Naatiq: Journal of Arabic Education
D O I : https://doi.org/10.33367/naatiq.v2i1.6829
I S S N (p r i n t) : 3063-7767
ejournal.uit–lirboyo.ac.id/index.php/naatiq/index | 17

Development of HOTS-Based Arabic Exam Questions According to Bloom's Taxonomy for Fifth Grade Students at Darul Hikmah Islamic School Tangerang

Syifa Amalia Nurkhalifah Fitrian[1], Siti Uriana Rahmawati[2], Raswan[3]
[1,2,3]Arabic Language Education, Syarif Hidayatullah State Islamic University, Jakarta, Indonesia
e-mail: syifaamalianf27@gmail.com[1] , uriana@uinjkt.ac.id[2] , raswan@uinjkt.ac.id[3]

*) syifaamalianf27@gmail.com

Abstract

This study aims to develop Higher Order Thinking Skills (HOTS)-based exam questions for Arabic language subjects in the fifth grade at Darul Hikmah Integrated Islamic School, Tangerang City, grounded in Bloom's Taxonomy framework. The background of this research stems from the need to improve the quality of Arabic language assessments, which have traditionally focused on memorization and basic understanding, by incorporating instruments that evaluate students' higher-order thinking abilities. There is a lack of empirically validated HOTS-based questions, particularly in Islamic elementary schools. The study employed the 4D development model (Define, Design, Develop, and Disseminate) and involved a sample of 20 fifth-grade students. The results of the study showed that: (1) the instrument was theoretically highly valid, receiving a validation score of 91.5% from two expert reviewers; (2) questions number 1, 2, 4, 5, and 9 were empirically valid, with point-biserial correlation values exceeding 0.468; (3) the instrument's reliability was categorized as moderate, with a KR-20 value of 0.59; and (4) the average difficulty index was 0.52, indicating good question quality. Therefore, the developed HOTS-based questions are feasible to be used as effective assessment tools to measure higher-order thinking skills in Arabic language learning at the elementary school level.

Keywords: HOTS; Arabic language; Test development;

مقدمة

في سياق تنفيذ منهج ٢٠١٣ في المدارس، سواء كانت الرسمية أو الدينية، يتوقع أن تحدث تغييرات كبيرة في تنفيذ أنشطة التعلم في جميع أنحاء إندونيسيا(M. Arif & Makalalag, 2022). إذ إن أنشطة التعليم والتعلم التي كانت تركز في السابق على المعلم أصبحت الآن تتمحور حول الطالب(de Quincey et al., 2019).

Copyright © 2024, This is an open access article under the CC BY-SA 4.0 license (https://creativecommons.org/licenses/by-sa/4.0/)
Published by Department of Arabic language Education

وبالتالي، يطلب من المعلمين أن يكونوا أكثر إبداعا وابتكارا في تقديم المواد. لاختبار مهارات التفكير لدى الطلاب، تم تصميم أسئلة تقييم نتائج التعلم، وخاصة في اللغة العربية، بأنماط متنوعة تتيح للطلاب الإجابة من خلال نفس عملية التفكير عالية المستوى، وفقا لتصنيف بلوم(Nafiati, 2021). وهذا يشير إلى تحول في نموذج التعليم، حيث يعزز من مشاركة الطلاب ويحسن من جودة التعلم بشكل عام(Maulana, 2022).

في إطار السعي لتطوير التعليم على المستوى الدولي، جاء تصميم المنهج الدراسي لعام ٢٠١٣ مع تحسينات جوهرية، أبرزها التركيز على معايير تقييم تعزز مهارات التفكير العليا لدى الطلاب(Lubis & Azizan, 2019). التقييم ليس مجرد أداة لقياس الأداء، بل هو عملية محورية تقدم تغذية راجعة دقيقة تساعد في توجيه المعلمين والطلاب نحو تحقيق نتائج تعليمية أفضل(Zulkifli, 2018). ولضمان جودة التعليم، توصى الأنظمة التعليمية بتبني تقييمات تركز على التفكير النقدي ومهارات التفكير العليا، باعتبارها أساسا لتطوير قدرات وتحفيز الإبداع الطلاب(Abdurrahman et al., 2019).

إضافة إلى ذلك، لتحقيق هدف إنتاج طلاب ذوي جودة عالية، ينبغي على المعلمين تضمين أسئلة تعزز مهارات التفكير العليا ضمن أدوات تقييم التعلم(Wenno & Karuna, 2021). ومع ذلك، لاحظ الباحثون أن العديد من المعلمين لا يزالون يفتقرون إلى تطوير تقييمات على شكل أسئلة اختبارية تساهم في تدريب الطلاب على مهارات التفكير العليا(Gozali et al., 2021). استنادا إلى هذه الملاحظة، يسعى الباحثون إلى تطوير أسئلة تستهدف مهارات التفكير العليا، وفقا لتصنيف بلوم.

من المتوقع أن يتمكن المعلمون من تصميم أسئلة تقييم تعتمد على مهارات التفكير العليا، بحيث لا يقتصر أداء الطلاب على المستويات الأدنى مثل C-1 (المعرفة)، C-2 (الفهم)، وC-3 (التطبيق)، بل يمتد ليشمل المستويات الأعلى مثل C-4 (التحليل)، C-5 (التقييم)، وC-6 (الإبداع)(Anisah & Lastuti, 2018). لتحقيق هذا الهدف، قام فريق تطوير المنهج لعام ٢٠١٣، من خلال الدورات التنشيطية، بتقديم مواد تدريبية تركز على أدوات التقييم المستندة إلى مهارات التفكير العلي(Deviana & Kusumaningtyas, 2019). لا يهدف هذا الجهد إلى تحسين جودة الأسئلة فحسب، بل يسعى أيضا إلى تعزيز جودة فهم الطلاب، وتطوير قدرتهم على التعامل مع الأسئلة التي تعزز مهارات التفكير العليا(Lubis & Azizan, 2019).

عدة نتائج بحثية تتعلق بمهارات التفكير العليا في تطبيق أسئلة اللغة العربية منها بحث إحسان (2019) الذي أجري في إندونسيا(Ichsan et al., 2019)،

حيث ذكر أن الأسئلة المطبقة تظهر أنها لا تزال تحتوي على أسئلة مهارة التفكير المنخفضة (*LOTS*)، وأوصى الباحثان بضرورة إجراء تحسينات، خاصة على الأسئلة رقم ٦ و١٥ التي لا تزال مصنفة على أنها C3، ورقم ١٤ لا تزال C2، والأسئلة رقم ٧-١٢ لا تزال مصنفة على أنها C1. وفي دراسة أخرى، أشار لطفية وحنيفة (2022) وأندرياني (Andriani, 2020) إلى أن أداة مهارات التفكير العليا التي تم تطويرها لمهارة القراءة باستخدام قويزيز كانت تتميز بالصلاحية والاتساق من حيث صلاحية المحتوى، وصلاحية البناء، والموثوقية، وملاءمة النموذج. تسلط هذه الدراسات الضوء على أهمية تحسين أدوات التقييم لتتماشى مع أهداف تعزيز مهارات التفكير العليا لدى الطلاب.

في إطار البحث الذي سيجريه الباحثون في الصف الخامس بمدرسة دار الحكمة الإسلامية الابتدائية في بيكاسي، تم جمع بيانات تشير إلى أن غالبية الطلاب يتمتعون بقدرات تفكير نقدي، سواء أثناء أنشطة التعلم أو عند الإجابة على أسئلة امتحانات اللغة العربية. ومع ذلك، أظهرت المعلومات أن معلمي اللغة العربية في المدرسة لم يطوروا بعد العديد من أدوات التقييم على شكل أسئلة اختبارية تهدف إلى تدريب الطلاب على مهارات التفكير العليا، وخاصة التفكير النقدي. لذلك، يسعى الباحثون إلى تطوير أسئلة امتحانات اللغة العربية النهائية للصف الخامس، بناء على مهارات التفكير العليا، بهدف تعزيز قدرات الطلاب النقدية وتطوير أدوات التقييم لتتناسب مع متطلبات التعليم الحديث.

