مقارنة تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية ونظرية استجابة البنود لدرس اللغة العربية في المدرسة المتوسطة الإسلامية الحكومية باتو

رسالة الماجستير

إعداد:

محمد جندي: ٢١٠١٠٤٢٢٠٠١٢

قسم تعليم اللغة العربية كلية الدراسات العليا

جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج

٢٠٢٣

استهلال

كُتِبَ عَلَيْكُمُ الْقِتَالُ وَهُوَ كُرْهٌ لَّكُمْ ۚ وَعَسٰٓى اَنْ تَكْرَهُوْا شَيْـًٔا وَّهُوَ خَيْرٌ لَّكُمْ ۚ وَعَسٰٓى اَنْ تُحِبُّوْا شَيْـًٔا وَّهُوَ شَرٌّ لَّكُمْ ۗ وَاللّٰهُ يَعْلَمُ وَاَنْتُمْ لَا تَعْلَمُوْنَ

﴿ البقرة : ٢١٦ ﴾

إهداء

أُهدي هذا البحث إلى:

أبي وأمي المحبوبين، الفاضل شمس والفاضلة فاتمي س. سوكولاني الذان ربياني وأعطياني رحمتهما وأنفقا لحياتي ودراستي حتى الآن و أدباني أحسن التربية.

شكر وتقدير

إن الحمد لله، نحمده ونستعينه ونستغفره، ونعوذ بالله من شرور أنفسنا وسيئات أعمالنا، من يهد الله فلا مضل له، ومن يضلل فلا هادي له، أشهد أن لا إله إلا الله وحده لا شريك له وأشهد أن محمدا عبده ورسوله، أما بعد.

فبتوفيق الله وبعنايته وبإذنه قد انتهى الباحث من كتابة هذه رسالة الماجستير تحت الموضوع **"مقارنة تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية ونظرية استجابة البنود لدرس اللغة العربية في المدرسة المتوسطة الإسلامية الحكومية باتو"** مع الرجاء لتكون هذه رسالة الماجستير نافعا للباحث ولجميع المهتمين باللغة العربية، و لن أَصِلَ إلى هذه النهاية إلا بفضله عز وجل، وكل من ساعدني بالمساعدات الكثيرة وأرشدني بالتوجيهات القيمة، فبهذا الواقع ينبغي للباحث تقديم الشكر والتقدير إلى:

١. الأستاذ الدكتور زين الدين الماجستير كمدير جامعة مولانا مالك إبراهيم الإسلامية الحكومية، والأستاذ الدكتور واحد موربي الماجستير كعميد كلية الدراسات العليا.
٢. الدكتور شهداء الماجستير كرئيس القسم والدكتور توفيق الرحمن الماجستير كسكرتير القسم في قسم تعليم اللغة العربية كلية الدراسات العليا.
٣. الدكتور نور قمري الماجستير والدكتور نور هادي الماجستير، الذان قد أرشدا الباحث بكل اهتمام وصبر وحكمة في كتابة هذه رسالة الماجستير.
٤. جميع أساتذتي في قسم تعليم اللغة العربية كلية الدراسات العليا جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج الذين قدموا العلوم والمعارف والتشجيع للباحث، وجزاهم الله خير الجزاء.
٥. والداي المحبوبين، الفاضل شمس والفاضلة فاتمي س. سوكولاني. و إلى أختي الكبيرة حكمة و إلى أخي الكبير محمد دلّع، وجميع أهلي الأحباء.

٦. أوانج الذي قد عرَّف نظرية استجابة البنود للباحث وساعدَ الباحث في المناقشات حول إنجاز هذا البحث.

٧. ذكراً وإمان، اللذين رافقا وساعدا الباحث خلال فترة دراسته في جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج.

٨. جميع أصحابي في فصل أ ماجستير تعليم اللغة العربية ٢٠٢١/٢٠٢٢ كلية الدراسات العليا جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج.

هذا، وأسأل الله أن تكون أعمالهم مقبولة ومثابة، وتكون هذه رسالة الماجستير نافعة ومفيدة للجميع، آمين يا رب العالمين.

باتو، ديسمبر ٢٠٢٣

الباحث،

محمد جندي

رقم التسجيل: ٢١٠١٠٤٢٢٠٠١٢

موافقة المشرف

بعد الاطلاع على رسالة الماجستير التي أعدها الطالب:

الاسم : محمد جندي

الرقم الجامعي : ٢١٠١٠٤٢٢٠٠١٢

العنوان : مقارنة تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية ونظرية استجابة البنود لدرس اللغة العربية في المدرسة المتوسطة الإسلامية الحكومية باتو

وافق المشرفان على تقديمها إلى لجنة المناقشة.

باتو، اكتوبر ٢٠٢٣ م

المشرف الأول،

الدكتور، الحاج، نور هادي، الماجستير

رقم التوزيف: ١٩٦٤٠١٠٣٢٠٠٣١٢١٠٠١

المشرف الثاني،

الدكتور، نور قمري، الماجستير

رقم التوزيف: ١٩٨٤٠٧٠٢٢٠١٦٠٨٠١١٠٩٥

اعتماد

رئيس قسم تعليم اللغة العربية

الدكتور، الحاج، شهداء، الماجستير

رقم التوزيف: ١٩٧٢٠١٠٦٢٠٠٥٠١١٠٠١

ه

اعتماد لجنة المناقشة

إن رسالة الماجستير بعنوان: مقارنة تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية ونظرية استجابة البنود لدرس اللغة العربية في المدرسة المتوسطة الإسلامية الحكومية باتو، التي أعدتها الطالب:

الاسم : محمد جندي

الرقم الجامعي : ٢١٠١٠٤٢٢٠٠١٢

قد قدّمتها الطالب أمام لجنة المناقشة في يوم الأربعاء، تاريخ ٢٩ نوفمبر ٢٠٢٣م، وقد تم التصحيح بناء على اقترحات لجنة المناقشة. وقررت اللجنة بقبولها شرطا للحصول على درجة الماجستير في تعليم اللغة العربية. وتتكون لجنة المناقشة من السادة:

الدكتور شهداء، الماجستير — مناقشا أساسيا

رقم التوظيف : ١٩٧٢٠١٠٦٢٠٠٥٠١١٠٠١

الدكتور عبد المنتقم الأنصاري، الماجستير — رئيسا مناقشا

رقم التوظيف : ١٩٨٦٠٩١٢٢٠١٥٠٣١٠٠٦

الدكتور نور هادي، الماجستير — مشرفا ومناقشا

رقم التوظيف : ١٩٦٤٠١٠٣٢٠٠٣١٢١٠٠١

الدكتور نور قمري، الماجستير — مشرفا ومناقشا

رقم التوظيف : ١٩٨٤٠٧٠٢٢٠١٦٠٨٠١١٠٩٥

اعتماد،

عميد كلية الدراسات العليا

الأستاذ الدكتور واحد [illegible]

رقم التوظيف: ١٩٦٩٠٣٠٣٢٠٠٠٠٣١٠٠٢

إقرار أصالة البحث

أنا المواقع أدناه:

الاسم : محمد جندي

الرقم الجامعي : ٢١٠١٠٤٢٢٠٠١٢

العنوان : مقارنة تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية ونظرية استجابة البنود لدرس اللغة العربية في المدرسة المتوسطة الإسلامية الحكومية باتو

أقر بأن هذه رسالة الماجستير الذي أعددتها لتوفير شرط من الشروط اللازمة للحصول على درجة الماجستير في قسم تعليم اللغة العربية كلية الدراسات العليا بجامعة مولانا ملك إبراهيم الإسلامية الحكومية مالانج. فقد حضرتها وكتبتها بنفسي وما زورتها من إبداع غيري أو تأليف أخر. وإذا احتج أحد استقبالا أنها من تأليفه وتبين أنها فعلا ليس من بحثي فأنا أتحمل المسؤولية على ذلك، ولن تكون المسؤولية على المشرف أو على كلية الدراسات العليا بجامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج.

كتبت هذا الإقرار بناء على رغبتي الخاصة ولا يجبرني أحد على ذلك.

باتو، اكتوبر ٢٠٢٣م

طالب،

محمد جندي

رقم التسجيل: ٢١٠١٠٤٢٢٠٠١٢

مستخلص البحث

جندي، محمد. ٢٠٢٣ م، مقارنة تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية ونظرية استجابة البنود لدرس اللغة العربية في المدرسة المتوسطة الإسلامية الحكومية باتو. رسالة الماجستير، قسم تعليم اللغة العربية، كلية الدراسات العليا، جامعة ملولانا مالك إبراهيم مالانج. المشرف الأول الدكتور نور هادي الماجستير والمشرف الثاني الدكتور نور قمري الماجستير.

الكلمات الأساسية: تحليل بنود الأسئلة، نظرية الاختبار التقليدية، نظرية استجابة البنود.

أحد الخطوات المهمة في تطوير الاختبار هو تحليل بنود الأسئلة، والذي يمكن أن يتم إجرائه كيفيا وكميا. استنادًا إلى الدراسة المكتبية، هناك نهجان رئيسيان في تحليل بنود الأسئلة الكمي، وهما نظرية الاختبار التقليدية ونظرية استجابة البنود. عمومًا، في سياق بحوث تعليم اللغة العربية، يعتمد العديد على نظرية الاختبار التقليدية، بينما لا يزال استخدام نظرية استجابة البنود غير متبع على نطاق واسع. تشير الدراسة الميدانية الأولية إلى أنه في عملية تطوير أدوات الاختبار لامتحان اللغة العربية في المدرسة المتوسطة الإسلامية الحكومية باتو، يتم تحليل عناصر الاختبار حاليًا بشكل كيفي ولا يشمل التحليل الكمي. لذلك، استنادًا إلى النتائج المستخلصة من الدراسة المكتبية والدراسة الميدانية الأولية، يهدف الباحث إلى مقارنة وتطبيق النهجين في تحليل بنود الأسئلة اللغة العربية في المدرسة.

