

تحليل بنود أسئلة الاختبار النهائي للغة العربية لطلاب الفصل الثامن في المدرسة الثانوية الإسلامية

باتاهان سيلاينغ

Sakiah Rangkuti

Zikra wahyuni

UIN Sjech M.Djamil Djambek Bukittinggi

sakiahrangkuti^{٨٨٠}@mail.com zikrawahyunimaiza@uinbukittinggi.ac.id

الملخص

هذا النوع من البحث هو بحث تقييمي باستخدام الأساليب الكمية. هدف هذا البحث هو التعرف على جودة بنود أسئلة الاختبار النهائي في درس اللغة العربية من حيث مستوى صعوبة الأسئلة، وقوة التمييز، ووظيفة المشتتات، والموثوقية، والصلاحية في الصف الثامن بالمدرسة الثانوية الإسلامية باتاهان سيلاينغ. وأظهرت نتائج البحث أن مستوى صعوبة الأسئلة هو: سهلة (٠٪)، متوسطة ١٢ بنداً (٣٠٪)، صعبة ٢٧ بنداً (٦٧,٥٪)، وصعبة جداً بند واحد (٢,٥٪). أما قوة تمييز الأسئلة، فكانت موزعة على الشكل التالي: جيدة جداً (٠٪)، جيدة ٢ بنداً (٥٪)، متوسطة ١٢ بنداً (٣٠٪)، ضعيفة ١٧ بنداً (٤٢,٥٪)، وضعيفة جداً ٨ بنود (٢٠٪). ومن ناحية فعالية المشتتات، تبين أن ٢٩ بنداً (٧٢,٥٪) ذات مشتتات فعالة تعمل بشكل جيد، بينما ١١ بنداً (٢٧,٥٪) لم تكن مشتتاتها فعالة. وأظهرت النتائج الخاصة بالصلاحية أن ١٤ بنداً (٣٥٪) فقط هي الصالحة، في حين بلغ عدد البنود غير الصالحة ٢٦ بنداً (٦٥٪)، وقيمة موثوقية الاختبار بلغت معامل ٠,٦ مما يعني أنها متوسطة وضمن معيار المقبول. وبناء على هذه النتائج، توصلت الباحثة إلى أن جودة الأسئلة بشكل عام تحتاج إلى مراجعة وتحسين، لأن معظم البنود لم تستوف المعايير الجيدة للاختبار.

الكلمات المفتاحية: تحليل بنود الأسئلة، الاختبار النهائي، اللغة العربية.

المُقدِّمة

يؤدي المعلم دوراً هاماً في عملية التعليم والتقييم، فهو المسؤول عن وضع الاختبارات لقياس مدى تحقق الأهداف التعليمية عند المتعلمين. وفي تعليم اللغة العربية، يتم التقييم عادةً من خلال اختبارات لقياس المهارات الأربع (الاستماع والكلام والقراءة والكتابة) التي تهدف إلى التعرف على مستوى اكتساب الطلبة للغة.^١

^١ Elviana, E. (٢٠٢٠). Analisis Butir Soal Evaluasi Pembelajaran Pendidikan Agama Islam Menggunakan Program Anates. Jurnal Mudarrisuna, ١٠(٢).

ويتكون الاختبار الجيد من مجموعة من البنود التي تقيس بشكل دقيق التحصيل الدراسي وتوفر معلومات موضوعية تميز بين الطلبة بحسب مستواهم الحقيقي. لذا فمن الضروري القيام بتحليل بنود الأسئلة بشكل دوري ومستمر للتأكد من فعاليتها وجودتها من الناحيتين النظرية والتطبيقية، والتعرف على نقاط القوة والضعف فيها بغرض تحسين عملية التعليم والتقييم.^٢

