Al-Lahjah : Jurnal Pendidikan, Bahasa Arab dan Kajian Linguistik Arab
Vol. 8 No. 2, Juli 2025, Halaman 1171-1181
E-ISSN: 2615-5656

Sidq wa Tsabat Bunud As'ilah Ikhtibar At-Taqyim At-Tajmi' An-Niha'i Li Maddah Al-Lugah Al-Arabiyah Bi Al-Madrasah Al-Mutawasitah Malang

Ulfatul Mufidha[1], Fikri Haikal Ibrahim[2], Abdul Gafur[3], Muhammad Ridho Anshori[4], Nur Qomari[5]
[1,2,3,4]Postgraduate Arabic Language Education, UIN Maulana Malik Ibrahim Malang, Indonesia
*Email: ulfaamufidha@gmail.com
Phone Number (WhatsApp): 0878 3686 9153

ABSTRACT
This study aims to analyze the validity and reliability of the final year summative exam (SAT) questions for the Arabic language subject for class VII of SMP Integral Malang. The subjects of this study were 25 students of class VII C, while the data were in the form of students' raw scores from the multiple-choice questions that had been completed. The method used in this study was a descriptive quantitative type, with data collection in the form of documentation. Data analysis was carried out using the IBM SPSS Statistic 25 application. The results of the analysis of the final year summative exam (SAT) questions for class VII of SMP Integral Malang showed that out of 33 multiple-choice questions, only 9 questions were valid, while 24 questions did not meet the validity criteria, with reference to the calculated r value> r table and Sig. <0.05. However, the reliability value using Cronbach's Alpha was 0.657, which was greater than the r table (0.396), indicating a high level of question reliability. This finding reveals that an instrument can remain reliable even though most of the questions are invalid. The results of this study provide an empirical basis for teachers to conduct evaluations and improvements in the preparation of questions and become an academic reference in developing evaluations of Arabic language learning at the junior high school level.

***Keywords**: Validity and Reliability; Test Items; Final Year Summative Exam; Arabic Language*

الملخص
يهدف هذا البحث إلى تحليل صدق وثبات بنود أسئلة اختبار التقييم التجميعي النهائي (SAT) لمادة اللغة العربية للصف السابع بالمدرسة المتوسطة بمدينة مالانج. يتكون موضوع البحث من طلاب الصف السابع فصل ج وعددهم ٢٥ طالبًا، وتمثلت البيانات في الدرجات الخام للطلاب الناتجة عن إجاباتهم على أسئلة الاختيار من متعدد. اعتمد البحث على المنهج الوصفي الكمي، وتم جمع البيانات من خلال التوثيق. تم تحليل البيانات باستخدام برنامج IBM SPSS Statistic 25 وأظهرت نتائج تحليل بنود أسئلة الاختبار أن من بين ٣٣ سؤالًا من نوع الاختيار من متعدد، كانت ٩ أسئلة فقط صادقة، بينما لم تستوفِ ٢٤ سؤالًا معيار الصدق، بناءً على قيمة (r) المحسوبة الأكبر من قيمة (r) الجدولية ومستوى الدلالة أقل من ٠,٠٥. ومع ذلك، بلغت قيمة الثبات باستخدام معامل Cronbach's Alpha (٠,٦٥٧)، وهي أكبر من قيمة (r) الجدولية (٠,٣٩٦)، مما يدل على مستوى عالٍ من الثبات للاختبار. تكشف هذه النتائج أن الأداة قد تكون ثابتة حتى لو لم تكن معظم بنودها صادقة. ويوفر هذا البحث أساسًا تجريبيًا للمعلمين لإجراء التقييم والتحسين في إعداد الأسئلة، كما يشكل مرجعًا أكاديميًا لتطوير تقييم تعليم اللغة العربية في مستوى المرحلة المتوسطة.

الكلمات الرئيسية: صدق; ثبات; بنود أسئلة الاختبار; التقييم التجميعي النهائي; اللغة العربية.

المقدمة

يُعَدُّ التقييم أحد العناصر الجوهرية التي لا ينبغي إهمالها في عملية التعليم، إلى جانب تقديم المادة العلمية بأسلوب جيد. التقويم هو المفهوم الرئيس الذي يشتمل على عناصر أخرى مثل القياس والاختبار بداخله (Ida & Musyarofah, 2021). يؤدي التقييم دورًا مهمًّا في قياس مدى تحقيق الأهداف التعليمية. وفي الممارسة العملية، غالبًا ما يتمثل التقييم في صورة اختبارات تتضمن بنود أسئلة تُستخدم لقياس كفاءة المتعلمين بصورة منهجية. ولا يقتصر دور الاختبار على كونه أداة قياس فحسب، بل يُعَدُّ جزءًا لا يتجزأ من عملية التعليم، إذ يوفر معلومات حول مدى فعالية التدريس ومستوى فهم المتعلمين (Faiz et al., 2022)

