Al-Lahjah : Jurnal Pendidikan, Bahasa Arab dan Kajian Linguistik Arab
Vol. 9 No. 1, Januari 2026, Halaman 1355-1371
E-ISSN: 2615-5656

Tahlil Bunud As'ilah Ikhtibariyah li al-Shaff al-Tsamin fi al-Madrasah al-Mutawassitah al-Islamiyyah al-Hukumiyyah 6 Lima Puluh Kota

Muhamad Ruli Aulia[1*], Rita Febrianta[2], Yelfi Dewi[3], Eka Rizal[4]
[1,2,3,4] Arabic Language, UIN Sjech M. Djamil Djambek Bukittinggi, Indonesia
*Email: mhdruliaulia130902@gmail.com
Phone Number (WhatsApp): 0838 4949 5845

ABSTRACT

The problem in this research lies in the Arabic language test questions at MTsN 6 Lima Puluh Kota, which have not yet been analyzed. This study aims to determine the quality of Arabic language test items for Grade VIII at MTsN 6 Lima Puluh Kota based on validity, reliability, difficulty level, discriminating power, and distractor effectiveness. The research employs a descriptive quantitative approach. The participants are students from two eighth-grade classes, 8¹ and 8², totaling 63 students. Data were collected through documentation and observation. The data analysis techniques are based on validity, reliability, difficulty level, discriminating power, and distractor effectiveness using Microsoft Excel. The results of the analysis of 30 test items show that 29 items (97%) have good validity, while one item (3%) is invalid. The reliability index is 0.897, indicating high reliability. Regarding the difficulty level, 2 items (7%) are classified as easy, 26 items (86%) as moderate, and 2 items (7%) as difficult. The discriminating power analysis shows that 1 item (3%) needs improvement, 4 items (13%) are fair, 14 items (54%) are good, and 9 items (30%) are very good. In terms of distractor effectiveness, 5 items (13%) are less effective and need revision, while 25 items (87%) are effective and suitable for use. It can be concluded that the Arabic test items for Grade VIII at MTsN 6 Lima Puluh Kota are of fairly good quality and can be reused.

***Keywords**: Arabic Language; Exam Questions; Item Analysis*

الملخص

المشكلة في هذا البحث توجد في أسئلة الاختبار اللغة العربية في المدرسة المتوسطة الإسلامية الحكومية ٦ ليما فولوه كوتا التي لم يتم تحليلها لاختبار اللغة العربية للصف الثامن. يهدف هذا البحث إلى معرفة جودة بنود أسئلة اختبار اللّغة العربيّة للصّف الثّامن في المدرسة المتوسّطة الإسلاميّة الحكوميّة ٦ ليما بولوه كوتا من حيث أبعاد تحليل الصلاحية والثبات ومستوى الصعوبة وقوة التمييز والفعالية المخادعة. وهذا البحث هو يستخدم هذا البحث التّحليل الكمي الوصفي. موضوع هذا البحث هو أسئلة اختبار اللّغة العربيّة للصّف الثّامن في المدرسة المتوسّطة الإسلاميّة الحكوميّة ٦ ليما بولوه كوتا. بينما كان المشاركون في هذا البحث طلاب الصف الثامن في فصلين دراسيين هما ٨١ و٨٢، بمجموع ٦٣ طالبًا وطالبة .استخدمت في عملية جمع البيانات أساليب التوثيق والملاحظة. تقنيات تحليل البيانات في هذا البحث من حيث أبعاد تحليل الصلاحية والثبات ومستوى الصعوبة وقوة التمييز والفعالية المخادعة باستخدام تطبيق Microsoft Excel. أظهرت نتائج هذا البحث على ٣٠ بنداً أن ٢٩ سؤالاً لها صلاحية جيدة (٩٧٪)، وسؤال واحد غير صلاحية (٣٪)، يظهر من الثبات السؤال أن هناك سؤال واحد (٣٪) له الثبات عالية بمؤشر ٠,٨٩٧، يظهر من مستوى الصعوبة أن هناك سؤالين (٧٪) لهما فئة سهلة, ٢٦ سؤالاً (٨٦٪) فئة متوسطة و٢ (٧٪) سؤالين (٧٪) فئة صعبة، يتضح من قوة التمييز أن هناك سؤال واحد (٣٪) لديه قوة تمييز تحتاج إلى تصحيح، و٤ أسئلة (١٣٪) لديها قوة تمييز كافية،

و١٤ سؤالاً (٥٤٪) لديها قوة تمييز جيدة و٩ أسئلة (٣٠٪) لديها قوة تمييز ممتازة. أما بالنسبة للفعالية المخادعة، فإن ٥ أسئلة (١٣٪) لديها الفعالية المخادعة غير مواتية تحتاج إلى مراجعة، و٢٥ سؤالاً (٨٧٪) لديها الفعالية المخادعة جيدة يمكن استخدامها. يمكن استنتاج أسئلة اختبار اللّغة العربيّة للصّف الثّامن في المدرسة المتوسّطة الإسلاميّة الحكوميّة ٦ ليما بولوه كوتا بجودة جيدة بما فيه الكفاية ويمكن إعادة استخدامها.

الكلمات الرئيسية: *تحليل البنود، أسئلة الاختبار، اللغة العربية*

المقدمة

التّعليم هو جهد موجه ومنهجي يهدف إلى رعاية الإمكانات الموجودة لدى الأفراد. ووفقاً للمعجم الإندونيسي الكبير، يشمل التّعليم عمليّة تحويل مواقف وسلوك الأفراد أو الجماعات لتعزيز النّضج البشري من خلال أنشطة التّعليم والتعلّم (Ida & Musyarofah, 2021). إن الغرض الرّئيسي من التّعليم هو بناء أساس للذكاء والمعرفة والشّخصية الجيّدة والمهارات اللازمة للعيش باستقلالية ومواصلة التّعليم. من ناحية أخرى، يعمل التّقييم كأداة لتحديد مدى تحقيق نتائج التّعلم للأهداف المحددة. وبدون المرور بعملية التّقييم، لا يمكننا معرفة ما إذا كان التعلّم ناجحًا أم لا. (Dewi et al., 2023).

ينقسم التعليم إلى تعليم رسمتي وغير رسمي. يتم تنفيذ التعليم الرسمي في مؤسسة تعليمية نسميها عادة مدرسة بها المعلمين والطلاب ومرافق وبنية تحتية ويحتاج المعلمون في التدريس إلى الوسائل التعليمية كأداة حتى تتم عملية التعليم بسلامة (Febrianta & Yusra, 2023).

هناك عنصران أساسيان في مجال التّعليم هما المعلّمون والطّلاب. يقوم المعلّمون بتوجيه الطّلاب في رحلة تعلّمهم، في حين أن الطّلاّب هم المتلقون للمعرفة. ووفقًا لما ذكره غوزا في مجلّة عبديماس أونواس، فإن معايير كفاءات المعلّمين تشمل أربع كفاءات رئيسيّة، وهي الكفاءة التّربوية، والكفاءة الشّخصية، والكفاءة الاجتماعيّة، والكفاءة المهنية (Fitria et al., 2019).