منهجية البحث

هذا البحث يستخدم بحث تطويري (بحث وتطوير) يهدف إلى إنتاج أدوات أسئلة صالحة وموثوقة وقابلة للستخدام في تقييم تعلم اللغة العربية. ونموذج التطوير المستخدم هو المنوذج رباعي الأبعاد (4D) الذي يتكون من أربع مراحل، وهي التعريف (Define)، والتصميم (Design)، والتطوير (Develop)، والنشر (Disseminate) (Stefano et al., 2021). كانت الموضوعات في هذا الدراسة طلاب الصف الخاس في مدرسة دار الحكمة الإبتدائية الإسلامية المتكاملة (SDIT) بمدينة تانجيرانج. كانت العينة المستخدمة في التجربة المحددة ٢٠ طالبًا. تم اختيارهم بشكل مقصود لتمثيل خصائص طلاب الصف الخامس بشكل عام المدرسة.

تتكون إجراءات البحث والتطوير من: ١) التعريف ٢) التصميم ٣) التطوير ٤) النشر

١. التعريف (Define)

وفقا لبرين (٢٠٢١)، تهدف مرحلة التعريف إلى تحديد الاحتياجات وتعريفها في عملية التعلم.

تعد هذه المرحلة خطوة أساسية لتحليل المشكلات وتحديدها من أجل جمع المعلومات المختلفة المتعلقة بالمنتج المراد تطويره(Z. U. Arif et al., 2022). استنادا إلى المشكلات القائمة، قام الباحثون بعدة خطوات في تحليل الاحتياجات، وهي:

أ) تحليل الواجهة الأمامية (Front-End Analysis)

يمكن إجراء هذا التحليل من خلال الدراسات الميدانية، والتي شملت في هذا البحث ملاحظات ميدانية ومقابلات مع معلمي اللغة العربية.

ب) تحديد المواد التعليمية المناسبة.

ج) تحديد أسئلة الامتحان النهائي في اللغة العربية التي سيتم بحثها وتطويرها

أسئلة الامتحان التي سيتم البحث والتطوير فيها هي أسئلة الامتحان النهائي لمادة اللغة العربية للصف الخامس.

د) تحديد الكفاءات الأساسية والكفاءات الرئيسية

٢. التصميم (Design)

تهدف مرحلة التصميم إلى إعداد نموذج أولي لأسئلة امتحان اللغة العربية المبنية على مهارات التفكير العليا (HOTS)(Sood et al., 2024). وتشمل الخطوات التي تم اتخاذها في هذه المرحلة فيما يلي:

أ) إعداد شبكة أسئلة امتحان اللغة العربية.

ب) تحديد عدد الأسئلة.

ج) صياغة أدوات اختبار اللغة العربية استنادًا إلى مؤشرات أسئلة مهارات التفكير العليا.

د) تحديد الأهداف من الاختبار.

٣. التطوير (Develop)

تهدف مرحلة التطوير إلى إنتاج المنتج المصمم وتنفيذه وفقا للإجراءات المعتمدة. تشمل هذه المرحلة التحقق من صلاحية الأداة المنتجة، وإجراء التعديلات اللازمة، بالإضافة إلى مراجعتها من قبل مجموعة من الخبراء لضمان الوصول إلى الشكل النهائي الصالح للاستخدام(Yusuf, 2023).

أ) التحقق من صحة الخبراء (Expert validation)

في هذه المرحلة، يتم التحقق من صلاحية الأداة المصممة من قبل مجموعة من الخبراء للحصول على التقييمات والملاحظات والاقتراحات للتحسين. وقد طلب من الخبراء تقديم نقدهم وتقييمهم للأداة من الجوانب التعليمية والفنية. في هذا البحث، شملت عملية التحقق مساهمة خبراء، بمن فيهم أساتذة في تعليم اللغة العربية.

D O I : https://doi.org/10.33367/naatiq.v2i1.6829
I S S N (p r i n t) : 3063-7767

ب) تصميم الأسئلة وفقا لاقتراحات ومراجعات الخبراء

بعد الانتهاء من عملية التحقق من صلاحية أداة أسئلة مهارات التفكير العليا من قبل الخبراء، الخطوة التالية بتعديل الأداة وتصميمها من جديد استنادا إلى النقد والملاحظات المقدمة من الخبراء.

ج) اختبار صحة السؤال (Question validity test)

يمكن قياس جودة الأداة من خلال مستوى صلاحيتها، حيث تشير الصلاحية إلى مدى ملاءمة الأداة للاستخدام. في عملية التحقق من صحة أسئلة الاختيار من متعدد، تتكون البيانات التي تم التحقق من صحتها من الرقمين ١ و٠. يتم إجراء هذا التحقق من الصحة باستخدام *Biserial Point Correlation*، والذي يتم حسابه باستخدام الصيغة التالية(Veronica et al., 2022):

$$\text{ر بوينت بايسيريال} = \frac{\text{م}_1 - \text{م}_{\text{صفر}} \sqrt{\text{ب ك}}}{\text{ع}_{\text{ص}}}$$

الملاحظة:

1 = م_1 متوسط الدرجات على المتغير (ص) للافراد الذين علاماتهم او تصنيفهم على المتغير (س) يساوي (1)

= $\text{م}_{\text{صفر}}$ متوسط الدرجات على المتغير (ص) للافراد الذين حصلوا على علامات على المتغير (س) يساوي صفر

= $\text{ع}_{\text{ص}}$ الانحراف المعياري للعلامات الكلية على المتغير (ص) بغض النظر عن التصنيف على المتغير (س)

ب = نسبة الافراد ضمن التصنيف (1)

ك = نسبة الافراد ضمن التصنيف (صفر)

$$\text{ع}_{\text{س}} = \sqrt{\frac{\text{مجـ ص}^2 - \frac{(\text{مجـ ص})^2}{\text{ن}}}{\text{ن} - 1}}$$

د) اختبار الموثوقية (Reability Test)

إجراء حسابات الموثوقية لأدوات الاختبار الموضوعية باستخدام صيغة KR20 التي اقترحها كودر (Kuder) وريكاردسون (Richadson).

D O I : https://doi.org/10.33367/naatiq.v2i1.6829
I S S N (p r i n t) : 3063-7767

$$r_{11} = \left(\frac{n}{n-1}\right)\left(\frac{S_t^2 - \sum p_i q_i}{S_t^2}\right)$$

هـ) تحليل مستوى الصعوبة

يهدف تحليل مستوى صعوبة عناصر الاختبار إلى تحديد مدى صعوبة أو سهولة أسئلة اللغة العربية بناء على مهارة التفكير العليا التي تم تطويرها واختبارها. يتراوح مقدار مؤشر الصعوبة يتراوح مؤشر الصعوبة لهذه الأسئلة بين ٠,٠٠ و ١,٠٠.

$$P = \frac{R}{T}$$

P = مؤشر الصعوبة

R = عدد الأفراد الذين نجحوا في الإجابة على البند

T = عدد أفراد المجموعة

٤. النشر (Disseminate)

في هذه المرحلة، تهدف العملية التي يتم تنفيذها إلى نشر نتائج البحث على جمهور واسع(Sihombing, 2024). تمت مرحلة النشر في هذا البحث من خلال النشر المحدود.

نتائج البحث ومناقشاتها

يهدف هذا البحث إلى تطوير أداة اختبار نهائية في اللغة العربية تعتمد على مهارات التفكير العليا لطلاب الصف الخامس في مدرسة دار الحكمة الإسلامية الابتدائية بيكاسي. نموذج التطوير المستخدم هو نموذج 4D، والذي يتكون من أربع مراحل، وهي التعريف والتصميم والتطوير والنشر. يتم شرح النتائج الإجمالية للدراسة في الشرح التالي:

١. التعريف (Define)

أ) تحليل الواجهة الأمامية (Front-End Analysis)

تم إجراء هذا التحليل من خلال الدراسات الميدانية التي شملت الملاحظات والمقابلات مع معلمي اللغة العربية ذوي الصلة.

ب) تحديد المادة التعليمية

المادة التعليمية التي تم اختيارها في بحث تطوير أسئلة اللغة العربية القائمة على مهارات التفكير العليا هي موضوع "في المقصف".

