم هذه الدراسة إسهاماً فريداً من خلال تركيزها على تحليل بنود أسئلة تقييم المهارات اللغوية العربية المتكاملة في سياق مدرسة ابتدائية إسلامية في إندونيسيا، وهو سياق بحثي لا يزال محدود التغطية في الأدبيات الدولية (Sukenti et al., 2020). أولاً، تملأ الدراسة فجوة تطبيقية من خلال تقييم كيفية ترجمة المبادئ النظرية لبناء الاختبارات (كالصدق والموثوقية) إلى ممارسات فعلية في الفصول الدراسية ذات الموارد المحدودة، مما يوفر رؤية واقعية تتجاوز الدراسات التي تركز على التحليل الكمي المجرد للبنود (Dunn & McCray, 2020). نياً، تسلط الضوء بشكل نقدي على التناقض بين النظرية والتطبيق في تقييم المهارات الشفهية (الاستماع والتحدث)، حيث تكشف أن التقييمات الختامية في المدرسة المدروسة ما تزال تركّز على المهارات الكتابية على الرغم من الإقرار همية المهارات الأربع، مما يعكس تحد ً منهجياً شائعاً في تعليم اللغات في البيئات المشابهة (Sukma et al., 2023). على سبيل المثال، بينما تؤكد الأدبيات على ضرورة استخدام مهام أدائية أصيلة لتقييم التحدث، توثق الدراسة قيود البنية التحتية التي تحول دون ذلك في الميدان (Hidayat et al., 2021). كما أن تحليل نموذج لبنود الأسئلة الفعلية المستخدمة في الصف الخامس يقدم دليلاً ملموساً على نقاط القوة والضعف في الصياغة العملية. و لتالي، لا تقتصر مساهمة هذه الدراسة على التشخيص الأكاديمي، بل تمتد إلى تقديم توصيات عملية قابلة للتطبيق لتحسين التوازن في التقييم اللغوي الشامل داخل المؤسسات التعليمية التي تشترك في ذات السياق والموارد.

تكتسي هذه الدراسة همية علمية وعملية لغة، إذ تسعى إلى تحليل جودة بنود أسئلة تقييم المهارات اللغوية العربية في مرحلة تعليمية أساسية، حيث تشير العديد من التقارير إلى وجود خلل في توازن تقييم المهارات اللغوية الأربع، مع تركيز غالب على القراءة والكتابة على حساب الاستماع والتحدث (Kremmel & Harding, 2020). تي هذا البحث استجابة لحاجة ماسة لتطوير أدوات تقييم أصيلة وشاملة في تعليم اللغة العربية، خاصة في المؤسسات الإسلامية التي تهدف إلى إتقان اللغة لأغراض التواصل والفهم الديني (Fidayani & Ammar, 2023). فمثلاً، أظهرت دراسة سابقة في مدرسة إسلامية

أن ٧٠٪ من أسئلة الاختبارات النهائية تركزت على الجانب الكتابي، مما يعكس فجوة تقييمية تؤثر على نمو الكفاءة التواصلية الشاملة لدى المتعلم (Thabtah et al., 2020). علاوة على ذلك، يسلط البحث الضوء على التحد ت العملية التي تواجه المدرسين، مثل نقص الوسائل السمعية والبصرية اللازمة لتقييم مهارتي الاستماع والتحدث بشكل فعال، وهي مشكلة متكررة في العديد من البيئات التعليمية ذات الموارد المحدودة (Yu et al., 2021). وعليه، فإن تحليل البنود الاختبارية في مدرسة "طريق الهدى" يمثل حالة دراسية نوعية يمكن أن تكشف عن الثغرات بين المبادئ النظرية للتقييم الجيد والممارسة الفعلية، مما يوفر أساساً لتقديم توصيات قابلة للتطبيق لتحسين جودة التقويم اللغوي وتطويره بما يتماشى مع المعايير التربوية الحديثة التي تؤكد على تقييم جميع المهارات بشكل متكامل.

تهدف هذه الدراسة إلى تحليل جودة بنود أسئلة تقييم المهارات اللغوية العربية المستخدمة في المدرسة الابتدائية الإسلامية طريق الهدى مالانج، وذلك من خلال تقييم مدى توافقها مع المبادئ العلمية لوضع الأسئلة اللغوية مثل الصدق والثبات ووضوح التعليمات وملاءمة شكل السؤال (Sukenti et al., 2020). يبرر هذا الهدف الحاجة الملحة لتحسين أدوات التقييم في المؤسسات التعليمية الإسلامية، حيث أشارت دراسات سابقة إلى وجود قصور في تقييم مهارتي الاستماع والتحدث مقارنة بمهارتي القراءة والكتابة في الاختبارات التجميعية (Ismail et al., 2022). على سبيل المثال، وجدت دراسة مماثلة أن ٧٠% من أسئلة الامتحان النهائي ركزت على القواعد والمفردات الكتابية، بينما تم تخصيص ١٠% فقط للأنشطة الشفهية (Essam et al., 2024). كما أن محدودية الموارد والمرافق في كثير من المدارس تعيق التطبيق العملي للتقييم الشامل للمهارات الأربع (Coombe et al., 2020). لذا، تفترض هذه الدراسة أن بنود الأسئلة في المدرسة المبحوثة تتمتع بصدق محتوى جيد وموثوقية مقبولة فيما يتعلق بمهارتي القراءة والكتابة، ولكنها تظهر ضعفاً في تقييم مهارتي الاستماع والتحدث بشكل متوازن وفعّال، مما يستدعي تطوير أدوات تقييم أصلية وتوفير بيئة داعمة (Rakhlin et al., 2021). وبناءً على ذلك، تسعى الدراسة للإجابة على سؤال رئيسي: كيف يمكن تحسين جودة بنود أسئلة تقييم المهارات اللغوية العربية لقياس المهارات الأربع بشكل متكامل ومتوازن في البيئة المدرسية الحالية؟

منهج البحث

تستخدم هذه الدراسة المنهج النوعي (الكيفي) بنوع دراسة الحالة. وقد تم اختيار هذا المنهج لأن الباحث يرغب في وصف وتحليل بنود الأسئلة مهارات اللغة العربية المستخدمة في مدرسة ثوريق الهدى الابتدائية الإسلامية بمدينة مالانج بشكل عميق. يتميز البحث النوعي بطابعه الوصفي ويركز

على المعنى الكامن وراء البيا ت التي تُدرس في سياقها. وقد تم اختيار دراسة الحالة لأن تركيز البحث ينحصر على موقع واحد وموضوع معين يتم دراسته بشكل مكثف (Assyakurrohim et al., 2022) تم إجراء هذا البحث في مدرسة ثوريق الهدى الابتدائية الإسلامية بمدينة مالانج. أما موضوع البحث فهو معلم مادة اللغة العربية، لإضافة إلى و ئق أسئلة الاختبارات المستخدمة في تقييم مهارات اللغة العربية لدى الطلاب.