تشير الكفاءة التّربوية إلى قدرة المعلّم على تيسير تعلّم الطّلاّب بفعاليّة، وتغطي عدة مجالات رئيسيّة. وتشمل هذه المجالات، الفهم المتين للمبادئ التّعليمية، النّظرة الثّاقبة لاحتياجات الطّلاّب ووجهات نظرهم، القدرة على تطوير المناهج الدّراسيّة والمنهج الدّراسي، تخطيط خبرات التّعلّم الجذّابة، تنفيذ إستراتيجيات التّعلّم التّعليميّة والحوارية، استخدام تكنولوجيا التّعلّم، تقييم نتائج التّعلّم، دعم المتعلّمين وتطويرهم لمساعدتهم على تحقيق إمكاناتهم المتنوّعة (Wardani & Saring Marsudi, 2017). ولذلك، فإن إحدى الكفاءات المهمّة للمعلّمين هي القدرة على التّقييم الفعال، سواء أثناء عملية التّعلّم أو في تقييم نتائج تعلم طلابهم.

يرى سوكاردي أن جودة التّعليم ترتبط ارتباطاً وثيقاً بأنشطة التّعلّم الّتي تتم في المدارس. وتشمل عملية التّعلّم ثلاثة عناصر رئيسيّة: أهداف التّعليم، وخبرات التّعليم والتّعلّم، ونتائج أنشطة التّعلّم أو التّقييم (Magdalena et al., 2021). التّقييم هو عمليّة جمع المعلومات لتقييم أداء الطّلاّب في التّعلّم. وتُستخدم هذه العملية بشكل عام لتقييم مخرجات التّعلّم وتغطي جوانب مختلفة، بما في ذلك الجوانب المعرفيّة والعاطفية

والنّفسية الحركية. التقييم هو عمليات تلخيصية، يلعب فيها الحكم على قيمة الشيء دورا كبيرا كما هو الحال في إعطاء التقديرات للطلاب ونقلهم من صف لآخر(Ali, 2011) .

التّقييم هو عمليّة جمع المعلومات لتقييم أداء الطّلاّب في التّعلّم. وتُستخدم هذه العملية بشكل عام لتقييم مخرجات التّعلّم وتغطي جوانب مختلفة، بما في ذلك الجوانب المعرفيّة والعاطفية والنّفسية الحركية. التقييم هو عمليات تلخيصية، يلعب فيها الحكم على قيمة الشيء دورا كبيرا كما هو الحال في إعطاء التقديرات للطلاب ونقلهم من صف لآخر (Dewi et al., 2024).

الاختبار هو أحد أدوات القياس في تقييم التّعلّم. ووفقًا لجرونلاند ولين، فإن الاختبار هو مجموعة من الأسئلة. تحليل الاختبار هو محاولة لتحديد جودة الاختبار الذي سيتم استخدامه. وترتبط هذه العملية ارتباطاً وثيقاً بتحليل البنود، وهو عبارة عن مراجعة الأسئلة في الاختبار من أجل الحصول على مجموعة أسئلة ذات نوعية جيدة. هناك نوعان من تحليل بنود أسئلة، وهما تحليل الصعوبة وتحليل المفاضلة، بالإضافة إلى النظر في جوانب الصلاحية والثبات (Arifianto et al., 2021).

استنادًا إلى الدرسة السابقة أجرتها تريميلن بالموضوع "تحليل فروق أسئلة الامتحان النهائي بين الصف المتميز والصف العادي في المدرسة المتوسطة الحكومية ٣ ليما فولوه كوتا". إن معادلة هذا البحث مع ذلك البحث هي أن كلا البحثين يقوم بتحليل البنود باستخدام نفس أسلوب التحليل مع اختلاف بسيط في البحث عنها. ويكمن الاختلاف في أهداف البحث، فهذا البحث يهدف إلى تحديد نوعية البنود بينما يهدف البحث السابق إلى تحديد الفرق في نوعية الأسئلة بين الصف المتميز والصف العادي.

أما بالنسبة للبحث الذي أجرتها تري إيدا أماليا "تحليل عناصر سؤال تقييم نهاية الفصل الدراسي الأول الذي أجراه برنامج ماجستير حديث القرآن الكريم للصف التاسع المتوسط في كاليونغو كندال للعام الدراسي ٢٠١٩/٢٠٢٠". معادلة هذا البحث هي نفسها معادلة تحليل بنود الأسئلة. ويكمن الاختلاف في أخذ مجتمع يشمل مدرستين وأنواع الأسئلة التي تم تحليلها، وهي أسئلة الاختيار من متعدد ومقالات لمادتي القرآن الكريم والحديث الشريف، وهذه الأسئلة من إعداد فريق من المعلمين (AMALIA, 2019).

وأما بالنسبة للبحث الذي أجرتها فيبريانا سري راهايو" تحليل أسئلة تقييم نهاية العام الدراسي في مادة اللغة العربية للصف الرابع في الصف الرابع في مقاطعة بايوماس معارف نيو باغيراجي سيلونجوك". يتمثل التشابه مع البحث الحالي في أن كلاهما يحلل جودة الأسئلة. والاختلاف مع البحث الحالي هو أن هذا البحث تحلل الأسئلة باستخدام تطبيق SPSS بينما البحث الحالي يستخدم تطبيقات Microsoft Excel، ومن حيث ما يتم تحليله فإن هذا البحث يحلل حتى قوة التفريق بين السؤال بينما البحث الحالي تحلل حتى تحليل فعالية مدقق الأسئلة والصيغة المستخدمة هناك اختلافات (Febrianta et al., 2022).

في تعليم اللّغة العربيّة هناك مكونات لغوية ومهارات لغوية مختلفة. وتشمل المكونات اللّغوية المفردات والنحو والصرف والبلاغة، بينما تنقسم المهارات اللغوية إلى أربع مهارات هي مهارات الاستماع والتحدث والقراءة والكتابة (Febrianta et al., 2022). في مواد اللّغة العربيّة، غالبًا ما يمثل التّقييم تحديًا في مواد اللّغة العربيّة، خاصةً في تقييم قدرات الطّلاّب في المهارات اللّغويّة مثل القراءة والكتابة والاستماع والمحادثة.

وبالإضافة إلى ذلك، تتطلب مواد اللّغة العربيّة المعقدة إعداداً دقيقاً لأسئلة الاختبارات لتكون ذات صلة حقيقية بالكفاءات المراد قياسها. وتعتبر الصّلاحيّة والثّبات جانبين رئيسيين يحددان جودة سؤال الاختبار.

في المدرسة المتوسّطة الإسلاميّة الحكوميّة ٦ ليما بولوه كوتا، تتمّ عملية إعداد أسئلة امتحان اللّغة العربيّة للصّف الثّامن في الفصل الدّراسي الأول من قبل المعلّمين بشكل مستقل. ومع ذلك، لم يتم إجراء تحليل رسمي لجودة الأسئلة التي تم إعدادها. وهذا أمر مثير للقلق لأنه بدون تحليل لجودة الأسئلة، يمكن أن تكون نتائج التقييم غير دقيقة ويصعب استخدامها كأساس لاتخاذ القرارات في عملية التّعلّم.

بناءً على الخلفية البحث أعلاه، فإن صيغ المشكلة في هذه الدراسة هي كيف جودة أسئلة اختبار اللغة العربية للصّف الثّامن في المدرسة المتوسّطة الإسلاميّة الحكوميّة ٦ ليما بولوه كوتا من حيث الصلاحية والثبات ومستوى الصعوبة وقوة التمييز والفعالية المخادعة. إذًا يهدف هذا البحث إلى معرفة جودة بنود أسئلة اختبار اللّغة العربيّة للصّف الثّامن في المدرسة المتوسّطة الإسلاميّة الحكوميّة ٦ ليما بولوه كوتا من حيث أبعاد تحليل الصلاحية والثبات ومستوى الصعوبة وقوة التمييز والفعالية المخادعة. وبالتالي، يُتوقَّع أن يُسهم هذا البحث إسهامًا كبيرًا في تحسين جودة إعداد فقرات أسئلة مادة اللغة العربية، خصوصًا في الصف الثامن، وأن يكون مرجعًا للمعلمين في تقييم وتحسين جودة الأسئلة المستخدمة في عملية التقييم التعليمي.