النهج المستخدم في هذا البحث هو النهج الوصفي الكمي مع عينة تتألف من ٣١٣ شخصًا. تشمل طرق جمع البيانات التوثيق والمقابلة. فيما يتعلق بتحليل البيانات، يتم استخدام نهجين: التحليل بناءً على نظرية الاختبار التقليدية والتحليل بناءً على نظرية استجابة البنود، والتي تتضمن تحليلًا رياضيًا يساعده البرنامج الحاسوبي. تتضمن التطبيقات المستخدمة Microsoft Excel 2013، SPSS v.26، R 4.3.0، وRstudio.

تشمل نتائج البحث ما يلي: ١) يظهر التحليل باستخدام نظرية الاختبار التقليدية أن ٨٤% من بنود الاختبار صحيحة، مع ثبات الأداة مرتفع يبلغ ٠,٨٦. ومع ذلك، فيما يتعلق بمستوى الصعوبة، هناك ٧١% من الأسئلة متوسطة، و٢٣% صعبة، وفقط ٦% سهلة. يحتاج التركيز أيضًا على درجة التمييز، حيث أن ٤٧% فقط لديها درجة التمييز عالية. ٢) يظهر التحليل باستخدام نظرية استجابة البنود بالنموذج اللوجستي ثلاثي المعاملات أن ٧٦% من الأسئلة تعتبر صحيحة وبتنوع ثبات بين مشاركي بالاختبار. على الرغم من أن افتراض أحادية البعد يتحقق، إلا أن افتراض عدم الترابط المحلي وثبات المعاملة لا تتحقق تمامًا ولكن يمكن التعامل معها من خلال حذف بعض البنود غير المتوافقة. هناك ٧٢% من الأسئلة ذات مستوى الصعوبة جيد، ولكن ٤٦% فقط من الأسئلة لديها درجة التمييز مرضية. و٧٠% من الأسئلة تظهر تخمين زائف جيدًا. يصل عدد البنود المعتبرة جيد جدا وجيد وفقًا لمعايير نظرية استجابة العنصر إلى ٧٠,٤٥%. ٣) هناك اختلاف في العملية بين نظرية الاختبار التقليدية ونظرية استجابة البنود، خاصة فيما يتعلق بالتقدير والتصنيف للمعاملات. ومع ذلك، لا يظهر تحديد نقاط الضعف في بنود الأسئلة فرقًا كبيرًا بين النهجين. لذلك، يمكن استخدام توجيه النظريتين لتحليل بنود الأسئلة بشكل أشمل. باستخدام النظرين، تشير نتائج التفسير إلى أن ٥٦% من بنود الاختبار تعتبر جيدة ويمكن إعادة استخدامها في المستقبل.

ABSTRACT

Jundi, Muhammad, 2023. *Comparison of the Analysis of Test Items Using Classical Test Theory and Item Response Theory for Arabic Language Lesson in the State Islamic Junior High School of Batu*. Thesis, Arabic Language Education Department, Postgraduate, Maulana Malik Ibrahim State Islamic University of Malang, advisors: 1) Dr. Nurhadi, M.A.; 2) Dr. Nur Qomari, M.Pd.
Keywords: Item Test Analysis, Classical Test Theory, Item Respons Theory.

One crucial step in test development is item analysis, which can be conducted both qualitatively and quantitatively. Based on the literature review, there are two main quantitative approaches to item analysis: classical test theory and item response theory. Generally, in the context of Arabic language education research, the classical test theory is more commonly applied, while the use of item response theory is not yet widespread. Preliminary field studies indicate that in the process of developing test instruments for Arabic language exams at MTs. N in Batu City, item analysis is currently conducted qualitatively and lacks quantitative analysis. Therefore, based on findings from the literature review and preliminary field studies, the researcher aims to compare and apply both approaches in analyzing Arabic language exam items at MTs. N in Batu City.

The research adopts a quantitative descriptive approach with a sample size of 313 participants. Data collection methods include documentation and interviews. Two data analysis methods are employed: analysis based on classical test theory and analysis based on item response theory, involving mathematical analysis assisted by computer software. The applications used include Microsoft Excel 2013, SPSS v.26, R 4.3.0, and Rstudio.

Research findings include the following: 1) Analysis using classical test theory shows that 84% of the items are valid, with a high instrument reliability coefficient of 0.86. However, in terms of difficulty, 71% of the items are considered moderate, 23% are difficult, and only 6% are easy. Attention is needed for item discrimination, as only 47% have a high discrimination level. 2) Analysis using item response theory with a 3-parameter logistic model indicates that 76% of the items are considered valid with varied reliability among test takers. While the unidimensionality assumption is met, the local independence and invariance of parameters are not fully satisfied but can be addressed by removing some non-compliant items. There are 72% of items with good difficulty, but only 46% have satisfactory discrimination. Meanwhile, 70% of items show good pseudo-guessing. Items classified as very good and good according to item response theory criteria reach 70.45%. 3) Differences in the process between classical test theory and item response theory exist, especially in the estimation and categorization of parameters. However, the identification of item weaknesses does not show significant differences between the two approaches. Therefore, a combination of both theories can be used for a more comprehensive item analysis. Using both theories, the interpretation results indicate that 56% of items are considered good and can be reused in the future.

ABSTRAK

Jundi, Muhammad, 2023. *Perbandingan Analisis Butir Soal menggunakan Teori Tes Klasik dan Teori Respons Butir pada Mata Pelajaran Bahasa Arab di MTs. N Batu.*Tesis, Program Studi Pendidikan Bahasa Arab, Program Pascasarjana, UIN Maulana Malik Ibrahim Malang, Pembimbing: 1) Dr. Nurhadi, M.A.; 2) Dr. Nur Qomari, M.Pd.

Kata Kunci: Analisis Butir Soal, Teori Tes Klasik, Teori Respons Butir.

Salah satu langkah penting dalam pengembangan tes adalah analisis butir soal, yang dapat dilakukan baik secara kualitatif maupun kuantitatif. Berdasarkan studi pustaka, terdapat dua pendekatan utama dalam analisis butir soal secara kuantitatif, yaitu teori tes klasik dan teori respons butir. Umumnya, dalam konteks penelitian pendidikan bahasa Arab, pendekatan yang banyak diterapkan adalah teori tes klasik, sedangkan penggunaan teori respons butir masih belum umum digunakan. Studi lapangan awal menunjukkan bahwa dalam proses pengembangan instrumen tes untuk ujian mata pelajaran bahasa Arab di MTs. N Kota Batu, analisis butir soal baru dilakukan secara kualitatif dan belum melibatkan analisis butir soal secara kuantitatif. Oleh karena itu, berdasarkan temuan dari studi pustaka dan studi lapangan awal ini, peneliti bertujuan untuk membandingkan dan menerapkan kedua pendekatan tersebut dalam menganalisis butir soal ujian bahasa Arab di MTs. N Kota Batu.

Pendekatan yang digunakan dalam penelitian ini adalah kuantitatif deskriptif dengan jumlah sampel yang terdiri dari 313 orang. Metode pengumpulan data yang digunakan adalah dokumentasi dan wawancara. Terkait dengan analisis data, digunakan dua metode: analisis berdasarkan teori ujian klasik dan analisis berdasarkan teori respons butir, yang melibatkan analisis matematis yang dibantu oleh perangkat lunak komputer. Aplikasi-aplikasi yang digunakan adalah Microsoft Excel 2013, SPSS v.26, R 4.3.0, dan Rstudio.

Temuan penelitian meliputi hal-hal berikut: 1) Hasil analisis menggunakan teori tes klasik menunjukkan bahwa 84% dari butir soal valid, dengan koefisien reliabilitas instrumen yang tinggi sebesar 0,86. Namun, pada tingkat kesukaran terdapat 71% soal menengah, 23% sulit, dan hanya 6% soal mudah. Daya pembeda soal juga perlu perhatian, karena hanya 47% yang memiliki tingkat daya pembeda tinggi. 2) Hasil analisis menggunakan teori respons butir dengan model 3 parameter logistik menunjukkan bahwa 76% soal dianggap valid dengan reliabilitas yang bervariasi di antara peserta tes. Meskipun asumsi unidimensi terpenuhi, asumsi independensi dan invariansi parameter tidak sepenuhnya terpenuhi, namun dapat diatasi dengan menghapus beberapa butir yang tidak memenuhi asumsi. Terdapat 72% soal dengan tingkat kesukaran yang baik, tetapi hanya 46% soal yang memiliki daya pembeda yang terkategori baik. Sementara, 70% soal memiliki tebakan semu yang baik. Butir-butir yang dianggap sangat baik dan baik menurut kriteria teori respons butir mencapai 70,45%. 3) Terdapat perbedaan dalam proses antara teori tes klasik dan teori respons butir, terutama pada parameter yang diestimasi dan kategorisasi. Meskipun demikian, identifikasi kelemahan butir soal tidak menunjukkan perbedaan signifikan antara kedua pendekatan tersebut. Oleh karena itu, kombinasi kedua teori dapat digunakan untuk melakukan analisis butir soal secara lebih komprehensif. Dengan menggunakan kedua teori, hasil interpretasi menunjukkan bahwa 56% dari butir soal dianggap baik dan dapat digunakan kembali di masa depan.