مصادر البيا ت تتكون مصادر البيا ت في هذا البحث من البيا ت الأولية، وهي و ئق فقرات أسئلة اختبارات مهارات اللغة العربية التي يستخدمها المعلمون في تلك المدرسة. أما البيا ت الثانوية فتشمل نتائج المقابلات مع معلم اللغة العربية، لإضافة إلى الو ئق الداعمة مثل خطة تنفيذ التعلم (RPP)، والمنهج الدراسي، أو جدول المواصفات (كيسي-كيسي) ذات الصلة. وقد استخدم الباحث تقنيتين رئيسيتين في جمع البيا ت، وهما: التوثيق، لجمع أسئلة اختبارات مهارات اللغة العربية التي استُخدمت سابقًا، ثم المقابلة التي أُجريت مع معلم اللغة العربية لمعرفة عملية إعداد الأسئلة، والمبادئ المستخدمة، والصعو ت التي تواجه في تقييم مهارات اللغة.

تحليل البيا ت ستخدام تقنية تحليل نموذج مايلز وهوبرمان، والتي تتكون من ثلاث مراحل، وهي: اختزال البيا ت، أي فرز البيا ت، واختيارها، وتلخيص البيا ت المهمة من و ئق الأسئلة ونتائج المقابلات (Saleh, 2017). عرض البيا ت (عرض البيا ت) يتم من خلال تنظيم البيا ت في شكل سردي، أو جداول، أو اقتباسات من الأسئلة. واستخلاص الاستنتاجات والتحقق منها يتم بهدف استنتاج جودة الأسئلة بناءً على مبادئ إعداد أسئلة مهارات اللغة العربية. وقد تم إجراء التحليل بمقارنة الأسئلة الموجودة مع النظر ت والمبادئ المثالية في إعداد أسئلة المهارات اللغوية، مثل الصدق، والثبات، ومناسبة شكل السؤال، وقابلية قياس الكفاءة.

نتائج البحث والمناقشة

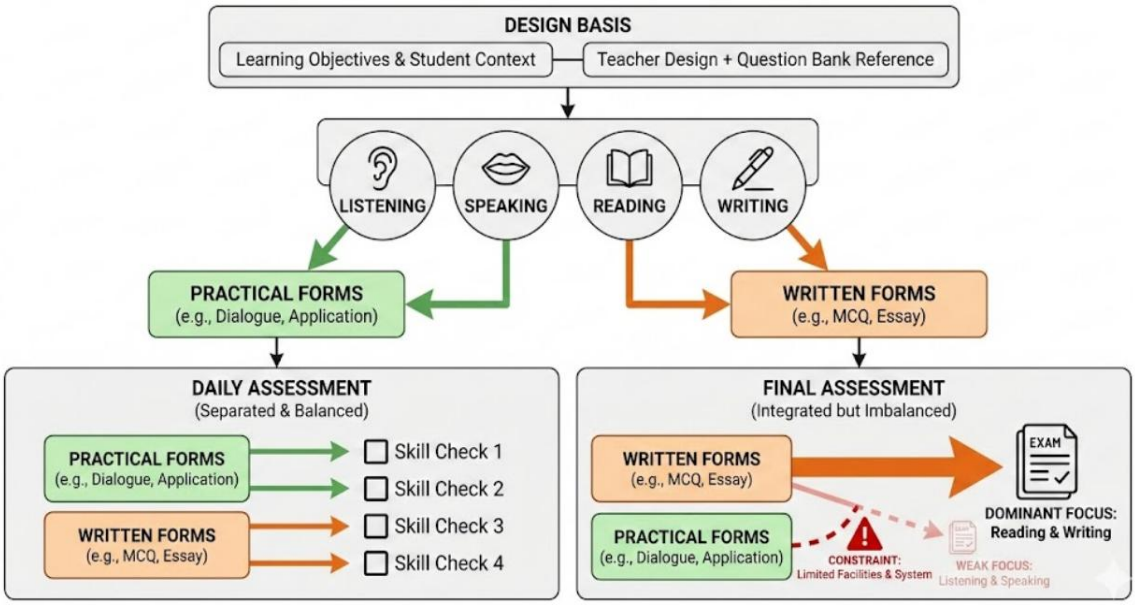
أشكال أسئلة مهارات اللغة المستخدمة في مدرسة الابتدائية الطاريق الهدى بمدينة مالانج

تُعَدُّ التقييمات جزءًا مهمًّا من عملية التعليم، بما في ذلك تعليم اللغة العربية. وللحصول على معلومات دقيقة حول مدى تمكّن الطلاب من المهارات اللغوية، ينبغي على المعلم تصميم أشكال من الأسئلة تتناسب مع خصائص كل مهارة. لذلك، سيتم في هذا القسم عرض نتائج البحث الميداني حول أشكال الأسئلة المستخدمة في تقييم مهارات اللغة العربية في مدرسة "طريق الهدى" الابتدائية الإسلامية بمدينة مالانج، والتي تشمل مهارات الاستماع، والكلام، والقراءة، والكتابة، لإضافة إلى

كيفية تكييف المعلم لهذه الأسئلة مع أهداف التعلم وظروف الطلاب وتطبيقها في الصف. وقد تم الحصول على هذه النتائج من خلال مقابلة مع معلم مادة اللغة العربية الذي يشارك مباشرة في إعداد وتنفيذ التقييم. وبناءً على نتائج المقابلة مع معلم اللغة العربية في مدرسة "طريق الهدى"، تبيّن أن المهارات اللغوية التي يتم تقييمها تشمل أربع مهارات رئيسية، وهي الاستماع، والكلام، والقراءة، والكتابة. ويتم تنفيذ تقييم هذه المهارات بما يتوافق مع الأهداف التعليمية التي حددها المعلم مسبقًا.