ج) أسئلة الامتحان النهائي في اللغة العربية التي سيتم بحثها وتطويرها

توجد أنواع عديدة من أدوات الاختبار المستخدمة في تقييم التعلم، وتُختار أداة الاختبار بما يتناسب مع الأهداف المراد قياسها. في هذه الدراسة، تم استخدام اختبار موضوعي يعتمد على أسئلة الاختيار من متعدد كأداة للتقييم.

د) تحديد الكفاءات الأساسية والكفاءات الرئيسية

الكفاءات الأساسية المستخدمة في هذا البحث هي الكفاءات الأساسية لمادة اللغة العربية للصف الخامس في الفصل الدراسي الثاني، والمتعلقة بموضوع "في المقصف".

٢. التصميم (Design)

أ. إعداد شبكة أسئلة امتحان اللغة العربية.

تتألف أداة اختبار مهارات التفكير العليا من خمسة عشر سؤالاً تغطي ثلاث مراحل معرفية، وهي مرحلة التحليل (C4)، ومرحلة التقييم (C5)، ومرحلة الإنشاء (C6)(Kiliańska-Przybyło, 2019).

ب. تحديد الأهداف من الاختبار.

تهدف أداة الاختبار التي تم تطويرها في هذه الدراسة إلى قياس مهارات التفكير العليا لدى طلاب الصف الخامس الابتدائي. ويتوقع أن يسهم هذا الاختبار في تعزيز عملية التعلم الفعال من خلال تقديم مسائل وسياقات معقدة تساهم في تدريب الطلاب على تطوير مهارات التفكير العليا(Pujawan et al., 2022).

ج. تحديد عدد السؤال

تأخذ عملية تحديد شكل الاختبار ومدته بعين الاعتبار عدة جوانب، بما في ذلك عدد المشاركين، والوقت المتاح، والمحتوى، والقدرات التي سيتم قياسها في الاختبار(Irvine, 2017). في هذا البحث، كان شكل الاختبار الذي تم اختياره في هذه الدراسة هو اختبار الاختيار من متعدد لأنه يسهل تحليل البيانات ويسمح بإكماله في وقت قصير نسبيا. تم تحديد مدة الاختبار بخمسة عشر سؤالا متعدد الاختيارات ومدته ٦٠ دقيقة.

D O I : https://doi.org/10.33367/naatiq.v2i1.6829
I S S N (p r i n t) : 3063-7767

د. صياغة أدوات اختبار اللغة العربية استنادا إلى مؤشرات أسئلة مهارات التفكير العليا.

في هذه المرحلة، يتم ترتيب عناصر الأسئلة وفقا لشبكات الأداة التي تم إعدادها. تتضمن بعض الجوانب التي يجب أخذها بعين الاعتبار عند كتابة عناصر أداة اختبار مهارات التفكير العليا ما يلي (١) ملاءمة البنود لمؤشرات مهارات التفكير العليا، (٢) وضوح بنود الأسئلة، (٣) الاتساق المنطقي في جميع خيارات الإجابة، (٤) دقة استخدام اللغة، (٥) وظيفة الأوصاف والصور والجداول في بنود الأسئلة(Naryatmojo, 2018).

٣. التطوير (Develop)

أ. التحقق من صحة الخبراء (Expert validation)

بعد إتمام الإجراءات في المرحلة السابقة، قام الباحثون بإجراء اختبار الصلاحية على أسئلة مهارات التفكير العليا بهدف الحصول على الملاحظات، والانتقادات، والاقتراحات، وكذلك لتحديد مدى صلاحية الأسئلة القائمة على مهارات التفكير العليا (Banfi et al., 2022). تم إجراء اختبار الصلاحية من قبل اثنين من خبراء صياغة الأسئلة، وهما السيدة لالاه علوية الماجستير، والدكتور زين المتقين الماجستير، وهما من أساتذة تعليم اللغة العربية في جامعة شريف هداية الله الإسلامية الحكومية في جاكرتا.

ومن نتائج الاستبانة من خبراء المادة عن صلاحيتها، حصل الباحثون على نسبة مئوية

$$\text{نسبة المئوية} = \frac{\sum (\text{نقاط قيمة جميع})}{(\text{عدد العناصر} \times \text{أعلى وزن})} \times ٪١٠٠$$

كما يلي:

$$\frac{\sum(٤٣)}{(} \times ٪١٠٠$$

نسبة المئوية = ٪٩٧

نسبة المئوية =

D O I : https://doi.org/10.33367/naatiq.v2i1.6829
I S S N (p r i n t) : 3063-7767

ومن نتائج الاستبانة من خبيرة المادة عن صحة عنصر الأسئلة، حصل الباحثون على

$$\text{نسبة المئوية} = \frac{\sum (\text{نقاط قيمة جميع})}{(\text{عدد العناصر} \times \text{أعلى وزن})} \times ٪١٠٠$$

$$\text{نسبة المئوية} = \frac{\sum(٣٨)}{\text{نسبة مئوية كما يلي}} \times ٪١٠٠$$

$$\text{نسبة المئوية} = ٪٨٦$$

بعد حصول الباحثين على نتائج الاستبانة من الخبراء، قامت بقياس اللياقة باستخدام مقياس ليكرت (Likert Scale) كما يلي(Vieira, 2016):

وبناءً على نتائج التحقق من صحة الأسئلة من الخبراء، حصلت أسئلة امتحان اللغة العربية بناء على مهارة التفكير العليا والتي تم تطويرها على درجة ٩١.٥٪ في الفئة "جيد جدا".

ب. تصميم الأسئلة وفقا لاقتراحات ومراجعات الخبراء

بعد إجراء التحقق من الصلاحية من قبل الخبراء على أداة أسئلة مهارة التفكير العليا التي تم إعدادها، فإن الخطوة التالية هي تعديل وتصميم أداة الأسئلة بناء على نقد وملاحظات الخبراء.(Gunawan, 2022)

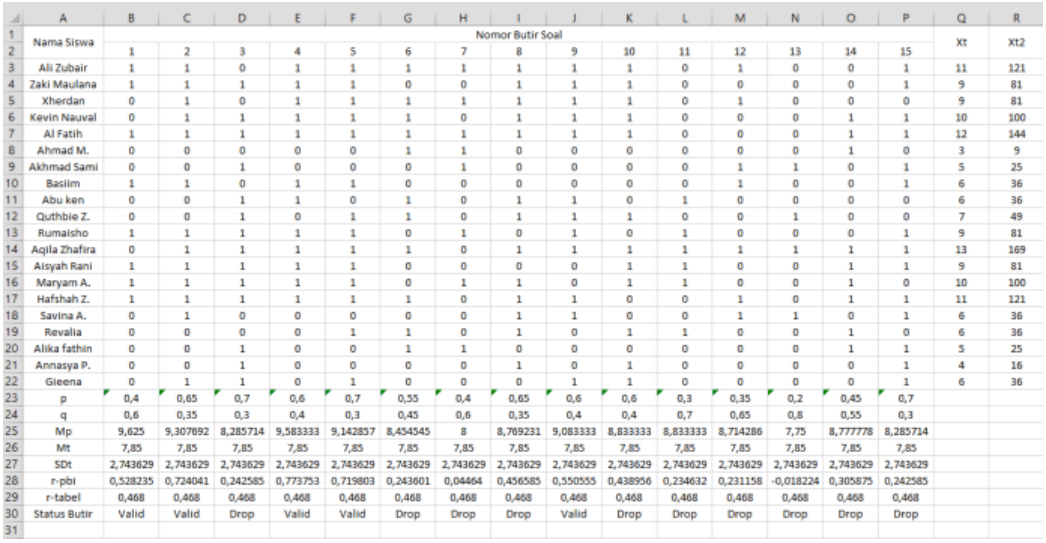
ج. اختبار صحة السؤال (Question validity test)

المرحلة التالية هي اختبار صلاحية أسئلة مهارة التفكير العليا في اللغة العربية التي تم تطويرها، ثم توزيعها إلى ٢٠ طالبا من الصف الخامس في مدرسة دار الحكمة الابتدائية الإسلامية بيكاسي.