المحتويات

استهلال ..... أ
إهداء ..... أ
شكر وتقدير ..... ب
تقرير المشرفين ..... د
إقرار أصالة البحث ..... هـ
اعتماد لجنة المناقشة ..... و
مستخلص البحث ..... ز
المحتويات ..... ي

الفصل الأول: الإطار العام

أ. خلفية البحث ١
ب. أسئلة البحث ٥
ج. أهداف البحث ٦
د. أهمية البحث ٦
هـ. تحديد البحث ٧
و. تحديد المصطلحات ٧
ز. الدرسات السابقة ٨

الفصل الثاني: الإيطار النظري

المبحث الأول: الاختبار وبنود الأسئلة ١٣
أ. تعريف الاختبار ١٣
ب. أنواع الاختبار ١٣
ج. اختبار اللغة ١٥
المبحث الثاني: تحليل بنود الأسئلة ١٧
أ. بنود الأسئلة ١٧
ب. معايرة بنود الأسئلة ١٨

ج. تحليل بنود الأسئلة ١٩
المبحث الثالث: نظرية الاختبار التقليدية ٢٠
أ. تعريف نظرية الاختبار التقليدية ٢٠
ب. تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية ٢١
المبحث الثالث: المدخل في تحليل بنود الأسئلة ٢٦
أ. تعريف نظرية استجابة البنود ٢٦
ب. تحليل بنود الأسئلة باستخدام نظرية استجابة البنود ٢٨
الفصل الثالث: منهجية البحث
أ. مدخل البحث ومنهجه ٣٤
ب. مجتمع البحث وعينته ٣٤
ج. البيانات ومصدرها ٣٥
د. أسلوب جمع البيانات ٣٥
هـ. أسلوب تحليل البيانات ٣٥
١. تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية ٣٥
٢. تحليل بنود الأسئلة باستخدام نظرية استجابة البنود ٣٦
الفصل الرابع: عرض البيانات وتحليلها
أ. **تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية** ٣٩
١. طبيعية البيانات ٤٠
٢. الصدق ٤٢
٣. الثبات ٤٥
٤. مستوى الصعوبة ٤٦
٥. درجة التمييز ٥١
٦. تفسير نتائج التحليل ٥٤
ب. **تحليل بنود الأسئلة باستخدام نظرية استجابة البنود** ٥٨
١. مطابقة النموذج ٥٩
٢. الصدق ٦١

٣. الثبات ٦٧
٤. الافتراضات ٦٨
٥. مستوى الصعوبة ٧١
٦. درجة التمييز ٧٤
٧. التخمين الزائف ٧٦
٨. تفسير نتائج التحليل ٧٩
ت. **تحليل بنود الأسئلة باستخدام نظرية استجابة البنود** ٩٤
١. تقدير الصدق ٩٤
٢. تقدير الثبات ٩٥
٣. المعاملات المقَدّرة ٩٦
٤. مقارنة نتائج التحليل باستخدام النظريتين ٩٧
الفصل الخامس: مناقشة نتائج البحث
أ. مناقشة تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية ١٠٣
ب. مناقشة تحليل بنود الأسئلة باستخدام نظرية استجابة البنود.....١٠٥
ج. مناقشة مقارنة تحليل بنود الأسئلة باستخدام النظريتين ١٠٨
الفصل الخامس: الخاتمة
أ. ملخص نتائج البحث.....١١٠
ب. التوصيات والاقتراحات.....١١١

الفصل الأول
الإطار العام

أ. خلفية البحث

يعد التقييم عملية أساسية في التعليم، ويهدف إلى تقييم قدرة الطلاب على تحقيق الأهداف التعليمية المحددة. ويمكن استخدام نتائج التقييم لتحسين عملية التعليم وتحديد المناطق التي يجب تحسينها في التدريس.[١] يجري التقييم لقياس تأثير التعليم على الطلاب، ويتعلق دائمًا بالأهداف التعليمية المحددة. لذلك، فإن التقييم هو عنصر أساسي في عملية التدريس لا يمكن تجاهله من قبل المدرس. فالمهمة التعليمية للمدرس لا تقتصر على التدريس فقط، بل تتضمن أيضًا إجراء التقييم للطلاب.[٢] وبإمكان نتائج التقييم استخدامها كمرجع للمدرس لتقييم نجاح الطلاب وتحسين جودة تدريسه.[٣] ويعتبر التقييم من العوامل المهمة لمساعدة المدرسين في اتخاذ القرارات الصحيحة في عملية التعليم.[٤] وبذلك، يمثل التقييم عنصر أساسي في عملية التدريس، ويساعد في تحسين جودة التعليم وتحقيق الأهداف التعليمية المحددة.

وتعد الاختبارات من أهم أدوات التقوم.[٥] الاختبارات هي واحدة من أكثر الأدوات المستخدمة في التعليم لقياس مستويات قدرة الطلاب.[٦] في معظم المؤسسات التعليمية الرسمية، يتم تقييم في التعليم من خلال الاختبارات الدورية التي تُستخدم كوسيلة لتقييم تقدم الطلاب في التعليم.[٧] يتم إجراء هذه الاختبارات في منتصف ونهاية كل فصل

---

[١] مهدي محمد عليمات, ''بناء اختبار محكي المرجع في الرياضيات وفق النموذج اللوجستي ثلاثي المعلمة,'' *مجلة جامعة القدس المفتوحة للأبحاث والدراسات التربوية والنفسية* ١٣.004-038-013-1182/10.33977/https://doi.org), (April 19, 2022) 38 .no , .

[٢] Kementrian Pendidikan dan Kebudayaan, "Permendiknas No.16, Tahun 2007, Tentang Standar Kualifikasi Akademik Dan Kompetensi Guru," 2007.

[٣] Djemari Mardapi, *Pengukuran Penilaian Dan Evaluasi Pendidikan*, 2nd ed. (Yogyakarta: Nuha Litera, 2016).

[٤] هاشم عواضة, *تقويم التعلم* (بيروت: دار العلم للملايين, ٢٠١٠); نورا الشامخ, *التقويم في التعليم* (الألوكة, ٢٠١٨) , https://www.alukah.net/books/files/book_10616/bookfile/taqweem.pdf.

[٥] عليمات, ''بناء اختبار محكي المرجع في الرياضيات وفق النموذج اللوجستي ثلاثي المعلمة''.

[٦] Evangelos Triantafillou, Elissavet Georgiadou, and Anastasios A. Economides, "The Design and Evaluation of a Computerized Adaptive Test on Mobile Devices," *Computers & Education* 50, no. 4 (May 2008): 1319–30, https://doi.org/10.1016/j.compedu.2006.12.005.

[٧] أحمد عودة, *القياس والتقويم في العملية التدريسية* (شركة الأمل للطباعة والنشر والتوزيع, ٢٠٠٥); رجاء محمود أبو علام, *تقويم التعلم* (دار المسيرة للطباعة والنشر, ٢٠٠٥).

دراسي لتوفير تمثيل دقيق لقدرات الطلاب في التعليم خلال هذه الفترة. يقوم المدرسون في كل مجال بإعداد أدوات الاختبار على شكل أسئلة، والتي تتناسب مع نوع الاختبار، سواء كان شفويًا أو كتابيًا. الغرض الرئيسي من الاختبارات هو قياس مستوى إنجاز الطلاب في مجال معين. تعد الاختبارات أداة فعالة لتقييم نتائج تعليم الطلاب وتوفير ملاحظات للطلاب والمدرسين. من خلال التقييم، يمكن للمدرسين تحديد المجالات التي يحتاج الطلاب فيها إلى دعم إضافي وتطوير استراتيجيات للتعامل مع هذه الاحتياجات. بشكل عام، تلعب الاختبارات الدور الحاسم في ضمان قدرة الطلاب على تحقيق أهداف التعليمية.

يجب أن تكون أداة الاختبار التي أعدها المدرس قادرة على تقديم معلومات صحيحة حول قدرات الطلاب. ومع ذلك، لا يزال هناك العديد من نقاط الضعف في عملية الصياغة مما أدى إلى اختبارات ليس لها قابلية قياس صحيحة.[٨] لا يمكن أن يوفر الاختبار غير الصحيح معلومات دقيقة حول قدرات الطلاب. غالبًا ما ينظر المدرسون إلى أعلى الدرجات وأدناها ومتوسطها من قياس نتائج تعلم الطلاب لتقييم إنجاز عملية التعلم. ومع ذلك، فإن هذه المعلومات ليست كافية لتقديم صورة دقيقة لقدرات الطلاب. يمكن للطلاب الحصول على درجات عالية لأن الأسئلة التي أعدها المدرس ذات مستوى منخفض من الصعوبة أو أن خيارات الإجابة ليست متجانسة بحيث يمكن للطلاب بسهولة تحديد اختيارات الإجابة الخاطئة.[٩] ولذلك، تتطلب تنفيذ التقييم في المدارس وجود أداة تقييم فعالة وهي الاختبار. فيجب للاختبار أن يتكون من بنود الأسئلة ذات جودة جيدة.

---

[٨] Rahmat Danni, Ajeng Wahyuni, and Tauratiya Tauratiya, "ITEM RESPONSE THEORY APPROACH: KALIBRASI BUTIR SOAL PENILAIAN AKHIR SEMESTER MATA PELAJARAN BAHASA ARAB," *Arabi : Journal of Arabic Studies* 6, no. 1 (June 30, 2021): 93–104, https://doi.org/10.24865/ajas.v6i1.320.

[٩] Dinar Pratama, "Analisis Kualitas Tes Buatan Guru Melalui Pendekatan Item Response Theory (IRT) Model Rasch," *Tarbawy : Jurnal Pendidikan Islam* 7, no. 1 (April 24, 2020): 61–70, https://doi.org/10.32923/tarbawy.v7i1.1187.

