

**Classical Test Theory in Analyzing Arabic Test Questions:
A Descriptive Study on Item Analysis Research in Indonesia/
نظرية الاختبار الكلاسيكية في تحليل الأسئلة العربية: الدراسة الوصفية على
بحوث تحليل بنود الأسئلة في إندونيسيا**

Muhammad Jundi

UIN Maulana Malik Ibrahim Malang
jundijundi10@gmail.com

ABSTRACT

This research aims to analyze the trends in item analysis research with a focus on the background, parameters used, and the utilization of assisting applications. The research method employed is descriptive qualitative, using a literature review approach involving data collection through documentation methods from Arabic language item analysis studies conducted in the last five years. Data analysis was conducted through qualitative analysis, examining each study and making small notes that were subsequently summarized. The research findings indicate that the background of item analysis research in Arabic language in Indonesia includes the absence of item analysis in schools, the presence of items that are either too easy or too difficult for students, and the desire to improve the quality of items. Commonly used parameters in item analysis include difficulty level and discrimination power, while distractor effectiveness is only used in a small number of studies. Furthermore, the research also shows that commonly used applications include Microsoft Excel, SPSS, and Anates. Therefore, future item analysis research is recommended to focus on the use of more advanced analysis methods and parameter measurements, as well as to utilize a wider range of applications.

Keywords: Arabic Subject, Tes Item Analysis, Classical Test Theory

ABSTRAK

Penelitian ini bertujuan untuk menganalisis tren penelitian dalam analisis butir soal dengan fokus pada latar belakang, parameter yang digunakan, dan pemanfaatan aplikasi pembantu. Metode penelitian yang digunakan adalah kualitatif deskriptif dengan studi literatur yang melibatkan pengumpulan data melalui metode dokumentasi dari penelitian-penelitian analisis butir soal bahasa Arab dalam Lima tahun terakhir. Analisis data dilakukan melalui analisis kualitatif dengan memeriksa setiap penelitian dan membuat catatan-catatan kecil yang kemudian disimpulkan. Hasil penelitian menunjukkan bahwa latar belakang penelitian analisis butir soal bahasa Arab di Indonesia meliputi belum adanya analisis butir soal di sekolah-sekolah, keberadaan soal-soal yang terlalu mudah atau terlalu sulit bagi siswa, serta

keinginan untuk meningkatkan kualitas butir soal. Parameter yang umum digunakan dalam analisis butir soal adalah tingkat kesulitan dan daya pembeda, sementara efektivitas pengecoh hanya digunakan dalam sebagian kecil penelitian. Selain itu, penelitian ini juga menunjukkan bahwa aplikasi yang sering digunakan adalah Microsoft Excel, SPSS, dan Anates. Oleh karena itu, penelitian analisis butir soal di masa depan disarankan untuk mengarahkan penggunaan metode analisis dan pengukuran parameter yang lebih mutakhir, serta memanfaatkan aplikasi yang lebih bervariasi.

Kata kunci: Mata Pelajaran Bahasa Arab, Analisis Butir Soal, Teori Tes Klasik

المقدمة

تحليل بنود الأسئلة هو عملية هامة جداً في تقييم جودة الاختبارات والامتحانات. الهدف من هذا التحليل هو تقييم مدى توافق أسئلة الاختبار مع الأهداف التعليمية المراد تحقيقه (Arikunto, 2009). هناك عدة تقنيات يمكن استخدامها في تحليل أسئلة الاختبار، مثل التحليل النوعي والكمي، بالإضافة إلى استخدام برامج الكمبيوتر لمعالجة البيانات (Farida, Syarifah, & Muhamad, 2021). في التحليل النوعي، يتم تقييم كل سؤال بشكل شامل عن طريق تحديد نقاط الضعف والقوة، وبالتالي يمكن الوصول إلى استنتاج حول مدى توافق الأسئلة مع الأهداف التعليمية. بينما في التحليل الكمي، يتم حساب الإحصائيات عن الأسئلة مثل مستوى الصعوبة والقدرة على التمييز ومعامل الاتساق (Pratama, 2020; Surapranata, 2009). وأخيراً، يمكن لاستخدام برامج الكمبيوتر تسريع وتسهيل عملية تحليل أسئلة الاختبار بسبب قدرتها على معالجة البيانات بسرعة ودقة. بإجراء تحليل أسئلة الاختبار الصحيح، يمكن الحصول على اختبارات وامتحانات ذات جودة عالية ومتوافقة مع الأهداف التعليمية المراد تحقيقها، وبالتالي يمكن زيادة فعالية عملية التعلم.

المدرسون وكتاب الاختبارات هم الأطراف المرتبطة بشكل أكبر بتنفيذ عملية التقييم. ويمكن للمدرسين تحسين جودة الاختبارات وزيادة فعالية التعليم باستخدام تحليل أسئلة الاختبار (Arifin, 2012). بينما يمكن لكتاب الاختبارات تقييم بنود الأسئلة المجرأة وتحسين تصميم الاختبار في المستقبل باستخدام تحليل بنود الأسئلة. بشكل عام، يعتبر تحليل بنود الأسئلة أمراً مهماً جداً لمساعدة المدرسين وكتاب الاختبارات على تطوير اختبارات فعالة ودقيقة في قياس فهم الطلاب للمواد التي يتعلمونها.

تحليل بنود الأسئلة هو أمر مهم جدا في عملية إعداد الاختبارات. يمكن اعتبار بنود الأسئلة العنصر الأكثر أهمية في الاختبارات، لأن بنود الأسئلة الجيدة والصحيحة يمكن أن تحدد نجاح أو فشل تقييم قدرات الطلاب. ولذلك، يجب إجراء تحليل دقيق لبنود الأسئلة للتأكد من أن بنود الأسئلة المستخدمة في الاختبار يمكنها قياس قدرات الطلاب بما يتوافق مع أهداف التعلم المحددة (Jundi & Dalle, 2020; Jundi & Solong, 2021; Laubaha dkk., 2022). يمكن أن يساعد تحليل بنود الأسئلة في تحديد نقاط القوة والضعف في كل بنود الأسئلة المستخدمة في الاختبار. يهدف عملية التحليل إلى تحسين جودة بنود الأسئلة وتحسين فعاليتها في تقييم قدرات الطلاب. عند إجراء تحليل بنود الأسئلة، يجب مراعاة بعض الأمور، مثل: مستوى صعوبة البنود، القدرة التمييزية للبنود، ووضوح البنود وملاءمته مع المناهج الدراسية والكفاءات التي يجب تحقيقها.

تم إجراء العديد من الأبحاث حول تقييم التعلم، سواء من حيث تقييم الإدارة وتنفيذ عملية التعلم أو من حيث تقييم نتائج تعليم الطلاب (Hasibuan & Jundi, 2023; Jundi & Ali, 2023). توجد العديد من الدراسات في مجال تقييم نتائج التعليم، وخاصة في موضوع تحليل بنود الأسئلة. وعلى الرغم من أن هناك العديد من الدراسات التي أجريت في مجال تحليل بنود الأسئلة في إندونيسيا، إلا أن المراجع المتاحة لا تزال محدودة وغالبا ما تعاني من التباؤ، حيث تدور المشكلات البحثية وطرق حلها في نفس الدوائر. ولا سيما في سياق تحليل بنود الأسئلة في اللغة العربية، هناك حاجة ملحة لتعزيز البحث في هذا المجال. مع زيادة الاهتمام والمشاركة من الطلاب في تعلم اللغة العربية، من المهم التأكد من أن أدوات التقييم المستخدمة يمكنها توفير نتائج دقيقة وموضوعية.