إن إعداد بنود الأسئلة في الاختبارات يجب أن يراعي معايير بناء الاختبارات الجيدة. فالاختبار التقييمي الجيد يجب أن يتصف بخصائص أساسية من أهمها أن يكون الاختبار محققًا لمعياري الصدق والثبات (Magdalena et al., 2021). كما يرى أريكُنتو أن الاختبار يُعَدُّ جيدًا إذا توفرت فيه عدة معايير أساسية، وهي الصلاحية، الثبات، الموضوعية، سهولة التطبيق (العملية)، والكفاءة الاقتصادية (Arikunto, 2018). وفقًا لفاضل (Fadli et al., 2023) تُظهر نتائج بحثهم أن الصدق والثبات يُعدّان من المؤشرات الرئيسية التي تُحدد جودة الأداة. ويمكن لعوامل متعددة أن تؤثر في عملية تقويم الصدق والثبات، مثل استجابات المشاركين، والظروف والمواقف في موقع البحث، بالإضافة إلى استخدام أدوات غير مناسبة. ومن خلال مراعاة مبادئ إعداد الاختبارات الجيدة، سيكون المعلم أكثر دقة ووضوحًا في تصميم أداة التقييم، مما يؤدي إلى إنتاج اختبار قادر على قياس قدرات الطلاب بدقة وموضوعية.

وعلى الرغم من أن المبادئ العلمية لإعداد الاختبارات قد نوقشت على نطاق واسع في الأدبيات، إلا أن الواقع يُظهِر أن العديد من الممارسين التربويين والمعلمين لا يزالون يواجهون صعوبات في تصميم أداة اختبار تستوفي الخصائص المثالية حقًّا. وفي تعليم اللغة العربية خاصةً، يتطلب إعداد اختبار فعّال تخطيطًا دقيقًا وتحليلًا معمقًا لمستوى الصعوبة، والقدرة التمييزية، ووظيفة المشتتات، حتى يكون الاختبار الناتج صالحًا وثابتًا. ومع ذلك، يواجه العديد من المعلمين تحديات في تطبيق هذه المبادئ بصورة مستمرة (Rezi & Aulia, 2020).

يُعَدُّ تحليل بنود الأسئلة خطوة منهجية تهدف إلى تقديم معلومات تفصيلية حول كل فقرة اختبارية تم إعدادها. وتُجرى هذه العملية للحصول على بيانات مهمة تُستخدم في تقييم نتائج تعلم الطلاب (Himawan & Suyata, 2022). وفي الوقت نفسه، يُعدّ التقويم جانبًا مهمًا في العملية التعليمية، إلا أن العديد من المعلمين لا يزالون يواجهون صعوبة في اختيار وتطوير الأداة المناسبة، وخاصة عند إعداد الاختبارات بشكل مستقل (Suryawati & Yulfikar, 2012). إن الأسئلة الامتحانية المستخدمة في المدرسة المتوسطة بمدينة مالانج يتم إعدادها بشكل مستقل من قِبَل المعلمين أنفسهم، ولضمان أن هذه الأسئلة تحقق معايير الاختبار الجيد من حيث الصلاحية والثبات، فلا بد من إجراء تحليل دقيق لها. ويهدف هذا التحليل إلى تقييم مدى قدرة الأسئلة على قياس كفاءة الطلاب بدقة وثبات، بحيث تعكس نتائج التقييم القدرة الحقيقية للطلاب على التعلم.

لقد أُجري العديد من الدراسات السابقة حول تحليل بنود أسئلة اللغة العربية، منها: أولًا، البحث الذي أجراه تانجونج وآخرون (Tanjung et al., 2024) بعنوان "تحليل بنود أسئلة التقييم النهائي لمادة اللغة العربية

للصف السابع بمدرسة المتوسطة المعارف راكيت باجانجارا جاوة الوسطى". وثانيا، البحث الذي أجرته فكرية (Fikriyah, 2021) بعنوان: "تحليل بنود أسئلة الاختبار النصفي لمادة اللغة العربية للصف السابع في الفصل الدراسي الأول بمدرسة متوسطة المحمدية الأولى بيوجياكرتا، السنة الدراسية ٢٠١٩/٢٠٢٠." وثالثا، البحث الذي أجراه فهمي وآخرون (Fahmi et al., 2022) بعنوان "تحليل بنود أسئلة اللغة العربية لطلاب المدرسة الثانوية معهد السلام كامبار رياو". ورابعا، البحث الذي أجراه ديانوفا و أنوار (Dianova & Anwar, 2024) بعنوان "تحليل بنود اختبار الصلاحية والثبات ومستوى الصعوبة والقدرة التمييزية للأسئلة التقييمية لمادة اللغة العربية بالمدرسة الإسلامية الابتدائية". وخامسا، فقد أجراه مولانا وسنوسي (Maulana & Sanusi, 2020) بعنوان: "تحليل فقرات أسئلة مادة اللغة العربية في اختبار نهاية المدرسة الموحد الإقليمي (UAMBD) للمدارس الابتدائية، السنة الدراسية ٢٠١٧-٢٠١٨."

ناقشت الأبحاث السابقة تحليل بنود أسئلة اللغة العربية في مراحل الابتدائية والمتوسطة والثانوية بمعايير تحليل واسعة مثل مستوى الصعوبة والقدرة التمييزية وفعالية المشتتات. أما هذه الدراسة، فتتناول جانبًا أكثر تحديدًا، وهو التركيز فقط على جوانب الصلاحية والثبات للأسئلة. وقد أُجري هذا البحث على بنود أسئلة اختبار مادة اللغة العربية في اختبار التقييم التجميعي النهائي (SAT) التي أُعدّت بشكل مستقل من قِبَل معلمي المدرسة المتوسطة بمدينة مالانج. وعليه، تهدف هذه الدراسة إلى تحليل بنود أسئلة اختبار التقييم التجميعي النهائي (SAT) لمادة اللغة العربية للصف السابع بالمدرسة المتوسطة بمدينة مالانج من ناحية الصلاحية والثبات، وذلك لمعرفة مدى جودة هذه الأسئلة وصلاحيتها للاستخدام كأداة تقييمية مناسبة.