تحليل بنود الأسئلة هو عملية منهجية لتقييم الأسئلة بهدف تحديد مستوى صعوبتها وقوة تمييزها وموثوقيتها وصلاحيتها ووظيفة المشتتات (البدائل الخاطئة). إن البنود المثالية ينبغي ألا تكون شديدة السهولة ولا بالغة الصعوبة، وأن تكون قادرة على تمييز الطلبة المتفوقين من الطلبة ذوي الأداء الضعيف.^٣ كما يجب أن تتصف الأسئلة بالموثوقية العالية، بحيث تعطي نتائج متسقة عند إعادة تطبيقها مرة أخرى، بالإضافة إلى توفر الصلاحية في قياس الأهداف التعليمية المحددة بشكل دقيق.^٤

في المدرسة الثانوية الإسلامية باتاهان سيلاينج، أجريت مقابلة أولية مع مدير المدرسة أفادت بأن الأسئلة المستخدمة حالياً في اختبار اللغة العربية لم تخضع سابقاً لتحليل علمي أو تقييم شامل للجودة. وعليه، لاحظت الباحثة أن الحاجة ماسة إلى إجراء دراسة منهجية لتحليل بنود اختبار اللغة العربية نهاية الفصل الدراسي للصف الثامن في المدرسة المذكورة، بهدف الكشف عن خصائص هذه بنود الأسئلة ومدى جودتها من حيث مستوى الصعوبة، وقوة التمييز، والموثوقية والصلاحية، ووظيفة المشتتات.

ومن هنا انطلقت الباحثة في القيام بهذا البحث للكشف عن جودة بنود الأسئلة بهدف المساهمة في تطوير جودة التعليم والتقييم، وإعطاء صورة علمية واضحة عن مدى تحقيق اختبارات اللغة العربية للأهداف المنشودة في العملية التربوية.

منهجية البحث

أ. نوع البحث ومنهجه

وفقاً لأهداف هذا البحث، فإن نوع البحث المستخدم هو البحث التقييمي (Evaluative Research). ويهدف هذا النوع من البحوث إلى جمع البيانات وتقديم معلومات دقيقة وموضوعية تتعلق بتنفيذ عملية التقييم التعليمي، معتمداً على المعايير العلمية المحددة. ويستعين الباحث في هذه الدراسة بالمنهج الوصفي التحليلي باستخدام الأساليب الكمية لتحليل بنود أسئلة اختبار مادة اللغة العربية في المدرسة الثانوية الإسلامية باتاهان سيلاينج.

ب. مجتمع البحث وعينته

^٢ Hughes, A. (٢٠٠٣). Testing for Language Teachers. Cambridge University Press.

^٣ Ebel, R. L., & Frisbie, D. A. (١٩٩١). Essentials of Educational Measurement. Prentice-Hall.

^٤ Fulcher, G., & Davidson, F. (٢٠٠٧). Language Testing and Assessment: An Advanced Resource Book. Routledge.

مجتمع البحث في هذه الدراسة هم جميع طلاب الفصل الثامن في المدرسة الثانوية الإسلامية باتاهان سيلابينغ للعام الدراسي ٢٠٢١/٢٠٢٢. أما العينة، فقد تم اختيارها بأسلوب أخذ العينات العشوائية البسيطة (Simple Random Sampling) وعددها ٢٨ طالبًا وطالبة، وهو ما يمثل جزءًا من مجتمع الدراسة، ويستوفي شروط البحث العلمية من حيث الحجم والخصائص الملائمة لإجراء التحليلات الإحصائية اللازمة.

ج. مصادر البيانات وأدوات جمعها

لضمان دقة وموثوقية النتائج، قام الباحث بجمع البيانات من مصادر متعددة وهي كالتالي:

١. المصادر الأولية

- نتيجة اختبار اللغة العربية النهائي لطلاب الصف الثامن، والتي تشمل إجابات الطلاب على أسئلة الاختبار الموضوعية (الاختيار من متعدد).

٢. المصادر الثانوية

- الوثائق والسجلات المدرسية التي تحتوي على معلومات عن تصميم الاختبار وآليته ومعايير التقويم المتبعة في المدرسة.