قامت عدة أبحاث سابقة بدراسة نظرية الاختبار الكلاسيكية لتحليل بنود أسئلة الاختبار. تفصل هذه الأبحاث في تطبيق هذه النظرية لدعم تحليل بنود أسئلة الاختبار (Purwati, Retnawati, 2021). وقد قامت بعض الأبحاث أيضا بتحليل بنود أسئلة الاختبار باستخدام النظرية الحديثة، التي تعتبر تحسينا للنظرية الكلاسيكية (Nurchahyo, 2017; Susanto & Retnawati, 2023). ومع ذلك، لم تظهر حتى الآن أبحاث تقوم بجمع وتحليل اتجاهات الأبحاث في تحليل بنود الاختبار باستخدام نظرية الاختبار الكلاسيكية، كما قام الباحث في هذه الدراسة.

إن الحاجة لتعزيز وتحديث اتجاهات البحث في تحليل بنود الأسئلة في اللغة العربية في إندونيسيا تستند إلى عدة أسباب. أولاً، قد شهد تطوير المناهج وتعليم اللغة العربية في إندونيسيا

تغيراً كبيراً في السنوات الأخيرة. ولذا، يتطلب البحث ذو الصلة والحديث لضمان أن أدوات التقييم المستخدمة تتوافق مع الاحتياجات والمتطلبات الحالية. ثانياً، غالباً ما تواجه الدراسات السابقة في تحليل بنود الأسئلة في اللغة العربية مشكلات مشابهة وتعاني من قيود منهجية محددة. وبالتالي، يتطلب تحليل الدراسات السابقة وضعها كميّار لتحديث اتجاهات البحث، مما يمكن من استخدام أساليب أكثر تطوراً وتجنب الدراسات المكررة، وتطوير نظريات أقوى في سياق بحث تحليل بنود الأسئلة في اللغة العربية في المستقبل. بناءً على الخلفية والأهداف المذكورة أعلاه، ستمحور مناقشة هذا البحث حول ثلاثة جوانب، وهي: تحليل الخلفية للدراسات السابقة، وتحديد المعلمات المستخدمة في التحليل، وتنويع استخدام البرامج في تحليل بنود الأسئلة.

منهجية البحث

يعتمد هذا البحث على نهج وصفي نوعي ونوع دراسة المراجع (Afifuddin & Saebani, 2012; Ainin, 2016; Sugiyono, 2016). وتستخدم طريقة جمع البيانات من خلال الوثائق، حيث يتم جمع الدراسات المتعلقة بتحليل أسئلة اللغة العربية لمدة خمس سنوات الماضية، من عام ٢٠١٨ حتى عام ٢٠٢٣، التي تتكون من ثلاثة عشر دراسة. ولجمع هذه الدراسات، يتم البحث باستخدام الكلمات الرئيسية "تحليل أسئلة اللغة العربية". بعد جمع جميع الدراسات، يقوم الباحث بعملية التصنيف للبحث عن الدراسات التي تركز على تحليل أسئلة الامتحانات المدرسية، وليس لأغراض الاختيار أو اختبار مهارات اللغة وما إلى ذلك. يتم ذلك للتأكد من أن الدراسات المستخدمة في هذا البحث تتناسب مع نطاق وأهداف البحث.

يتم إجراء تحليل البيانات بطريقة وصفية نوعية باستخدام طريقة تحليل المحتوى. يقوم الباحث بقراءة ودراسة كل دراسة تم جمعها بعناية، ثم يقوم بتدوين ملاحظات صغيرة حول الاكتشافات والنتائج ذات الصلة (Zed, 2008). يتم ملخص هذه الملاحظات في سرديات تصف بشمولية الاكتشافات من الدراسات التي تم تحليلها. من خلال استخدام طريقة تحليل المحتوى، يمكن للباحث تحديد الأنماط والاتجاهات في تحليل أسئلة اللغة العربية في إندونيسيا خلال الفترة الزمنية المدروسة. من خلال النهج الوصفي النوعي وطريقة تحليل المحتوى، يهدف هذا البحث إلى توفير فهم شامل لاستخدام نظرية الاختبار الكلاسيكية في تحليل أسئلة اللغة العربية في إندونيسيا. من المتوقع

أن تسهم طريقة جمع البيانات التي تستخدم وثائق الدراسات الحديثة وتحليل البيانات بشكل نوعي في تحقيق اكتشافات عميقة وتقديم إسهامات مهمة في تطوير هذا البحث.

نتائج البحث ومناقشتها

يفهم اختبار اللغة كمجموعة من الأسئلة أو المهام التي يجب على الطلاب الإجابة عليها أو التعامل معها، بهدف قياس مستواهم في اتقان مهارات لغوية محددة، وشرح تقدم أدائهم ومقارنته مع أداء الطلاب الآخرين. قدرة الشخص على استخدام اللغة العربية تتعلق بكيفية استخدامها في التواصل اليومي، مما يمكنه من التواصل مع الآخرين دون الحاجة لمعرفة جميع القواعد النظرية للغة العربية (Hasibuan, Jundi, & Elbaghdadi, 2023; Jundi, Hasibuan, & Kadir, 2023). التواصل باللغة العربية ليس مجرد معرفة كيفية بناء الجمل أو اختيار الكلمات بشكل صحيح، بل يتعلق أيضاً بقدرة الشخص على التحدث وفهم اللغة بشكل واقعي. اختبار اللغة العربية مرتبط بشكل وثيق بالقدرة على التحدث والفهم، وليس فقط بمعرفة اللغة العربية. في علم اللغة، هناك فرق بين القدرة على استخدام اللغة بشكل مجرد وبين القدرة العملية على استخدامها (Jundi & Nabila, 2023; Kadir dkk., 2022). هذه القدرة العملية تمكن الشخص من استخدام اللغة في سياقات حقيقية، سواء كان ذلك شفهاً أو كتابياً. وبالتالي، يشمل اختبار اللغة اختبار القدرة النظرية والقدرة العملية على استخدامها في التواصل (Andriani, t.t.).

إن قدرة الفرد على استخدام اللغة العربية بشكل تقليدي يمكن تقسيمها إلى ثلاثة جوانب رئيسية، وهي الأصوات والمفردات والتركيب وأيضاً أربع مهارات رئيسية (Jundi, 2023; Jundi, Fitriani, & Aquil, 2023; Jundi, Kasan, & Kadir, 2021). هذه المهارات الأربع تشمل مهارة الاستماع، وهي القدرة على فهم اللغة المنطوقة. ومهارة القراءة، التي تتضمن القدرة على فهم اللغة المكتوبة. ومهارة الكلام، وهي القدرة على التعبير عن النفس شفهاً. وأخيراً، مهارة الكتابة، والتي تركز على القدرة على التعبير عن النفس بالكتابة. غالباً ما يتم تصميم اختبارات اللغة العربية لقياس قدرة الفرد على استخدام اللغة العربية بشكل عام (Jundi, Adam, Ali, Kadir, & Laubaha, 2022; Jundi & Hasibuan, 2023). ولذلك، يمكن أن تتكون الاختبارات من عدة أنواع تشمل اختبارات الاستماع واختبارات القراءة واختبارات الكلام واختبارات الكتابة، حسب تفاصيل الهدف من كل اختبار (Hula, 2021).