منهج البحث

يستخدم هذا البحث منهجًا وصفيًا كميًا، إذ إن المنهج الكمي هو أسلوب يُستخدم لجمع البيانات وتحليلها وعرضها في شكل أرقام، ويُدعَم بتقديمها من خلال الجداول والرسوم البيانية والمخططات والصور أو وسائل التصوير الأخرى، بهدف تعزيز صلاحية البيانات المدروسة(Arikunto, 2002). في هذا البحث، يُستخدم المنهج الوصفي الكمي لوصف جودة بنود أسئلة مادة اللغة العربية المستخدمة في المدرسة المتوسطة بمدينة مالانج من خلال تحليل جانبي الصلاحية والثبات لها.

يتكون مجتمع البحث من جميع طلاب الصف السابع في المدرسة المتوسطة، وأما العينة المأخوذة فهي صف واحد فقط وهو الصف السابع (فصل ج) وعددهم ٢٥ طالبًا. وتتمثل بيانات هذا البحث في الدرجات الخام للطلاب الناتجة عن بنود الأسئلة التي أجابوا عليها، ثم تُعالج هذه البيانات للبحث عن مستوى صدق وثبات الأسئلة لمعرفة جودة الأسئلة المستخدمة في المدرسة المتوسطة بمدينة مالانج. وأُستخدمت تقنية جمع البيانات من خلال التوثيق، وذلك بتوثيق أوراق أسئلة الاختيار من متعدد وأوراق إجابات الطلاب. ثم تم تحليل البيانات باستخدام التحليل الكمي، حيث يتم إعطاء كل إجابة صحيحة درجة ١ والإجابة الخاطئة درجة ٠ بنظام تسجيل ثنائي (Purwono et al., 2019) وباستخدام برنامج IBM SPSS STATISTICS 25 لمعرفة نتائج مستوى الصدق والثبات لبنود أسئلة مادة اللغة العربية في اختبار التقييم التجميعي النهائي(SAT) .

نتائج البحث ومناقشتها

نتائج البحث

تم إجراء تحليل لبنود الأسئلة بهدف تحديد البنود الجيدة وتلك التي تقل جودتها. وقد ذكر (Arikunto, 2008) أن تحليل الأسئلة هو إجراء يتم بصورة منهجية للحصول على معلومات محددة حول كل بنود اختبارية تم إعدادها. وفي السياق ذاته، أوضح (Sudjana, 1989) أن تحليل بنود الأسئلة، أو تحليل البنود، هو نشاط يُجرى لفحص الأسئلة في الاختبار من أجل الحصول على مجموعة أسئلة ذات جودة عالية وكافية.

وقبل استخدام الأسئلة، ينبغي تحليلها أولًا لمعرفة مدى جودتها. وتُعدّ الأسئلة ذات جودة إذا استوفت خصائص التقويم، أي إذا كانت تتمتع بدرجة عالية من الصدق والثبات (Novia et al., 2020). فكلما ارتفعت درجة الصدق والثبات في الأداة، كانت البيانات الناتجة عنها أكثر دقة (Rasmuin et al., 2021). وكما أوضح (Anshari et al., 2024)، يُستخدم اختبار الصدق للتأكد من أن كل فقرة تقيس الجانب المناسب، بينما تشير القيمة العالية للثبات إلى اتساق الأداة ككل. وبناءً على ذلك، يمكن للمعلم أن يحدد الفقرات الضعيفة ويُجري التحسينات اللازمة، بحيث تصبح الأداة ليست متسقة فحسب، بل دقيقة أيضًا في قياس كفاءة الطلاب.

وفي هذا البحث، استخدم الباحث أسئلة اختبار التقييم التجميعي النهائي (SAT) لمادة اللغة العربية للصف السابع (الفصل ج) بالمدرسة المتوسطة بمدينة مالانج كموضوع للدراسة. وكانت الأسئلة على شكل اختيار من متعدد بعدد ٣٣ بنود، لكل منها أربع خيارات للإجابة. ومن خلال هذا التحليل تم الحصول على معلومات مفيدة لإعداد أسئلة ذات جودة عالية، وفي هذه المناقشة سيتم الكشف عن البنود الصالحة وغير الصالحة، وكذلك البنود الثابتة وغير الثابتة.

بعد جمع البيانات، أجرى الباحث تحليلًا لبنود أسئلة اختبار التقييم التجميعي النهائي (SAT) لمادة اللغة العربية للصف السابع (الفصل ج) بالمدرسة المتوسطة بمدينة مالانج. وتم إجراء هذا التحليل باستخدام برنامج IBM SPSS Statistics 25، الذي نتجت عنه بيانات إحصائية تتعلق بمستوى صلاحية وثبات بنود الأسئلة.