يتم إعطاء درجة ١ إذا أجاب الطالب بشكل صحيح، وتستخدم درجة ٠ إذا أجاب الطالب بشكل خاطئ.

Naatiq: Journal of Arabic Education
D O I : https://doi.org/10.33367/naatiq.v2i1.6829
I S S N (p r i n t) : 3063-7767
ejournal.uit–lirboyo.ac.id/index.php/naatiq/index | 26

صورة ١



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Nama Siswa	Nomor Butir Soal															Xt	Xt2
2		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
3	Ali Zubair	1	1	0	1	1	1	1	1	1	1	0	1	0	0	1	11	121
4	Zaki Maulana	1	1	1	1	1	0	0	1	1	1	0	0	0	0	1	9	81
5	Xherdan	0	1	0	1	1	1	1	1	1	1	0	1	0	0	0	9	81
6	Kevin Nauval	0	1	1	1	1	1	0	1	1	1	0	0	0	1	1	10	100
7	Al Fatih	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	12	144
8	Ahmad M.	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	3	9
9	Akhmad Sami	0	0	1	0	0	0	1	0	0	0	0	1	1	0	1	5	25
10	Basiim	1	1	0	1	1	0	0	0	0	0	0	1	0	0	1	6	36
11	Abu ken	0	0	1	1	0	1	0	1	1	0	1	0	0	0	0	6	36
12	Quthbie Z.	0	0	1	0	1	1	0	1	1	1	0	0	1	0	0	7	49
13	Rumaisho	1	1	1	1	1	0	1	0	1	0	1	0	0	0	1	9	81
14	Aqila Zhafira	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	13	169
15	Aisyah Rani	1	1	1	1	1	0	0	0	0	1	1	0	0	1	1	9	81
16	Maryam A.	1	1	1	1	1	0	1	1	0	1	1	0	0	1	0	10	100
17	Hafshah Z.	1	1	1	1	1	1	0	1	1	0	0	1	0	1	1	11	121
18	Savina A.	0	1	0	0	0	0	0	1	1	0	0	1	1	0	1	6	36
19	Revalia	0	0	0	0	1	1	0	1	0	1	1	0	0	1	0	6	36
20	Alika fathin	0	0	1	0	0	1	1	0	0	0	0	0	0	1	1	5	25
21	Annasya P.	0	0	1	0	0	0	0	1	0	1	0	0	0	0	1	4	16
22	Gleena	0	1	1	0	1	0	0	0	1	1	0	0	0	0	1	6	36
23	p	0,4	0,65	0,7	0,6	0,7	0,55	0,4	0,65	0,6	0,6	0,3	0,35	0,2	0,45	0,7		
24	q	0,6	0,35	0,3	0,4	0,3	0,45	0,6	0,35	0,4	0,4	0,7	0,65	0,8	0,55	0,3		
25	Mp	9,625	9,307692	8,285714	9,583333	9,142857	8,454545	8	8,769231	9,083333	8,833333	8,833333	8,714286	7,75	8,777778	8,285714		
26	Mt	7,85	7,85	7,85	7,85	7,85	7,85	7,85	7,85	7,85	7,85	7,85	7,85	7,85	7,85	7,85		
27	SDt	2,743629	2,743629	2,743629	2,743629	2,743629	2,743629	2,743629	2,743629	2,743629	2,743629	2,743629	2,743629	2,743629	2,743629	2,743629		
28	r-pbi	0,528235	0,724041	0,242585	0,773753	0,719803	0,243601	0,04464	0,456585	0,550555	0,438956	0,234632	0,231158	-0,018224	0,305875	0,242585		
29	r-tabel	0,468	0,468	0,468	0,468	0,468	0,468	0,468	0,468	0,468	0,468	0,468	0,468	0,468	0,468	0,468		
30	Status Butir	Valid	Valid	Drop	Valid	Valid	Drop	Drop	Drop	Valid	Drop	Drop	Drop	Drop	Drop	Drop		
31																		

استنادا إلى نتائج صلاحية بنود الأسئلة باستخدام معامل الارتباط النقطي الثنائي (*Biserial Point Correlation*)، مع اعتماد قيمة $\alpha = 5\%$ وقيمة معامل الارتباط 'r' برودكت مومنت تساوي (*r pruduct moment*) ٠,٤٦٨، تم الحصول على النتائج التالية:

بنود الأسئلة التي تم اعتبارها "صحيحة" هي البنود رقم واحد، اثنان، أربعة، خمسة، وتسعة، حيث كانت قيمة الارتباط النقطي الثنائي (*Biserial Point Correlation*) أكبر من ٠,٤٦٨.

د. اختبار الموثوقية (Reability Test)

يتم التعبير عن مستوى موثوقية الاختبار من خلال قيمة معامل الارتباط الذي يتراوح من +١,٠٠ إلى -١,٠٠.

Naatiq: Journal of Arabic Education
D O I : https://doi.org/10.33367/naatiq.v2i1.6829
I S S N (p r i n t) : 3063-7767
ejournal.uit–lirboyo.ac.id/index.php/naatiq/index | 27

صورة ٢

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Nama Siswa	Nomor Butir Soal															Skor Total
2		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
3	Ali Zubair	1	1	0	1	1	1	1	1	1	1	0	1	0	0	1	11
4	Zaki Maulana	1	1	1	1	1	0	0	1	1	1	0	0	0	0	1	9
5	Xherdan	0	1	0	1	1	1	1	1	1	1	0	1	0	0	0	9
6	Kevin Nauval	0	1	1	1	1	1	0	1	1	1	0	0	0	1	1	10
7	Al Fatih	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	12
8	Ahmad M.	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	3
9	Akhmad Sami	0	0	1	0	0	0	1	0	0	0	0	1	1	0	1	5
10	Baslim	1	1	0	1	1	0	0	0	0	0	0	1	0	0	1	6
11	Abu ken	0	0	1	1	0	1	0	1	1	0	1	0	0	0	0	6
12	Quthbie Z.	0	0	1	0	1	1	0	1	1	1	0	0	1	0	0	7
13	Rumaisho	1	1	1	1	1	0	1	0	1	0	1	0	0	0	1	9
14	Aqila Zhafira	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	13
15	Aisyah Rani	1	1	1	1	1	0	0	0	0	1	1	0	0	1	1	9
16	Maryam A.	1	1	1	1	1	0	1	1	0	1	1	0	0	1	0	10
17	Hafshah Z.	1	1	1	1	1	1	0	1	1	0	0	1	0	1	1	11
18	Savina A.	0	1	0	0	0	0	0	1	1	0	0	1	1	0	1	6
19	Revalia	0	0	0	0	1	1	0	1	0	1	1	0	0	1	0	6
20	Alika fathin	0	0	1	0	0	1	1	0	0	0	0	0	0	1	1	5
21	Annasya P.	0	0	1	0	0	0	0	1	0	1	0	0	0	0	1	4
22	Gleena	0	1	1	0	1	0	0	0	1	1	0	0	0	0	1	6
23	k	15															
24	Jumlah	8	13	14	12	14	11	8	13	12	12	6	7	4	9	14	
25	p	0,4	0,65	0,7	0,6	0,7	0,55	0,4	0,65	0,6	0,6	0,3	0,35	0,2	0,45	0,7	
26	q	0,6	0,35	0,3	0,4	0,3	0,45	0,6	0,35	0,4	0,4	0,7	0,65	0,8	0,55	0,3	
27	pq	0,24	0,2275	0,21	0,24	0,21	0,2475	0,24	0,2275	0,24	0,24	0,21	0,2275	0,16	0,2475	0,21	3,3775
28	Varians Total	7,5275															
29	KR-20	0,590691															
30																	