عند تحليل بنود الأسئلة كميًا، يوجد نهجان يمكن استخدامهما: نظرية الاختبار الكلاسيكية (Classical Test Theory/CTT) ونظرية استجابة البنود (Item Response Theory/IRT). ينظر نهج نظرية الاختبار الكلاسيكية إلى الاختبار كمجموعة من العناصر المنفصلة والمستقلة التي يتم حسابها بشكل منفصل لقياس مستوى الصعوبة وقدرة التمييز والموثوقية (Pratama, 2020; Purwati dkk., 2021). نظرية الاختبار الكلاسيكية (CTT) هي نظرية القياس التي تستخدم مفهوم النقاط الكلية لقياس جودة الاختبار، بينما نظرية استجابة البنود (IRT) هي نظرية القياس التي تركز على خصائص البنود واستجابة الأفراد للبنود الاختبارية. تستخدم نظرية الاختبار الكلاسيكية أساليب إحصائية بسيطة مثل المتوسط والانحراف المعياري، بينما تستخدم نظرية استجابة البنود نموذجًا رياضيًا معقدًا لنمذجة احتمالية الاستجابة في كل مستوى من مستويات القدرة. قد توفر نظرية استجابة البنود معلومات أكثر تفصيلاً حول مستوى القدرة وجودة البنود، وتأخذ في الاعتبار أيضاً فائدة كل بنود في تحليل القياس.

نظرية الاختبار الكلاسيكية مبنية على استخدام نموذج رياضي بسيط يربط بين درجة الملاحظة (X)، والدرجة الحقيقية (T)، ودرجة الخطأ (E) (Mardapi, 2016). مدخل نظرية الاختبار الكلاسيكية في تحليل بنود الأسئلة مستند إلى الأسس الفلسفية للإيجابية المنطقية التي تؤكد على استخدام المنهج العلمي في قياس النفس. تفترض نظرية الاختبار الكلاسيكية أن كل فرد لديه "نقطة حقيقية" تمثل القدرة الفعلية في بناء يتم قياسه. السمة الرئيسية لنظرية الاختبار الكلاسيكية هي التركيز على تحليل الإحصائيات للأسئلة وتحديد جودة الأجزاء بناء على معايير مثل مستوى الصعوبة والقدرة التمييزية. يعتبر هذا النهج أن الأجزاء الجيدة يجب أن تكون لديها قدرة تمييزية عالية، أي أنها قادرة على التفريق بين الأفراد ذوي المستويات المرتفعة والمنخفضة من القدرة.

من بين العديد من الأبحاث التي تم جمعها بنجاح، تم تقليص عددها وفقاً للمعايير المحددة، وهي دراسات تحليل بنود أسئلة الاختبار اللغة العربية لاحتياجات الامتحان. فيما يلي توضيح مفصل لنتائج التحليل التي أجريت على ثلاث عشرة دراسة تم جمعها.

جوانب	بحث
-------	-----

برنامج	معاملة	خلفية	
مايكروسوفت إكسيل وأناتيس	مستوى الصعوبة والقدرة التمييزية وفعالية الخيارات	لم يتم تنفيذ تحليل بنود أسئلة اختبار في المدرسة التي تمت فيها الدراسة، خاصة في مادة اللغة العربية.	١
مايكروسوفت إكسيل وأناتيس و SPSS	مستوى الصعوبة والقدرة التمييزية	تنفيذ إجراءات تحليل بنود الأسئلة للحصول على أسئلة عالية الجودة.	٢
مايكروسوفت إكسيل	تصنيف بنود الأسئلة إلى ستة مستويات من التفكير.	تقييم أداء الطلاب، تقييم قدراتهم، وتحديد الصعوبات في فهم المواد، توفير أساس لتحسين طرق التدريس واستراتيجياتهم.	٣
مايكروسوفت إكسيل و SPSS	مستوى الصعوبة والقدرة التمييزية	للتعرف على نواقص الأسئلة التي تم تطويرها.	٤
مايكروسوفت إكسيل و SPSS	مستوى الصعوبة والقدرة التمييزية	النتائج المنخفضة التي حصل عليها الطلاب في الامتحان النهائي ناتجة عن نقص في تحليل بنود الأسئلة، الذي لم يمكن التعرف به على الصعوبات التي يواجهها الطلاب.	٥
مايكروسوفت إكسيل وأناتيس و SPSS	مستوى الصعوبة والقدرة التمييزية وفعالية الخيارات	تحليل جودة أسئلة امتحان مادة اللغة العربية التي تتطلب مهارات التفكير على مستوى عال.	٦
-	تصنيف بلوم: العاطفي، المعرفي، والنفسحركي	من الضروري إجراء تحليل للأسئلة في كتب تعلم اللغة العربية الصادرة حديثاً.	٧
مايكروسوفت إكسيل	الصدق والثبات	تحديد مدى درجة الصدق بنود الأسئلة لامتحان النهائي.	٨

٩	كثير من الطلاب حصلوا على درجات منخفضة في الاختبار النهائي مادة اللغة العربية ولم يصلوا إلى معيار النجاح الأدنى (KKM). لذا، أثار هذا اهتمام الباحث لتقييم جودة أسئلة الاختبار التي تم إعدادها.	الصدق والثبات ومستوى الصعوبة والقدرة التمييزية	مايكروسوفت إكسيل وأناتيس
١٠	المعلمون غالباً ما يقومون بإعداد أسئلة الاختبار بشكل عجل، مما يؤدي إلى العديد من بنود الأسئلة التي لا تستطيع إنتاج بيانات دقيقة وموضوعية حول أداء الطلاب.	الصدق والثبات والقدرة التمييزية	SPSS و
١١	الأسئلة في الامتحان لديها مستوى صعوبة عالي، حتى أن بعض الأسئلة فيها لا تقيس الجوانب التي يجب أن تقاس.	الصدق ومستوى الصعوبة	مايكروسوفت إكسيل
١٢	تنفيذ إجراءات تحليل بنود الأسئلة للحصول على أسئلة عالية الجودة.	الصدق والثبات ومستوى الصعوبة	وأناتيس
١٣	لتحليل أسئلة التفسير في مادة النحو	الصدق والثبات ومستوى الصعوبة والقدرة التمييزية	مايكروسوفت إكسيل و SPSS

من خلال شرح مفصل لنتائج تحليل الأبحاث التي تم جمعها، تم قطع عددها وفقاً للمعايير المحددة، وهي دراسات تحليل بنود الأسئلة في اللغة العربية لأغراض الامتحان. يتم مناقشة النتائج بعد ذلك باستخدام النظريات المتاحة. سيتم التركيز في النقاش أدناه على ثلاثة جوانب، وهي: خلفية البحوث، والمعاملات المستخدمة، تقدير الصدق والثبات، واستخدام برامج مساعدة في تحليل الاختبار.

١. خلفية البحوث

في شرح الواقع التجريبي المتعلق بالمشاكل التي تعد خلفية لتحليل بنود الأسئلة للغة العربية، قام بتحليل الدراسات التي أُجريت خلال السنوات الخمس الماضية. قام الباحث بجمع الدراسات المتعلقة بتحليل بنود الأسئلة في امتحان اللغة العربية باستخدام محرك بحث Google Scholar. تم التركيز في البحث على عناوين الدراسات التي أُجريت في المدارس واستخدمت كأداة لامتحان الطلاب، وليس كأداة لاختبار الاختيار أو اختبار مهارات اللغة الأجنبية مثل اختبار TOAFL. باستخدام هذه الطريقة، وجد الباحث أربعة عشر دراسة ثم قام بتحليل الخلفية لكل دراسة. بناء على التحليل الذي تم إجراؤه، يمكن تصنيف المشاكل التي تشكل خلفية لتحليل بنود الأسئلة لاختبار اللغة العربية في إندونيسيا في الفئات التالية.