الصدق

الصدق هي ترجمة لمصطلح validity، وتشير إلى مدى قدرة أداة القياس (مثل الاختبار) على القياس بدقة وصدق وفقًا للهدف الذي وضعت من أجله. يُقال إن الاختبار يتمتع بصلاحية عالية إذا أدت الأداة وظيفتها القياسية بدقة وأنتجت بيانات تتماشى مع هدف القياس المحدد (Ramadhan et al., 2024). وللصلاحية درجات تتراوح من منخفضة إلى عالية، فكلما ارتفعت قيمة الصلاحية لأداة ما زادت درجة موثوقية نتائج قياسها (Susetyo, 2011). وتُعد الصلاحية مؤشرًا يدل على مدى صحة ودقة الأداة في قياس ما ينبغي قياسه. فكلما ارتفعت درجة الصلاحية لأداة ما زادت إمكانية الاعتماد عليها، وعلى العكس من ذلك إذا انخفضت صلاحيتها كانت الأداة أقل موثوقية (Azwar, 2011).

وتُعتبر الأداة صالحة إذا كانت قيمة (r) المحسوبة أكبر من (r) الجدولية وقيمة الدلالة (Sig) أقل من ٠,٠٥.وقد تم تحليل صلاحية كل بند في هذا التحليل باستخدام تقنية الارتباط الثنائي. ثم تم تفسير قيمة معامل الصلاحية لكلبند بناءً على المعايير التي حددها (Arikunto, 2009)، كما هو موضح في الجدول (١):

الجدول ١: جدول تفسير الصدق

معيارالمقياس	التصنيف
٠,٠٠ – ٠،٢٠	منخفض جدًا
٠،٢١ – ٠،٤٠	منخفض
٠،٤١ – ٠،٦٠	متوسط
٠،٦١ – ٠،٨٠	مرتفع
٠،٨١ – ١،٠٠	مرتفع جدًا

بناءً على تحليل البيانات الذي قام به الباحث بمساعدة برنامج IBM SPSS Statistics 25باستخدام تقنية الارتباط الثنائي، تم الحصول على نتائج صلاحية بنود أسئلة اختبار التقييم التجميعي النهائي (SAT) لمادة اللغة العربية في المدرسة المتوسطة بمدينة مالانج كما هو موضح في الجدول أدناه:

الجدول ٢: جدول قيم الصدق

رقم السؤال	معامل ارتباط بيرسون/ (r) المحسوبة	القيمة (r) الجدولة ٥٪	قيمة كبيرة	معلومة	تفسير
.١	٠،٠٩٨	٠،٣٩٦	٠،٦٤٠	غير صدق	-
.٢	٠،٣٥٤	٠،٣٩٦	٠،٠٨٣	غير صدق	-
.٣	٠،٦٥٣	٠،٣٩٦	٠،٠٠٠	صدق	مرتفع
.٤	٠،٢٩١	٠،٣٩٦	٠،١٥٩	غير صدق	-
.٥	٠،٢٧٢	٠،٣٩٦	٠،١٨٩	غير صدق	-
.٦	٠،١٤٤	٠،٣٩٦	٠،٤٩٣	غير صدق	-
.٨	٠،١٠١	٠،٣٩٦	٠،٦٣٣	غير صدق	-
.٩	٠،٦١٢	٠،٣٩٦	٠،٠٠١	صدق	مرتفع
.١٠	٠،٢٧١	٠،٣٩٦	٠،١٩٠	غير صدق	-
.١١	٠،٢٢٨	٠،٣٩٦	٠،٢٧٢	غير صدق	-
.١٢	٠،٦٣٧	٠،٣٩٦	٠،٠٠١	صدق	مرتفع
.١٣	٠،٤٤١	٠،٣٩٦	٠،٠٢٧	صدق	متوسط

.١٤	٠،٢١٢	٠،٣٩٦	٠،٣٠٩	غير صدق	-
.١٥	٠،٥٣٨	٠،٣٩٦	٠،٠٠٦	صدق	متوسط
.١٦	٠،٣٨٥-	٠،٣٩٦	٠،٠٥٨	غير صدق	-
.١٧	٠،٣٥٩	٠،٣٩٦	٠،٠٧٨	غير صدق	-
.١٨	٠،٣٨٥	٠،٣٩٦	٠،٠٥٨	غير صدق	-
.١٩	٠،١٧١-	٠،٣٩٦	٠،٤١٣	غير صدق	-
.٢٠	٠،٣٩٠	٠،٣٩٦	٠،٠٥٤	غير صدق	-
.٢١	٠،١٧٩	٠،٣٩٦	٠،٣٩١	غير صدق	-
.٢٢	٠،٣٣٤	٠،٣٩٦	٠،١٠٢	غير صدق	-
.٢٣	٠،٤٤٥	٠،٣٩٦	٠،٠٢٦	صدق	متوسط
.٢٤	٠،٤٩٩	٠،٣٩٦	٠،٠١١	صدق	متوسط
.٢٥	٠،٠٩٨	٠،٣٩٦	٠،٦٤٠	غير صدق	-
.٢٦	٠،٠٥٤	٠،٣٩٦	٠،٧٩٧	غير صدق	-
.٢٧	٠،٥٩٠-	٠،٣٩٦	٠،٠٠٢	غير صدق	-
.٢٨	٠،٣٨٠	٠،٣٩٦	٠،٠٦١	غير صدق	-
.٢٩	٠،٠٧٥	٠،٣٩٦	٠،٧٢١	غير صدق	-
.٣٠	٠،٢٤٣	٠،٣٩٦	٠،٢٤٢	غير صدق	-
.٣١	٠،١٥٦	٠،٣٩٦	٠،٤٥٦	غير صدق	-
.٣٢	٠،٣٠١	٠،٣٩٦	٠،١٤٤	غير صدق	-
.٣٣	٠،٥٢٧	٠،٣٩٦	٠،٠٠٧	صدق	متوسط
.٣٤	٠،٧٠١	٠،٣٩٦	٠,٠٠٠	صدق	مرتفع