الجدول ١

نظرة عامة على تقييم اللغة العربية في مدرسة طارق الهدى (مالانج)



أشكال أسئلة مهارات اللغة المستخدمة في مدرسة طريق الهدى الإسلامية الابتدائية مالانج أوضح المعلم، بناءً على نتائج المقابلة، أن مهارات اللغة العربية التي يتم تقييمها في المدرسة تشمل أربع مهارات رئيسية: الاستماع (استماع)، والكلام (كلام)، والقراءة (قراءة)، والكتابة (كتابة). وتُختبر هذه المهارات بما يتماشى مع أهداف التعلم، مما يدل على مرونة ووعي المعلم همية إتقان المهارات اللغوية بشكل شامل. وهذا يتفق مع النظر ت التي تؤكد أن المهارات اللغوية ينبغي تقييمها وفقًا لطبيعة النشاط التواصلي الخاص بها(Abd-Elmoneim et al., 2023) يرى "براون" أن هذه المهارات الأربع تُعدُّ ركائز رئيسية في إتقان اللغة، ويجب تقييم كل مهارة منها بطرق مختلفة تتناسب مع خصائصها.

أما عن أشكال الأسئلة المستخدمة فهي متنوعة، وتتناسب مع نوع المهارة اللغوية المراد تقييمها وأهداف التعلم المنشودة. وعادةً ما تُستخدم أسئلة من نوع الاختيار من متعدد، والأسئلة المقالية، لإضافة إلى التطبيق العملي، لا سيما في مهارتي الكلام والاستماع. وتُظهر البيا ت التي تم الحصول عليها من المعلم أن المعلم يكيّف شكل السؤال حسب نوع المهارة اللغوية التي يرغب في قياسها. وتنوّع

أشكال الأسئلة بين الاختيار من متعدد والمقالي والتطبيقي يدل على فهم لخصائص كل مهارة لغوية. ويتفق هذا مع رأي "ويديويوكو" (٢٠١٥) الذي يرى أن التقييم الجيد هو الذي يعكس عملية التعلم ونتائجها بشكل أصيل، وليس مجرد تقييم معرفي على الورق (Puad & Ashton, 2023).

أما فيما يخص تنفيذ أسئلة مهارات اللغة العربية في المدرسة، فإن التقييم اليومي يتم فصل كل مهارة فيه، بينما يُدمج التقييم النهائي الفصلي بين المهارات. ويُظهر ذلك وجود عدم توازن في تنفيذ التقييم، خاصة في التقييم الختامي، حيث ما تزال الأسئلة المستخدمة تهيمن عليها الأشكال الكتابية، مثل الاختيار من متعدد والمقالية، وتميل إلى قياس مهارتي القراءة (القراءة) والكتابة (الكتابة) فقط. في حين أن مهارتي الاستماع والكلام لا تزالان غير مفعَّلتين لشكل الكافي، رغم أن النظر ت تُبيّن أن مهارة الاستماع هي القدرة على فهم المعنى من خلال عملية استماع نشطة وموجَّهة، وأن مهارة الكلام هي القدرة على التعبير عن الأفكار شفهيًا، ويجب تقييمها من خلال التطبيق العملي مثل الحوار أو الخطاب (Qiao & Zhao, 2023).

تُعدُّ أسئلة مهارات اللغة العربية في مدرسة "طريق الهدى" الابتدائية من إعداد معلمي المدرسة أنفسهم، حيث يطبقون منهجًا مرً في إعداد الأسئلة. وتُصمَّم الأسئلة بشكل ذاتي (أصيلة)، مع الاستفادة أيضًا من بنك الأسئلة كمرجع. وتُعدُّ هذه الممارسة من الأساليب الموصى بها في التقييم الحديث، لأن بنك الأسئلة يوفر التنوع، والمعايير، والكفاءة، ومع ذلك لا بد من تكييفه مع السياق وخصائص الطلاب. إن استخدام بنك الأسئلة إلى جانب تطوير الأسئلة ذاتيًا يوفّر فرصة لبناء أدوات تقييم صادقة و بتة وتلبي احتياجات الطلاب (Pittman et al., 2023)، ومع ذلك، أكد المعلم أيضًا على أن تكييف الأسئلة مع أهداف التعلم أمر لغ الأهمية، لتجنب عدم الاتساق بين المحتوى والنواتج التعليمية. وبمعنى آخر، فإن معلمي مدرسة "طريق الهدى" قد أظهروا فهمًا لأهمية سياقية الأسئلة، مع الحفاظ في الوقت ذاته على المعايير من خلال الرجوع إلى بنك الأسئلة.

وقد ثبت تنوع أدوات اختبار مهارات اللغة العربية في المدرسة من خلال تنفيذها بما يتماشى مع أهداف التعلم. كما أُتيحت للمتعلمين فرصة للتطبيق العملي داخل الصف. وعلاوة على ذلك، أقرّ المعلم ن محدودية المرافق ونظام المدرسة تمثل عوائق في تطبيق تقييم مهارات الكلام والاستماع. وهذا يؤكد أن التقييم لا يُعدّ مجرد أداة لقياس النتائج، بل هو أيضًا وسيلة لتحسين جودة التعليم وتوفير الوسائل، كما هو مذكور في النظرية التي تؤكد أن للتقييم دورًا استراتيجيًا في تحسين الخدمات التعليمية وتطوير المنهج (Golden & Kohlbeck, 2020).

وبناءً على نتائج المقابلة المتعلقة شكال أسئلة مهارات اللغة المستخدمة في مدرسة "طريق الهدى" في مالانج، فإن تقييم مهارات اللغة العربية يشمل أربع مهارات رئيسية: الاستماع، والكلام، والقراءة، والكتابة. وتتنوع أشكال الأسئلة المستخدمة، مثل الاختيار من متعدد، والأسئلة المقالية، والتطبيق العملي، وفقًا لنوع المهارة وأهداف التعلم. يتم إعداد الأسئلة من قبل المعلم استنادًا إلى أهداف التعلم والمحتوى وظروف الطلاب، مع الرجوع إلى بنك الأسئلة كمصدر مرجعي. يُنفذ التقييم اليومي بشكل منفصل لكل مهارة، بينما يميل التقييم الختامي إلى دمج المهارات في اختبار واحد يركّز أكثر على مهارتي القراءة والكتابة. أما مهار الاستماع والكلام فلا تزالان غير مفعّلتين لشكل الكافي، خاصة في التقييمات الختامية، وذلك بسبب محدودية المرافق والنظام. وبوجه عام، فإن إعداد وتنفيذ التقييمات يتماشى مع مبادئ التقييم التربوي، خاصة من حيث صدق المحتوى وملاءمة شكل الأسئلة. ومع ذلك، لا بد من تعزيز الجانب المتعلق لبنية التحتية وتكامل النظام لضمان تقييم جميع المهارات بشكل متوازن وأصيل.