الأسئلة صعبة جداً أو سهلة جداً - في البحث الثاني، ذُكر أن نتائج اختبار الطلاب كانت عالية، مما يشير إلى احتمالية أن الأسئلة المطروحة في الاختبار كانت سهلة جداً بالنسبة للطلاب. وبناء على ذلك، استنتج الباحث أنه يجب إجراء تحليل لبنود الأسئلة. بالمقابل، في الأبحاث الرابعة والثالثة عشرة والرابعة عشرة، تبين أن هناك العديد من الطلاب الذين ليسوا كفيين في الاختبار، مما يشير إلى احتمالية أن الأسئلة التي طُرحت كانت صعبة جداً بالنسبة لهم. وبالإضافة إلى ذلك، كشفت البحث الثالثة عشرة أيضاً أن هناك بعض البنود التي لا تقيس الكفايات التي يجب قياسها في الاختبار (Nurfauziah, 2018; مطرفي, 2019).

لم يتم تنفيذ تحليل بنود الأسئلة - في البحث الأول والبحث السادس، يشرح أنه لم يتم إجراء تحليل لبنود الأسئلة لامتحان اللغة العربية. يقوم فريق إعداد الأسئلة فقط بإعداد أسئلة الامتحان دون إجراء تحليل لبنود تلك الأسئلة (Ahmad, 2022; Farida dkk., 2021). في حين ذلك، في البحث السابع، يذكر أن المعلمين يشعرون بالشكوك حول جودة الأسئلة، ولكن لم يتم تنفيذ تحليل بنود الأسئلة في المؤسسة (Sholihah, 2021). تحليل بنود السؤال هو أحد عمليات تطوير الأسئلة التي غالباً ما تتم نسيانها وعدم تنفيذها، على الرغم من أن هذه العملية مهمة جداً لضمان أن الأسئلة التي تم إعدادها تحمل جودة عالية.

الحاجة إلى تحسين جودة الأسئلة - في الأبحاث الثالث والثامن والثاني عشر، يتعين إجراء تحليل لبنود الأسئلة لتحديد الأسئلة التي تتمتع بجودة جيدة. وبالتالي، يمكن استخدام الأسئلة عالية الجودة في الامتحان. بالإضافة إلى ذلك، إذا تم إجراء تحليل على الأسئلة التي

تم استخدامها بالفعل، فمن المأمول أن تكون نتائج التحليل هذه دليلاً لإعداد الأسئلة في المستقبل (Arumningtyas, 2021; Dian, 2022; Mubarrirroh, 2020). وفي البحث الخامس، يشرح أن الأسئلة تم إعدادها بعجالة، وبالتالي يتعين إجراء تحليل لتحسين جودة الأسئلة. تحديد قدرات التفكير في الأسئلة- يشرح البحث التاسع أن الأسئلة تم إعدادها عشوائياً دون مراعاة مستوى التفكير المدرج في دليل الأسئلة وأهداف التعلم (Maspar, 2021). في حين أن البحوث العاشرة والحادية عشرة تهدف إلى تحديد قدرات الطلاب ومستوى التفكير الذي يتضمنه كل سؤال (Hilda Yuni Mirandi, 2019; Indah Rahmi Nurfauziah, 2019). معرفة مستوى التفكير في بنود الأسئلة هي أمر مهم للغاية.

تماماً كما هو الحال في تخطيط التعليم في مرحلة تحديد أهداف التعليم (Adam, Jundi, Ali, Laubaha, & Kadir, 2023). يجب على المعلم أن يفهم الكفايات التي يفرضها المنهج التعليمي لكي يتمكن من صياغة أهداف التعليم التي يرغب في تحقيقها. ينبغي أن يتم تحديد أهداف التعليم بعناية، لأن هذا يتعلق ارتباطاً وثيقاً بعملية ونتائج التعليم. الأهداف التي ترغب في تحقيقها ستحدد المواد المطلوبة، ومصادر التعليم المستخدمة، والأساليب والاستراتيجيات المعتمدة، وكذلك التقويم الذي سيتم تقديمه. لذلك، في صياغة أهداف التعليم، من المهم للغاية مراعاة مستوى التفكير الذي يفرضه المنهج.

٢. المعاملات

في تحليل أجزاء الأسئلة باستخدام نظرية الاختبار الكلاسيكية، يعتبر مستوى الصعوبة أحد المعايير المستخدمة لتقييم جودة أجزاء الأسئلة. يقيس مستوى الصعوبة مدى صعوبة أو سهولة جزء السؤال على الطلاب المشاركين في الاختبار. يعتمد قياس مستوى الصعوبة على نسبة الطلاب الذين يجيبون بشكل صحيح. إذا كان هناك الكثير من الطلاب الذين يمكنهم الإجابة بشكل صحيح، فإن جزء السؤال يصنف كسهل. وعلى العكس، إذا كان هناك عدد قليل من الطلاب الذين يمكنهم الإجابة بشكل صحيح، فإن جزء السؤال يعتبر صعباً. يعد قياس مستوى الصعوبة أمراً هاماً لضمان أن نطاق صعوبة أجزاء الأسئلة يتناسب مع احتياجات الاختبار ويميز بين الطلاب ذوي مستويات قدرات مختلفة. ترمز مستوى صعوبة الأسئلة بـ " p_i ". إذا كانت قيمة " p_i " تقترب من الصفر، فإن السؤال يعتبر صعباً جداً. بينما

إذا كانت قيمة " p_i " تقترب من الواحد، فإن السؤال يعتبر سهلاً جداً، وبالتالي يحتاج إلى إلغاء أو مراجعة. (Retnawati, 2016)

مستوى صعوبة بنود الأسئلة، والتي يتم التعبير عنها بنسبة تتراوح بين الصفر والواحد، ويكون القيمة متمركزة حول $p_i = 0.50$ ، يعتبر لديها معايير متوسطة تفصل بين البنود السهلة والبنود الصعبة. يعتبر البنود سهلة إذا كانت قيمة $p_i > 0.50$ ، ويعتبر البنود صعبة إذا كانت قيمة $p_i < 0.50$. (Naga, 1992). وهناك وجهات نظر أخرى تشير إلى أن مؤشر صعوبة الفقرة الجيد يتراوح عادة بين 0,3 و 0,7. في هذا النطاق، يتم الحصول على معلومات عن قدرات الطلاب بشكل أقصى. إذا كانت قيمة المؤشر $p_i < 0.30$ ، يعتبر بنود الأسئلة صعبة، إذا كانت تتراوح قيمة p_i بين 0,30 و 0,70، يعتبر مستوى صعوبة بنود الأسئلة متوسطاً، وإذا كانت قيمة $p_i > 0.70$ ، يعتبر بنود الأسئلة سهلة. (Allen, 1979)

$$p_i = \frac{\sum B}{N} \dots\dots\dots$$

p_i	:	مستوى الصعوبة
$\sum B$:	عدد المشاركين في الاختبار الذين أجابوا على الأسئلة بشكل صحيح
N	:	عدد المشاركين في الاختبار

في تحليل بنود الأسئلة باستخدام نظرية الاختبار الكلاسيكية، يعتبر القدرة التمييزية معياراً يستخدم لقياس قدرة جزء السؤال على تمييز بين الطلاب في الاختبار الذين لديهم مستوى قدرات عالي والطلاب الذين لديهم مستوى قدرات منخفض. يعتمد قياس القدرة التمييزية على اختلاف نسبة الطلاب الذين يجيبون بشكل صحيح بين مجموعة الطلاب ذوي المستوى العالي ومجموعة الطلاب ذوي المستوى المنخفض. إذا كان لدى جزء السؤال قدرة تمييزية عالية، فهذا يعني أنه يمكنه تمييز الطلاب بوضوح بين المستويات المختلفة للقدرة.