أظهرت نتائج تحليل ٣٣ بنود أسئلة الاختيار من متعدد لاختبار التقييم التجميعي النهائي (SAT) لمادة اللغة العربية في المدرسة المتوسطة بمدينة مالانج أن هناك ٩ بنود أسئلة صالحة؛ منها ٤بنود مصنفة ضمن الفئة العالية و٥ بنود ضمن الفئة المتوسطة، وهي الأسئلة ذات الأرقام: ٣، ٩، ١٢، ١٣، ١٥، ٢٣، ٢٤، ٣٣ و٣٤. بينما تم تصنيف ٢٤ بنود بأنها غير صالحة، وهي ذات الأرقام: ١، ٢، ٤، ٥، ٦، ٨، ١٠، ١١، ١٤، ١٦، ١٧، ١٨، ١٩، ٢٠، ٢١، ٢٢، ٢٥، ٢٦، ٢٧، ٢٨، ٢٩، ٣٠، ٣١ و٣٢. وقد تم تحديد حالة صلاحية بنود الاختبار بناءً على معيارين، هما: يجب أن تكون قيمة الدلالة أقل من ٠,٠٥ أو أن تكون قيمة (r) المحسوبة أكبر من قيمة (r) الجدولية (Sugiyono, 2013). ويُظهر هذا أن غالبية بنود الأسئلة التي تم اختبارها في المدرسة المتوسطة بمدينة مالانج تقع ضمن فئة الاختبار غير الصالح.

الثبات

إن مصطلح الثبات مشتق من الكلمة الإنجليزية reliable والتي تعني: موثوق به أو جدير بالثقة. وتُعتبر أداة القياس موثوقة إذا كانت قادرة على إنتاج بيانات ثابتة أو مستقرة من وقت لآخر، بحيث تكون نتائج قياسها موثوقة (Ainin et al., 2006). ووفقًا لـ (Azwar, 2011) فإن مفهوم الثبات يشير إلى درجة الثقة في نتائج القياس، إذ يُعد القياس موثوقًا إذا أعطى نتائج متشابهة نسبيًا عند إجرائه عدة مرات على نفس مجموعة الأفراد، ما دامت جوانب القياس لدى الأفراد لم تتغير.

ويهدف اختبار الثبات إلى التأكد من أن الأداة موثوقة بدرجة كافية وجديرة بأن تُستخدم كوسيلة لجمع البيانات نظرًا لجودتها العالية (Arikunto, 2018). ويتراوح معامل الثبات عمومًا بين ٠,٠٠ إلى ١,٠٠، وله معايير محددة تُستخدم كمرشد لتفسير نتائج حساب الثبات للاختبار باستخدام برنامج IBM SPSS Statistics (Mutholib, 2013).

وقد أوضح (Arikunto, 2009) أن القيم المستخدمة لتفسير الثبات يمكن عرضها في الجدول الآتي:

الجدول ٣: تفسير فئات الثبات

معيار المقياس	التصنيف
٠،٠٠ – ٠،٢٠	منخفض جدًا
٠،٢١ – ٠،٤٠	منخفض
٠،٤١ – ٠،٦٠	متوسط
٠،٦١ – ٠،٨٠	مرتفع
٠،٨١ – ١،٠٠	مرتفع جدًا

تم إجراء اختبار الثبات في هذا البحث باستخدام برنامج IBM SPSS Statistics 25 وفيما يلي نتائج الثبات التي تم الحصول عليها:

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.657	.674	34

الصورة ١: ملخص معالجة الحالات

Case Processing Summary

		N	%
Cases	Valid	25	100.0
	Excluded[a]	0	.0
	Total	25	100.0

a. Listwise deletion based on all variables in the procedure.

الصورة ٢: إحصاءات الثبات

بناءً على نتائج تحليل البيانات، تم اختبار ثبات الأسئلة في اختبار نهاية السنة لمادة اللغة العربية للصف السابع في المدرسة المتوسطة الرحمة بمدينة مالانج باستخدام طريقة Cronbach's Alpha. وتم تحديد الثبات بناءً على قيمة Cronbach's Alpha فإذا كانت قيمة Cronbach's Alpha أكبر من قيمة (r) الجدولية فإن الأسئلة تُعتبر موثوقة، أما إذا كانت القيمة أقل من قيمة (r) الجدولية فتُعتبر غير موثوقة (Joko Widiyanto, 2018)

أما النتيجة التي تم الحصول عليها استنادًا إلى الجدول أعلاه باستخدام طريقة Cronbach's Alpha فهي ٠,٦٥٧، وهي أكبر من قيمة (r) الجدولية البالغة ٠,٣٩٦. ويدل هذا على أن بنود الأسئلة في اختبار التقييم التجميعي النهائي (SAT) لمادة اللغة العربية للصف السابع (الفصل ج) في المدرسة المتوسطة بمدينة مالانج موثوقة وتقع ضمن الفئة العالية.