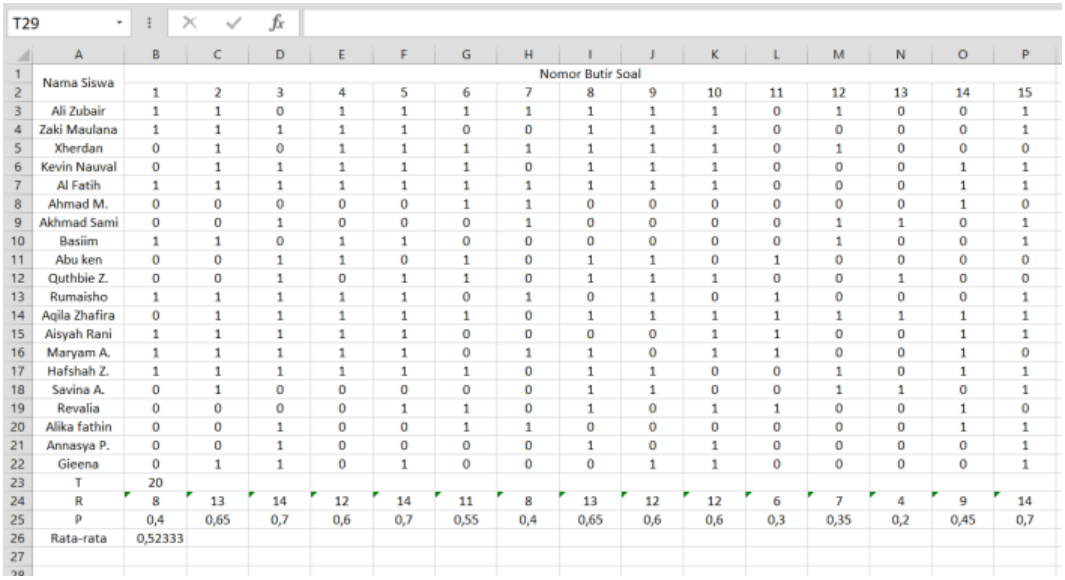
استنادا إلى نتائج اختبار الموثوقية الذي تم حسابه باستخدام معادلة KR20 عبر برنامج Microsoft Excel، تم الحصول على درجة موثوقية أسئلة اختبار مهارات التفكير العليا في مادة اللغة العربية بمقدار ٠,٥٩، وهي تقع ضمن الفئة "المتوسطة"(Sugiyono, 2019).

هـ. تحليل مستوى الصعوبة

يتراوح مقدار مؤشر الصعوبة في أسئلة مهارات التفكير العليا للغة العربية التي سيتم تحليلها بين ٠,٠٠ و ١,٠٠.

Naatiq: Journal of Arabic Education
D O I : https://doi.org/10.33367/naatiq.v2i1.6829
I S S N (p r i n t) : 3063-7767
ejournal.uit–lirboyo.ac.id/index.php/naatiq/index | 28

صورة ٣



Nama Siswa	Nomor Butir Soal														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Ali Zubair	1	1	0	1	1	1	1	1	1	1	0	1	0	0	1
Zaki Maulana	1	1	1	1	1	0	0	1	1	1	0	0	0	0	1
Xherdan	0	1	0	1	1	1	1	1	1	1	0	1	0	0	0
Kevin Nauval	0	1	1	1	1	1	0	1	1	1	0	0	0	1	1
Al Fatih	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1
Ahmad M.	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0
Akhmad Sami	0	0	1	0	0	0	1	0	0	0	0	1	1	0	1
Basiim	1	1	0	1	1	0	0	0	0	0	0	1	0	0	1
Abu ken	0	0	1	1	0	1	0	1	1	0	1	0	0	0	0
Quthbie Z.	0	0	1	0	1	1	0	1	1	1	0	0	1	0	0
Rumaisho	1	1	1	1	1	0	1	0	1	0	1	0	0	0	1
Aqila Zhafira	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1
Aisyah Rani	1	1	1	1	1	0	0	0	0	1	1	0	0	1	1
Maryam A.	1	1	1	1	1	0	1	1	0	1	1	0	0	1	0
Hafshah Z.	1	1	1	1	1	1	0	1	1	0	0	1	0	1	1
Savina A.	0	1	0	0	0	0	0	1	1	0	0	1	1	0	1
Revalia	0	0	0	0	1	1	0	1	0	1	1	0	0	1	0
Alika fathin	0	0	1	0	0	1	1	0	0	0	0	0	0	1	1
Annasya P.	0	0	1	0	0	0	0	1	0	1	0	0	0	0	1
Gieena	0	1	1	0	1	0	0	0	1	1	0	0	0	0	1
T	20														
R	8	13	14	12	14	11	8	13	12	12	6	7	4	9	14
P	0,4	0,65	0,7	0,6	0,7	0,55	0,4	0,65	0,6	0,6	0,3	0,35	0,2	0,45	0,7
Rata-rata	0,52333														

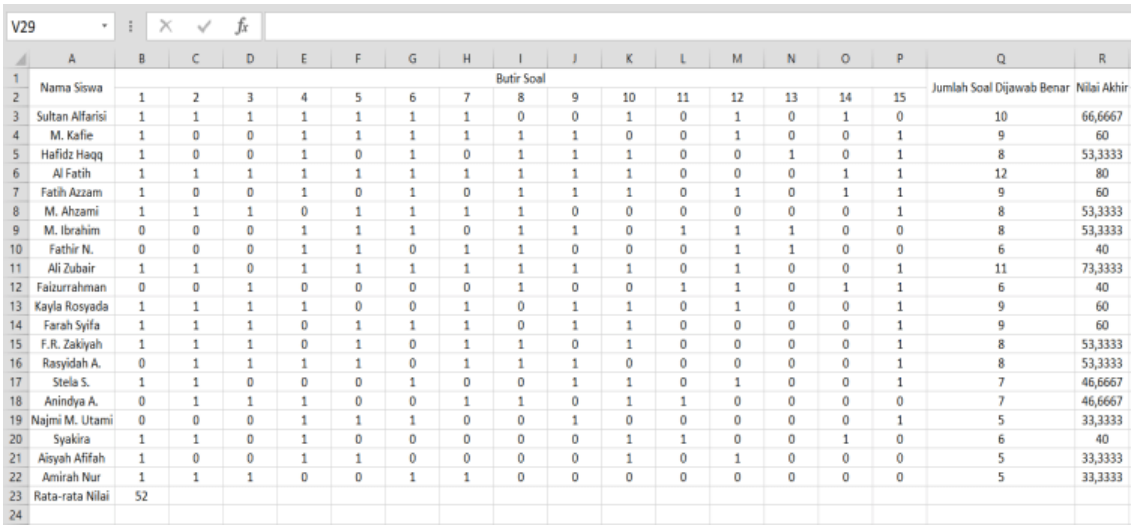
بناء على نتائج تحليل مستوى الصعوبة باستخدام Microsoft Excel، كما هو موضح في الصورة ٤.٧، تبين أن من بين ١٥ سؤالا التي أجاب عليها ٢٠ طالبا، كان "١٣ سؤالا في الفئة المتوسطة" و"سؤالين في الفئة الصعبة". وتم الحصول على متوسط من خمسة عشر سؤالا ٠,٥٢، مع فئة "متوسطة"(Fanani & Kusmaharti, 2018).

وبذلك، فإن مستوى صعوبة أسئلة امتحان اللغة العربية بناء على مهارات التفكير عالي المستوى التي تم تطويرها يتمتع بجودة جيدة.

٤. النشر (Disseminate)

تعد هذه المرحلة المرحلة النهائية من البحث في تطوير نموذج 4D، والتي تهدف إلى نشر أداة اختبار اللغة العربية بناء على مهارات التفكير العليا والتي تم التحقق من صلاحيتها من خلال خبيرين والتحقق التجريبي. في هذا البحث، قام الباحثون بتحديد نطاق البحث التطويري ليصل فقط إلى مرحلة النشر المحدود لطلاب الصف الخامس في المدرسة دار الحكمة الابتدائية الإسلامية بيكاسي.