منهج البحث

يستخدم هذا البحث التّحليل الكمي الوصفي. التّحليل الكمي الوصفي هو تحليل إحصائي يستخدم لوصف البيانات الكمية وتلخيصها وتحليلها. والبيانات الكمية هي البيانات التي يمكن قياسها أو حسابها باستخدام الأرقام، مثل العمر والوزن والطول وما إلى ذلك. والغرض من التّحليل الوصفي الكمي هو تقديم وصف واضح ومفصل للبيانات التي تم جمعها، بحيث يمكن أن يسهل تفسيرها واتخاذ القرارات بناءً على البيانات الموجودة (Sudirman et al., 2023).

في هذه الدراسة، تم استخدام المنهج الكمي الوصفي لوصف جودة أسئلة اللغة العربية المستخدمة في اختبار اللّغة العربيّة للصّفّ الثّامن في الفصل الدّراسي الأوّل في المدرسة المتوسّطة الإسلاميّة الحكوميّة ٦ ليما فولوه كواتا من حيث صلاحية، الثّبات، مستوى الصعوبة، قوة التمييز، فعالية المخادعة. في هذا البحث تم إجراء هذا البحث من خلال اتخاذ مكان أو موقع في المدرسة المتوسّطة الإسلاميّة الحكوميّة ٦ ليما فولوه كواتا.

مصادر البيانات المستخدمة في هذه الدراسة هي أسئلة الاختبارات، ومفاتيح الإجابة، وأوراق الإجابة الفصل الدّراسي الأوّل للصّف الثّامن في العام الدّراسي ٢٠٢٤/٢٠٢٥. وأدوات هذا البحث الملاحظة والتوثيق. أما طريقة تحليل البيانات باستخدام برنامج Microsoft Excel.

نتائج البحث ومناقشتها

استُخدمت هذه الدراسة لتحديد جودة بنود اختبار الفصل الدراسي الأول المواد اللغة العربية للصف الثامن في المدرسة المتوسّطة الإسلاميّة الحكوميّة ٦ ليما بولوه كوتا في العام الدّراسي ٢٠٢٤/٢٠٢٥ من حيث

الصلاحية والثّبات ومستوى الصعوبة والتمييز والفعالية المخادعة. البيانات المستخدمة في شكل بنود اختبار الفصل الدراسي الاول لمواد اللغة العربية للصف الثامن في المدرسة المتوسّطة الإسلاميّة الحكوميّة ٦ ليما بولوه كوتا في العام الدّراسي ٢٠٢٤/٢٠٢٥، حيث تم تحليل الأسئلة في شكل ٣٠ سؤالاً موضوعياً في شكل اختيار من متعدد، وبلغ مجموع الطلاب الذين تم استخدامهم كبيانات ٦٣ طالباً تم أخذهم من صفين هما الصف ٨١ بإجمالي ٣٢ طالباً والصف ٨٢ بإجمالي ٣١ طالباً.

نتائج البحث

الصلاحية

الصّلاحيّة مشتقة من كلمة" الصّلاحيّة"، وهي تشير إلى مدى دقة ودقة أداة القياس (الاختبار) في أداء وظيفتها. ويُعتبر الاختبار ذا صلاحيّة عالية إذا كان قادراً على أداء هذه الوّظيفة بدقة، بحيث ينتج نتائج قياس تتوافق مع الغرض من القياس الّذي يتم إجراؤه. بعبارة أخرى، يجب أن تعكس نتائج القياس بالضبط الوّقائع أو الظّروف الحقيقية لما يتم قياسه(Ramadhan et al., 2024).

تحليل صلاحية بنود أسئلة باستخدام صيغة الارتباط العزمي للمنتج. الصيغة المستخدمة هي:

$$rxy = \frac{\sum xy - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

الشرح:

rxy: معامل الصلاحية

n: العديد من الموضوعات

x: مجموع درجات البنود

y: مجموع الدرجات الكلية للبنود

يتم حساب الصلاحية باستخدام معادلة ارتباط عزم الضرب (rxy). عدد الطلبة في الصفين ٨٢ و٨٣ هو ٦٣ شخصاً، بناءً على عدد أفراد البحث، أي ٦٣ شخصاً، وبالنظر إلى معادلة r-tabel عند مستوى الدلالة ٥٪، فإن n-٢ من ٦٣ شخصاً هو ٦١ شخصاً، وبالتالي فإن عدد r-tabel هو ٠,٢٥. نتائج البحث على تحليل البنود بناءً على المعيار إذا كان r-tabel أكبر من ٠,٢٥ فيتم إعلان صلاحية، أما إذا كان r-tabel أقل من ٠,٢٥ فيتم إعلان غير صلاحية.

فيما يلي نتائج اختبار صلاحية جودة بنود أسئلة اختبار اللّغة العربيّة للصّف الثّامن في المدرسة المتوسّطة الإسلاميّة الحكوميّة ٦ ليما بولوه كوتا.

الجدول ١. نتائج مخرجات المشكلة بناءً على مستوى الصلاحية

رقم السؤال	r- hitung (rxy)	r- tabel	الوصف
١	٠,٣٩	٠,٢٥	صلاحية
٢	٠,٦٠	٠,٢٥	صلاحية
٣	٠,٤٦	٠,٢٥	صلاحية

رقم السؤال	r- hitung (rxy)	r- tabel	الوصف
٤	٠,٥٠	٠,٢٥	صلاحية
٥	٠,٤٣	٠,٢٥	صلاحية
٦	٠,٢٩	٠,٢٥	صلاحية
٧	٠,٦٧	٠,٢٥	صلاحية
٨	٠,٤٢	٠,٢٥	صلاحية
٩	٠,٥٩	٠,٢٥	صلاحية
١٠	٠,٥٣	٠,٢٥	صلاحية
١١	٠,٦١	٠,٢٥	صلاحية
١٢	٠,٦٣	٠,٢٥	صلاحية
١٣	٠,٥٣	٠,٢٥	صلاحية
١٤	٠,٥٠	٠,٢٥	صلاحية
١٥	٠,٤٥	٠,٢٥	صلاحية
١٦	٠,٦٠	٠,٢٥	صلاحية
١٧	٠,٥٩	٠,٢٥	صلاحية
١٨	٠,٥٠	٠,٢٥	صلاحية
١٩	٠,٥٨	٠,٢٥	صلاحية
٢٠	٠,٣٩	٠,٢٥	صلاحية
٢١	٠,٦١	٠,٢٥	صلاحية
٢٢	٠,٤٩	٠,٢٥	صلاحية
٢٣	٠,٥٨	٠,٢٥	صلاحية
٢٤	٠,٦١	٠,٢٥	صلاحية
٢٥	٠,٤٢	٠,٢٥	صلاحية
٢٦	٠,٣٠	٠,٢٥	صلاحية
٢٧	٠,١١	٠,٢٥	غير صلاحية
٢٨	٠,٦١	٠,٢٥	صلاحية
٢٩	٠,٣٧	٠,٢٥	صلاحية
٣٠	٠,٤٨	٠,٢٥	صلاحية

الجدول ٢. نسبة تحليل صلاحية السؤال

الرقم	جودة صلاحية	سؤالا	نسبة مئوية
١	صلاحية	١-٢٦ و٢٨-٣٠	٩٧٪
٢	غير صلاحية	٢٧	٣٪

استنادًا إلى الجدول أعلاه، فإن نتائج تحليل ٣٠ بنود أسئلة الاختيار من متعدد حول الاختبار للفصل الدراسي الاول في المدرسة المتوسّطة الإسلاميّة الحكوميّة ٦ ليما بولوه، يمكن ملاحظة أن هناك سؤالاً واحداً (٣٪) من أسئلة الاختيار من متعدد تم الإعلان عن غير صلاحية والأسئلة التي تم الإعلان عن صلاحية هي ٢٩ سؤالاً (٩٧٪).