مناقشة البحث

بناءً على نتائج تحليل صدق ٣٣ سؤالاً من بنود أسئلة الاختيار من متعدد في اختبار التقييم التجميعي النهائي (SAT) لمادة اللغة العربية للصف السابع بالمدرسة المتوسطة بمدينة مالانج، يتضح أن معظم الأسئلة لا تستوفي معايير الصدق المحددة. تشير هذه النتيجة إلى أن نسبة الأسئلة القادرة على قياس مهارات الطلاب بدقة ووضوح لا تزال منخفضة نسبيًا. ويشير هذا المستوى المنخفض من الصدق إلى أن عددًا كبيرًا من الأسئلة لا يمثل أهداف التعليم تمثيلًا كافيًا، سواء من حيث المحتوى أو البنية أو سياق استخدامها.

وقد وُجدت نتائج مشابهة في البحث الذي أجرته فتحية (٢٠١٩)، حيث بيّنت نتائج الدراسة أن غالبية بنود أسئلة الاختيار من متعدّد لاختبار نهاية الفصل الأول للعام الدراسي ٢٠١٨/٢٠١٩ لمادة اللغة العربية لطلاب الصف الحادي عشر – قسم العلوم الاجتماعية في مدرسة "رَوْضَة العُلُوم" الثانوية الإسلامية بـ"باجَكْ

– مالانغ" لم تستوفِ معايير الصدق. إذ تبيّن أن ١٢ بندا من أصل ١٥ بنود لم تكن صالحة (غير صادقة)، بينما لم تُعتبر صادقة إلا ثلاث بنود فقط. وهذا يدل أيضًا على أن معظم بنود أسئلة اختبار اللغة العربية في المدرسة نفسها غير صادقة. ورغم وجود اختلاف في عدد الأسئلة ونسبة البنود الصادقة بين الدراستين، فإن كليهما تُظهران بوجه عام ضعف جودة صدق أداة تقويم تعلم اللغة العربية، مما قد يؤثر على دقة قياس قدرات الطلاب.

وقد تبيّن أن بنود أسئلة اختبار التقييم التجميعي النهائي (SAT) لمادة اللغة العربية في المدرسة المتوسطة بمدينة مالانج موثوقة رغم أن معظم البنود لا تحقق معايير الصلاحية. يتماشى هذا الشرط مع رأي (Sugiyono, 2004) الذي ينص على أن الأداة الموثوقة ليست بالضرورة صالحة، وأن الموثوقية شرط مطلق في اختبار الصلاحية. هذا يعني أن أداة القياس يمكن أن تُظهر مستوى عالٍ من الاتساق، ولكنها ليست بالضرورة قادرة على قياس الجوانب التي يجب قياسها بدقة. على سبيل المثال، يمكن أن ينتج اختبار يتكون من أسئلة سهلة للغاية إجابات موحدة من المشاركين (جميعها صحيحة)، على الرغم من أن الاختبار لا يعكس بدقة القدرات الفعلية للطلاب. في مثل هذه الحالة، يُقال إن الاختبار موثوق لأنه متسق، ولكنه غير صالح لأنه ليس على الهدف. بمعنى آخر، تُظهر الموثوقية فقط مستوى اتساق إجابات المشاركين، وليس ملاءمة محتوى السؤال للكفاءة المراد قياسها. لذلك، على الرغم من أن غالبية الأسئلة غير صالحة، تظل قيمة الموثوقية عالية لأن المشاركين يجيبون على الأسئلة بشكل متسق.

وتتفق هذه النتيجة مع ما توصلت إليه فتيحة (Fathiyah, 2019) في بحثها، حيث كشفت أن نتيجة تحليل ثبات بنود أسئلة الاختيار من متعدد لاختبار نهاية الفصل الأول للعام الدراسي ٢٠١٨/٢٠١٩ في مادة اللغة العربية لطلاب الصف الحادي عشر – قسم العلوم الاجتماعية في مدرسة "رَوْضَة العُلُوم" الثانوية الإسلامية بـ"باجَكْ – مالانغ" بلغت (٠,٥٤)، مما يدل على أنها تقع في فئة "المتوسط"، أي أن الاختبار يُعد موثوقًا (ثابتًا). وتتوافق نتائج فتحية هذه مع نتائج هذه الدراسة، إذ إنّ غالبية بنود الأسئلة لم تستوفِ معايير الصدق، حيث كانت ١٢ فقرة من أصل ١٥ فقرة غير صادقة، بينما لم تكن صادقة سوى ثلاثة بنود فقط. ومع ذلك، فإن تحليل الثبات يُظهر أن مستوى ثبات الاختبار ما يزال جيدًا. ويدل هذا على أنه رغم عدم صدق غالبية البنود أو عدم دقتها في قياس الكفاءة المستهدفة، فإن الاختبار ككل ما يزال يُظهر اتساقًا في نتائج القياس. وتُبيّن هذه الحالة أن أداة التقويم يمكن أن تكون ذات درجة عالية من الثبات (الموثوقية)، حتى وإن كانت معظم بنودها غير صادقة.