صورة ٤



Nama Siswa	Butir Soal															Jumlah Soal Dijawab Benar	Nilai Akhir
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
Sultan Alfarisi	1	1	1	1	1	1	1	0	0	1	0	1	0	1	0	10	66,6667
M. Kafie	1	0	0	1	1	1	1	1	1	0	0	1	0	0	1	9	60
Hafidz Haqq	1	0	0	1	0	1	0	1	1	1	0	0	1	0	1	8	53,3333
Al Fatih	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	12	80
Fatih Azzam	1	0	0	1	0	1	0	1	1	1	0	1	0	1	1	9	60
M. Ahzami	1	1	1	0	1	1	1	1	0	0	0	0	0	0	1	8	53,3333
M. Ibrahim	0	0	0	1	1	1	0	1	1	0	1	1	1	0	0	8	53,3333
Fathir N.	0	0	0	1	1	0	1	1	0	0	0	1	1	0	0	6	40
Ali Zubair	1	1	0	1	1	1	1	1	1	1	0	1	0	0	1	11	73,3333
Faizurrahman	0	0	1	0	0	0	0	1	0	0	1	1	0	1	1	6	40
Kayla Rosyada	1	1	1	1	0	0	1	0	1	1	0	1	0	0	1	9	60
Farah Syifa	1	1	1	0	1	1	1	0	1	1	0	0	0	0	1	9	60
F.R. Zakiyah	1	1	1	0	1	0	1	1	0	1	0	0	0	0	1	8	53,3333
Rasyidah A.	0	1	1	1	1	0	1	1	1	0	0	0	0	0	1	8	53,3333
Stela S.	1	1	0	0	0	1	0	0	1	1	0	1	0	0	1	7	46,6667
Anindya A.	0	1	1	1	0	0	1	1	0	1	1	0	0	0	0	7	46,6667
Najmi M. Utami	0	0	0	1	1	1	0	0	1	0	0	0	0	0	1	5	33,3333
Syakira	1	1	0	1	0	0	0	0	0	1	1	0	0	1	0	6	40
Aisyah Afifah	1	0	0	1	1	0	0	0	0	1	0	1	0	0	0	5	33,3333
Amirah Nur	1	1	1	0	0	1	1	0	0	0	0	0	0	0	0	5	33,3333
Rata-rata Nilai	52																

جدول ٥

الفصيلة	القيمة
جيد جدا	٪٧٦ - ٪١٠٠
جيد	٪٦٦ - ٪٧٥
منقص	٪٥١ - ٪٦٥
مردود	٪٠ - ٪٥٠

وقد حصلت نتائج اختبار قدرات الطالب بعد إجراء اختبار على خمسة عشر سؤالا من أسئلة الاختيار من متعدد منتظم في اللغة العربية بناء على مهارات التفكير العليا على معدل ٥٢ في الفئة " منقص".

تظهر نتائج هذه الدراسة أن تطوير أدوات تقويم مستندة إلى مهارات التفكير العليا (HOTS) في تعليم اللغة العربية أمرٌ ممكن وفعال .(Saepurrohman et al., 2023)، حتى في المرحلة الابتدائية. ويكتسب هذا العمل أهميته في ظل الدعوة المتزايدة نحو إصلاح التعليم من خلال دمج مهارات القرن الحادي والعشرين، والتي من أبرزها التفكير النقدي وحل المشكلات (Yusella et al., 2023).

ويتوافق هذا التوجه مع نظرية بلوم (Bloom) في تصنيفه الشهير للأهداف المعرفية، حيث تقع مهارات التفكير العليا في المستويات الثلاثة العليا من الهرم: التحليل، التقييم، والإبداع، وهي التي استهدفتها أدوات التقويم المطوّرة في هذه الدراسة (Mustofa et al., 2022).

علاوة على ذلك، تنسجم هذه الدراسة مع مبادئ نظرية البنائية(Constructivism) ، كما عرضها جان بياجيه وليف فيجوتسكي، حيث تؤكد هذه النظرية أن التعلم الفعّال لا يتم عبر التلقين والحفظ، بل من خلال انخراط المتعلم في مواقف تعليمية تتيح له بناء المعرفة بنفسه (Kouicem, 2020). وهذا ما تحققه أسئلة HOTS التي تحفّز الطالب على التفكير والتفسير والاستنتاج (Adzidzah & Yudiawan, 2024). فالتقويم في هذا السياق لا يهدف فقط إلى قياس الفهم، بل إلى تطويره أيضًا، كما أشار إلى ذلك بلاك وويليام (Black & Wiliam) في نظريتهما حول "التقويم من أجل التعلم (Assessment for Learning) "

وتتضح أهمية هذا النوع من التقويم من منظور نظرية الذكاءات المتعددة لهوارد غاردنر (Gardner's Multiple Intelligences)، حيث تبيّن أن الطلاب يمتلكون قدرات متنوعة، منها اللغوية، والمنطقية-الرياضية، والذكاء الذاتي والاجتماعي. وعليه، فإن الأسئلة التي تتطلب تحليل النصوص وفهم السياقات والاستدلال تدفع إلى تفعيل عدة أنواع من الذكاء، مما يجعل عملية التقويم أكثر شمولية وعدلا في قياس قدرات الطلاب المتنوعة.

ومن جهة التحليل السيكومتري، فإن استخدام معامل KR20 لتحليل الثبات، ومعامل الارتباط النقطي الثنائي لتحليل الصدق التجريبي، يدل على الالتزام بجودة الأداة من حيث الخصائص السيكومترية (Andrich & Marais, 2019)، مما ينسجم مع معايير نظرية القياس الكلاسيكي (Classical Test Theory) التي تؤكد على ضرورة التحقق من صدق وثبات أدوات التقويم لضمان دقة النتائج (Muñiz, 2014; Devi, 2017).

أما على صعيد التعليم العربي خصوصا، فإن نتائج هذه الدراسة تفتح نقاشا حول ضرورة تطوير محتوى اللغة العربية بحيث لا يكون فقط وسيلة لحفظ المفردات والقواعد، بل أداة للتفكير والتحليل. وهذا يتماشى مع المقاربة التواصلية (Communicative Approach) في تعليم اللغات، التي ترى أن اللغة تكتسب بفعالية عندما تستخدم في سياقات حياتية واقعية وتفاعلية، لا عبر الحفظ المجرد كما تتماشى نتائج الدراسة مع ما تنادي به نظرية التعلم القائم على المهام(Task-Based Language Learning)، والتي تركز على تقديم أنشطة حقيقية معقدة تنمي مهارات اللغة والتفكير في آن واحد، مثل تفسير المعاني، واستنتاج العبرة من النص، وتقييم المواقف (Alsubhi et al., 2023). وهذا هو جوهر ما تحققه الأسئلة المبنية على مهارات التفكير العليا.

ورغم قوة المنهجية والنتائج، فإن محدودية العينة تبقى قيدا منهجيا يستوجب التوسيع في الدراسات المستقبلية. كما أن غياب التقييم الطولي يجعل من الصعب الحكم على استدامة أثر هذه الأسئلة في تطوير التفكير لدى المتعلمين.

تشير نتائج هذه الدراسة إلى أن تطوير أدوات تقويم تستند إلى مهارات التفكير العليا في اللغة العربية للمرحلة الابتدائية ليس أمرا ممكنا فحسب، بل هو أيضا ضروري في ضوء الاتجاهات التربوية الحديثة. لقد أصبح من المسلم به في ميدان التربية أن التعلّم لا يجب أن يقتصر على استيعاب المعرفة، بل يجب أن يتعداه إلى تنمية قدرات التفكير النقدي، وحل المشكلات، واتخاذ القرار .(Saepurrohman et al., 2023; Haniefa, 2022) ومن هذا المنطلق، تسعى هذه الدراسة إلى مواءمة أدوات التقويم مع هذه الغايات العليا التي ينادي بها منهج ٢٠١٣ في إندونيسيا.

لقد أظهرت النتائج أن أداء الطلاب في الإجابة على أسئلة تقويم مهارات التفكير العليا كان جيدا نسبيا، وهو مؤشر على قابلية الطلاب في هذه المرحلة العمرية للتفاعل مع مهام معرفية معقدة، شرط أن تصمم بطريقة تراعي خصائصهم النمائية واللغوية (Jamil, 2024). وهذا يتحدى الافتراض الشائع بأن طلاب المرحلة الابتدائية غير قادرين على التفكير على مستوى عالٍ، ويوجّه المعلمين إلى ضرورة إعادة النظر في أساليب التدريس والتقويم المعتمدة.