أما أدوات جمع البيانات المستخدمة فهي كالتالي:

- الملاحظة: لجمع البيانات عن وضعية الاختبارات والأسئلة وطريقة تقديمها.

- المقابلة: تم إجراؤها مع مدير المدرسة وبعض مدرسي اللغة العربية لمعرفة إجراءات وضبط بنود الاختبار ومصادر الأسئلة.

- التوثيق: وثائق الاختبار والنتائج ومحاضر الدرجات التي تتضمن بيانات تفصيلية عن أداء الطلاب وجودة الاعتبارات التقويمية.

د. إجراءات تحليل البيانات

تم إجراء تحليل البيانات من خلال الطرق الإحصائية الكمية وذلك باستخدام برنامج التحليل الإحصائي SPSS. وقد شمل التحليل الجوانب الآتية:

- مستوى صعوبة الأسئلة (Difficulty Level)

- قوة التمييز (Item Discrimination)

- فعالية المشتتات (Distractor Effectiveness)

- صلاحية الأسئلة (Validity of Test Items)

- موثوقية الاختبار (Reliability of Test)

وفي نهاية هذه الإجراءات، تم تصنيف بنود الاختبار وفق معايير الجودة المتعارف عليها في القياس والتقويم التعليمي، وتقديم نتائج تفصيلية لهذه التحليلات للاستفادة منها في تحسين الاختبارات المستقبلية بالمدرسة.

بهذا التفصيل في قسم المنهجية، يكون القسم قد شمل بصورة واضحة نوع ومنهج الدراسة، تفصيل مجتمع البحث وعينته، تحديد مصادر وأدوات جمع البيانات، وكذلك الإجراءات الدقيقة لتحليل البيانات مما يمنح مصداقية وتكاملاً منهجياً وعلمياً للبحث.

نتائج البحث ومناقشتها

في هذا الفصل قامت الباحثة بتحليل خصائص أسئلة الاختبار النهائي الذي صممه مدرس اللغة العربية لطلاب الصف الثامن بالمدرسة الثانوية الإسلامية باتاهان سيلابنغ، وذلك من خلال قيامها بدراسة الجوانب التالية: متوسط صعوبة الأسئلة، قوة تمييز الأسئلة، فعالية استخدام المشتتات، الصلاحية والموثوقية.

١. مستوى صعوبة الأسئلة:

إن مستوى صعوبة السؤال هو مؤشر أساسي يحدد مدى ملاءمة السؤال وقدرته على قياس التحصيل الدراسي بشكل عادل ومتوازن. فالأسئلة المثلى يجب أن تكون متوسطة الصعوبة، بحيث يتمكن معظم الطلاب من الإجابة على جزء منها دون أن تكون سهلة جداً أو شديدة الصعوبة، وهو ما يضمن تحقيق أهداف الاختبار بشكل أفضل.

تستخدم الباحثة الصيغة التالية لحساب صعوبة الأسئلة الموضوعية:

$$[P = \frac{N}{P}]$$

وبتحليل الجدول، تبين التالي:

- سؤال واحد فقط (أي ما نسبته ٢,٥٪) صُنِّف ضمن فئة "صعب جداً"، وهو السؤال رقم ١، وكونه في هذه الفئة يعني بوضوح أنه لم يكن مناسباً لقياس مستوى التحصيل، وبالتالي يوصى بشدة بحذفه من الاختبارات المستقبلية .

- وجاءت ١٢ سؤالاً (٣٠٪) في المرتبة المتوسطة المناسبة، وهي أسئلة رقم ٢، ٣، ٥، ٦، ١٠، ١٧، ١٩، ٢٠، ٢٣، ٢٤، ٢٨، ٣٤ وهي أسئلة يمكن استخدامها كما هي أو مع تعديلات طفيفة تُحسِّن من قوتها وجودتها.