بحسب وزارة التربية الوطنية، فإن معايير الاختبار الجيد هي الاختبارات التي مرت بنود الأسئلة فيها بمراحل التحليلين.[١٠] فالهدف الأساسي من تحليل الأسئلة هو معرفة جودة بنود الأسئلة ونواقصها لتصحيحها وتحسينها.[١١] هناك مدخلان لتحليل جودة بنود الأسئلة، وهما الأسلوب الكيفي والكمي. يتضمن التحليل الكيفي تقييم بنود أسئلة الاختبار بمراعة على تقنية الكتابة واستخدام اللغة وتناسب المواد. عند المدخل الكيفي، هناك تقنيتا التحليل وهما تقنية الوسيط وتقنية اللوحة. من ناحية أخرى، يمكن رؤية التحليل الكمي من خصائص الاختبار الداخلية التي تم الحصول عليها من نتائج القياسات التجريبية للمتقدمين للاختبار. يتضمن التحليل الكمي استخدام بيانات الاختبار لتقييم جودة بنود الأسئلة، مثل مستوى الصعوبة وقوة التمييز والموثوقية.[١٢] للتحليل الكمي، يجب التحقق من صحة الاختبار قبل الاستخدام.[١٣] اختيار الطريقة المناسبة لتحليل بنود الاختبار يمكن أن يساعد على تحسين جودة الاختبار وتعزيز تطوير التعلم بشكل أفضل.

عند تحليل بنود الأسئلة كميًا، يوجد نهجان يمكن استخدامهما: نظرية الاختبار التقليدية (Classical Test Theory/CTT) ونظرية استجابة البنود (Item Responsee Theory/IRT).[١٤] في اللغة الإندونيسية، تُسمى نظرية الاختبار التقليدية بـ "Teori Tes

---

[١٠] Dyah Febria Wardhani and Aminuddin Prahatama Putra, "Pengembangan Instrumen Tes Standar Kognitif pada Mata Pelajaran IPA Kelas 7 SMP di Kabupaten Banjar," in *Seminar Nasional XIII Pendidikan Biologi FKIP UNS 2016* (Seminar Nasional XIII Pendidikan Biologi FKIP UNS 2016, Sebelas Maret University, 2016), https://www.neliti.com/publications/175334/; Kementrian Pendidikan dan Kebudayaan, "Panduan Penulisan Soal Pilihan Ganda" (Depdiknas, 2007).

[١١] Farida, Syarifah, and Muhamad, "تحليل بنود الأسئلة في الإمتحان النهائي لمادة اللغة العربية في الفصل العاشر بالمدرسة العالية 'الإسلام' كموجا," *AL-MU'ARRIB: JOURNAL OF ARABIC EDUCATION* 1, no. 2 (November 12, 2021): 1–15, https://doi.org/10.32923/al-muarrib.v1i2.2064.

[١٢] Sumarna Supranata, *Analisis Validitas, Reliabilitas, dan Interpretasi Hasil Tes: Implementasi Kurikulum 2004*, 4th ed. (Bandung: Remaja Rosdakarya, 2009); Dinar Pratama, "ANALYSIS OF CLASICAL TEST THEORY (CTT) APPROCH ON ACADEMIC ABILITY TEST INSTRUMENT," *JISAE: Journal of Indonesian Student Assessment and Evaluation* 5, no. 2 (2019): 43–54, https://doi.org/10.21009/jisae.v5i2.12701.

[١٣] Pratama, "Analisis Kualitas Tes Buatan Guru Melalui Pendekatan Item Response Theory (IRT) Model Rasch."

[١٤] Igor Himelfarb, "A Primer on Standardized Testing: History, Measurement, Classical Test Theory, Item Response Theory, and Equating," *Journal of Chiropractic Education* 33, no. 2 (October 1, 2019): 151–63, https://doi.org/10.7899/JCE-18-22.

Klasik/TTK"، وتُسمى نظرية استجابة البنود بـ "Teori Respons Butir/TRB". ينظر نهج نظرية الاختبار التقليدية إلى الاختبار كمجموعة من البنود المنفصلة والمستقلة التي يتم حسابها بشكل منفصل لقياس مستوى الصعوبة وقوة التمييز والموثوقية.[١٥] بينما تنظر نهج نظرية استجابة البنود إلى الخصائص الفردية لبنود الاختبار وتحسب قدرة المجيب بناءً على القدرة المقاسة بالبنود التي تم الإجابة عليها.[١٦] على الرغم من أن كلا النهجين لهما نقاط القوة والضعف الخاصة بهما، فإن نهج نظرية الاختبار التقليدية عمومًا تم استخدامه ودراسته على نطاق واسع في تطوير أدوات التقييم. ومن ناحية أخرى، لم يتم استخدام نهج نظرية استجابة البنود على نطاق واسع بسبب التعقيد الفني المرتبط بتحليل البيانات. في الدراسة المكتبية الذي تمت من خلال موقع جوجل الأكاديمية، تم اكتشاف أنه خلال فترة الخمس سنوات الماضية، تم إجراء ٣٧ دراسة أو أكثر تحليلية لبنود اختبار اللغة العربية باستخدام نظرية الاختبار التقليدية.[١٧] في حين أن هناك اثنين فقط من الدراسات التي تطبق نظرية استجابة البنود في تحليل بنود الاختبار. بناءً على مراعاة مزايا وعيوب كل من هاتين النظريتين، يتم تحليل هاتين النهجين بشكل مقارن بهدف الحصول على نتائج تحليلية مثلى.

إن المدرسة الثناوية الإسلامية الحكومية باتو واحدة من المدارس الحكومية الإسلامية المتوسطة الموجودة في المدينة. وتقوم المدرسة بتنفيذ التقييمات الدورية للطلاب، بما في ذلك مادة اللغة العربية. ولأغراض اختبارات اللغة العربية، يعد أسئلة الاختبارات من

---

[١٥] Pratama, "ANALYSIS OF CLASICAL TEST THEORY (CTT) APPROCH ON ACADEMIC ABILITY TEST INSTRUMENT"; Heni Purwati et al., "Analisis Karakteristik Butir Soal Ujian Nasional Matematika SMP/MTs Berdasarkan Pendekatan Teori Tes Klasik," *Jurnal Sains Dan Edukasi Sains* 4, no. 2 (Agustus 2021), https://doi.org/10.24246/juses.v4i2p46-51.

[١٦] أحمد سالم سعيد القرني, "أثر طرق تقدير القدرة للأفراد على دقة معادلة الاختبارات في نظرية الاستجابة للمفردة," *المجلة العربية للعلوم التربوية والنفسية* ٦ (October 1, 2022): 1–22, https://doi.org/10.21608/jasep.2022.265536; مصطفى علي خلف علي, "استخدام نظرية الاستجابة للمفردة الاختبارية في بناء مقياس الاتجاه نحو مهنة التدريس," *مجلة البحث في التربية وعلم النفس* ٣٤ (January 1, 2019): 1–47, https://doi.org/10.21608/mathj.2019.82386; محمد دبوس, "استخدام نظرية االستجابة للفقرة في بناء فقرات اختبار محكي المرجع في الرياضيات بفقرات ثنائية التدريج ومتعددة التدريج وفق النموذج اللوجستي ثنائي المعلم," *مجلة جامعة النجاح لألبحاث (العلوم اإلنسانية)* ٣٠ (2016), https://journals.najah.edu/media/journals/full_texts/7_B2fCogF.pdf.

[١٧] Muhammad Jundi, "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 الأسلة في إندونيسيا," *ATHLA: Journal of Arabic Teaching, Linguistic and Literature* 4, no. 2 (December 2023).

قبل فريق من مدرسي اللغة العربية في المدرسة. يقوم كل مدرّس بإعداد مجموعة من الأسئلة المطوَّرة بناءً على إرشادات الأسئلة الموجَّهة والتي تم تحديدها لاختبار المدرسة وفقًا لقرار وزير الشؤون الدينية رقم ١٨٣ الصادر عن المديرية العامة للتعليم الإسلامي بوزارة الشؤون الدينية في جمهورية إندونيسيا. يشمل الأسئلة جميع المواد التعليمية التي تم دراستها من الصف السابع إلى الصف التاسع، بدءًا من فصل التعريف في الفصل الأول للصف السابع وصولاً إلى فصل الحفاظ على البيئة في الفصل الثاني للصف التاسع.

تم وتحليل الأسئلة بشكل نوعي لضمان توافقها مع أهداف التعليم والكفايات الأساسية ومؤشرات التعلم وشبكة الأسئلة.على الرغم من ذلك، لم يتم إجراء تحليل كمي لأسئلة الاختبار حتى الآن. ويتعين إجراء تحليل كمي ونوعي على أسئلة الاختبار، حيث يمكن أن يوفر التحليل الكمي معلومات هامة حول صدق الأسئلة وثباتها ومستوى صعوبتها وقدرتها على التمييز وغير ذلك. ويمكن استخدام هذه المعلومات لتحسين جودة الأسئلة وضمان توافقها مع الأهداف التعليمية، والتي يمكن أن تساعد على تحسين استراتجيات التعليم وتطويرها. ويمكن من خلال ذلك تحسين جودة تعليم اللغة العربية في المدرسة الثناوية الإسلامية الحكومية باتو.

بناءً على الخلفية التي تم تفصيلها أعلاه، فمن المهم إجراء بحث حول مقارنة تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية ونظرية استجابة البنود لدرس اللغة العربية بالمدرسة المتوسطة الإسلامية الحكومية باتو.