الثبات

الثّبات هي مستوى اتساق أداة القياس في تقديم نفس النّتائج في القياسات المتكررة. ووفقًا لسوجييونو، الثّبات هي ثابت أداة القياس الّتي تُظهر مدى توفير الأداة لنتائج متسقة عند استخدامها في نفس الموقف. في تقييم التّعلّم، تُظهر الموثوقي ثبات نتائج الاختبار بمرور الوقت أو عبر مجموعات مختلفة من المتعلّم (Zuhri et al., 2024) .

تم إجراء اختبار الموثوقية باستخدام صيغة KR-20 باستخدام تطبيق Microsoft Excel في حساب موثوقية الاختبار. الصيغة المستخدمة هي:

$$kr20 = \frac{k}{k-1}(1 - \Sigma\frac{pi.qi}{S^2})$$

الشرح:

k : عدد الأسئلة

piqi: التباين الإجمالي للدرجات

pi: نسبة الإجابات الصحيحة

qi: نسبة الإجابات الخاطئة

S^2: الاستجابة الكلية لأحذية فاريان

يتم اختبار موثوقية الاختبار لتحديد مستوى ثبات أو اتساق الاختبار. في هذه الدراسة استخدم الباحثون معادلة kr20. تتراوح درجات اختبار KR-20 من ٠ إلى ١، حيث أن ٠ تعني عدم الموثوقية و١ تعني الثبات الكاملة. وكلما كانت الدرجة أقرب إلى ١، كلما كان الاختبار أكثر ثبات. يعتمد ما يشكل درجة KR-20 المقبولة على نوع الاختبار. بشكل عام، عادة ما تعتبر الدرجات التي تزيد عن ٠,٥ معقولة.

بناءً على التحليل الذي تم إجراؤه، فإن سؤال الاختبار يحتوي على r11 أكثر من ٠,٥، أي ٠,٩٠، وهي فئة الثبات التي يمكن أن تكون عالية جداً، وفيما يلي جدول الشرح:

الجدول ٣. تحليل أسئلة اختبار النهائي من حيث الثبات

مجموع الأسئلة	r11	الفئة
٣٠ سؤالا	٠,٨٩٧	عالية

استناداً إلى نتائج الجدول أعلاه، يمكن ملاحظة أن الأسئلة الثلاثين التي تم تحليلها حصلت على معدل موثوقية مرتفع يبلغ ٠,٨٩٧ وهو ما يندرج ضمن فئة الموثوقية العالية.

مستوى الصعوبة

السؤال الجيد هو سؤال متوازن، ليس سهلاً للغاية وليس صعبًا للغاية. فالمسائل السهلة للغاية لا تشجع الطلاب على بذل المزيد من الجهد لحلها. ومن ناحية أخرى، يمكن للمسألة التي تكون صعبة للغاية أن تجعل الطلاب يشعرون بالإحباط ويفقدون الرغبة في المحاولة مرة أخرى لأنها تعتبر فوق طاقتهم (Rahman & Nasryah, 2019). مؤشر الصعوبة هو رقم يعكس مستوى صعوبة السؤال. تتراوح قيمة مؤشر الصعوبة من (٠,٠٠) إلى (١,٠). يعطي هذا الرقم فكرة عن مدى صعوبة أو سهولة السؤال. على سبيل المثال، يشير السؤال الذي تبلغ قيمة مؤشر صعوبته (٠,٠) إلى أن السؤال صعب للغاية، بينما يشير المؤشر (١,٠) أن السؤال سهل للغاية. في سياق التقييم، يُرمز إلى مؤشر الصعوبة هذا بالحرف P (P بحرف P كبير)، والذي يرمز إلى كلمة" نسبة".

معادلة إيجاد مستوى الصعوبة (P) هي:

$$P = \frac{B}{JS}$$

الشرح:

P: مستوى الصعوبة

B: عدد الطلاب الذين أجابوا على السؤال بشكل صحيح

JS: إجمالي عدد الطلاب المتقدمين للاختبار

يهدف اختبار الصعوبة إلى تحديد مقدار مستوى الصعوبة في السؤال. واستناداً إلى نتائج تحليل أسئلة امتحان اللغة العربية للصف الثامن في المدرسة المتوسّطة الإسلاميّة الحكوميّة ٦ ليما بولوه، تم تفسير نتائج التحليل إلى ثلاثة معايير وفقاً لمؤشر الصعوبة وهي الصعب والمتوسط والسهل. فيما يلي نتائج اختبار صلاحية جودة بنود أسئلة اختبار اللّغة العربيّة للصّف الثّامن في المدرسة المتوسّطة الإسلاميّة الحكوميّة ٦ ليما بولوه كوتا.

الجدول ٤. تحليل أسئلة اختبار النهائي من حيث مستوى الصعوبة

رقم السؤال	المجموع الصحيح	مستوى الصعوبة	التفسير
١	٤٧	٠,٧٥	السهلة
٢	٢٠	٠,٣٢	المتوسطة
٣	٣٣	٠,٥٢	المتوسطة
٤	٢٥	٠,٤٠	المتوسطة
٥	٣٣	٠,٥٢	المتوسطة
٦	٥٢	٠,٨٣	السهلة

رقم السؤال	المجموع الصحيح	مستوى الصعوبة	التفسير
٧	٢٣	٠,٣٧	المتوسطة
٨	٣١	٠,٤٩	المتوسطة
٩	٣٧	٠,٥٩	المتوسطة
١٠	٢٧	٠,٤٣	المتوسطة
١١	٣٠	٠,٤٨	المتوسطة
١٢	٢٩	٠,٤٦	المتوسطة
١٣	٣٠	٠,٤٨	المتوسطة
١٤	٢٨	٠,٤٤	المتوسطة
١٥	٢٩	٠,٤٦	المتوسطة
١٦	٣٠	٠,٤٨	المتوسطة
١٧	٢٧	٠,٤٣	المتوسطة
١٨	٢٧	٠,٤٣	المتوسطة
١٩	٢٦	٠,٤١	المتوسطة
٢٠	٣٠	٠,٤٨	المتوسطة
٢١	٢٩	٠,٤٦	المتوسطة
٢٢	٣٦	٠,٥٧	المتوسطة
٢٣	٢٢	٠,٣٥	المتوسطة
٢٤	٢٧	٠,٤٣	المتوسطة
٢٥	١٩	٠,٣٠	الصعبة
٢٦	٣٠	٠,٤٨	المتوسطة
٢٧	٢١	٠,٣٣	المتوسطة
٢٨	٢٠	٠,٣٢	المتوسطة
٢٩	١٩	٠,٣٠	الصعبة
٣٠	٣٢	٠,٥١	المتوسطة