وتتوافق هذه النتيجة أيضًا مع نتائج دراسة أخرى أجراها ترمذي وآخرون (Tarmizi et al., 2021)، والتي أظهرت أن درجة ثبات بنود أسئلة الاختيار من متعدد في موضوع "الصحة أمر مهم" تقع ضمن الفئة العالية، حيث بلغت قيمة معامل الثبات (٠،٨٣). وتدل هذه القيمة على أن فقرات الأسئلة تتمتع باتساق داخلي قوي، مما يسمح بالاستنتاج أن الأداة المستخدمة تُعد موثوقة (ثابتة). وبناءً على ذلك، فإن نتائج التقييم المستخلصة من استخدام هذه الأداة تُعتبر مستقرة نسبيًا وجديرة بالثقة. وتُؤكّد هذه النتيجة بشكل أكبر على أهمية اختبار الثبات بوصفه أحد المؤشرات الرئيسية في تحديد جودة أداة التقويم، وخصوصًا في ضمان أن تكون البيانات الناتجة متسقة وصالحة للاستخدام كأساس في اتخاذ القرارات في العملية التعليمية.

الخاتمة

أظهرت نتائج التحليل لصلاحية وثبات بنود أسئلة اختبار التقييم التجميعي النهائي (SAT) لمادة اللغة العربية للصف السابع في المدرسة المتوسطة بمدينة مالانج أن من بين ٣٣ بنود أسئلة اختيار من متعدد، تبيّن أن ٩ بنود فقط صالحة بينما الـ ٢٤ بنود الأخرى غير صالحة بناءً على معيار أن تكون قيمة (r) المحسوبة أكبر من قيمة (r) الجدولية وقيمة الدلالة أقل من ٠,٠٥. ومع ذلك، فقد أظهر اختبار الثبات باستخدام طريقة Cronbach's Alpha قيمة قدرها ٠,٦٥٧ وهي أكبر من قيمة (r) الجدولية البالغة ٠,٣٩٦، مما يُصنف ضمن الفئة العالية للثبات. وتوضح هذه النتيجة أن أداة القياس قد تكون موثوقة حتى وإن كانت العديد من فقراتها غير صالحة، لأن ما يُقاس هنا هو مستوى الاتساق وليس دقة القياس للكفاءة المستهدفة. يوفر هذا البحث صورة عملية يمكن أن تكون مادة تقييم للمعلمين لتحسين جودة الأسئلة، كما يثري المراجع الأكاديمية المتعلقة بإعداد أدوات تقييم تعلم اللغة العربية في المرحلة المتوسطة. لذا يُوصى بأن يقوم المعلم بإجراء تحليل شامل لبنود الأسئلة قبل استخدامها. ويُؤمل أن تشمل الدراسات اللاحقة جوانب أخرى مثل مستوى الصعوبة، قوة التمييز، وفعالية المشتتات، مع إشراك المنهج النوعي لفهم أسباب انخفاض صلاحية الأسئلة.

المراجع

Ainin, M., Tohir, M., & Asrori, I. (2006). *Evaluasi dalam Pembelajaran Bahasa Arab*. Malang: Misykat.

Anshari, M. I., Nasution, R., Irsyad, M., Alifa, A. Z., & Zuhriyah, I. A. (2024). Analisis Validitas dan Reliabilitas Butir Soal Sumatif Akhir Semester Ganjil Mata Pelajaran PAI. *EDUKATIF : JURNAL ILMU PENDIDIKAN, 6*(1), 964–975. https://doi.org/10.31004/edukatif.v6i1.5931

Arikunto, S. (2002). *Prosedur Penelitian Suatu Pendekatan Praktek*. Jakarta: Rineka Cipta.

Arikunto, S. (2008). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.

Arikunto, S. (2009). *Dasar-Dasar Evaluasi Pendidikan* (Revisi). Jakarta: Bumi Aksara.

Arikunto, S. (2018). *Dasar-Dasar Evaluasi Pendidikan, Edisi 3*. Jakarta: Bumi Aksara.

Azwar, S. (2011). *Penyusunan Skala Psikologi.* (p. 5). Yogyakarta: Pustaka Pelajar.

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), 1–13. https://doi.org/10.47134/jbdi.v1i3.2863

Fadli, R., Hidayati, S., Cholifah, M., Siroj, R. A., & Afgani, M. W. (2023). Validitas dan Reliabilitas pada Penelitian Motivasi Belajar Pendidikan Agama Islam Menggunakan Product Moment. *JIIP - Jurnal Ilmiah Ilmu Pendidikan, 6*(3), 1734–1739. https://doi.org/10.54371/jiip.v6i3.1419

Fahmi, B., Rizqi, S., & Harmeilinda, N. E. (2022). Analisis Butir Soal Bahasa Arab Siswa MAS Pondok Pesantren Assalam Kampar Riau. *Ta'lim Al-'Arabiyyah: Jurnal Pendidikan Bahasa Arab & Kebahasaaraban, 6*(1), 95–105. https://doi.org/10.15575/jpba.v6i1.16193

Faiz, A., Putra, N. P., & Nugraha, F. (2022). Memahami Makna Tes, Pengukuran (Measurement), Penilaian (Assessment), dan Evaluasi (Evaluation) dalam Pendidikan. *Jurnal Education and Development, 10*(3), 4492–4495. https://doi.org/10.31004/obsesi.v5i2.972

Fathiyah, S. F. Al. (2019). Analisis Butir Soal Pelajaran Bahasa Arab di MA Roudlotul Ulum Pagak Malang. *Tarbiyatuna: Jurnal Pendidikan Ilmiah, 4*(1), 76–100. https://ejournal.kopertais4.or.id/mataraman/index.php/tarbiyatuna/article/view/3657