ب. أسئلة البحث

استنادا على خلفية البحث، يريد الباحث أي يقوم بمقارنة تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية ونظرية استجابة البنود لدرس اللغة العربية بالمدرسة المتوسطة الإسلامية الحكومية باتو. فالباحث يأتي بأسئلة البحث:

١. كيف يتم إجراءات تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية لاختبار درس اللغة العربية في المدرسة المتوسطة الإسلامية الحكومية باتو؟

٢. كيف يتم إجراءات تحليل بنود الأسئلة باستخدام نظرية استجابة البنود لاختبار درس اللغة العربية في المدرسة المتوسطة الإسلامية الحكومية باتو؟
٣. كيف يتم مقارنة إجراءات تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية ونظرية استجابة البنود لاختبار درس اللغة العربية في المدرسة المتوسطة الإسلامية الحكومية باتو؟

ج. أهداف البحث

بناءً على أسئلة البحث التي قدمها الباحث، فالباحث يأتي بالهدف:
١. لوصف إجراءات تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية لاختبار درس اللغة العربية في المدرسة المتوسطة الإسلامية الحكومية باتو
٢. لوصف إجراءات تحليل بنود الأسئلة باستخدام نظرية استجابة البنود لاختبار درس اللغة العربية في المدرسة المتوسطة الإسلامية الحكومية باتو
٣. لوصف مقارنة إجراءات تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية ونظرية استجابة البنود لاختبار درس اللغة العربية في المدرسة المتوسطة الإسلامية الحكومية باتو

د. أهمية البحث

وأما أهمية البحث التى ترد الوصول في هذا البحث على النحو التالي:
١. نظريا: أن يكون هذا البحث زيادة المعلومات والخبرات لدي الباحث والقارئ حول بتحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية ونظرية استجابة البنود لدرس اللغة العربية بالمدرسة المتوسطة الإسلامية الحكومية باتو

٢. عمليا: أ) للباحث: أن يكون هذا المرجع والمدخل في تحليل الأسئلة باستخدام نظرية الاختبار التقليدية ونظرية استجابة البنود وتوفير المعلومات لتأليف بنود الأسئلة الجيدة.

ب)للمدرّس: أن يكون هذا البحث تسهيلا لمدرّسي اللغة العربية على تحليل الأسئلة وقدرات الطلاب بحيث يمكن استخدامها كمرجع لتحسين وتطوير الأسئلة في المستقبل.

ت)للمدرسة: أن يكون هذا مرجعًا للمدرسة فيما يتعلق بتنفيذ التقييم في كل مادة دراسية وتطوير أدواته

هـ. تحديد البحث

يكون الحدود في هذا البحث يحتوي على:

١. الحدود الموضوعية: أخذ الباحث تركيز موضع بحثه حول تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية ونظرية استجابة البنود على اختبار اللغة العربية النهائي للصف التاسع بالمدرسة المتوسطة الإسلامية الحكومية باتو

٢. الحدود الزمـانـية: حدد الباحث زمان بحثه حول بتحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية ونظرية استجابة البنود لدرس اللغة العربية بالمدرسة المتوسطة الإسلامية الحكومية باتو في السنة الدراسية ٢٠٢٣/٢٠٢٤

و. تحديد المصطلحات

١. تحليل بنود الأسئلة: المقصود بتحليل بنود الأسئلة في هذه الدراسة هو دراسة وتحليل بنية الأسئلة بشكل كمي واستناداً إلى البيانات التجريبية، حيث يتم استخدام اختبار الأسئلة كمصدر للتحليل. وبناءً على ذلك، يتم استنتاج نتائج التحليل استناداً إلى سياق المشاركين الفعلي. يجب التنويه أن هذا التحليل ليس من النوع الكيفي أو الداخلي، وأن الدراسة ليست تركز

على مناقشة محتوى الأسئلة أو مدى توافقها مع المواد وأهداف التعليم.

٢. نظرية الاختبار التقليدية (CTT): نظرية الاختبار التقليدية المشار إليها في هذه الدراسة تعد جزءاً من نظرية الاختبار التقليدية المستخدمة لتحليل بنود الاختبار ذات المعاملات المحددة، منها مستوى الصعوبة، ودرجة التمييز، وفعالية الخيارات.

٣. نظرية استجابة البنود (IRT): نظرية استجابة البنود المشار إليها في هذه الدراسة تمثل جزءًا من النظرية المستخدمة لتحليل البنود الاستدلالية بمعايير محددة، ومنها: درجة الصعوبة، ودرجة التمييز، والتخمين الزائف.

٤. درس اللغة العربية: التعليم للغة العربية المشار إليه في هذا البحث هو التعليم الرسمي في الفصول الدراسية. والأسئلةُ التي ستتمُّ تحليلُها هي أسئلةُ الامتحان النهائي.

ز. الدراسات السابقة

لأجل تجنب تكرار البحث حول نفس الموضوع، لقد حلل الباحث الدراسات السابقة فيما يتعلق بتحليل بنود الأسئلة لدرس اللغة العربية وهي:

١. هدف البحث الأول لتحليل بنود أسئلة اختبار نهاية الفصل في مادة الاقتصاد للمدرسة لعالية الحكومية باستخدام نموذج راش.[١٨] وتمت دراسة على طلاب الصف الحادي عشر في مدارس المتوسطة في مدينة يوجياكرتا. استُخدمت طريقة العينة العشوائية النسبية لاختيار ٣٠٦ طلابا. تضمّنت البيانات إجابات الطلاب

[١٨] Fitri Alfarisa and Dian Purnama, "ANALISIS BUTIR SOAL ULANGAN AKHIR SEMESTER MATA PELAJARAN EKONOMI SMA MENGGUNAKAN RASCH MODEL," *Jurnal Pendidikan Ekonomi* 11, no. 2 (2019).

على اختبار مادة الاقتصاد المؤلف من ٤٠ سؤالًا متعددة الاختيارات. تمّ التحقق من صدق الاختبار وثباته، حيث أثبت مؤشر Aiken V صدقه، وكانت ثباته ٠,٧٧ بواسطة KR-20. تم تحليل النتائج باستخدام برنامج Winstep بناءً على نموذج راش. من بين الأسئلة ٤٠، كان ٣٩ منها متوافقة مع النموذج، وتراوحت درجات صعوبة الأسئلة بين −٢,٥١ و ١,٧٧. أظهرت أعلى معلومات اختبار قيمة ١٨,٣٣٧٩، وكان متوسط قدرة الطلاب 0.2 (θ) لوغاريتم، مع خطأ قياسي للاختبار ٠,٢١.

٢. هدف البحث الثاني تحليل بنود أسئلة مادة اللغة العربية لطلاب المدرسة العالية الإسلامية الأهلية بمعهد السلام كامبار رياو.[١٩] توصل الباحث إلى النتائج التالية: أظهرت نتائج اختبار التمييز وجود ٧ أسئلة في الفئة العالية لدرجة التمييز الشديدة، و٩ أسئلة في الفئة العالية، و١٠ أسئلة في الفئة المتوسطة، و١٣ أسئلة في الفئة المنخفضة، و١١ سؤالًا في الفئة الضعيفة لدرجة التمييز. وبخصوص مستوى الصعوبة، أظهرت النتائج أن ٤٣ سؤالًا في الفئة السهلة مع نسبة إحصائية ٤٣:٧، وهناك ٧ أسئلة في الفئة المتوسطة. وفيما يتعلق بالصدق، تبين أن هناك ٢٤ سؤالًا بحالة صادقة و٢٦ سؤالًا بحالة غير صادقة. وأخيرًا، أظهرت نتائج اختبار الثبات أن الأسئلة كانت ثابتة بشكل كبير، وذلك من خلال معامل ألفا لكرونباخ البالغ قيمته ٠,٨٦٨.

٣. هدف البحث الثالث توصيف أسئلة اختبار مهارات التفكير النقدي باستخدام تحليل نظرية استجابة البنود في موضوع السوائل الثابتة (*Static Fluid*).[٢٠] اختبر النسخة الأولى على ١١ طالبًا من الصف العاشر للعلوم، وبمراجعة نتائج الاختبار

---

[١٩] Bahrudin Fahmi, Syahrul Rizqi, and Nurul Elmira Harmeilinda, "Analisis Butir Soal Bahasa Arab MAS Pondok Pesantren Assalam Kampar Riau," *Ta'lim al-'Arabiyyah: Jurnal Pendidikan Bahasa Arab & Kebahasaaraban* 6, no. 1 (June 1, 2022): 95–105, https://doi.org/10.15575/jpba.v6i1.16193.

[٢٠] Popon Komala, Ida Kaniawati, and Ridwan Efendi, "KARAKTERISASI SOAL TES KETERAMPILAN BERPIKIR KRITIS MENGGUNAKAN ANALISIS ITEM RESPONSE THEORY PADA MATERI FLUIDA STATIS," *WaPFi (Wahana Pendidikan Fisika)* 5, no. 2 (September 11, 2020): 100–109, https://doi.org/10.17509/wapfi.v5i2.7755.

وعمل تعديلات، اختبرنا الأسئلة على فئات مختلفة من الصفوف العاشرة والحادي عشر للعلوم. أظهرت النتائج أن ١٥ من أصل ١٧ سؤالًا صادقة لفئات ذوي القدرات المتوسطة والعالية، وتمثل ١٣ سؤالًا مؤشراً قويًا للقدرات المتوسطة، و٤ أسئلة تعكس القدرات العالية. بناءً على ذلك، يمكننا القول إن الأسئلة التي أعددناها قادرة على قياس مهارات التفكير النقدي في موضوع السوائل الثابتة لفئات ذوي القدرات المتوسطة والعالية.