تحليل مبادئ صياغة أسئلة مهارات اللغة العربية

في صياغة أسئلة مهارات اللغة العربية، هناك بعض المبادئ الأساسية التي يجب مراعاتها لضمان فعالية التقييم وقياس قدرات الطلاب بشكل موضوعي. فيما يلي تحليل لهذه المبادئ:

الصدق (الصلاحية)

الصدق هو مقياس يدل على مدى قدرة فقرة الاختبار على قياس ما يُراد قياسه وفقًا لأهداف التعلم أو الكفاءة الأساسية. فالسؤال الصادق يعني أن السؤال دقيق وملائم في قياس قدرة أو معرفة المتعلم. ويعكس الصدق أيضًا دقة القياس وملاءمة محتوى السؤال للمادة التي يتم تدريسها (Dianova & Anwar, 2024). ويمكن اختبار صدق فقرة الاختبار تجريبيًا من خلال النظر في العلاقة الارتباطية بين درجات فقرة الاختبار والدرجة الكلية للاختبار، فكلما زادت العلاقة الارتباطية، زادت صلاحية الاختبار (Ni et al., 2024). وفيما يلي البيا ت التي يمكن استخلاصها من نتائج المقابلة في المدرسة الابتدائية الإسلامية طريق الهدى مالانج.

بناءً على نتائج المقابلة، أوضح المعلم أن المعلم عند إعداد الأسئلة يعتمد على المنهج الدراسي المعتمد، وكذلك نتائج التدريب الذي حصل عليه المعلم. وتعتبر هذه الإرشادات أساسية لضمان أن الأسئلة تظل في إطار التعلم المعياري. وهذا يتوافق مع مبدأ الصدق حسب ما تم ذكره أعلاه، ويدعمه النظرية التي تشترط التوافق بين أهداف التعلم ومحتوى السؤال، كما أوضحه أريكونتو، ويؤكده مبدأ

مطابقة المادة عند إعداد أسئلة اللغة العربية بحيث تكون مناسبة للمؤشرات والكفاءات (Ashfia & Ridlo, 2024)

الموثوقية

الموثوقية تشير إلى مدى ثبات أو استقرار نتائج القياس من أداة الاختبار. فالموثوقية العالية تعني أن الأسئلة تنتج نتائج متسقة إذا تم اختبارها بشكل متكرر أو تقييمها من قبل مقيمين مختلفين. ومع ذلك، فإن الموثوقية العالية لا تضمن لضرورة الصدق، لأنه من الممكن أن تكون الأسئلة متسقة ولكنها لا تقيس الجانب المناسب (Dianova & Anwar, 2024).

بناءً على نتائج المقابلة، أوضح المعلم أن المعلم يسعى إلى مواءمة مستوى صعوبة الأسئلة مع قدرات الطلاب، حتى يتمكن جميع الطلاب من فهم الأسئلة والإجابة عليها بشكل جيد، دون الشعور لإرهاق المفرط. كما يقوم المعلم أيضًا جراء تقييم أو مراجعة الأسئلة بعد تنفيذ الاختبار. وهذا يتوافق مع النظرية التي تعكس الجهود لز دة ثبات نتائج القياس، حسب تعريف الموثوقية من قبل براون ومبدأ الثبات في تطوير اختبارات اللغة العربية (Muhammad Taufiq Ismail, 2016). كما أن تقييم ومراجعة الأسئلة خطوة مهمة للحفاظ على الموثوقية، حيث يهدف التقييم إلى تحديد نقاط ضعف فقرات الاختبار، مثل مستوى الصعوبة غير المناسب أو التعليمات غير الواضحة، حتى يمكن إجراء المراجعة لتحسين جودة الأسئلة (Dinanti, 2024).

وضوح التعليمات

وضوح التعليمات يعني مدى وضوح وسهولة فهم التوجيهات أو الإرشادات المقدمة في السؤال. فالتعليمات الواضحة مهمة جدًا حتى يفهم المتعلم ما هو مطلوب منه دون ارتباك. والتعليمات غير الواضحة قد تؤدي إلى سوء الفهم وتؤثر على نتائج قياس قدرة المتعلم. كما يشمل وضوح التعليمات استخدام لغة بسيطة ومباشرة تخدم هدف السؤال (Ummah, 2019).

بناءً على نتائج المقابلة، أكد المعلم أن تعليمات الأسئلة تتوافق مع أهداف التعلم. فعلى سبيل المثال، إذا كان السؤال يهدف إلى قياس مهارة الكلام، فإن السؤال مصمم لفعل لتقييم قدرة المتعلم على الكلام، وليس فقط معرفته عن الكلام. وهذا يتوافق مع مبدأ وضوح التعليمات الذي يتطلب لغة بسيطة وغير مبهمة، كما أوضحه هيوز وقاعدة بناء السؤال التي تتطلب أن يكون السؤال موجزًا وواضحًا وحازمًا (Ramadhan & Firdaus, 2022).

ملاءمة شكل السؤال

ملاءمة شكل السؤال تتعلق بنوع وتنسيق السؤال المستخدم في قياس كفاءة معينة. فيجب أن يكون شكل السؤال مناسبًا لهدف القياس والمادة التي سيتم اختبارها. على سبيل المثال، الأسئلة الاختيارية من متعدد مناسبة لقياس المعرفة الواقعية، بينما الأسئلة المقالية أكثر ملاءمة لقياس القدرة على التحليل أو التعبير الكتابي. كما تضمن ملاءمة شكل السؤال أن الأسئلة تستوعب مستو ت الصعوبة والقدرات المختلفة للمتعلمين (Nuswowati et al., 2010).

بناءً على نتائج المقابلة، اعترف المعلم بوجود قيود في البنية التحتية، و لنسبة للتقويم اليومي (PH)، فإن الأسئلة المستخدمة تعتبر فعالة لأنها قادرة على قياس المهارات التي يتم تدريسها بشكل مفصل. أما لنسبة للتقويم النهائي (الاختبارات النهائية)، فإن فعاليتها ما زالت محدودة، لأن المهارات التي يتم قياسها تقتصر على القراءة والكتابة، بينما مهارات الاستماع والكلم لم تُستوعب بشكل كافٍ. وهذا يتوافق مع مبدأ أن شكل السؤال يجب أن يكون مناسبًا للمهارة التي يتم قياسها (Ashfia & Ridlo, 2024) ومع ذلك، لا تزال هناك معوقات في عدم ملاءمة شكل السؤال لمهارات الكلام والاستماع، والتي تتطلب وسائط خاصة مثل الصوت أو الممارسة المباشرة.