حساب القدرة التمييزية هو قياس مدى قدرة بنود الأسئلة على تمييز الطلاب الذين يتقنون الكفايات عن الطلاب الذين لا يتقنون الكفايات أو يتقنونها بشكل غير كاف، بناءً على معايير محددة. كلما ارتفع معامل القدرة التمييزية لبنود الأسئلة، زادت قدرتها على التمييز

بين الطلاب الذين يتقنون الكفايات والطلاب الذين يتقنونها بشكل غير كاف (Arifin, 2012). يعتبر مؤشر قدرة التمييز جيدا إذا كان لديه قيمة تساوي أو تفوق ٣,٠٠. مؤشر قدرة التمييز لبنود الأسئلة ذات قيمة صغيرة يؤدي إلى عدم قدرتها على تمييز الطلاب ذوي القدرات العالية عن الطلاب ذوي القدرات المنخفضة (Retnawati, 2016).

$$DB = \frac{\sum TB}{\sum T} -$$

$$\frac{\sum RB}{\sum R}$$

$\sum TB$:	عدد المشاركين الذين أجابوا بشكل صحيح في المجموعة ذوي الكفاءة العالية
$\sum RB$:	عدد المشاركين الذين أجابوا بشكل صحيح في المجموعة ذوي الكفاءة المنخفضة
$\sum T$:	عدد المشاركين ذوي الكفاءة العالية
$\sum R$:	عدد المشاركين ذوي الكفاءة المنخفضة

فيمكن تفسير النتائج من خلال معلمتين هامتين: مستوى الصعوبة والقدرة التمييزية. يقيس مستوى الصعوبة مدى صعوبة أو سهولة بنود الأسئلة بالنسبة للمشاركين في الاختبار، بينما يقيس القدرة التمييزية قدرة بنود الأسئلة على التمييز بين المشاركين ذوي مستويات قدرات مختلفة. يتم تحديد مستوى الصعوبة استنادا إلى نسبة المشاركين الذين أجابوا بشكل صحيح، في حين يتم تحديد القدرة التمييزية عن طريق مقارنة نسبة المشاركين ذوي القدرات العالية والمنخفضة. من المهم إيجاد توازن مناسب بين مستوى الصعوبة والقدرة التمييزية لضمان أن تقدم بنود الأسئلة معلومات ذات صلة وتستطيع التمييز بين المشاركين ذوي مستويات قدرات مختلفة.

في استخدام معايير تحليل أسئلة الاختبار، تقوم جميع الدراسات التي تم جمعها في هذا البحث بقياس مستوى الصعوبة كإحدى المعايير الرئيسية. يوفر قياس مستوى الصعوبة معلومات حول مدى صعوبة أو سهولة سؤال معين بالنسبة للمستجيبين للإجابة عليه بشكل صحيح. بالإضافة إلى ذلك، تستخدم ٨٣,٣٣٪ من الدراسات معلمة التمييز. توفر هذه المعلمة معلومات عن مدى قدرة سؤال معين على التمييز بين المستجيبين الذين لديهم مستوى

فهم عال والمستجيبين الذين لديهم مستوى فهم منخفض. هذا أمر مهم في قياس جودة أسئلة الاختبار، حيث يعتبر سؤالاً يمتلك قدرة تمييز عالية أفضل في التفريق بين قدرات المستجيبين. بالإضافة إلى معلمة مستوى الصعوبة والتمييز، هناك جزء صغير من الدراسات التي تستخدم المعلمة الثالثة، وهي فعالية الخيارات. في نظرية الاختبار الكلاسيكية، تشير فعالية الخيارات إلى قدرة الخيارات الخاطئة على جذب المستجيبين الذين ليس لديهم فهم جيد وتوجيههم لاختيار إجابة خاطئة. يمكن أن توفر هذه المعلمة معلومات حول مدى قدرة الخيارات الخاطئة على إرباك المستجيبين وتقليل صدق أسئلة الاختبار.

يجب أن نولي اهتماماً خاصاً لضمان أن مستوى صعوبة والقدرة التمييزية في أسئلة الامتحان يكونان في حالة جيدة (Fahmi, Rizqi, & Harneilinda, 2022; Retnawati, 2012; Sumaryanta, 2021). يعني ذلك أن مستوى الصعوبة يجب أن يكون متناسباً، حيث يكون عدد الأسئلة الصعبة متوازناً مع عدد الأسئلة السهلة، ويجب أن يكون هناك مزيد من الأسئلة المتوسطة (Fikriyah, 2021). نسبتها هي ٢٧٪ للأسئلة السهلة، و ٢٧٪ للأسئلة الصعبة، و ٤٦٪ للأسئلة المتوسطة (Arikunto, 2009; Purwati dkk., 2021). أما بالنسبة للقدرة التمييزية، يجب أن تكون لدى جميع الأسئلة القدرة التمييزية جيدة، حيث يعد ذلك مفتاحاً لقياس مدى قدرة الطلاب على فهم مواد التعلم.

٣. تقدير الصدق والثبات

الصدق هو مقياس مدى قدرة الاختبار أو أداة القياس على قياس ما يجب أن يتم قياسه. هناك عدد طرق لقياس درجة الصدق للاختبار منها الصدق المحتوى وصدق التوقع وصدق البناء وصدق القياس وغير ذلك. يعد قياس درجة الصدق أمراً هاماً بالنسبة للاختبار بسبب أن الصدق يشير إلى مدى قدرة الاختبار على توفير معلومات دقيقة وملائمة حول الخصائص المرغوبة. في هذا السياق، يشير الصدق المذكور إلى الصدق الإمبريقي، الذي يستخدم تقنيات إحصائية لاختبار العلاقة بين الاختبار والبناء الذي يتم قياسه. إحدى التقنيات الإحصائية الشائعة المستخدمة في الصدق الإمبريقي هي تحليل الارتباط، حيث يستخدم الصيغة الخاصة بارتباط بيرسون لحساب الصدق. من خلال استخدام هذه التقنية،

يمكن قياس مدى ارتباط الاختبار بالبناء المقاس وضمان أن الاختبار يقدم نتائج قابلة للإعتماد وذات مغزى. (Arifin, 2012).

أما الثبات هو مفهوم يشير إلى مدى موثوقية أو دقة قياس أو اختبار في تقديم نتائج متسقة وثابتة (Azwar, 2014). قياس ثبات فحوصات بنود الاختبار ضروري لأنه يقدم معلومات حول مدى يمكن الاعتماد على بنود الاختبار في قياس البنية المرغوبة. بعبارة أخرى، يشير ثبات بنود الاختبار إلى مدى قدرة الاختبار على تقديم نتائج متسقة وقابلة للثقة. في نظرية الاختبار التقليدية، يُحتسب الثبات على مستوى الأداة ويفترض أن يكون خطأ القياس المعياري واحداً لجميع المشتركين في الاختبار. يمكن حساب ثبات بنود الاختبار عن طريق حساب معامل الاتساق الداخلي. واحدة من الصيغ الأكثر شيوعاً لحساب معامل الاتساق الداخلي لاختبار هي صيغة ألفا كرونباخ (Retnawati, 2017). هذا المعامل يوفر إشارة حول مدى ترابط بنود الاختبار ويعكس مستوى الاتساق الداخلي للفحص.