من ناحية المنهج، فإن اعتماد نموذج 4D (التحديد، والتصميم، والتطوير، والنشر) منح البحث بنية قوية ومتكاملة، سمحت بإنتاج أداة تقويم مدروسة من حيث الصياغة والمحتوى والخصائص السيكومترية. كما أن الجمع بين التحقق النظري والتجريبي يعكس حرص الباحثين على ضمان صلاحية وموثوقية الأداة، مما يُعزز من قابلية استخدامها في بيئات تعليمية مماثلة إلا أن النتائج أظهرت أن الموثوقية العامة للأداة كانت متوسطة (KR20 = 0.59)، ما يستدعي إجراء تحسينات مستقبلية في صياغة البنود أو في عددها وتنوعها، لضمان تجانسها الداخلي ورفع مستوى الثبات. كذلك، فإن صلاحية خمسة بنود فقط من أصل عشرة وفق التحليل الإحصائي تشير إلى الحاجة لإعادة تصميم البنود الأخرى أو مراجعتها من حيث الوضوح والارتباط بمستويات التفكير العليا كما ينبغي الإشارة إلى أن استخدام حجم عينة صغير (٢٠ طالبا فقط) يعد من القيود المنهجية المهمة في هذه الدراسة، حيث يحد من إمكانية تعميم النتائج على جمهور أوسع. ومن المهم أن تجرى دراسات لاحقة باستخدام عينات أكبر وأكثر تنوعا من حيث الخلفية الاجتماعية والجغرافية، كما يمكن إدماج التقييم الطولي لمعرفة مدى استمرارية تأثير هذا النوع من التقويم على أداء الطلاب وتطور مهاراتهم (Qomariyah & Bahrodin, 2024).

من جهة أخرى، تكمن أهمية هذا البحث في فتح آفاق جديدة في مجال تعليم اللغة العربية في المدارس الإسلامية في إندونيسيا، والتي غالبا ما تركز على الجوانب التقليدية من الحفظ والفهم،

دون إيلاء اهتمام كاف لتطوير المهارات العقلية العليا. وهنا يظهر الإسهام الحقيقي للدراسة، إذ إنها تقدم نموذجا عمليا يمكن أن يحتذى به في تصميم أدوات تقويم تراعي مهارات القرن الحادي والعشرين، وتعيد صياغة وظيفة اللغة من أداة حفظ إلى أداة تفكير وبناء معرفة (Ilmiani & Delima, 2021).

وختاما، فإن هذه الدراسة لا تقدم نتائج فقط، بل تفتح أيضا نقاشا تربويا أوسع حول دور التقويم في العملية التعليمية، وتظهر أنه لا يمكن تحقيق أهداف المنهج الحديث ما لم تتطور أدوات التقويم لتكون شريكة في توجيه التعلم لا مجرد وسيلة لقياسه.

بناء ذلك كله، يمكن القول إن نتائج الدراسة لا تقف عند حدود فاعلية أداة تقييم معينة، بل تتجاوزها لتؤكد أن دمج مهارات التفكير العليا في تعليم اللغة العربية هو توجه مستند إلى أصول نظرية قوية، ويتماشى مع الاتجاهات العالمية في إصلاح التعليم، ومع حاجات المتعلم في هذا العصر.

الخاتمة

تشمل مهارات التفكير العليا القدرة على التفكير النقدي والمنطقي والتأملي والمعرفي والإبداعي، أو التعامل مع مشكلات جديدة أو غير مألوفة، أو طرح أسئلة صعبة، أو مواجهة عدم اليقين. في تصتيف بلوم المنقح لأندروسون وكراثو ول، مؤشرات تقييم مهارات التفكير العليا هي التجليل على المستوى المعرفي (**c4**)، والتقييم على المستوى المعرفي (**c5**)، والإبداع على المستوى المعرفي (**c6**)، يتم إعداد أسئلة من خلال تحليل الطفاءات الأساسية، وتجميع الشبكات، واختيار المحفزات السياية، وكتابة بنود الأسئلة، ووضع إرشادات التقييم. بالنسبة للمعلمين، ليس من السهل وضه أسئلة تشجع على التفكير عالي المستوى، لذا يازم التدريب المستمر. ستدعم الأسئلة عالية الجودة تنمية الإبداع والتفكير النقدي لدى الطلاب، وتخريج خريجين أكفاء.

المصادر والمراجع

Abdurrahman, A., Setyaningsih, C. A., & Jalmo, T. (2019). Implementating multiple representation-based worksheet to develop critical thinking skills. *Journal of Turkish Science …*. http://www.tused.org/index.php/tused/article/view/235

Adzidzah, N., & Yudiawan, A. (2024). HOTS-Based Formative Assessment: The Key to Improving The Quality of Learning. Journal of Quality Assurance in Islamic Education. https://doi.org/10.47945/jqaie.v4i2.1670

Alsubhi, A., Bin Mohamed Adnan, M. A., Yusof, A. bin, Awae, F., & Abuhassna, H. (2023). Unlocking the Future of Arabic languages teaching: Exploring Communicative Language Teaching Methodology. International Journal of Academic Research in Business & Social Sciences. https://doi.org/10.6007/ijarbss/v13-i10/19015

Andriani, R. (2020). MODEL PEMBELAJARAN KETERAMPILAN MEMBACA PEMAHAMAN BERBASIS HOTS (Higher Order of Thinking Skill). *Jurnal Riset Pendidikan Dasar Dan Karakter*. https://ojs.adzkia.ac.id/index.php/pdk/article/view/9

Andrich, D., & Marais, I. (2019). Reliability and Validity in Classical Test Theory (pp. 41–53). Springer, Singapore. https://doi.org/10.1007/978-981-13-7496-8_4

Anisah, A., & Lastuti, S. (2018). Pengembangan bahan ajar berbasis HOTS untuk meningkatkan kemampuan pemecahan masalah matematis mahasiswa. *Kreano, Jurnal Matematika Kreatif-Inovatif*. https://journal.unnes.ac.id/nju/index.php/kreano/article/view/16341

Arif, M., & Makalalag, C. (2022). Pengembangan Kurikulum 2013: Pembelajaran Bahasa Arab Di Madrasah Aliyah. *Jurnal Ilmiah Pendidikan Dan Keislaman*. http://jipkis.stai-dq.org/index.php/home/article/view/31

Arif, Z. U., Khalid, M. Y., Zolfagharian, A., & ... (2022). 4D bioprinting of smart polymers for biomedical applications: Recent progress, challenges, and future perspectives. *Reactive and Functional ...*. https://www.sciencedirect.com/science/article/pii/S138151482200219X

Assafi, M. N., Hossain, M. M., Chileshe, N., & ... (2022). Development and validation of a framework for preventing and mitigating construction delay using 4D BIM platform in Bangladeshi construction sector. *Construction ...*. https://doi.org/10.1108/CI-08-2021-0160

Banfi, F., Brumana, R., Salvalai, G., & Previtali, M. (2022). Digital twin and cloud BIM-XR platform development: From scan-to-BIM-to-DT process to a 4D multi-user live app to improve building comfort, efficiency and .... *Energies*. https://www.mdpi.com/1996-1073/15/12/4497

de Quincey, E., Briggs, C., Kyriacou, T., & ... (2019). Student centred design of a learning analytics system. *Proceedings of the 9th ...*. https://doi.org/10.1145/3303772.3303793

Devi, R. S. (2017). Reliability and Validity of a Psychological Test. 4(1), 185–194. http://www.i-scholar.in/index.php/SMU/article/download/138346/126567

Deviana, T., & Kusumaningtyas, D. I. (2019). Analisis kebutuhan penyusunan perangkat pembelajaran tematik berbasis HOTS (higher of order thinking skills) pada kurikulum 2013 di SD Muhammadiyah 05 Batu. *Edumaspul: Jurnal ...*. https://ummaspul.e-journal.id/maspuljr/article/view/141