جدول ١

تحليل فقرة السؤال بناءً على مبادئ صياغة بنود الأسئلة

التقييم	التطبيق على فقرة السؤال	الجوانب الفرعية	المبدأ الرئيسي
مناسب	يقيس السؤال القدرة على ترجمة جملة بسيطة إلى العربية، وهو هدف تعليمي مناسب لمستوى الصف الخامس الابتدائي.	صلاحية المحتوى	**الصدق (الصلاحية)**
مناسب	البدائل (أ، ب، ج، ح) متشابهة في البناء وتتضمن مشتتات شائعة (كلمات: مَجَلَّةٌ، مَكْتَبٌ، غُرْفَةُ الْمُذَاكَرَةِ، غُرْفَةُ الْإِسْتِقْبَالِ)، مما يجعلها تقيس الفهم الدقيق للمفردات والتركيب.	صلاحية البناء	
عالي	السؤال موضوعي (اختيار من متعدد جابة واحدة صحيحة)، لذا فإن نتائج التطبيق المتكرر لنفس المجموعة ستكون متسقة في الغالب.	الاتساق	**الثبات (الموثوقية)**

عالي	لا يوجد مجال للتفسير الذاتي في التصحيح. الإجابة صحيحة أو خاطئة بوضوح.	الموضوعية	
واضحة جداً	صيغة السؤال واضحة ومباشرة) :الصيغة العربية للجملة هي.(.....	البساطة والوضوح	**وضوح الإرشادات**
واضحة جداً	يعرف الطالب بوضوح أنه مطالب ختيار الترجمة الصحيحة للجملة المعطاة من بين البدائل الأربعة.	عدم الغموض	
ملائم جداً	شكل الاختيار من متعدد هو الأنسب لقياس مهارة الترجمة والتعرف على المفردات والتركيب الصحيح في هذا المستوى.	تناسب الشكل مع الهدف	**ملاءمة شكل السؤال**
ملائم جداً	الجملة بسيطة (في غرفة الدراسة هناك مجلة) والبدائل لا تتطلب تفكيراً معقداً، بل دقة في الملاحظة، مما يناسب طلاب الصف الخامس الابتدائي.	تناسب مستوى الصعوبة	

الملاحظة :فقرة السؤال هذه جيدة وتتوافق مع معايير صياغة البنود من حيث الصدق والثبات ووضوح الإرشادات وملاءمة شكل السؤال لمستوى الطلاب المستهدفين.

مراجعة موجزة عن الأبحاث ذات الصلة

تركز الأبحاث المتعلقة بتحليل جودة أسئلة اختبار مهارات اللغة العربية بشكل كبير على أهمية الصحة والموثوقية ووضوح التعليمات وملاءمة شكل الأسئلة كجوانب رئيسية في إعداد أدوات التقييم. على سبيل المثال، تشير بعض الدراسات إلى أن صحة محتوى الأسئلة يجب أن تستند إلى المناهج الدراسية وأهداف التعلم حتى تقيس الأسئلة لفعل المهارات المطلوبة، مثل التحدث والاستماع والقراءة والكتابة للغة العربية. لإضافة إلى ذلك، تتأثر موثوقية الأسئلة بشكل كبير بملاءمة مستوى الصعوبة مع قدرات الطلاب واتساق نتائج التقييم التي يمكن الحصول عليها من خلال التقييم والمراجعة الدورية للأسئلة.

تدعم هذه الدراسة أيضًا بحث محمد أ ندا موسى، "تحليل بنود أسئلة اللغة العربية في مدرسة متوسطة نوية نور بمدينة ريباري". ويؤكد هذا البحث على أهمية التحليل الشامل لبنود الأسئلة لتحسين جودة امتحا ت اللغة العربية. وتشمل الجوانب التي تم تحليلها صحة الأسئلة وموثوقيتها

وملاءمة محتواها وهيكلها ولغتها بحيث تتوافق مع أهداف التعلم وقدرات الطلاب. تدعم هذه الدراسة أهمية التقييم النوعي والكمي للأسئلة للتأكد من أن الأسئلة تقيس كفاءة الطلاب بدقة (موسى وموتيا، بدون ريخ). كما تدعمها دراسة يمنى حنيفة ''تحليل أسئلة اللغة العربية لطلاب الصف الحادي عشر في مدرسة إسلامية سيدوارجو ستخدام بر مج أ تيس''. استخدمت هذه الدراسة بر مج *Anates* لقياس صحة وموثوقية ومستوى صعوبة وقدرة التمييز وجودة الخداع في أسئلة اللغة العربية. وأظهرت النتائج أن معظم الأسئلة صحيحة وموثوقة، ولكن هناك بعض الأسئلة التي تحتاج إلى تحسين (Aalst, 2000).

المناقشة

تظهر النتائج أن تقييم مهارات اللغة العربية في المدرسة الابتدائية الإسلامية "طريق الهدى" مالانج يشمل المهارات الأربع الأساسية: الاستماع والكلام والقراءة والكتابة، وذلك عبر استخدام أشكال متنوعة من الأسئلة كالاختيار من متعدد والمقالي والأداء العملي، مما يعكس مراعاة لخصائص كل مهارة (Qodri & Sanjaya, 2024). وقد أعدت الأسئلة بناء على أهداف التعلم والقدرات الطلابية مع الاستعانة ببنك الأسئلة، مما يدل على سعي نحو المعيارية والملاءمة. ومع ذلك، لوحظ وجود تفاوت في التوازن التطبيقي؛ إذ نفذ التقييم اليومي لكل مهارة على حدة، بينما ركز التقييم التجميعي النهائي بشكل كبير على مهارتي القراءة والكتابة، مما أدى إلى تقييم غير كاف لمهارتي الاستماع والتحدث (Panadero et al., 2019). وبشكل عام، حققت الأسئلة مستو ت مقبولة من الصدق والموثوقية ووضوح التعليمات، لكن قيود البنية التحتية والوسائل شكلت عائقاً رئيسياً أمام تحقيق تقييم متكامل وعملي للمهارات الشفوية.