من المهم أن نلاحظ أنه قبل إجراء تحليل لهذه المعلمات، تعطى أولوية دائماً لحساب صدق الأسئلة بأكملها. صدق الأسئلة هي مقياس مدى قدرة الأسئلة على قياس ما هو مقصود بالفعل. يتم ذلك باستخدام مجموعة متنوعة من الأساليب والتقنيات، مثل صدق المحتوى، صدق البناء، وصدق المعايير، لضمان أن الأسئلة المستخدمة في هذا البحث لديها جودة عالية ويمكن الاعتماد عليها في قياس قدرات المستجيبين في اللغة العربية. استناداً إلى الأبحاث التي تم جمعها، الأسلوب الشائع المستخدم لحساب الصحة هو صيغة ارتباط بيرسون، بينما يتم استخدام ألفا كرونباخ لقياس الموثوقية. باستخدام هذه المعلمات ومع إعطاء الأولوية لصدق الأسئلة، من المتوقع أن توفر هذه الدراسات فهماً أكثر شمولاً لتحليل أسئلة اللغة العربية في إندونيسيا، وتحسين جودة أدوات التقييم في هذا السياق.

٤. الاستفادة من البرامج

في استخدام البرامج، تستخدم الدراسات المجموعة عدة برامج مختلفة لدعم تحليل بنود الأسئلة. أحد البرامج الشائعة هو برنامج مايكروسوفت إكسيل (Sumardi, 2002). مايكروسوفت إكسيل هو برنامج جداول بيانات يحتوي على مجموعة متنوعة من الميزات لإدارة وتحليل البيانات. في سياق هذا البحث، يستخدم مايكروسوفت إكسيل لتجميع

البيانات وحساب المعايير والمعلومات المتعلقة بتحليل بنود الأسئلة. يوفر هذا التطبيق سهولة للباحث في ترتيب البيانات في شكل جداول وحساب الإحصاءات الأساسية اللازمة للتحليل.

بالإضافة إلى ذلك، يتم استخدام تطبيق آخر بشكل شائع وهو الحزمة الإحصائية للعلوم الاجتماعية (SPSS) (Janna & Herianto, 2023). إنه هو تطبيق مصمم خصيصاً لإجراء التحليلات الإحصائية. يتم استخدام SPSS بشكل مكمل مع مايكروسوفت إكسيل. بعد معالجة وتجميع البيانات باستخدام مايكروسوفت إكسيل، يمكن للباحث إجراء تحليلات إحصائية أكثر تعقيداً باستخدام SPSS. يوفر هذا التطبيق مجموعة متنوعة من الطرق الإحصائية، مثل اختبار t وتحليل التباين والانحدار وتحليل العوامل، التي يمكن استخدامها لتحليل بيانات البحث بشكل أعمق. عموماً، يتم استخدام هذا التطبيق لاختبار الصدق والثبات في تحليل بنود الأسئلة (Anggraini, Aprianti, Setyawati, & Hartanto, 2022).

علاوة على ذلك، يوجد أيضاً تطبيق يسمى أناتيس (Anates) والذي يستخدم على نطاق واسع لتحليل بنود الأسئلة. أناتيس هو تطبيق مصمم خصيصاً لمساعدة تحليل بنود الأسئلة بسهولة. من خلال أناتيس، يمكن للباحث تحديد المعلومات المرغوبة في تحليل بنود الأسئلة بسرعة، مثل مستوى الصعوبة والقدرة التمييزية وفعالية الخيارات الخاطئة. يوفر هذا التطبيق أدوات مفيدة لتصور وتحليل بيانات بنود الأسئلة، بحيث يمكن للباحث الحصول على فهم أعمق لخصائص وجودة بنود الأسئلة المستخدمة في البحث. من خلال استخدام هذه البرامج المختلفة، يمكن للباحث استغلال الميزات والوظائف المقدمة لتحليل بيانات بنود الأسئلة بشكل أكثر فعالية وكفاءة. يوفر استخدام مايكروسوفت إكسيل و SPSS و أناتيس نهجاً شاملاً لمعالجة وتحليل بيانات البحث، بحيث يمكن أن تكون نتائج البحث حول تحليل بنود الأسئلة باللغة العربية في إندونيسيا دقيقة وموثوقة أكثر.

الخلاصة

الدراسة تكشف عن تنوع في مستوى صعوبة أسئلة اللغة العربية المستخدمة في الامتحانات في إندونيسيا. تشير بعض الدراسات إلى وجود اتجاه لوجود أسئلة سهلة جداً أو صعبة جداً، بينما

تركز الدراسات الأخرى على ضرورة تحسين جودة الأسئلة. يلزم إجراء تحليل لبنود الأسئلة لضمان تناسب مستوى صعوبة الأسئلة مع قدرات المشاركين في الامتحان. بالإضافة إلى ذلك، من المهم مراعاة معايير التقييم وأهداف التعلم الموضحة في مواصفات الأسئلة. يمكن لتحليل بنود الأسئلة الجيد أن يوفر معلومات قيمة حول قدرات الطلاب ومستوى التفكير المدرج في كل سؤال .

الدراسة تشير إلى أنه في بعض الدراسات التي أجريت في إندونيسيا، لم يتم تنفيذ تحليل بنود الأسئلة لامتحان اللغة العربية. ينتج عن ذلك عدم اليقين بشأن جودة الأسئلة التي تم إعدادها ومدى توافقها مع معايير التقييم وأهداف التعلم. يمكن أن تكون نتائج تحليل بنود الأسئلة مرجعا لإعداد أسئلة الامتحانات بشكل أفضل في المستقبل. لذلك، من المهم تطبيق تحليل بنود الأسئلة كخطوة هامة لضمان جودة وصحة امتحان اللغة العربية في إندونيسيا. بذلك، يمكن اتخاذ إجراءات تحسين لضمان وجود أسئلة ذات جودة وملائمة لاحتياجات التعلم. بالنظر إلى اتجاهات البحث خلال السنوات الخمس الماضية التي تم توضيحها أعلاه ، فإنه من الضروري تنفيذ أساليب تحليل الأسئلة الكمية المتقدمة أكثر ، واستخدام عينة كافية ، وتطبيق أدوات تحليل متنوعة للأسئلة. وبالتالي ، يمكن للبحث في مجال تحليل الأسئلة أن ينتج نتائج أكثر صلة وموثوقية وقابلية للتطبيق في المستقبل.