٤. هدف البحث الرابع إجراء تحليل جودة الاختبار صنعه المدرّس من خلال نموذج راش لنظرية استجابة البنود.[٢١] يأتي هذا البحث بطابع كمّي ووصفي، حيث يركز على دراسة استجابة ٦٧ طالباً للاختبار، الذي يتضمن خمس إجابات بديلة. تم اختيار الاختبارات من الفصل الدراسي النهائي للعام الأكاديمي ٢٠١٨/٢٠١٩، وتم تحليل البيانات باستخدام نهج راش IRT باستخدام برنامج QUEST. أظهرت النتائج أن ٢٨ بنداً من أصل ٣٠ يتوافق مع نموذج راش، وتم تصنيف مستويات الصعوبة إلى فئات مختلفة، حيث تراوحت بين صعوبة كبيرة (٢٥٪) وصعوبة (٢١,٤٪) ومعتدلة (٧,١٤٪) وسهولة (٤٦,٤٪)، دون وجود أسئلة سهلة جداً. تراوحت مستويات الصعوبة بين −٢,٩٤ و ٤,١٨. كما بلغت قيمة موثوقية تقدير البند ٠,٩٤ بفئة ممتازة، فيما بلغت قيمة موثوقية تقدير الحالة ٠,٣٨ بفئة ضعيفة.

٥. هدف البحث الخامس تحليل خصائص أسئلة امتحان الرياضيات للمدارس المتوسطة والمدارس المتوسطة الإسلامية بناءً على نهج نظرية الاختبار التقليدية.[٢٢] أظهرت النتائج أن متوسط صعوبة نبد المجموعة العالية كان متوسطًا، وصعوبة البند في المجموعة المنخفضة كانت عالية. بالإضافة، كان مؤشر درجة التمييز للمجموعتين غير مرضٍ، مما يشير إلى عدم قدرة البند على تمييز بين قدرات الطلاب المتفاوتة. وربما تكون الأسباب تعود إلى استخدام التخمين من قبل

[٢١] Pratama, "Analisis Kualitas Tes Buatan Guru Melalui Pendekatan Item Response Theory (IRT) Model Rasch."

[٢٢] Purwati et al., "Analisis Karakteristik Butir Soal Ujian Nasional Matematika SMP/MTs Berdasarkan Pendekatan Teori Tes Klasik."

المشاركين في الإجابات، مما أثر على الموضوعية. أما بخصوص موثوقية البنود، فقد كانت قيم المؤشر قريبة من +١,٠٠، مما يشير إلى وجود موثوقية عالية.

٦. هدف البحث السادس تحليل نظرية الاختبار التقليدية (CTT) لأداة اختبار القدرات الأكاديمية.[٢٣] تبيَّن أنَّ مستوى الصعوبة في مختلف مجالات المشكلات لم يكن متوازناً بين الأسئلة السهلة والمتوسطة والصعبة. وتمَّ توزيع الأسئلة المتوسطة بنسبة ٢٥% من الأسئلة السهلة، و ٥١% من الأسئلة المتوسطة، و ٢٤% من الأسئلة الصعبة. كما بيَّن مؤشر قوة التمييز وجود ٣٧% من الأسئلة التي تميز بين قدرات الممتحنين، و ١٨% منها لا تميز. فيما يتعلق بالخيارات الخادعة، فقد كانت ٧٩,٦% منها فعَّالة، و ٢٠,٣% غير فعَّالة. أما تحليل الصدق من ١٠٠ سؤال، فأظهر أن ٥٤% منها لديها الصدق أكبر من ١,٥٩، في حين أن ٤٦% لديها قيمة أقل من ذلك. وبالنسبة لمعامل الثبات الكلي، فقد تم قياسه بقيمة ٠,٨١، مما يعني أنَّ ٨٠% من اختلافات درجات الممتحنين مع الآخرين تعود إلى قدراتهم الفعلية ولا تتأثر بعوامل أخرى كمصدر للخطأ في القياس.

٧. هدف البحث السابع تحليل أسئلة تقييم نهاية الفصل لمادة اللغة العربية باستخدام نظرية استجابة البنود.[٢٤] تمت الدراسة على ١٧٦ طالبًا من الصف الحادي عشر، حيث تم جمع البيانات من ٤٠ سؤالاً من نوع الاختيارات المتعددة. أظهرت النتائج صحة ٤٠ سؤالًا (١٠٠٪) من حيث الصحة الاستدلالية مع وجود عوامل تحميل، وكانت موثوقية الاختبار ٠,٨٨٤، وتبين أن ٣٣ سؤالًا (٨٢,٥٪) مناسبة للحفظ في بنك الأسئلة، بينما ٧ سؤالًا (١٧,٥٪) بحاجة للمراجعة. الاختبار مناسب للاستخدام من قبل الطلاب ذوي القدرات المتوسطة إلى المنخفضة، وذلك ضمن النطاق من −٣,٥ إلى +١,٥ على مقياس اللوغاريتم.

---

[٢٣] Pratama, "ANALYSIS OF CLASICAL TEST THEORY (CTT) APPROCH ON ACADEMIC ABILITY TEST INSTRUMENT."

[٢٤] Danni, Wahyuni, and Tauratiya, "ITEM RESPONSE THEORY APPROACH."

بعد استعراض الأبحاث التي تم تلخيصها أعلاه، يمكن استخلاص بعض النقاط الرئيسية. التشابه بين هذا البحث والبحوث السابقة يكمن في تركيز التحليل الكمي لبنود الاختبار باستخدام نهجين، ألا وهما نظرية الاختبار التقليدية ونظرية استجابة البنود. يتماشى ذلك مع البحوث السابقة التي اعتمدت أيضًا هاتين النظريتين، كما في البحوث السابقة رقم ٢، ٥، و ٦ لنظرية الاختبار التقليدية ، ورقم ١، ٣، ٤، و ٧ لنظرية استجابة البنود. على الرغم من ذلك، تكمن الفارق الرئيسي في مجال البحث، حيث يقوم هذا البحث بتحليل بنود اختبار مادة اللغة العربية بشكل خاص. على النقيض، تناولت معظم البحوث السابقة تحليل بنود الاختبار لمواد دراسية أخرى غير اللغة العربية. علاوة على ذلك، لم تستخدم أي من البحوث السابقة كليهما نهجي التحليل الكمي لبنود الاختبار في آن واحد. ولذلك، يمكن استنتاج أن هذا البحث يقدم إسهامًا جديدًا ويستحق أن يكون موضوع دراسة مستقبلية.

الفصل الثاني
الإطار النظري

المبحث الأول: الاختبار

أ. تعريف الاختبار

المصطلح "اختبار" ينحدر من اللغة اللاتينية "تيستوم" والذي يشير إلى صحن أو وعاء من الطين. في سياق علم النفس، يستخدم مصطلح "اختبار" لوصف طريقة تحقيق فرد. يمكن أيضًا تعريف الاختبار كأداة أو إجراء يُستخدم لاستخلاص أو قياس شيء ما في سياق معين، باستخدام أساليب وقواعد محددة مسبقًا.[٢٥] يمكن أن يكون الاختبار عبارة عن أسئلة، ولذلك يجب أن تستوفي أنواع الأسئلة وصياغة الأسئلة وخيارات الإجابة معايير صارمة.[٢٦] الاختبار هو أحد الأدوات المستخدمة لإجراء قياسات وجمع معلومات حول خصائص كائن ما، بهدف جمع البيانات المتعلقة به.[٢٧] في تعريف آخر، يمكن أيضًا تفسير الاختبار على أنه سلسلة من الأسئلة التي يجب الإجابة عليها لقياس مستوى قدرات الفرد أو الكشف عن جانب محدد للشخص المختبر.[٢٨] لذا، في سياق هذا البحث، يُعرف الاختبار على أنه أداة قياس تحتوي على سلسلة من المهام أو الأسئلة التي يجب على الطلاب أن يقوموا بحلها لقياس مستوى قدراتهم في فهم واستيعاب المواد التعليمية التي تم تقديمها.

ب.أنواع الاختبار

هناك مجموعة واسعة من التنوعات وأنواع الاختبارات المستخدمة في ميدان التعليم. في سياق هذا البحث، سيتم توضيح تقسيم أنواع الاختبارات بناءً على المعايير ذات الصلة مع نطاق هذا البحث، وهي:

---

[٢٥] A. Muri Yusuf, *Asesmen Dan Evaluasi Pendidikan* (Jakarta: Prenada Media, 2017).

[٢٦] Zainal Arifin, *Evaluasi Pembelajaran* (Jakarta: Direktorat Jenderal Pendidikan Islam Kementerian Agama RI, 2012).

[٢٧] Djemari Mardapi, *Pengukuran Penilaian Dan Evaluasi Pendidikan* (Yogyakarta: Nuha Medika, 2012).

[٢٨] S. E. P. Widoyoko, *Teknik Penyusunan Instrumen Penelitian* (Yogyakarta: Pustaka Pelajar, 2012), https://www.semanticscholar.org/paper/Teknik-Penyusunan-Instrumen-Penelitian-Widoyoko/3732ac3bd6dde089e841c7353f71c2f44bc0a702.

بناءً على كيفية إعدادها، يمكن تقسيم الاختبارات إلى نوعين،[٢٩] وهما:

١. اختبار صناعة المعلم: الاختبار الذي يتم إعداده بواسطة المعلم نفسه الذي سيستخدم هذا الاختبار. يتم استخدام هذا النوع من الاختبارات عادة للاختبارات اليومية، والاختبارات التكوينية، والاختبارات التكميلية. يتم تحضير اختبار المعلم لقياس مستوى استيعاب الطلاب للمواد التي تم تقديمها.

٢. اختبار المعيار: الاختبار المعياري أو الاختبار القياسي هو اختبار يتمتع بمستوى عالٍ من الصدق والثبات بناءً على التجارب التي أجريت على عينة كبيرة وممثلة بشكل كاف. يتم تقييم الاختبار المعياري بشكل متكرر لمجموعة كبيرة من الطلاب، وعناصره ذات صلة ولها قوة تمييز عالية. بالإضافة إلى ذلك، تم تصنيف الاختبار المعياري حسب فئات العمر والصفوف الدراسية. عادةً ما يتم تحليل الاختبار المعياري إحصائياً واختباره تجريبيًا من قبل الخبراء، لذا يمكن القول إنه صالح (صحيح) للاستخدام بشكل عام.