يعود سبب هذا التفاوت في التركيز بين المهارات في التقييمات النهائية إلى عدة عوامل مترابطة. أولاً، تعود الأسباب العملية إلى محدودية الإمكا ت المادية والتقنية المتاحة في المدرسة، مثل نقص أجهزة التسجيل الصوتي المناسبة أو الغرف المعزولة لتقييم المحادثة، مما يجعل تقييم المهارات الشفوية أكثر تعقيداً وتكلفة من الناحية اللوجستية (Su & Jiang, 2024). نياً، هناك سبب منهجي مرتبط بضغوط الوقت وسهولة التصحيح، حيث إن الأسئلة الكتابية (كالاختيار من متعدد) توفر للمعلم معياراً موضوعياً وسريعاً للتصحيح مقارنة مع تقييم الأداء العملي الذي يتطلب وقتاً أطول وحكماً ذاتياً أكبر (Darling-Hammond et al., 2020). وأخيراً، قد يعكس هذا الواقع ثيراً تراكمياً لمنهجية التدريس نفسها، التي

ربما تولي أولوية أقل للتطبيقات التواصلية الحية في الفصل الدراسي، مما ينعكس تلقائياً على أولو ت التقييم (Rahman et al., 2021).

تترتب على هذا القصور في تقييم المهارات الشفوية عدة عواقب تربوية. فعلى مستوى التحصيل الفردي، قد لا يحصل الطلاب على صورة دقيقة عن مستوى كفاءتهم التواصلية الحقيقية، مما يحد من فرص تطوير ثقتهم وقدرتهم على استخدام اللغة في مواقف حقيقية (Bella & Huda, 2022). وعلى مستوى المنهج التعليمي، قد تؤدي هذه الممارسة إلى تضييق نطاق التعلم، حيث يميل الطلاب والمعلمون على حد سواء إلى التركيز على الجوانب التي يتم تقييمها، وهي المهارات القرائية والكتابية، على حساب تنمية الطلاقة الشفوية والفهم السمعي (Jauharoh et al., 2021). وبشكل أوسع، يمكن أن يساهم ذلك في إنتاج متعلمين يتمتعون بمهارات غير متوازنة، قادرين على الأداء في الاختبارات النظرية ولكنهم يعانون من ضعف في التواصل الفعال، مما يقلل من تحقيق الغاية الأساسية من تعليم اللغة كأداة للحياة والتواصل (Coombe et al., 2020).

تتفق نتائج هذه الدراسة مع ما أشارت إليه عدة دراسات سابقة حول عدم التوازن في تقييم المهارات اللغوية، لكنها تظهر أيضاً بعض الفروق الدقيقة. فمثلاً، وجدت دراسة فكريه (٢٠٢١) أن أسئلة اللغة العربية في مرحلة مختلفة ركزت بشكل كبير على القراءة والكتابة، وهو ما يتوافق مع ما تم ملاحظته في هذه الحالة. ومع ذلك، بينما أشارت دراسة فراديسي وزملاؤه (٢٠٢٥) إلى وجود خلل في توزيع مستوى الصعوبة وتمييز الأسئلة كإشكالية رئيسية، فإن هذه الدراسة وجدت أن جودة البنود من حيث الصعوبة والتمييز كانت مقبولة، لتركيزها على القيود الهيكلية واللوجستية كعامل محوري (Graff Zivin et al., 2020). كما أن دراسة عاشفيا ورضلو (٢٠٢٤) أكدت على أهمية ملاءمة شكل السؤال للمهارة، وهو ما التزم به المعلمون في التقييمات اليومية لكنه تعثر في التقييمات الختامية بسبب العوائق المذكورة، مما يضيف بعداً تطبيقياً لفجوة بين التنظير والممارسة في سياقات محددة.

بناء على ما سبق، تقدم الدراسة عددا من التوصيات العملية لمعالجة القصور القائم. أولاً، على مستوى السياسة المدرسية والتخطيط، ينبغي تخصيص ميزانية لتعزيز البنية التحتية الداعمة للتقييم العملي، مثل تجهيز معمل لغة بسيط مزود جهزة تسجيل وسماعات لتقييم الاستماع، وتحديد مساحة مخصصة للمقابلات الشفوية *(Oral Proficiency Interviews)* لتقييم الكلام (Mohapatra & Laures-Gore, 2021). نياً، على المستوى المنهجي، يحتاج المعلمون إلى تدريب على تطوير وتصميم أدوات تقييم أصيلة (Authentic Assessment) للمهارات الشفوية، كاستخدام سجلات الملاحظة (Observation

Checklists) أثناء المناقشات الصفية أو مشاريع العروض التقديمية القصيرة، التي تقيس الكفاءة التواصلية في سياق ذي معنى (Wahyuni et al., 2021). وأخيراً، يقترح إجراء مراجعة دورية لمواصفات الاختبارات التجميعية (جدول المواصفات) لضمان تمثيل جميع المهارات الأربع بنسب عادلة تتناسب مع أهداف التعلم المعلنة، مما يجعل التقييم أداة دافعة لتعلم متوازن وشامل.

الخاتمة

أظهر هذا البحث أن عملية إعداد وتنفيذ تقييم مهارات اللغة العربية في مدرسة "ثوريق الهدى" الابتدائية الإسلامية بمدينة مالانج قد شملت المهارات الأربع الأساسية، وهي الاستماع، والتحدث، والقراءة، والكتابة. وقد تصميم شكل الأسئلة بما يتناسب مع خصائص كل مهارة وأهداف التعلم، كما تم إعدادها من قبل المعلم لاستناد إلى المنهج الدراسي وظروف الطلاب.

ويُجرى التقييم اليومي بشكل منفصل، بينما يغلب على التقييم التجميعي التركيز على مهارتي القراءة والكتابة فقط، مما يدل على أن مهارتي الاستماع والتحدث لم تُقيما بشكل كافٍ. ومن حيث مبادئ إعداد الأسئلة، فإن غالبية الأسئلة استوفت معايير الصلاحية، والموثوقية، ووضوح التعليمات، وملاءمة الشكل. ومع ذلك، فإن نقص الوسائل والأنظمة يُعد عائقًا أمام تحقيق تقييم متوازن وشامل لمهارات اللغة.

لذلك، من الضروري تعزيز وسائل التقييم كاستخدام الوسائط السمعية لمهارة الاستماع، وتوفير غرف للتدريب العملي لمهارة التحدث، إلى جانب تدريب المعلمين على إعداد أسئلة قائمة على المهارات اللغوية سلوب أصيل وتكاملي.