المراجع

- Adam, M. Z., Jundi, M., Ali, I., Laubaha, S. A., & Kadir, S. D. (2023). Studi Komparatif Kompetensi Dasar KMA 183 dan KMA 164 pada Mata Pelajaran Bahasa Arab Kelas VII MTs pada Kompetensi Ranah Kognitif. *Assuthur: Jurnal Pendidikan Bahasa Arab*, 2(1).
- Afifuddin, & Saebani, B. A. (2012). *Metodologi Penelitian Kualitatif*. Bandung: Pustaka Setia.
- Ahmad, F. (2022). تحليل جودة بنود أسئلة اللغة العربية لاختبار اخر الفصل الدراسي لطلاب الصف العاشر في الثانوية الإسلامية سومبانج بانويوماس العام الدراسي: ((Sains Al-Qur`an)) مدرسة معارف نهضة العلماء ٢٠٢١/٢٠٢٠ (Skripsi, IAIN Purwokerto). IAIN Purwokerto. Diambil dari <https://repository.uinsaizu.ac.id/14037/>
- Ainin, M. (2016). *Metodologi Penelitian Bahasa Arab*. Malang: Bintang Sejahtera.
- Allen, M. J. (1979). *Introduction to measurement theory*. Monterey, Calif: Brooks/Cole Pub. Co.
- Andriani, R. S. (t.t.). LANGUAGE TESTS. *Jurnal Educate*, 3(2). Diambil dari <https://media.neliti.com/media/publications/241552-tes-kebahasan-9bda9d9f.pdf>
- Anggraini, F. D. P., Aprianti, A., Setyawati, V. A. V., & Hartanto, A. A. (2022). Pembelajaran Statistika Menggunakan Software SPSS untuk Uji Validitas dan Reliabilitas. *Jurnal Basicedu*, 6(4), 6491–6504. <https://doi.org/10.31004/basicedu.v6i4.3206>
- Arifin, Z. (2012). *Evaluasi Pembelajaran*. Jakarta: Direktorat Jenderal Pendidikan Islam Kementerian Agama RI.
- Arikunto, S. (2009). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Arumningtyas, A. (2021). تحليل بنود أسئلة الإمتحان النهائي لمادة اللغة العربية في الفصل الثامن في المدرسة (Diploma, IAIN Ponorogo). IAIN Ponorogo. Diambil dari <http://etheses.iainponorogo.ac.id/16334/>
- Azwar, S. (2014). *Reliabilitas dan Validitas* (4 ed.). Yogyakarta: Pustaka Pelajar.
- Dian, V. I. (2022). مستوى صعوبة الختبار النهائي لمادة اللغة العربية بمدرسة دار الصلاح الثانوية الإسلامية جمبر ٢٠٢٣/٢٠٢٢ (Undergraduate, UIN KH Achmad Shiddiq Jember). UIN KH Achmad Shiddiq Jember. Diambil dari <http://digilib.uinkhas.ac.id/16073/>
- Fahmi, B., Rizqi, S., & Harmeilinda, N. E. (2022). Analisis Butir Soal Bahasa Arab MAS Pondok Pesantren Assalam Kampar Riau. *Ta'lim al-'Arabiyyah: Jurnal Pendidikan Bahasa Arab & Kebahasaan*, 6(1), 95–105. <https://doi.org/10.15575/jpba.v6i1.16193>

- Farida, Syarifah, & Muhamad. (2021). تحليل بنود الأسئلة في الإمتحان النهائي لمادة اللغة العربية في الفصل العاشر بالمدرسة العالية “الإسلام” كموجا. *AL-MU'ARRIB: JOURNAL OF ARABIC EDUCATION*, 1(2), 1–15. <https://doi.org/10.32923/al-muarrib.v1i2.2064>
- 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), 128–140. <https://doi.org/10.18196/mht.v3i2.10501>
- Hasibuan, R., & Jundi, M. (2023). Crafting Success: Creative Management Techniques for Darul Ulum Sipaho Islamic Boarding School's Arabic Language Learning Program Implementation. *Kitaba*, 1(3), 135–146. <https://doi.org/10.18860/kitaba.v1i3.24218>
- Hasibuan, R., Jundi, M., & Elbaghdadi, Z. A. (2023). Al Samātu Al Lughowiyyatu Lil Arobiyyati Fil Ma'had Al Islāmi Dirōsatu Hālati Fi Ma'hadi Al 'Ashri Bil Indōnīsiyyā. *Lughawiyyat: Jurnal Pendidikan Bahasa Dan Sastra Arab*, 6(2), 127–144. <https://doi.org/10.38073/lughawiyyat.v6i2.1088>
- Hilda Yuni Mirandi, 140202172. (2019). تحليل اختبار المدرس النهائي بالمدرسة الثانوية الإسلامية الحكومية (دراسة تحليلية في بنود الأسئلة) *Meulaboh* (Skripsi, UIN Ar-Raniry Banda Aceh). UIN Ar-Raniry Banda Aceh. Diambil dari <http://library.ar-raniry.ac.id/>
- Hula, I. R. (2021). EVALUASI DAN TES BAHASA ARAB: Tinjauan Teori. *Language Development Center*, 12(7). <https://doi.org/10.31219/osf.io/uwt8x>
- Indah Rahmi Nurfauziah, -. (2019). ANALISIS KUALITAS TES BAHASA ARAB BERBASIS HOTS (HIGHER ORDER THINKING SKILL) & LOTS (LOWER ORDER THINKING SKILL) (Studi Analisis Di Mts Al-Musyawahar Lembang) (Other, Universitas Pendidikan Indonesia). Universitas Pendidikan Indonesia. Diambil dari <http://repository.upi.edu>
- Janna, N. M., & Herianto, H. (2023). Konsep Uji Validitas Dan Reliabilitas Dengan Menggunakan SPSS. <https://doi.org/10.31219/osf.io/v9j52>
- Jundi, M. (2023). تصميم برنامج أندرويد لتعزيز إتقان المفردات لدى طلاب المدرسة العالية. *Lahjah Arabiyah: Jurnal Bahasa Arab Dan Pendidikan Bahasa Arab*, 4(2), 140–158. <https://doi.org/10.35316/lahjah.v4i2.140-158>
- Jundi, M., Adam, M. Z., Ali, I., Kadir, S. D., & Laubaha, S. A. (2022). Himpunan Mahasiswa Jurusan dan Matrikulasi Bahasa Arab bagi Mahasiswa Baru. *Jurnal Naskhi: Jurnal Kajian Pendidikan dan Bahasa Arab*, 4(1).
- Jundi, M., & Ali, M. (2023). Assessing The Classroom Learning in Arabic Matriculation Program: Tutor and Student Perspective. *Jurnal Al-Maqayis*, 10(2), 21–36.
- Jundi, M., & Dalle, M. (2020). ASPEK AFEKTIF DALAM PEMBELAJARAN DARING BAHASA ARAB BERDASARKAN KMA 183 TAHUN 2019. *Muhadasah: Jurnal Pendidikan Bahasa Arab*, 2(2), 204–226.