Fikriyah, N. (2021). Analisis Butir Soal Ulangan Tengah Semester Mata Pelajaran Bahasa Arab Kelas VII Semester Ganjil SMP Muhammadiyah 1 Yogyakarta Tahun Ajaran 2019/2020. *Maharaat: Jurnal Pendidikan Bahasa Arab, 3*(2). https://doi.org/10.18196/mht.v3i2.10501

Himawan, R., & Suyata, P. (2022). Developing Hots Questions: Evaluating Persuasive Speech Text Learning in Grade IX of Junior High School. *Gramatika: Jurnal Penelitian Pendidikan Bahasa Dan Sastrs Indonesia*, *8*(1), 50–64. https://doi.org/10.22202/jg.2022.v8i1.4810

Ida, F. F., & Musyarofah, A. (2021). Validitas dan Reliabilitas dalam Analisis Butir Soal. *AL-MU'ARRIB: JOURNAL OF ARABIC EDUCATION*, *1*(1), 34–44. https://doi.org/10.32923/al-muarrib.v1i1.2100

Joko Widiyanto. (2018). *SPSS for Windows untuk Analisis Data Statistik dan Penelitian* (p. 108). Surakarta: BP-FKIP UMS.

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. *BINTANG : Jurnal Pendidikan Dan Sains*, *3*(2), 198–214. https://doi.org/ 10.36088/bintang.v3i2.1291

Maulana, D., & Sanusi, A. (2020). ANALISIS BUTIR SOAL BAHASA ARAB UJIAN AKHIR MADRASAH BERSAMA DAERAH (UAMBD) MADRASAH IBTIDAIYAH TAHUN 2017-2018. *Ta'lim Al-'Arabiyyah: Jurnal Pendidikan Bahasa Arab & Kebahasaaraban*, *4*(1), 12–24. https://doi.org/10.15575/jpba.v4i1.8054

Mutholib, A. (2013). Analisis Butir Soal Bahasa Arab Ujian AKhir Madrasah Berstandar Nasional (UAMBN) Bagi Siswa Madrasah Ibtidaiyah (MI) Tahun Pelajaran 2012-2013. *Arabia*, *5*(2), 141–161. https://doi.org/10.21043/arabia.v5i2.1387

Novia, T., Wardani, A., Canda, C., Nurdi, N., & Nurmasyitah, N. (2020). Analisis Validitas dan Reliabilitas Butir Soal UTS Fisika Kelas X SMA Swasta Muhammadiyah 4 Langsa. *GRAVITASI: Jurnal Pendidikan Fisika Dan Sains*, *3*(01), 19–22. https://doi.org/10.33059/gravitasi.jpfs.v3i01.2256

Purwono, Hasyim, F., Ulya, A. U., Purnasari, N., & Juniatmoko, R. (2019). *Metodologi Penelitian (Kuantitatif, Kualitatif dan Mixed Method)* (pp. 33–40). Jakarta: Guepedia Group.

Ramadhan, M. F., Siroj, R. A., & Afgani, M. W. (2024). Validitas and Reliabilitas. *Journal on Education*, *6*(2), 10967–10975. https://doi.org/10.31004/joe.v6i2.4885

Rasmuin, R., Jais, E., & Wahyuni, A. A. S. M. (2021). Reliabilitas Tes Buatan Guru Mata Pelajaran Matematika Menurut Teori Tes Klasik Tingkat SMP Di Kota Baubau. *Jurnal Akademik Pendidikan Matematika*, 191–196. https://doi.org/10.55340/japm.v7i2.465

Rezi, M., & Aulia, A. (2020). Tahapan Penyusunan dan Analisis Tes Bahasa Arab. *Jurnal Ilmiah Al-Furqon*, *5*(2), 57. https://doi.org/10.69880/alfurqan.v7i2.9

Sudjana, N. (1989). *Penilaian hasil proses belajar mengajar*. Bandung: PT. Remaja Rosdakarya.

Sugiyono. (2004). *Metode Penelitian Bisnis*. Bandung: Alfabeta.

Sugiyono. (2013). *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. Bandung: Alfabeta.

Suryawati, & Yulfikar. (2012). Kualitas Tes dan Hasil Belajar Matematika Siswa Kelas VIII SMP Negeri 9 Banda Aceh Tahun Pelajaran 2011/2012. *Jurnal Peluang*, *1*(1). https://jurnal.usk.ac.id/peluang/article/view/1300

Susetyo, B. (2011). *Menyusun Tes Hasil Belajar*. Bandung: Cakara.

Tanjung, M. A. H. R., Fahmi, A. A., Rahmanita, F., Habibah, I. F., & Qomari, N. (2024). Analisis Butir Soal Penilaian Akhir Tahun Pelajaran Bahasa Arab Kelas VII MTS Al-Ma'arif Rakit Bajanegara Jawa Tengah. *Mantiqu Tayr: Journal of AArabic Language*, *4*(1), 2774–6372. https://doi.org/10.252117/mantiqutayr.v4i1.4038

Tarmizi, P., Setiono, P., Amaliyah, Y., & Agrian, A. (2021). Analisis Butir Soal Pilihan Ganda Tema Sehat Itu Penting Kelas V SD Negeri 04 Kota Bengkulu. *ELSE (Elementary School Education Journal) : Jurnal Pendidikan Dan Pembelajaran Sekolah Dasar*, *4*(2), 124. https://doi.org/10.30651/else.v4i2.7090