الجدول ٥. مؤشر المشكلات حسب مستوى الصعوبة

الرقم	جودة مستوى الصعوبة	سؤالا	نسبة مئوية
١	السهلة	١,٦	٧٪
٢	المتوسطة	٢,٣,٤,٥,٧,٨,٩,١٠,١١,١٢,١٣,١٤,١٥,١٦, ١٧,١٨,١٩,٢٠,٢١,٢٢,٢٣,٢٤	٨٦٪
٣	الصعبة	٢٥,٢٩	٧٪

استنادًا إلى الجدول أعلاه، تم توزيع بنود الأسئلة الثلاثين على أساس مستوى الصعوبة باستخدام تطبيق Microsoft Excel مع النتائج التي تم الحصول عليها. وبلغت الأسئلة المصنفة على أنها ذات مستوى صعب سؤالين (٧٪) بمؤشر (٠,٠٠-٠,٣٠)، أما الأسئلة المصنفة على أنها ذات مستوى متوسط فبلغت ٢٦ سؤالاً (٨٦٪) بمؤشر (٠,٣١-٠,٧٠)، وبلغت الأسئلة المصنفة على أنها سهلة سؤالين (٧٪) بمؤشر (٠,٧١-١,٠٠).

قوة التمييز

تتمثل القوة التمييز للسؤال في قدرة السؤال على التمييز بين المتعلمين ذوي التحصيل العالي والمتعلمين ذوي التحصيل المتدني. هذا الجانب مهم جدًا في قياس نتائج تعلم الطلاب، ويوفر معلومات قيمة للمعلمين لتحسين جودة العناصر الموجودة. من خلال تحليل القدرة على التمييز، يمكن للمعلمين تحديد ما إذا كان العنصر قد حقق معيارًا جيدًا أو يحتاج إلى مراجعته أو حتى ينبغي رفضه وعدم استخدامه مرة أخرى. (Dony Handriawan, 2021)

معادلة تحديد مؤشر التمييز هي:

$$D = \frac{BA}{JA} - \frac{BB}{JB} = PA - PB$$

الشرح:

D : قوة التمييز

BA : عدد المشاركين في المجموعة العليا الذين أجابوا على السؤال بشكل صحيح

JA : عدد المشاركين في المجموعة العليا

BB : عدد المشاركين في المجموعة العليا الذين أجابوا على السؤال بشكل غير صحيح

JB : عدد المشاركين في المجموعة الدنيا

PA : عدد المشاركين في المجموعة العليا الذين (تذكر p كرمز لمؤشر الصعوبة)

PB : نسبة المشاركين من المجموعة الدنيا الذين أجابوا بشكل صحيح

يعتبر تصنيف القوة التمييزية (٠,٠٠ - ٠,٢٠) فئة ضعيفة، والقوة التمييزية (٠,٢٠ - ٠,٤٠) فئة كافية، والقوة التمييزية (٠,٤٠ - ٠,٧٠) فئة جيدة، والقوة التمييزية (٠,٧٠ - ١,٠٠) فئة جيدة جدًا بينما يتم تجاهل القيم السلبية.

فيما يلي نتائج اختبار قوة التمييز جودة بنود أسئلة اختبار اللّغة العربيّة للصّف الثّامن في المدرسة المتوسّطة الإسلاميّة الحكوميّة ٦ ليما بولوه كوتا.

الجدول ٦. مؤشر المشكلات حسب قوة التمييز

الرقم	قوة التمييز	بنون الأسئلة	نسبة مئوية
١	سلبي	-	-
٢	ضعيف (<٠,٢٠)	٢٧	٪٣
٣	مقبول (٠,٢٠-٠,٣٩)	٦,٢٥,٢٦,٢٩	٪١٣
٤	جيد (٠,٤٠-٠,٦٩)	١,٢,٣,٤,٥,٨,١٠,١٤,١٥,١٨,١٩,٢٠,٢٢,٢٣,٢٨,٣٠	٪٥٤
٥	جيد جدا (٠,٧٠-١,٠٠)	٧,٩,١١,١٢,١٣,١٦,١٧,٢١,٢٤	٪٣٠

استنادًا إلى نتائج تحليل ٣٠ سؤالاً، يمكن ملاحظة أن الأسئلة التي تتمتع بمعايير قوة التمييز الضعيفة أو غير الكافية هي سؤال واحد (٣٪)، والأسئلة ذات المعايير الكافية هي ١٤ سؤالاً (١٣٪)، والأسئلة ذات المعايير الجيدة هي ١٦ سؤالاً (٥٤٪)، والأسئلة ذات المعايير الممتازة هي ٩ أسئلة (٣٠٪).

فعالية المخادعة

تشير فعالية أداة المخادعة إلى قدرة أداة المخادعة على خداع المتعلمين الأقل مهارة في اختيار بديل الإجابة. يُظهر تحليل فعالية المخادعة أن هذه الإجابات لها اختلافات دقيقة جدًا مع الإجابة الصحيحة، لذا فإنها غالبًا ما توقع الطّلاّب في فخ عند الإجابة عن أسئلة الاختبار(Radja et al., 2023).

معادلة تحديد الفعالية المخادعة هي:

$$ED = \frac{P}{N} x100\%$$

ED: مؤشر فعالية المخادعة

A: عدد الطلاب الذين اختاروا المخادعة

N: عدد الإجابات البديلة

عند تقييم فعالية المخادعة، نحتاج إلى الرجوع إلى معيار محدد مسبقاً. وتنص هذه المعايير على أن المُشتت يعتبر قد أدى وظيفته بنجاح إذا قام ٥٪ على الأقل من جميع المتقدمين للاختبار باختياره. يتم استخدام تحليل فعالية المخادعة لتحديد فعالية المخادعة. تتكون أسئلة اختبار اللغة العربية من ٤ اختيارات للإجابة بحيث يكون هناك ٣ إجابات بديلة تصبح المخادعة.

الجدول ٧. فئة مؤشر التعرض لأسئلة اختبار اللغة العربية

رقم السؤال	أ	ب	ج	د	التفسير
١	٪٧٤,٦	٪٧,٩	٪١٢,٧	٪٤,٨	غير فعالة
٢	٪١٦	٪٣٨	٪٣٢	٪١٤	فعالة
٣	٪١١	٪٢١	٪١٦	٪٥٢	فعالة
٤	٪٤,٨	٪٣٤,٩	٪٣٦,٥	٪٢٣,٨	غير فعالية
٥	٪٥٢	٪٢١	٪١٣	٪١٤	فعالية
٦	٪٤,٨	٪٨٢,٥	٪٦,٣	٪٦,٣	غير فعالية