بناءً على تنسيقها، ينقسم الاختبار إلى نوعين، وهما:

١. الاختبار الهدف: الاختبار الهدف غالبًا ما يُشار إليه أيضًا باسم اختبار الثنائية (dichotomously scored item) لأن إجاباته تكون إما صحيحة أو خاطئة وتسجل بين ١ أو ٠. يُسمى الاختبار الهدف لأن تقييمه موضوعي. أي شخص يقوم بتصحيح إجابات الاختبار الهدف سيحصل على نفس النتيجة لأن إجابات المفتاح واضحة وثابتة. يتطلب الاختبار الهدف من الطلاب اختيار الإجابة الصحيحة من بين الإجابات المقدمة، وتقديم إجابات قصيرة، واستكمال الأسئلة أو البيانات غير المكتملة. الاختبار الهدف مناسب جدًا لتقييم القدرات التي لا تتطلب عمليات عقلية مرتفعة، مثل التذكر، والتعرف، والفهم، وتطبيق المبادئ. يتكون الاختبار الهدف من عدة أشكال، مثل صح أو خطأ، اختيار من متعدد، ربط، واستكمال أو إجابة قصيرة.

---

[٢٩] Arifin, *Evaluasi Pembelajaran*.

٢. الاختبار الوصفي: يُستخدم الاختبار بشكل وصفي لقياس الأنشطة التعلم التي يصعب قياسها بواسطة الاختبار الهدف. يُسمى الاختبار بصيغة الوصف لأنه يتطلب من الطلاب وصف وتنظيم وعرض الإجابات بكلماتهم الخاصة بأشكال وتقنيات وأساليب مختلفة تختلف واحدة عن الأخرى. يُشار أحيانًا إلى الوصف بصفة فردية، لأنه في تنفيذه يتأثر غالبًا بعوامل الفردية للمعلم. من حيث الأسئلة التي يتم طرحها، يمكن تقسيم اختبار الوصف إلى نوعين، وهما وصف محدود (restricted respons items) ووصف موسع (extended respons items).

ج. اختبار اللغة

يُفهم 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. لذلك، يجب على متعلمي اللغة أن يكتسبوا فهمًا واسعًا للمفردات وإتقان قواعد اللغة بشكل كافٍ ليكونوا قادرين على التواصل بسلاسة في اللغة التي يدرسونها.[٣٠]

التواصل باللغة العربية ليس مجرد معرفة كيفية بناء الجمل أو اختيار الكلمات بشكل صحيح، بل يتعلق أيضًا 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

---

[٣٠] Muhammad Jundi, "تصميم برنامج أندرويد لتعزيز إتقان المفردات لدى طلاب المدرسة العالية *Lahjah Arabiyah: Jurnal Bahasa Arab Dan Pendidikan Bahasa Arab* 4, no. 2 (July 31, 2023): 140–58, https://doi.org/10.35316/lahjah.v4i2.140-158; Muhammad Jundi, "Persepsi Mahasiswa terhadap Pelaksanaan Perkuliahan Naḥwu dengan Sistem Daring," *Al-Ma'rifah: Jurnal Budaya, Bahasa, dan Sastra Arab* 17, no. 2 (October 31, 2020): 105–18, https://doi.org/10.21009/almakrifah.17.02.01; Muhammad Jundi, Yuslin Kasan, and Suleman D. Kadir, "Maf'ul Mutlaq Bahasa Arab Dan Cognate Object Bahasa Inggris: Analisis Kontrastif Dan Sintaksis Pada Juz 'Amma," *Lisanan Arabiya: Jurnal Pendidikan Bahasa Arab* 5, no. 2 (December 31, 2021): 153–78, https://doi.org/10.32699/liar.v5i2.2190.

العملية على استخدامها.[٣١] هذه القدرة العملية تمكن الشخص من استخدام اللغة في سياقات حقيقية، سواء كان ذلك شفهيًا أو كتابيًا.[٣٢] وبالتالي، يشمل اختبار اللغة اختبار القدرة النظرية والقدرة العملية على استخدامها في التواصل.[٣٣]

إن قدرة الفرد على استخدام اللغة العربية بشكل تقليدي يمكن تقسيمها إلى أربع مهارات رئيسية. 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 شفهياً. وأخيرًا، مهارة الكتابة، والتي تركز على القدرة على التعبير عن النفس بالكتابة. غالبًا ما يتم تصميم اختبارات اللغة العربية لقياس قدرة الفرد على استخدام اللغة العربية بشكل عام. ولذلك، يمكن أن تتكون الاختبارات من عدة أنواع تشمل اختبارات الاستماع واختبارات القراءة واختبارات الكلام واختبارات الكتابة، حسب تفاصيل الهدف من كل اختبار.[٣٤]

كما هو الحال مع الاختبارات أو الامتحانات بشكل عام، يجب أن تستند اختبارات اللغة أيضًا إلى أهداف التعلم المحددة. يتم استخدام الاختبارات في التقييم أو التقويم في سياق التعليم بهدف تقييم إلى أي مدى حقق المتعلمون الأهداف المحددة خلال عملية التعلم. ولذا، يجب أن يتم تحديد مؤشرات تحقيق التعلم بعناية فائقة وبدقة، نظرًا لأن لهذه المؤشرات تأثيراً كبيرًا على إعداد تصميم الأسئلة.[٣٥] كذلك، ينبغي أن يتم مراعاة الخطة

---

[٣١] Muhammad Jundi and Nabila Nabila, "Transformational-Generative Theory Perspective in Teaching The Four Language Skills of Arabic," *Al-Fusha : Arabic Language Education Journal* 5, no. 2 (June 30, 2023): 72–81, https://doi.org/10.36835/alfusha.v5i2.1193.

[٣٢] Muhammad Jundi and Yuslin Kasan, "GAYA DAN MAKNA BAHASA TULISAN: KAJIAN DESKRIPTIF CHAT MAHASISWA KEPADA DOSEN," *MARDIBASA: Jurnal Pembelajaran Bahasa Dan Sastra Indonesia* 1, no. 2 (September 2021), https://doi.org/10.21274/jpbsi.2021.1.2.290-315.

[٣٣] Rina Suci Andriani, "LANGUAGE TESTS," *Jurnal Educate* 3, no. 2, accessed September 1, 2023, https://media.neliti.com/media/publications/241552-tes-kebahasan-9bda9d9f.pdf.

[٣٤] Ibnu Rawandhy Hula, "EVALUASI DAN TES BAHASA ARAB: Tinjauan Teori," *Language Development Center* 12, no. 7 (December 9, 2021), https://doi.org/10.31219/osf.io/uwt8x.

[٣٥] Muhammad Jundi and Najamudin Petta Solong, "Analisis Kesesuaian Indikator Dan Kompetensi Dasar Bahasa Arab KMA 183," *Tatsqifiy: Jurnal Pendidikan Bahasa Arab* 2, no. 1 (January 28, 2021): 61–70, https://doi.org/10.30997/tjpba.v2i1.3642; Muhammad Zikran Adam et al., "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, no. 1 (July 2023); Siti Aliyya Laubaha Laubaha et al., "Kesesuaian Materi Berdasarkan Teori Dan

الرئيسية للأسئلة، والتي تعتبر دليلًا لتطوير بنود الاختبار، بعناية لضمان صحة وارتباطها الفعّال.

المبحث الثاني: تحليل بنود الأسئلة

أ. بنود الأسئلة

الاختبار وبنود الأسئلة لهما علاقة وثيقة. فالبند هو المكون الأساسي في الاختبار، حيث تشكل كل بند سؤالًا أو مهمة يجب على المشاركين في الاختبار الإجابة عليها. التوازن المناسب لبنود الأسئلة يمكن أن يسمح بقياس شامل وموضوعي لمهارات ومعرفة وفهم المشاركين في الاختبار. جودة وخصائص بنود الأسئلة تؤثر في صحة وثقة الاختبار بشكل عام، وبالتالي فإن التصميم واختيار بنود الأسئلة المناسبة يعدان أمرًا بالغ الأهمية في تطوير اختبار فعال وموثوق به. والأسئلة لها أنواع مختلفة ومتنوعة. يمكن أن تشمل هذه الأنواع مختلف الأشكال والأهداف في القياس. بعض أمثلة أنواع الأسئلة تتضمن الاختيار من متعدد، الإكمال القصير، الإنشاء، تحليل الحالة، حل المشكلات، وغيرها. لكل نوع من الأسئلة خصائصه وفائدته الخاصة في قياس الفهم والمهارات وقدرات الفرد.[٣٦]

في سياق هذا البحث، النوع المقصود من الأسئلة هو الأسئلة ذات الاختيارات المتعددة. الأسئلة ذات الاختيارات المتعددة هي أحد أشكال الاختبار التي تنطوي على الاستجابة المختارة وتستخدم على نطاق واسع لأغراض متنوعة.[٣٧] استخدام الأسئلة ذات الخيارات المتعددة على نطاق واسع يرجع إلى فاعليتها في قياس مختلف أنواع المعرفة والنتائج التعليمية المعقدة.[٣٨] أحد مزايا الأسئلة ذات الخيارات المتعددة هو ملائمتها للامتحانات

Praktek Pada Bahan Ajar Bahasa Arab," *Naskhi: Jurnal Kajian Pendidikan Dan Bahasa Arab* 4, no. 2 (October 24, 2022): 10–21, https://doi.org/10.47435/naskhi.v4i2.1198.