المصادر والمراجع

Aalst, J. van. (2000). An introduction to physics education research. *Canadian Journal of Physics*, *78*(1), 57–71. https://doi.org/10.1139/p00-005

Abd-Elmoneim, D. M., Ghandour, H. H., Elrefaie, D. A., & Khodeir, M. S. (2023). Development of an Arabic test for assessment of semantics for the Arabic-speaking children: the Arabic semantic test. *The Egyptian Journal of Otolaryngology*, *39*(1), 49. https://doi.org/10.1186/s43163-023-00405-3

Al-Rawafi, A., Sudana, D., Lukmana, I., & Syihabuddin, S. (2021). Students' apologizing in Arabic and English: An interlanguage pragmatic case study at an Islamic boarding school in Indonesia. *Indonesian Journal of Applied Linguistics*, *10*(3). https://doi.org/10.17509/ijal.v10i3.31740

Almelhes, S. (2024). Enhancing Arabic Language Acquisition: Effective Strategies for Addressing Non-Native Learners' Challenges. *Education Sciences*, *14*(10), 1116.

https://doi.org/10.3390/educsci14101116
Aloudah, N. M. (2022). Qualitative research in the Arabic language. When should translations to English occur? A literature review. *Exploratory Research in Clinical and Social Pharmacy*, *6*, 100153. https://doi.org/10.1016/j.rcsop.2022.100153
Ashfia, A., & Ridlo, U. (2024). *E-ISSN : 2792-0876 Optimalisasi Higher Order Thinking Skill (HOTS) dalam Kurikulum Merdeka : Strategi dan Konsep Penyusunan Soal Bahasa Arab di MTs Pembangunan Jakarta*. *5*(1), 330–342. https://doi.org/10.37274/mauriduna.v5i2.1189
Assyakurrohim, D., Ikhram, D., Sirodj, R. A., & Afgani, M. W. (2022). Case Study Method in Qualitative Research. *Jurnal Pendidikan Sains Dan Komputer*, *3*(01), 1–9.
Baroroh, U., & Hamani, T. (2022). Development of Authentic Assessment in Islamic Religious Education in Elementary School. *Nazhruna: Jurnal Pendidikan Islam*, *5*(3), 940–955. https://doi.org/10.31538/nzh.v5i3.2380
Bella, S., & Huda, M. M. (2022). The Use Of Youtube Media In Improving Listening And Speaking Skills In UIN Kiai Haji Achmad Siddiq Jember. *Journal of Arabic Language Teaching*, *2*(1), 43–56. https://doi.org/10.35719/arkhas.v2i1.1275
Coombe, C., Vafadar, H., & Mohebbi, H. (2020). Language assessment literacy: what do we need to learn, unlearn, and relearn? *Language Testing in Asia*, *10*(1), 3. https://doi.org/10.1186/s40468-020-00101-6
Darling-Hammond, L., Flook, L., Cook-Harvey, C., Barron, B., & Osher, D. (2020). Implications for educational practice of the science of learning and development. *Applied Developmental Science*, *24*(2), 97–140. https://doi.org/10.1080/10888691.2018.1537791
Dianova, F. R., & Anwar, N. (2024). Analisis Butir Uji Validitas, Reliabilitas, Tingkat Kesukaran, dan Daya Pembeda Soal Sumatif Bahasa Arab SD Islam. *Jurnal Bahasa Daerah Indonesia*, *1*(3), 13. https://doi.org/10.47134/jbdi.v1i3.2863
Dinanti, S. D. (2024). *Shaut Al- ' Arabiyah Analisis Butir Soal Bahasa Arab d i Madrasah Ibtida ' iyyah Bengkulu*. *12*(2), 518–530.
Dobrinić, D., Miler, M., & Medak, D. (2025). Mapping the Green Urban: A Comprehensive Review of Materials and Learning Methods for Green Infrastructure Mapping. *Sensors*, *25*(2), 464. https://doi.org/10.3390/s25020464
Dunn, K. J., & McCray, G. (2020). The Place of the Bifactor Model in Confirmatory Factor Analysis Investigations Into Construct Dimensionality in Language Testing. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.01357
Essam, M., Deif, M. A., & Elgohary, R. (2024). Deciphering Arabic question: a dedicated survey on Arabic question analysis methods, challenges, limitations and future pathways. *Artificial Intelligence Review*, *57*(9), 251. https://doi.org/10.1007/s10462-024-10880-6
Fidayani, E. F., & Ammar, F. M. (2023). The Use of Azhari Curriculum in Arabic Language Learning at Islamic Boarding School. *Nazhruna: Jurnal Pendidikan Islam*, *6*(1), 25–45. https://doi.org/10.31538/nzh.v6i1.2866
Fulcher, G. (2012). Assessment Literacy for the Language Classroom. *Language Assessment Quarterly*, *9*(2), 113–132. https://doi.org/10.1080/15434303.2011.642041
Golden, J., & Kohlbeck, M. (2020). Addressing cheating when using test bank questions in online Classes. *Journal of Accounting Education*, *52*, 100671. https://doi.org/10.1016/j.jaccedu.2020.100671

Graff Zivin, J., Song, Y., Tang, Q., & Zhang, P. (2020). Temperature and high-stakes cognitive performance: Evidence from the national college entrance examination in China. *Journal of Environmental Economics and Management*, *104*, 102365. https://doi.org/10.1016/j.jeem.2020.102365

Hidayat, W., Lawahid, N. A., & Mujahidah, M. (2021). roblems and Constraints of Authentic Assessment among Children s Early Education Teachers. *Pacific Early Childhood Education Research Association*, *15*(2), 87–109. https://doi.org/10.17206/apjrece.2021.15.2.87

Ismail, S. M., Rahul, D. R., Patra, I., & Rezvani, E. (2022). Formative vs. summative assessment: impacts on academic motivation, attitude toward learning, test anxiety, and self-regulation skill. *Language Testing in Asia*, *12*(1), 40. https://doi.org/10.1186/s40468-022-00191-4

Jauharoh, E., Anam, W., & Huda, M. M. (2021). The Use of Expressions in Improving Kalam Skill in Learning Arabic for MTSN 2 Kediri Students. *Asalibuna*. https://jurnalfaktarbiyah.iainkediri.ac.id/index.php/asalibuna/article/view/586

Kaya, M. H., & Adiguzel, T. (2021). Technology Integration Through Evidence-Based Multimodal Reflective Professional Training. *Contemporary Educational Technology*, *13*(4), ep323. https://doi.org/10.30935/cedtech/11143

Kremmel, B., & Harding, L. (2020). Towards a Comprehensive, Empirical Model of Language Assessment Literacy across Stakeholder Groups: Developing the Language Assessment Literacy Survey. *Language Assessment Quarterly*, *17*(1), 100–120. https://doi.org/10.1080/15434303.2019.1674855

Li, M., & Zhang, X. (2021). A meta-analysis of self-assessment and language performance in language testing and assessment. *Language Testing*, *38*(2), 189–218. https://doi.org/10.1177/0265532220932481

Mohapatra, B., & Laures-Gore, J. (2021). Moving toward accurate assessment of working memory in adults with neurogenically based communication disorders. *American Journal of Speech-Language Pathology*, *30*(3), 1292–1300. https://doi.org/10.1044/2021_AJSLP-20-00305

Muhammad Taufiq Ismail. (2016). *ANALISIS BUTIR SOAL PELAJARAN BAHASA ARAB SUMATIF AKHIR SMESTER GANJIL TAHUN AJARAN 2022/2023 KELAS XI SEKOLAH MENENGAH ATAS AL-FATTAH SIDOARJO*. *09*, 1–23.