رقم السؤال	أ	ب	ج	د	التفسير
٧	٣٧٪	١٧٪	٣٣٪	١٣٪	فعالية
٨	١٤٪	٣٢٪	٤٨٪	٦٪	فعالية
٩	٥٩٪	١١٪	١٩٪	١١٪	فعالية
١٠	٤٣٪	٢٤٪	١٣٪	٢١٪	فعالية
١١	١٦٪	٤٨٪	٢٣٪	١٣٪	فعالية
١٢	٨٪	٢٥٪	٢١٪	٤٦٪	فعالية
١٣	٤٨٪	١٩٪	٢١٪	١٢٪	فعالية
١٤	٤٤٪	١٩٪	٢٢٪	١٤٪	فعالية
١٥	٢٥٪	٤٦٪	٢١٪	٨٪	فعالية
١٦	١١٪	٤٨٪	١٩٪	٢٢٪	فعالية
١٧	٢٤٪	١٧٪	٤٣٪	١٦٪	فعالية
١٨	٤٤٪	٢٤٪	١٦٪	١٦٪	فعالية
١٩	١٧٪	٢٧٪	٤٠٪	١٦٪	فعالية
٢٠	٤٦٪	١٦٪	٣٢٪	٦٪	فعالية
٢١	١٩٪	١٧٪	١٧٪	٤٦٪	فعالية
٢٢	٢٢٪	١٣٪	١٠٪	٥٦٪	فعالية
٢٣	٣٣٪	٢١٪	٣٢٪	١٤٪	فعالية
٢٤	٤٤٪	٣٢٪	٢١٪	٣٪	غير فعالية
٢٥	٢٦٪	١٩٪	٢٦٪	٢٩٪	فعالية
٢٦	٣٢٪	٤٨٪	٨٪	١٣٪	فعالية
٢٧	٢٩٪	٢٢٪	٣٣٪	١٦٪	فعالية
٢٨	٣٢٪	٢٤٪	٣٣٪	١١٪	فعالية
٢٩	٣٠٪	٣٧٪	١٧٪	١٦٪	فعالية
٣٠	١٤٪	٢٢٪	٥١٪	١٣٪	فعالية

الجدول أعلاه هو نتيجة فعالية المخادعة أسئلة الاختيار من متعدد في اختبار اللغة العربية للفصل الدراسي الأول من أصل ٣٠ سؤالاً. يعمل ما مجموعه ٢٦ عنصرًا بفعالية و٤ عناصر غير فعالية. ويعتمد تفسير عدم فعالية هذه العناصر على معادلة مؤشر المتتبع حيث يتم اختيار المتتبع الجيد بنسبة ٥٪ على الأقل من إجمالي عدد الطلاب الذين أدوا الاختبار. يعمل النموذج بفعالية إذا كانت جميع النماذج تعمل بفعالية. توجد المسائل ذات المشغلات الفعالة في الأعداد ٢، ٣، ٥، ٥، ٦، ٦، ٧، ٧، ٨، ٨، ٩، ١٠، ١٠، ١١، ١٢، ١٢، ١٣، ١٣، ١٤، ١٥، ١٦، ١٦، ١٧، ١٨، ١٨، ١٩، ٢٠، ٢١، ٢١، ٢٢، ٢٣، ٢٣، ٢٥، ٢٦، ٢٧، ٢٨٢٩، ٣٠. ينتمي السؤال

إلى فئة المشغلات غير الفعالة إذا كان هناك مشغل واحد أو اثنين فقط يعمل بشكل صحيح. عناصر السؤال التي تحتوي على مشغلات غير فعالة هي رقم ١، و٤، و٦، و٢٤.

الجدول ٨. توزيع أسئلة اختبار اللغة العربية على أساس فعالية المخادعة

جودة فعالية المخادعة	سؤالا	نسبة مئوية
فعالية	٢,٣,٥,٧,٨,٩,١٠,١١,١٢,١٣,١٤,١٥,١٦,١٧,١٨,١٩ ٢٠,٢١,٢٢,٢٣,٢٥,٢٦,٢٧,٢٨,٢٩,٣٠	٪٨٧
غير فعالية	١,٤,٦,٢٤	٪١٣

يوضح الجدول أعلاه النسبة المئوية لفعالية المخادعة أسئلة الاختيار من متعدد في اختبار اللغة العربية للصف الثامن في الفصل الدراسي الأول في المدرسة المتوسّطة الإسلاميّة الحكوميّة ٦ ليما بولوه كوتا في العام الدراسي ٢٠٢٤/٢٠٢٥، ويبين الجدول أن الأسئلة ذات النماذج الفعالية ذات النسبة المئوية الأكبر من تلك ذات النماذج غير الفعالية. تبلغ نسبة عناصر الأسئلة ذات المخادعة الفعالية ٨٧٪ بينما تبلغ نسبة عناصر الأسئلة ذات المخادعة غير الفعالية ١٣٪ فقط. وبالتالي، فإن المخادعة الفعالية ستجعل العديد من الطلاب ينخدعون بالإجابة الخاطئة لأنهم يعتقدون أنها الإجابة الصحيحة.

مناقشة البحث

الصلاحيّة

تم إجراء اختبار الصلاحية باستخدام صيغة الارتباط العزمي للمنتج، باستخدام تطبيق *Microsoft Excel* في حساب صلاحية الاختبار. أفراد البحث، أي ٦٣ شخصاً، وبالنظر إلى معادلة *r-tabel* عند مستوى الدلالة ٥٪، فإن ٢-n من ٦٣ شخصاً هو ٦١ شخصاً، وبالتالي فإن عدد *r-tabel* هو ٠,٢٥. نتائج البحث على تحليل البنود بناءً على المعيار إذا كان *r-tabel* أكبر من ٠,٢٥ فيتم إعلان صلاحية، أما إذا كان *r-tabel* أقل من ٠,٢٥ فيتم إعلان غير صلاحية.

بناءً على نتائج تحليل الأسئلة، أظهرت نتائج التحليل أنّ ٢٩ بندًا من أصل ٣٠ سؤالا (بنسبة ٩٧٪) صلاحية، حيث تجاوزت قيمة *r-hitung* فيها قيمة (0,25) *r-tabel*، بينما يوجد سؤالا واحد فقط (البند رقم ٢٧) غير صلاحية لكون قيمة الارتباط فيه أقل من القيمة الجدولية. بالنسبة للسؤال رقم ٢٧ غير الصلاحية، يمكن أن يكون ذلك بسبب عدة عوامل. هذا يتوافق مع النظرية التي طرحها غرونلوند في كتاب زينال عارفين، والتي تنص على أن هناك ثلاثة عوامل تؤثر على صحة نتائج الاختبار، وهي: عامل الأدوات المستخدمة في الاختبار، وعامل الإدارة والتصحيح، وعامل إجابات الطلاب(Zainal, 2012).

بناءً على ما سبق، يمكن استنتاج أن بنود أسئلة اختبار الفصل الدراسي الأول المواد اللغة العربية للصف الثامن في المدرسة المتوسّطة الإسلاميّة الحكوميّة ٦ ليما بولوه كوتا في العام الدّراسي ٢٠٢٤/٢٠٢٥، تتمتع بجودة عالية من حيث صلاحية الأسئلة. هذه الصلاحية تقتصر على صلاحية الأسئلة من حيث نتائج إجابات المشاركين في اختبار، ولم يتم النظر إليها من حيث صلاحية المحتوى والبناء.

الثبات

تم إجراء اختبار الثبات باستخدام صيغة KR-20 باستخدام تطبيق Microsoft Excel في حساب ثبات الاختبار. تفسير نتائج حساب الثبات باستخدام المعايير التالية: تتراوح درجات اختبار KR-20 من ٠ إلى ١، حيث أن ٠ تعني عدم الثبات و١ تعني الثبات الكاملة. وكلما كانت الدرجة أقرب إلى ١، كلما كان الاختبار أكثر ثبات. يعتمد ما يشكل درجة KR-20 المقبولة على نوع الاختبار. بشكل عام، عادة ما تعتبر الدرجات التي تزيد عن ٠,٥ معقولة.

بناءً على نتائج تحليل الأسئلة، بلغت قيمة الثبات التي تم الحصول عليها باستخدام معادلة KR-20 نحو ٠,٨٩٧، وهي تقع ضمن فئة الثبات العالي جدًا. إن أداة الاختبار التي تتمتّع بصلاحية جيدة في كل بند من بنودها ستتمتّع أيضًا بدرجة عالية من الثبات. هذا يتوافق مع نظرية سوهارسيمي أريكونتو التي تنص على أن درجة صلاحية الاختبار يمكن أن تشير إلى درجة ثبات الاختبار(Arikunto, 2013).