- وأخيراً، جاءت الأغلبية العظمى من الأسئلة (٢٧ سؤالاً، بنسبة ٦٧,٥٪) ضمن فئة الصعب، وهذه المجموعة الكبيرة من الأسئلة الصعبة تستدعي من مدرس المادة إعادة النظر فيها وإجراء تعديلات جوهرية لضمان عدم إجحاف الطلاب والتأكد من مدى مناسبتها لأهداف الدرس ومستوى الطلاب.

وفقاً لهذا التحليل، يمكن الاستنتاج بشكل عام أن الاختبار يتسم بمستوى صعوبة مرتفع، وهو الأمر الذي يوجب على مدرس المادة العمل على تعديلات جوهرية عاجلة لتحسين توازن مستوى صعوبة الأسئلة في الاختبار القادم.

٢. قوة تمييز الأسئلة:

قوة التمييز تعني مدى قدرة السؤال على إظهار الفروق بين الطلاب الذين يحققون مخرجات تعلم عالية (المتفوقون) وأولئك الذين يعانون من مشكلات في استيعاب المادة الدراسية (الضعفاء).

أظهر تحليل قوة التمييز التالي:

- سؤالان فقط (٥٪)، هما رقما ٥ و ١٧، في فئة "جيد" يعدان مثاليين ويُنصح بالإبقاء عليهما في صيغتهما الحالية دون أي تعديل نظرًا لفعاليتهم القوية.

- بينما توجد ١٢ سؤالاً (٣٠٪) جاءت في فئة "متوسط"، وهي الأسئلة ذات الأرقام ٤، ٨، ٢٢، ٢٣، ٢٥، ٢٩، ٣٣، ٣٤، ٣٥، ٣٦، ٣٧، ٣٨ وهي مناسبة لكنها بحاجة إلى تعديلات دقيقة تعزز من فعاليتها في تحديد مستوى الطلاب.

- وظهرت ١٧ سؤالاً (٤٢،٥٪) في فئة "قبیح" وهي الأسئلة رقم ١، ٢، ٧، ٩، ١٠، ١٢، ١٣، ١٤، ١٥، ١٦، ١٨، ٢٤، ٢٧، ٢٨، ٣١، ٣٢، ٤٠، وهذه الأسئلة بحاجة إلى مراجعة جذرية في محتواها أو صياغتها لضمان تحقيق الهدف الفعلي من بينهم.

- أما أكثر الأسئلة ضرراً هي الفئة الرابعة، فئة "قبیح جداً" بواقع ٨ أسئلة (٢٠٪)، وهي أرقام ٣، ٦، ١٩، ٢٠، ٢١، ٢٦، ٣٠، ٣٩، ويوصى باستبعادها تماماً من الاختبارات المقبلة.

تكشف هذه الأرقام عن ضعف شديد في قدرة الاختبار على أداء وظيفته في تمييز الفروق الفردية بين الطلاب، مما يستوجب مراجعة حازمة للعديد من بنود الاختبار.

٣. فعالية استخدام المشتتات:

المشتتات الجيدة عنصر حيوي هام في الاختبارات متعددة الاختيارات، إذ تساعد بشكل فعال في قياس مدى إدراك الطالب وقدرته على التمييز بين الصواب والخطأ بدقة. وجدت الدراسة:

- ٢٩ سؤالاً (٧٢،٥٪) كانت مشتتاتها جيدة ومُتقنة، مما يدل على استخدام فعال للمشتتات، وهذه الأسئلة بأرقام ١، ٣، ٤، ٦، ٧، ٨، ٩، ١١، ١٢، ١٤، ١٥، ١٦، ١٧، ١٨، ٢٠، ٢١، ٢٤، ٢٥، ٢٧، ٢٨، ٢٩، ٣١، ٣٢، ٣٤، ٣٥، ٣٦، ٣٨، ٣٩، ٤٠.

- و١١ سؤالاً (٢٧،٥٪) مشتتاتها ضعيفة وتحتاج لتحسين جذري.