- Jundi, M., Fitriani, L., & Aquil, A. (2023). COLLABORATIVE LEARNING: BOOSTING QAWAID MASTERY WITH STAD MODEL. *Thariqah Ilmiah: Jurnal Ilmu-Ilmu Kependidikan & Bahasa Arab*, 11(1), 35–48. <https://doi.org/10.24952/thariqahilmiah.v11i1.7639>
- Jundi, M., & Hasibuan, R. (2023). Enhancing Arabic Language Proficiency among Students: A Case Study of Language Matriculation Strategies at Al-Hasyimiyah Darul Ulum Sipaho Islamic Boarding School. *Kilmatuna: Journal Of Arabic Education*, 3(2), 114–129. <https://doi.org/10.55352/pba.v3i2.619>
- Jundi, M., Hasibuan, R., & Kadir, S. D. (2023). Probing the Fusion of Arabic and Gorontalo Dialect in the Daily Linguistic Landscape of Students at Al-Falah Islamic Boarding School. *Kitabina: Jurnal Bahasa & Sastra Arab*, 4(02), 34–46. <https://doi.org/10.2345/kitabina.v4i02.20255>
- Jundi, M., Kasan, Y., & Kadir, S. D. (2021). Maf'ul Mutlaq Bahasa Arab dan Cognate Object Bahasa Inggris: Analisis Kontrastif dan Sintaksis pada Juz 'Amma. *Lisanan Arabiya: Jurnal Pendidikan Bahasa Arab*, 5(2), 153–178. <https://doi.org/10.32699/liar.v5i2.2190>
- Jundi, M., & Nabila, N. (2023). Transformational-Generative Theory Perspective in Teaching The Four Language Skills of Arabic. *Al-Fusha : Arabic Language Education Journal*, 5(2), 72–81. <https://doi.org/10.36835/alfusha.v5i2.1193>
- Jundi, M., & Solong, N. P. (2021). Analisis Kesesuaian Indikator dan Kompetensi Dasar Bahasa Arab KMA 183. *Tatsqifiy: Jurnal Pendidikan Bahasa Arab*, 2(1), 61–70. <https://doi.org/10.30997/tjpb.v2i1.3642>
- Kadir, S. D., Jundi, M., Laubaha, S. A., Ali, I., Adam, M. Z., & Kasan, Y. (2022). Korelasi Bahasa dan Pikiran dalam Perspektif Al-Qur'an. *Fikroh: Jurnal Pemikiran Dan Pendidikan Islam*, 15(1), 1–17. <https://doi.org/10.37812/fikroh.v15i1.294>
- Laubaha, S. A. L., Bahri, R. B. H., Kadir, S. D., Adam, muhammad Z., Ali, I., & Jundi, M. (2022). Kesesuaian Materi Berdasarkan Teori dan Praktek pada Bahan Ajar Bahasa Arab. *Naskhi: Jurnal Kajian Pendidikan Dan Bahasa Arab*, 4(2), 10–21. <https://doi.org/10.47435/naskhi.v4i2.1198>
- Mardapi, D. (2016). *Pengukuran Penilaian dan Evaluasi Pendidikan* (2 ed.). Yogyakarta: Nuha Litera.
- Maspar, F. G. (2021). تحليل بنود الأسئلة في الكتاب المدرسي لوزارة الشؤون الدينية للمدرسة الثانوية الفصل الحادي عشر على ضوء تصنيف بلوم (Masters, Universitas Islam Negeri Maulana Malik Ibrahim). Universitas Islam Negeri Maulana Malik Ibrahim. Diambil dari <http://etheses.uin-malang.ac.id/31182/>
- Mubarrirroh, A. (2020). تحليل بنود الأسئلة للإمتحان النهائي على درس المطالعة بمعهد المودة 2019-2020 . Dirosatuna: *Journal of Islamic Studies*, 3(1), 53–63. <https://doi.org/10.31538/drstn.v3i1.777>

- Naga, D. S. (1992). *Pengantar Teori Sekor pada Pengukuran Pendidikan*. Jakarta: Gunadarma.
- Nurchahyo, F. (2017). Aplikasi IRT dalam Analisis Aitem Tes Kognitif. *Buletin Psikologi*, 24, 64. <https://doi.org/10.22146/buletinpsikologi.25218>
- Nurfauziah, A. M. (2018). تحليل بنود الاختبار النهائي لمادة اللغة العربية في الفصل الخامس بالمدارس الابتدائية في بانوانجي جاوى الشرقية (Undergraduate, Universitas Islam Negeri Maulana Malik Ibrahim). Universitas Islam Negeri Maulana Malik Ibrahim. Diambil dari <http://etheses.uin-malang.ac.id/11899/>
- Pratama, D. (2020). Analisis Kualitas Tes Buatan Guru Melalui Pendekatan Item Response Theory (IRT) Model Rasch. *Tarbawy : Jurnal Pendidikan Islam*, 7(1), 61–70. <https://doi.org/10.32923/tarbawy.v7i1.1187>
- Purwati, H., Retnawati, H., Jailani, & Retnowati, T. H. (2021). Analisis Karakteristik Butir Soal Ujian Nasional Matematika SMP/MTs Berdasarkan Pendekatan Teori Tes Klasik. *Jurnal Sains dan Edukasi Sains*, 4(2). <https://doi.org/10.24246/juses.v4i2p46-51>
- Retnawati, H. (2012). *Analisis Butir Soal dengan Pendekatan Teori Tes Klasik dan Teori Respons Butir*. Training dipresentasikan pada Pelatihan Analisis Butir dan Pemanfaatan Hasil Ujian Bagi Guru di SMKN 2 Tarakan Kalimantan Timur, Tarakan, Kalimantan Timur.
- Retnawati, H. (2016). *Validitas, Reliabilitas, dan Karakteristik Butir*. Yogyakarta: Parama Publishing.
- Retnawati, H. (2017). *Reliabilitas Instrumen Penelitian*. Lokakarya dipresentasikan pada Workshop Pembelajaran Berbasis Riset dan Analisis Data, Bukittinggi. Bukittinggi.
- Sholihah, M. (2021). تحليل بنود أسئلة في الامتحان النهائي لمادة اللغة العربية للصف العاشر بالمدرسة الثانوية ٢٠٢٠/٢٠٢١ الإسلامية الحكومية الأولى فونوروجو العام الدراسي (Diploma, IAIN Ponorogo). IAIN Ponorogo. Diambil dari <http://etheses.iainponorogo.ac.id/16389/>
- Sugiyono, S. (2016). *Metode Penelitian Pendidikan; Pendekatan Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Sumardi, Y. (2002). PENGGUNAAN MICROSOFT EXCEL DALAM ANALISIS DATA EKSPERIMEN PADA PEMBELAJARAN FISIKA. *Cakrawala Pendidikan: Jurnal Ilmiah Pendidikan*, 2.
- Sumaryanta. (2021). *Teori Tes Klasik & Teori Respon Butir: Konsep & Contoh Penerapannya*. Cirebon: Confident. Diambil dari https://pep.pps.uny.ac.id/sites/pep.pps.uny.ac.id/files/Buku%20Tes%20Teori%20Klsik%20_%20Teori%20Respon%20Butir%20ok.pdf
- Surapranata, S. (2009). *Analisis Validitas, Reliabilitas, Dan Interpretasi Hasil Tes: Implementasi Kurikulum 2004* (4 ed.). Bandung: Remaja Rosdakarya.

Susanto, H. P., & Retnawati, H. (2023). Kalibrasi Instrumen Literasi Matematika Siswa Menggunakan IRT dan Aplikasinya untuk Estimasi Skor. *Edumatica : Jurnal Pendidikan Matematika*, 13(01), 23–36.
<https://doi.org/10.22437/edumatica.v13i01.23135>

Zed, M. (2008). *Metode Penelitian Kepustakaan*. Jakarta: Yayasan Obor Indonesia.

مطرفي, ١٣٢٣٣٠٢٠٣٧. تحليل بنود الأسئلة في الامتحان النهائي لدرس اللغة العربية في الفصل السابع في رسالة المدرسة الأندلسي المتوسطة كباسين بانيوماس السنة الدراسية ٢٠٢٠٢/٢٠٢١ رسالة (Skripsi). Jember.
Diambil dari <https://repository.uinsaizu.ac.id/5821/>