بناءً على ما سبق، يمكن استنتاج أن بنود أسئلة اختبار الفصل الدراسي الأول المواد اللغة العربية للصف الثامن في المدرسة المتوسّطة الإسلاميّة الحكوميّة ٦ ليما بولوه كوتا في العام الدّراسي ٢٠٢٤/٢٠٢٥، هي أسئلة جيدة من حيث ثباتها.

مستوى الصّعوبة

تم إجراء اختبار مستوى الصّعوبة باستخدام تطبيق Microsoft Excel في حساب مستوى الصّعوبة الاختبار. تفسير نتائج حساب مستوى الصّعوبة باستخدام المعايير التالية: تصنيف مستوى صعوبة الأسئلة (٠,٠٠-٠,٣٠) على أنه صعب، ومستوى صعوبة الأسئلة (٠,٣١-٠,٧٠) على أنه متوسط، ومستوى الصعوبة (٠,٧١-٠,١٠٠) على أنه سهل.

بناءً على نتائج تحليل الأسئلة، توزيع مستوى الصعوبة يظهر أن معظم الأسئلة تقع في فئة متوسطة (٢٦ سؤالاً أو ٨٦٪)، بينما الأسئلة السهلة والصعبة تتكون كل منهما من ٢ سؤال فقط (٧٪). هذه النسبة تعتبر مثالية نسبياً، لأن معظم الأسئلة تقع في مستوى الصعوبة المتوسط، مما يساعد على التمييز بدقة بين الطلاب ذوي المستوى العالي والمنخفض(Arikunto, 2013). ومع ذلك، يجب مراجعة الأسئلة السهلة جداً أو الصعبة جداً حتى لا تؤثر على قدرة الاختبار على التمييز.

وبالتالي، يمكن استنتاج أن بنود أسئلة اختبار الفصل الدراسي الأول المواد اللغة العربية للصف الثامن في المدرسة المتوسّطة الإسلاميّة الحكوميّة ٦ ليما بولوه كوتا في العام الدّراسي ٢٠٢٤/٢٠٢٥، تعتبر أسئلة ذات مستوى صعوبة جيد، حيث أن ٨٦٪ من الأسئلة تتميز بمستوى صعوبة متوسط.

قوة التمييز

تم إجراء اختبار قوة التمييز باستخدام تطبيق Microsoft Excel في حساب قوة التمييز الاختبار. تفسير نتائج حساب قوة التمييز باستخدام المعايير التالية: صنيف قوة التمييزية (٠,٠٠ - ٠,٢٠) فئة ضعيفة، وقوة التمييزية (٠,٢٠ - ٠,٤٠) فئة كافية، وقوة التمييزية (٠,٤٠ - ٠,٧٠) فئة جيدة، وقوة التمييزية (٠,٧٠ - ١,٠٠) فئة جيدة جدًا بينما يتم تجاهل القيم السلبية.

بناءً على نتائج تحليل الأسئلة، أظهرت نتائج تحليل قوة التمييز أن معظم البنود تقع ضمن فئة جيدة (٥٤٪) وجيدة جدًا (٣٠٪). بينما توجد بنود ذات قوة تمييز مقبولة (١٣٪)، وبند واحد فقط (٣٪) ذو قوة تمييز ضعيفة (رقم ٢٧). وتدل البنود ذات قوة التمييز العالية على فعاليتها في التفريق بين الطلاب ذوي القدرة العالية والمنخفضة، بينما تشير البنود ذات قوة التمييز المنخفضة إلى ضعف في إعطاء معلومات دقيقة عن مستوى الطلاب(Sudijono, 2013). وهذا يتماشى مع نتائج الصلاحية، حيث إن السؤال رقم ٢٧ أيضًا أظهر ضعفًا ملحوظًا .

وبالتالي، يمكن استنتاج أن بنود أسئلة اختبار الفصل الدراسي الأول المواد اللغة العربية للصف الثامن في المدرسة المتوسّطة الإسلاميّة الحكوميّة ٦ ليما بولوه كوتا في العام الدّراسي ٢٠٢٤/٢٠٢٥، قوة التمييز بين الطلاب ذوي القدرات العالية والطلاب ذوي القدرات المنخفضة، وذلك لأن هذه الأسئلة تتمتع بقوة تمييزية جيدة، حيث أن غالبية الأسئلة أو ما نسبته ٥٤٪ تتمتع بقدرة تمييزية جيدة.

فعالية المخادعة

تم إجراء اختبار فعالية المخادعة باستخدام تطبيق Microsoft Excel في حساب فعالية المخادعة الاختبار. تفسير نتائج حساب فعالية المخادعة باستخدام المعايير التالية: عند تقييم فعالية المخادعة، نحتاج إلى الرجوع إلى معيار محدد مسبقاً. وتنص هذه المعايير على أن المُشتت يعتبر قد أدى وظيفته بنجاح إذا قام ٥٪ على الأقل من جميع المتقدمين للاختبار باختياره(Zainal, 2012).

بناءً على نتائج تحليل الأسئلة، أظهرت النتائج أن ٢٦ بندًا (٨٧٪) تحتوي على مخادعات فعالة، في حين أن ٤ بنود فقط (١٣٪) كانت مخادعاتها غير فعالة. وتكمن أهمية المخادعات الفعالة في قدرتها على دفع الطلاب الذين لم يتقنوا المادة لاختيار الإجابة الخاطئة، مما يعزز وظيفة السؤال في التمييز. أما المخادعات غير الفعالة، فهي غالبًا ما يتم تجاهلها من قبل الطلاب، مما يقلل من جودة السؤال(Santosa & Badawi, 2022).

وبالتالي، يمكن استنتاج أن بنود أسئلة اختبار الفصل الدراسي الأول المواد اللغة العربية للصف الثامن في المدرسة المتوسّطة الإسلاميّة الحكوميّة ٦ ليما بولوه كوتا في العام الدّراسي ٢٠٢٤/٢٠٢٥، تعتبر أسئلة ذات فعالية المخادعة يعني فعالة، حيث أن ٨٧٪ من الأسئلة تتمتع بفعالية المخادعة يعني فعالة.

الآثار المترتبة على نتائج البحث

تتضمن نتائج هذا البحث عدة آثار مهمة. أولاً، بالنسبة لمعلمي اللغة العربية في المدرسة المتوسطة الإسلامية الحكومية ٦ ليما بولوه كوتا، يمكن الاستفادة من هذه النتائج كأساس لتحسين البنود التي لم ترقَ بعد إلى معايير الجودة المطلوبة، ولا سيما البند رقم ٢٧ الذي أظهر ضعفاً في جانبي الصلاحية وقوة التمييز في آنٍ واحد. ثانياً، بالنسبة للمدرسة، يمكن توظيف نتائج هذا التحليل أساساً لبناء بنك أسئلة موحد ومعياري، إذ يمكن حفظ البنود ذات الجودة الجيدة (٨٣٪) واستخدامها مجدداً في الاختبارات اللاحقة. ثالثاً، فيما يتعلق بتطوير ممارسات التقويم في المدارس الإسلامية، تُبرز هذه الدراسة أهمية إجراء تحليل بنود الأسئلة بصورة منتظمة في كل فصل دراسي بوصفه جزءاً من ضبط جودة أدوات التقييم، بما يُسهم في الرفع المستمر

والمتواصل من جودة الأسئلة. رابعاً، بالنسبة للباحثين اللاحقين، يمكن أن تكون هذه الدراسة مرجعاً أولياً لبحوث أكثر شمولاً وعمقاً، كإضافة تحليل صلاحية المحتوى وصلاحية البناء اللتين لم تشملهما هذه الدراسة.