Ni, U., Novikasari, I., Islam, U., Prof, N., & Zuhri, K. H. S. (2024). *Analisis Butir Soal Akhir Semester I Mata Pelajaran Bahasa Indonesia Kelas II Madrasah Ibtidaiyah*. *7*(1).

Nuswowati, M., Binadja, A., Efti, K., & Ifada, N. (2010). Pengaruh Validitas Dan Reliabilitas Butir Soal Ulangan Akhir Semester Bidang Studi Kimia Terhadap Pencapaian Kompetensi. *Jurnal Inovasi Pendidikan Kimia*, *4*(1), 566–573.

Panadero, E., Fraile, J., Fernández Ruiz, J., Castilla-Estévez, D., & Ruiz, M. A. (2019). Spanish university assessment practices: examination tradition with diversity by faculty. *Assessment & Evaluation in Higher Education*, *44*(3), 379–397. https://doi.org/10.1080/02602938.2018.1512553

Pittman, R. T., Chang, H., Lindner, A., Binks-Cantrell, E., & Joshi, M. (2023). What do classroom teachers of varying backgrounds know about English spelling? *Annals of Dyslexia*, *73*(3), 415–439. https://doi.org/10.1007/s11881-023-00286-4

Puad, L. M. A. Z., & Ashton, K. (2023). A critical analysis of Indonesia's 2013 national curriculum: Tensions between global and local concerns. *The Curriculum Journal*, *34*(3), 521–535. https://doi.org/10.1002/curj.194

Qiao, H., & Zhao, A. (2023). Artificial intelligence-based language learning: illuminating the impact on speaking skills and self-regulation in Chinese EFL context. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1255594

Qodri, M., & Sanjaya, B. (2024). Evaluation of the Implementation of Arabic Language Learning for Postgraduate Masters Students at UIN STS Jambi / Evaluasi Pelaksanaan Pembelajaran Bahasa Arab Pada Mahasiswa Magister Pascasarjana UIN STS Jambi. In *Loghat Arabi : Jurnal Bahasa Arab dan Pendidikan Bahasa Arab* (Vol. 5, Issue 1, p. 105). Institut Agama Islam (IAI DDI) Polewali Mandar. https://doi.org/10.36915/la.v5i1.228

Rahman, K. A., Seraj, P. M. I., Hasan, M. K., Namaziandost, E., & Tilwani, S. A. (2021). Washback of assessment on English teaching-learning practice at secondary schools. *Language Testing in Asia*, *11*(1), 12. https://doi.org/10.1186/s40468-021-00129-2

Rakhlin, N. V., Aljughaiman, A., & Grigorenko, E. L. (2021). Assessing language development in Arabic: The Arabic language: Evaluation of function (ALEF). *Applied Neuropsychology: Child*, *10*(1), 37–52. https://doi.org/10.1080/21622965.2019.1596113

Ramadhan, R., & Firdaus, F. N. (2022). Analisis Butir Soal Ujian Tengah Semester Bahasa Arab Kelas XII di SMA Al-Izzah IIBS Malang. *Tsaqofiya : Jurnal Pendidikan Bahasa Dan Sastra Arab*, *4*(1), 126–135. https://doi.org/10.21154/tsaqofiya.v4i1.49

Saleh, S. (2017). Penerbit Pustaka Ramadhan, Bandung. *Analisis Data Kualitatif*, *1*, 180.

Soliman, R., & Khalil, S. (2024). The teaching of Arabic as a community language in the UK. *International Journal of Bilingual Education and Bilingualism*, *27*(9), 1246–1257. https://doi.org/10.1080/13670050.2022.2063686

Su, Y. E., & Jiang, Y. (2024). Challenges with computing scalar and ad-hoc implicatures in Mandarin-speaking 4–8-year-old autistic children. *Journal of Communication Disorders*, *110*. https://doi.org/10.1016/j.jcomdis.2024.106427

Sukenti, D., Tambak, S., & Charlina, C. (2020). Developing Indonesian language learning assessments: Strengthening the personal competence and Islamic psychosocial of teachers. *International Journal of Evaluation and Research in Education (IJERE)*, *9*(4), 1079. https://doi.org/10.11591/ijere.v9i4.20677

Sukma, E., Ramadhan, S., Aldiyah, M. P., & Sihes, A. J. (2023). Challenges in Implementing Indonesian Language Teaching Materials in Elementary Schools. *Lnternational Electronic Journal of Elementary Education*. https://doi.org/10.26822/iejee.2024.327

Thabtah, F., Hammoud, S., Kamalov, F., & Gonsalves, A. (2020). Data imbalance in classification: Experimental evaluation. *Information Sciences*, *513*, 429–441. https://doi.org/10.1016/j.ins.2019.11.004

Ummah, M. S. (2019). VALIDITAS TES DAN KUALITAS BUTIR SOAL. *Sustainability (Switzerland)*, *11*(1), 1–14.

Wahyuni, L. G. E., Dewi, N. L. P. E. S., & Paramartha, A. A. G. Y. (2021). *Authentic Assessment Practice*. https://doi.org/10.2991/assehr.k.210407.258

Wong, H. M., Kwek, D., & Tan, K. (2020). Changing Assessments and the Examination Culture in Singapore: A Review and Analysis of Singapore's Assessment Policies. *Asia Pacific Journal of Education*, *40*(4), 433–457. https://doi.org/10.1080/02188791.2020.1838886

Yu, M. H., Reynolds, B. L., & Ding, C. (2021). Listening and Speaking for Real-World Communication: What Teachers Do and What Students Learn From Classroom Assessments. *Sage Open*, *11*(2). https://doi.org/10.1177/21582440211009163