وفقاً لذلك، تعد فعالية المشتتات مقبولة إلى حد كبير، لكن لا يزال هناك مجال واسع للتحسين والتطوير.
٤. صلاحية الأسئلة:

الصلاحية تمثل مدى نجاح الاختبار في قياس الأهداف المحددة. استخدمت الباحثة دلالة إحصائية ٥٪، وعينة مكونة من ٢٨ طالباً، لذا فإن قيمة الارتباط الجدولية I هي ٠،٣٨٨. مثال على الصلاحية وعدمها:

- رقم ٥ صالح (٠،٤٢١ < ٠،٣٨٨).

- رقم ١٠ غير صالح (٠،٣٠١ > ٠،٣٨٨).

والمقارنة كشفت للأسف أن ٦٥٪ من الأسئلة متعددة الاختيار غير صالحة بشكل عام! وهو إنذار هام يوجب التدخل العاجل لتحسين جودة الأسئلة وصحتها.
٥. الموثوقية (الثبات):

بلغ معامل الموثوقية ٠،٦٠، وهذه القيمة تشير إلى ظروف اختبار موثوقة نسبياً، حيث تقع ضمن فئة "العالية". بالتالي فإن الاختبار يمتلك ثباتاً طيباً ومقبولاً مما يعني إمكانية استخدامه مستقبلاً مع إجراء التعديلات والتحسينات المذكورة سابقاً.

في ختام هذا التحليل، تؤكد نتائج الدراسة بضرورة العمل الشامل على تطوير وتحسين بنود اختبارات اللغة العربية المستخدمة في المدرسة، من خلال تطبيق هذه التوصيات المهمة تعزيزاً للفعالية والجودة العالية بما يضمن تحقيق الأهداف المنشودة من العملية التربوية والتعليمية بأبهى صورة.

الخاتمة

بناءً على نتائج البحث والمناقشة التي أجرتها الباحثة، فإن استنتاجات هذا البحث هي على النحو التالي:

أ. فيما يتعلق بمستوى صعوبة الأسئلة، فقد توصلت النتائج إلى أن هناك بنداً واحداً (٢،٥٪) صعب جداً لا يُوصى باستخدامه مرة أخرى، و١٢ بنداً (٣٠٪) في الفئة المتوسطة وهي بنود جيدة وقابلة للاستعمال، و٢٧ بنداً (٦٧،٥٪) تقع ضمن الفئة الصعبة التي تحتاج إلى مراجعة وتعديل قبل الاستخدام التالي.

ب. أما بالنسبة لقوة التمييز، فقد تبين وجود ٢ بنداً (٥٪) في فئة جيدة، في حين أن ١٢ بنداً (٣٠٪) ضمن الفئة المتوسطة وهي تحتاج إلى تحسين ومراجعة، و١٧ بنداً (٤٢،٥٪) ضمن الفئة الضعيفة، و٨ بنود (٢٠٪) في فئة ضعيفة جداً يجب حذفها تماماً وعدم استخدامها في الاختبارات القادمة.

ج. فيما يخص فعالية المشتتات، أظهرت النتائج أن ٢٩ بنداً (٧٢،٥٪) من المشتتات تعمل بشكل صحيح وفعال، بينما هناك ١١ بنداً (٢٧،٥٪) لا تؤدي المشتتات فيها وظيفتها بشكل فعال وتحتاج للمراجعة.

د. من جانب الصلاحية، تبين أن هناك ١٤ بنداً (٣٥٪) فقط صالحة للاستخدام، بينما ٢٦ بنداً (٦٥٪) غير صالحة ولا تستوفي معايير جودة البنود.

هـ. بالنسبة لموثوقية الأسئلة، فقد بلغت قيمة معامل الموثوقية ٠،٦؛ ما يشير إلى مستوى موثوقية متوسط مقبول ضمن المعايير الإحصائية للاختبارات التعليمية.