الخاتمة

بوجه عام، تُظهِر جودة أسئلة اللغة العربية المستخدمة في اختبار اللّغة العربيّة للصّفّ الثّامن في الفصل الدّراسي الأوّل في المدرسة المتوسّطة الإسلاميّة الحكوميّة ٦ ليما فولوه كواتا خصائص تُعتَبَر صالحة للاستخدام في تقييم التعلم. تعكس هذه الأسئلة تحقيق الجوانب المهمة في تحليل أدوات التقييم، مثل الصدق، والثبات، ومستوى الصعوبة، وقوة التمييز، وفعالية المشتتات. وتُعزّز هذه النتائج أهمية تحليل مفردات الأسئلة كخطوة منهجية لضمان جودة التقييم في البيئة المدرسية. وتُساهم هذه الدراسة في تطبيق منهج التحليل الكمي على أسئلة اختبار نهاية الفصل الدراسي التي لم تُحلَّل من قبل. وتتمثل الجِدّة في هذه الدراسة في تركيز التحليل الشامل على خمسة أبعاد من جودة الأسئلة في مؤسسة تعليمية واحدة. أما ميزة هذه الدراسة فهي في منهجها المنظم والمبني على بيانات ميدانية تجريبية، بينما تكمن محدوديتها في عدد العيّنات القليل واعتمادها على شكل واحد من الاختبار في فصل دراسي واحد. ومن الموصى به في الدراسات القادمة أن يُجرى تحليل الأسئلة في فترة زمنية أوسع، وأن يشمل أشكالًا مختلفة من التقييم، بما في ذلك التقييم التكويني والختامي، حتى تكون النتائج أكثر شمولًا ويمكن استخدامها لتحسين نظام تقييم تعلم اللغة العربية بشكل مستدام.

المراجع

Ali, M. S. (2011). *Ittijahat wa Tathbiqat Hadithah fi al-Manahij wa Thuruq al-Tadris*. Dar al-Masirah.

AMALIA, T. (2019). *Analisis butir soal tes penilaian akhir semester I buatan MGMP mata pelajaran Al-Qur'an Hadis kelas IX MTs di Kaliwungu Kendal tahun pelajaran 2019/2020*.

Arifianto, M. L., Ainin, M., Ahsanuddin, M., Irhamni, I., Fitria, N., Nikmah, K., & Anwar, M. S. (2021). *Evaluasi Pembelajaran dan Pengembangan Tes Interaktif Bahasa Arab*. Tonggak Media.

Arikunto, S. (2013). Dasar-dasar evaluasi pendidikan edisi kedua. *Jakarta: Bumi Aksara*.

Dewi, Y., Pranata, K., Zakir, S., Kamal, M., & S. (2023). Evaluasi Kegiatan Keagamaan Dalam Memperkuat Nilai-Nilai Religius Siswa Menggunakan Model Cipp (Context, Input, Procces, Product) Di Sdn 14 Atts Bukit Tinggi. *Jurnal Ilmiah Pendidikan Kebudayaaan Dan Agama*, *1*(1), 102–114.

Dewi, Y., Rizal, E., Rahmadani, N., Rahmi, U., & Aprison, W. (2024). Analisis butir soal Asesmen Akhir Sumatif (AAS) mata pelajaran Akidah Akhlak Kelas XI di MAN 1 Rokan Hulu. *Idarah Tarbawiyah: Journal of Management in Islamic Education*, *5*(6), 627–636.

Dony Handriawan, N. (2021). Evaluasi Pembelajaran Bahasa Arab. *Sanabil Publishing*, 234.

Febrianta, R., Abdiyah, H., Rizal, E., & Yusra, O. (2022). Fa'aliyyah Istikhdam Lu'bah al-Awraq al-Mumazzaqah li Tarqiyah Maharah al-Qira'ah lada Thullab fi al-Madrasah al-Islamiyyah al-Syaikh Ibrahim Harun. *'ARABIYYA: JURNAL STUDI BAHASA ARAB*, *11*(02), 319–332.

Febrianta, R., & Yusra, O. (2023). *Ta'tsīr Istikhdām Wasīlah Shundūq al-Qirā'ah 'alā Mahārah al-Qirā'ah al-Thullāb bi Madrasah al-Tsanawiyah al-Hukūmiyah 2 Padang Panjang*. *13*(1), 131.

Fitria, H., Kristiawan, M., & Rahmat, N. (2019). Upaya meningkatkan kompetensi guru melalui pelatihan penelitian tindakan kelas. *Abdimas Unwahas*, *4*(1).

Ida, F. F., & Musyarofah, A. (2021). *Validitas dan reliabilitas dalam analisis butir soal. Al-Mu'Arrib: Journal of Arabic Education, 1 (1), 34–44*.

Magdalena, I., Fauziah, S. N., Faziah, S. N., & Nupus, F. S. (2021). *Analisis Validitas, Reliabilitas, Tingkat Kesulitan dan Daya Beda Butir Soal Ujian Akhir Semester Tema 7 Kelas III SDN Karet 1 Sepatan*.

Radja, S. P., Bano, V. O., & Ina, A. T. (2023). Analisis Kualitas Butir Soal Hasil Belajar Peserta Didik Berdasarkan Tingkat Kesukaran, Daya Beda Dan Efektivitas Pengecoh Di Sman 1 Pandawai. *Jurnal Edusavana*, *1*(1), 33.

Rahman, A. A., & Nasryah, C. E. (2019). Evaluasi Pembelajaran. In *Uwais Inspirasi Indonesia*.

Ramadhan, M. F., Siroj, R. A., & Afgani, M. W. (2024). Validitas and Reliabilitas. *Journal on Education*, *6*(2), 10967–10975.

Santosa, S., & Badawi, J. A. (2022). Analisis Butir Soal Pilihan Ganda Tema Pertumbuhan dan Perkembangan Makhluk Hidup Kelas III Madrasah Ibtidaiyah. *Jurnal Basicedu*, *6*(2), 1678–1686.

Sudijono, A. (2013). *Pengantar evaluasi pendidikan*.

Sudirman, M. L. K., Sriwahyuningrum, A., Cahaya, I. M. E., Astuti, N. L. S., Setiawan, J., Tandirerung, W. Y., Rahmi, S., Nusantari, D. O., Indrawati, F., & Fittriya, N. L. (2023). Metodologi Penelitian (Suci Haryati, Ed.; Vol. 1). *CV. Media Sains Indonesia*.

Wardani, W. K., & Saring Marsudi, S. H. (2017). *Analisis Kompetensi Pedagogik Guru Dalam Proses Pembelajaran di SD Muhammadiyah 16 Surakarta*. Universitas Muhammadiyah Surakarta.

Zainal, A. (2012). *Evaluasi Pembelajaran: Prinsip, Teknik, Prosedur*.

Zuhri, N. Z., Sopian, A., Sauri, S., & Nurbayan, Y. (2024). Analisis Validitas dan Reliabilitas Soal Bahasa Arab Melalui Website OpExams Pembuat Soal Berbasis AI. *Jurnal Pendidikan Modern*, *9*(2), 87–91.