Fanani, A., & Kusmaharti, D. (2018). Pengembangan pembelajaran berbasis HOTS (higher order thinking skill) di sekolah dasar kelas V. *Jurnal Pendidikan Dasar*. http://journal.unj.ac.id/unj/index.php/jpd/article/view/JPD.91.01

Gozali, I., Lie, A., Tamah, S. M., & Jemadi, F. (2021). HOTS questioning ability and HOTS perception of language teachers in Indonesia. *Indonesian Journal of Applied Linguistics*, *11*(1). https://doi.org/10.17509/ijal.v11i1.34583

Gunawan, I. (2022). *Metode Penelitian Kualitatif: teori dan praktik*. books.google.com. https://books.google.com/books?hl=en&lr=&id=AqSAEAAAQBAJ&oi=fnd&pg=PP1&dq=pendekatan+fenomenologi+metode+penelitian+kualitatif&ots=m2VurnbZrR&sig=PoKJ9cdPSflyteb05zNlztweYW8

Ichsan, I. Z., Sigit, D. V., Miarsyah, M., Ali, A., Arif, W. P., & Prayitno, T. A. (2019). HOTS-AEP: Higher order thinking skills from elementary to master students in environmental learning. *European Journal of Educational Research*, *8*(4). https://doi.org/10.12973/eu-jer.8.4.935

Ilmiani, A. M., & Delima, D. (2021). *Innovation in Learning Arabic Reading Skills using Higher Order Thinking Skills*. 9(1), 99–110. https://doi.org/10.23971/ALTARIB.V9I1.2603

Irvine, J. (2017). A comparison of revised Bloom and Marzano's new taxonomy of learning. *Research in Higher Education Journal*, *33*.

Jamil, M. A. (2024). *Development of arabic language tests using wondershare quiz creator at stit bustanul ulum, central lampung*. At-Turots, 735–746. https://doi.org/10.51468/jpi.v6i2.765

Kiliańska-Przybyło, G. (2019). Lia Litosseliti (ed.), Research Methods in Linguistics . London: Bloomsbury Academic, 2018. *Theory and Practice of Second Language …*. https://www.ceeol.com/content-files/document-975069.pdf

Kouicem, K. (2020). Constructivist Theories Of Piaget And Vygotsky: Implications For Pedagogical Practices. 13(3), 359–372. https://www.asjp.cerist.dz/en/downArticle/117/13/3/125613

Li, H., Yan, G., Luo, W., Liu, T., Wang, Y., Liu, R., & ... (2021). Mapping fetal brain development based on automated segmentation and 4D brain atlasing. *Brain Structure and …*. https://doi.org/10.1007/s00429-021-02303-x

Lubis, M. A., & Azizan, N. (2019). *Pembelajaran Tematik SD/MI Implementasi Kurikulum 2013 Berbasis HOTS (Higher Order Thinking Skills)*. books.google.com. https://books.google.com/books?hl=en&lr=&id=il3EEAAAQBAJ&oi=fnd&pg=PA1&dq=hight+order+thinking+skil+keterampilan+membaca&ots=ypYVIMFFew&sig=AucbuXiHCUqWvOUpkGQPzYyg96I

Lutfiyatun, E., & Haniefa, R. (2022). Pengembangan Hots Online Assessment Dengan Quizizz Berkarakter Untuk Keterampilan Membaca Pemahaman. *JURNAL Al-AZHAR INDONESIA SERI …*. https://jurnal.uai.ac.id/index.php/SH/article/view/1587

Maulana, R. (2022). Analisis Capaian Pembelajaran Bahasa Arab dengan Taksonomi Bloom Revisi. *Jurnal PTK Dan Pendidikan*. https://jurnal.uin-antasari.ac.id/index.php/ptkpend/article/view/7621

Muñiz, J. (2014). Classical Test Models. 1–7. https://doi.org/10.1002/9781118445112.STAT06386

Mustofa, S., Desrani, A., & Ritonga, A. W. (2022). HOTS in Arabic Learning: A Study of The Implementation of HOTS on Students' Critical Thinking Ability. Al-Ta'rib, 10(2), 133–144. https://doi.org/10.23971/altarib.v10i2.4088

Nafiati, D. A. (2021). Revisi taksonomi Bloom: Kognitif, afektif, dan psikomotorik. *Humanika*, *21*(2). https://doi.org/10.21831/hum.v21i2.29252

Naryatmojo, D. L. (2018). Penggunaan Taksonomi Bloom dalam Pembelajaran Keterampilan Menyimak Bermuatan Pendidikan Karakter Profetik Untuk Mengukur Keberhasilan Hasil Belajar …. *Proceeding Seminar Nasional Pertemuan* …. https://proceeding.unikal.ac.id/index.php/pibsi40/article/view/126

Pujawan, I. G. N., Rediani, N. N., Antara, I. G. W. S., Putri, N. N. C. A., & Bayu, G. W. (2022). REVISED BLOOM TAXONOMY-ORIENTED LEARNING ACTIVITIES TO DEVELOP SCIENTIFIC LITERACY AND CREATIVE THINKING SKILLS. *Jurnal Pendidikan IPA Indonesia*, *11*(1). https://doi.org/10.15294/jpii.v11i1.34628

Qomariyah, L., & Bahrodin, A. (2024). Developing An Adventure Game For Teaching Arabic To Develop Critical Thinking Skills Of Elementary School Students/ تطوير اللعبة المغامرات حبا لتعليم اللغة العربيةفي تنمية مهارة التفكير الناقذ لدى طلاب المرحلة الإبتدائية. Ijaz Arabi, 7(3). https://doi.org/10.18860/ijazarabi.v7i3.26098

Saepurrohman, A., Sunarya, Y., & Majid, M. N. (2023). Higher Order Thinking (HOTS) Skills Assessment Model in Arabic Language Skills Learning in Madrasah. 2(1), 1–15. https://doi.org/10.15575/ta.v2i1.24150

Sihombing, B. (2024). Model Pengembangan 4D (Define, Design, Develop, dan Disseminate) dalam Pembelajaran Pendidikan Islam. *Journal of Islamic Education El* …. http://journal.marwah-madani-riau.id/index.php/JIEE/article/view/135

Sood, S. K., Lamba, Y. S., & Singh, A. K. (2024). Progress and prospects of 3d and 4d printing in sustainable development: A scientometric review. *IEEE Transactions on* …. https://ieeexplore.ieee.org/abstract/document/10663245/

Stefano, M. Di, Paulsen, J., Jost, D., & ... (2021). 4D nucleome modeling. *… in Genetics &development*. https://www.sciencedirect.com/science/article/pii/S0959437X2030143X

Sugiyono. (2019). Pengertian Metode Penelitian Kuantitatif dan Kualitatif. In *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*.

Veronica, A., Abas, M., Hidayah, N., Sabtohadi, J., & ... (2022). *Metodologi Penelitian Kuantitatif*. Get Press.

Vieira, P. C. (2016). T-test with likert scale variables. *Available at SSRN 2770035*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2770035

Wenno, E. C., & Karuna, K. (2021). Hots (High Order Thinking Skill) Dalam Tes Bahasa Jerman. *J-EDu: Journal-Erfolgreicher ….* https://ojs3.unpatti.ac.id/index.php/jedu/article/view/3176

Yusella, Y., Suciptaningsih, O. A., & Degeng, M. D. K. (2023). Development Of Hots-Based Assessment in the Curriculum to Increase Critical Thinking Skills of Elementary School Students. JIIP (Jurnal Ilmiah Ilmu Pendidikan), 6(7), 4991–4998. https://doi.org/10.54371/jiip.v6i7.2036

Yusuf, M. (2023). Development of Arabic Language Teaching Materials With 4D Model for the Second Semester at STAI Al-Furqan Makassar. *Bulletin of Science Education*. https://attractivejournal.com/index.php/bse/article/view/662

Zulkifli, M. (2018). Analisis Bentuk Evaluasi Kurikulum 2013 Mata Pelajaran Bahasa Arab Di MI. *… -Madrasah: Jurnal Ilmiah Pendidikan Madrasah ….* https://jurnal.stiq-amuntai.ac.id/index.php/al-madrasah/article/view/29