وبالتالي، توصي الباحثة بضرورة إجراء تعديلات ومراجعات للبنود التي ثبت ضعفها وعدم صلاحيتها وفق المعايير العلمية لتحليل بنود الأسئلة، لضمان تطوير جودة الاختبارات في المستقبل.

المراجع

- Aiken, L. R. (٢٠٠٣). *Psychological testing and assessment* (١١th ed.). Allyn & Bacon.
- Aji Nugroho, M. (٢٠١٦). Studi analisis butir soal latihan buku ajar bahasa Arab Madrasah Tsanawiyah. *Jurnal Arabia*, 8(٢), ٨.
- Al-Ani, W. T. (٢٠١٣). Blended learning approach using Moodle and student's achievement at Sultan Qaboos University in Oman. *Journal of Education and Learning*, 2(٣), ٩٦-١١٠.
- Alderson, J. C., Clapham, C., & Wall, D. (١٩٩٥). *Language test construction and evaluation*. Cambridge University Press.
- Alruwais, N., Wills, G., & Wald, M. (٢٠١٨). Advantages and challenges of using e-assessment. *International Journal of Information and Education Technology*, 8(١), ٣٤-٣٧.
- Bachman, L. F., & Palmer, A. S. (١٩٩٦). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.
- Brown, H. D., & Abeywickrama, P. (٢٠١٠). *Language assessment: Principles and classroom practices* (٢nd ed.). Pearson Education.
- Crocker, L., & Algina, J. (٢٠٠٨). *Introduction to classical and modern test theory*. Cengage Learning.
- Dörnyei, Z., & Taguchi, T. (٢٠٠٩). *Questionnaires in second language research: Construction, administration, and processing* (٢nd ed.). Routledge.
- Ebel, R. L., & Frisbie, D. A. (١٩٩١). *Essentials of educational measurement* (٥th ed.). Prentice-Hall.

- Elviana, E. (٢٠٢٠). Analisis butir soal evaluasi pembelajaran pendidikan agama Islam menggunakan program Anates. *Jurnal Mudarrisuna*, 10(٢), ٦٠.
- Fulcher, G., & Davidson, F. (٢٠٠٧). *Language testing and assessment: An advanced resource book*. Routledge.
- Haladyna, T. M., & Rodriguez, M. C. (٢٠١٣). *Developing and validating test items*. Routledge.
- Hughes, A. (٢٠٠٣). *Testing for language teachers* (٢nd ed.). Cambridge University Press.
- Ilyas, M. (٢٠١٨). Analisis butir soal ujian akhir semester bahasa Arab MTs kelas VII keluaran dari Kemenag Banyumas di MTs MWI Kebarongan, Kemranjen, Banyumas tahun ajaran ٢٠١٧-٢٠١٨ [Skripsi]. IAIN Purwokerto.
- Kunnan, A. J. (Ed.). (٢٠٠٠). *Fairness and validation in language assessment: Selected papers from the 19th Language Testing Research Colloquium, Orlando, Florida*. Cambridge University Press.
- Maulana, D. (٢٠٢٠). Analisis butir soal bahasa Arab ujian akhir madrasah bersama daerah Madrasah Ibtidaiyah Tahun ٢٠١٧-٢٠١٨. *Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban*, 2(١), ١٣.
- McNamara, T. (٢٠٠٠). *Language testing*. Oxford University Press.
- Ningsih Pasi, S., & Yusrizal. (٢٠١٨). Analisis butir soal ujian bahasa Indonesia buatan guru MTsN di Kabupaten Aceh Besar. *Jurnal Master Bahasa*, 6(٢), ١.
- أنوجراه، ر. (٢٠١٤). تقويم بنود الأسئلة في اختبار اللغة العربية الشامل الوطني على مستوى المدرسة الثانوية الإسلامية للسنة الدراسية ٢٠١٤-٢٠١٥ م (في المدرسة الثانوية الإسلامية الحكومية ٣ مالانق). [Skripsi Sarjana] مالانق.