[٣٦] Arifin, *Evaluasi Pembelajaran*.

[٣٧] Rizki Nor Amelia and Kriswantoro Kriswantoro, "Implementation of Item Response Theory for Analysis of Test Items Quality and Students' Ability in Chemistry," *JKPK (Jurnal Kimia dan Pendidikan Kimia)* 2, no. 1 (May 1, 2017): 1, https://doi.org/10.20961/jkpk.v2i1.8512.

[٣٨] M.D. Miller, R.L. Linn, and N.E. Grondlund, *Measurement and Assessment in Teaching*, 10th ed. (New Jersey: Pearson Education, 2009).

التي تشارك فيها عدد كبير من المشاركين وتحتاج إلى إعلان نتائج سريعة. بالإضافة إلى ذلك، يزداد معدل الموثوقية نظرًا للعدد الكبير من البنود في الاختبار.[٣٩]

ومع ذلك، توجد بعض العيوب في الأسئلة ذات الخيارات المتعددة. أولاً، لا يحظى الطلاب بالحرية في الكتابة وتنظيم أفكارهم والتعبير عنها بكلماتهم وجملهم الخاصة. ثانيًا، فهي غير مناسبة لقياس قدرات حل المشكلات. ثالثًا، فهي تتسم بارتفاع احتمالية التخمين. رابعًا، يتطلب تطوير اختبار جيد وقتًا أطول مقارنةً بأنواع أخرى من الاختبارات. وأخيرًا، من الصعب تحديد خيارات الإجابة التي تكون حقاً متسقة ومنطقية وتعمل كمشتتات فعالة.[٤٠]

ب. معايرة بنود الأسئلة

تعتبر معايرة البنود هي عملية تقدير معاملات البند بناءً على نموذج نظرية استجابة البنود.[٤١] في إطار نظرية استجابة البنود، تتضمن المعايرة تقدير معاملات نظرية استجابة البنود، والتي تتألف من معاملات البند ومعاملات الأفراد.[٤٢] يهدف الغرض من المعايرة هو تقدير معاملات البند ومراقبة قدرة البند على التمييز بين مستويات السمة الكامنة.[٤٣] في سياق تحليل بنود الاختبار، فإن نهج نظرية الاستجابة للبند في معايرة البنود لديه فوائد محتملة.[٤٤] في البداية، ليست لدى جميع البنود معاملات، لذا فإن المعايرة ضرورية للحصول

---

[٣٩] Sumarna Surapranata, *Panduan Penulisan Tes Tertulis (Penilaian Berbasis Kelas)* (Bandung: Remaja Rosdakarya, 2005).

[٤٠] Surapranata.

[٤١] J. Umar, G. N. Masters, and J.P. Keeves, "Item Banking," in *Advances in Measurement in Educational Research and Assessment* (Oxford: Elsevier Science Ltd., 1999), 207–18; Wendy Yen and Anne Fitzpatrick, "Item Response Theory," 2006; Robert L. Brennan, *Educational Measurement* (Praeger Publishers, 2006).

[٤٢] American Educational Research Association, National Council on Measurement in Educatio, and American Psychological Association, *Standards for Educational and Psychological Testing* (Washington DC: American Educational Research Association, 2014), https://www.testingstandards.net/uploads/7/6/6/4/76643089/9780935302356.pdf.

[٤٣] Craig S. Wells, Michael J. Subkoviak, and Ronald C. Serlin, "The Effect of Item Parameter Drift on Examinee Ability Estimates," *Applied Psychological Measurement* 26 (2002): 77–87, https://doi.org/10.1177/0146621602026001005.

[٤٤] Dina Huriaty, "METODE KALIBRASI DAN DESAIN TES BERDASARKAN TEORI RESPONS BUTIR (IRT)," *Math Didactic: Jurnal Pendidikan Matematika* 1, no. 3 (2015).

على المعاملات المستخدمة في النظام. تشمل معايرة البنود تحديد معاملات مثل مستوى الصعوبة وقوة التمييز والتخمين.

ج. تحليل بنود الأسئلة

تحليل بنود الأسئلة هو طريقة تستخدم لتحقق صحة بنود الاختبار. تحليل بنود الأسئلة هو عملية هامة جدًا في تقييم جودة الاختبارات والامتحانات. الهدف من هذا التحليل هو تقييم مدى توافق أسئلة الاختبار مع الأهداف التعليمية المراد تحقيقها تقييم تأثير كل بند على الاختبار.[٤٥] يمكن تعديل أو إزالة البنود التي يتبين أنها تحتوي على أخطاء أو غير فعالة.[٤٦] هناك عدة تقنيات يمكن استخدامها في تحليل أسئلة الاختبار، مثل التحليل النوعي والكمي، بالإضافة إلى استخدام برامج الكمبيوتر لمعالجة البيانات.[٤٧] 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.[٤٨] وأخيرًا، 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 التقويم. ويمكن للمدرسين تحسين جودة الاختبارات وزيادة فعالية التعليم باستخدام تحليل أسئلة الاختبار[٤٩]. بينما يمكن لكُتّاب الاختبارات تقييم بنود الأسئلة المجراة وتحسين تصميم

---

[٤٥] Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan* (Jakarta: Bumi Aksara, 2009)

[٤٦] Amrita Kumari and S B Bhattacharya, "Item Analysis of Diagnostic Test in English Language Skills of Secondary School Students," n.d.

[٤٧] Farida, Syarifah, and Muhamad, "تحليل بنود الأسئلة في الإمتحان النهائي لمادة اللغة العربية في الفصل العاشر بالمدرسة العالية 'الإسلام' كموجا".

[٤٨] Surapranata, *Analisis Validitas, Reliabilitas, Dan Interpretasi Hasil Tes: Implementasi Kurikulum 2004*; Pratama, "Analisis Kualitas Tes Buatan Guru Melalui Pendekatan Item Response Theory (IRT) Model Rasch."

[٤٩] Arifin, *Evaluasi Pembelajaran*.

الاختبار في المستقبل باستخدام تحليل بنود الأسئلة. بشكل عام، يعتبر تحليل بنود الاختبار أمرًا مهمًا جدًا لمساعدة المدرسين وكتاب الاختبارات على تطوير اختبارات فعالة ودقيقة في قياس فهم الطلاب للمواد التي يتعلمونها.

المبحث الثالث: نظرية الاختبار التقليدية (Classical Test Theory)

أ. تعريف نظرية الاختبار التقليدية (Classical Test Theory)

واحدة من أقدم نظريات القياس في قياس السلوك هي نظرية الدرجة الصحيحة التقليدية. تعتبر نظرية الاختبار التقليدية سهلة التطبيق وتوفر نموذجًا مفيدًا لشرح كيف يمكن أن تؤثر أخطاء القياس على الدرجات الملاحظة. تؤكد نظرية الاختبار التقليدية على الدرجات النقية المستمدة من الاختبار، والتي تعكس قدرة الفرد. من خلال استخدام هذه الدرجات النقية، يمكن إجراء تحليلات وتفسيرات مختلفة وفقًا لاحتياجات البحث.[٥٠] يشار في كثير من الأحيان إلى نموذج الاختبار التقليدي بأنه "نموذج ضعيف" لأن افتراضات هذا النموذج يمكن تلبيتها بسهولة من خلال بيانات الاختبار. تتضمن هذه الافتراضات أن يكون الأداة لديها بعد واحد فقط، وأن تكون درجات خطأ القياس غير تفاعلية مع الدرجات الصحيحة، وأن تكون درجات خطأ القياس غير مترابطة مع الدرجات الصحيحة، وأن تكون درجات خطأ القياس غير مترابطة بين الاختبارات المختلفة لنفس المشاركين. بالإضافة إلى ذلك، يعتبر متوسط درجة خطأ القياس مساويًا للصفر.[٥١]

نظرية الاختبار التقليدية مبنية على استخدام نموذج رياضي بسيط يربط بين درجة الملاحظة (X)، والدرجة الحقيقية (T)، ودرجة الخطأ (E).[٥٢]مدخل نظرية الاختبار التقليدية في تحليل بنود الأسئلة مستند إلى الأسس الفلسفية للإيجابية المنطقية التي تؤكد على استخدام المنهج العلمي في قياس النفس. تفترض نظرية الاختبار التقليدية أن كل فرد

---

[٥٠] Bambang Sumintono and Wahyu Widhiarso, *Aplikasi Model Rasch Untuk Penelitian Ilmu-Ilmu Sosial (Edisi Revisi)*, 2014.

[٥١] Ronald K. Hambleton and Russell W. Jones, "An NCME Instructional Module on: Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test Development," *Educational Measurement: Issues and Practice* 12, no. 3 (October 25, 2005): 38–47, https://doi.org/10.1111/j.1745-3992.1993.tb00543.x.

[٥٢] Mardapi, *Pengukuran Penilaian Dan Evaluasi Pendidikan*, 2016.

لديه "نقطة حقيقية" تمثل القدرة الفعلية في بناء يتم قياسه. السمة الرئيسية لنظرية الاختبار التقليدية هي التركيز على تحليل الإحصائيات للأسئلة وتحديد جودة النود بناءً على معايير مثل مستوى الصعوبة والقوة التمييزية. يعتبر هذا النهج أن البنود الجيدة يجب أن تكون لديها الصدق، والثبات، ومستوى االصعوبة المناسبة أي ليس سهلا جدا وليس صعبا جدا على طلبة، ودرجة التمييز عالية أي أنها قادرة على التفريق بين الأفراد ذوي المستويات المرتفعة والمنخفضة من القدرة.

ب. تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية

١. الصدق

الصدق هو مقياس لمدى قدرة الاختبار أو أداة القياس على قياس ما يجب أن يتم قياسه.[٥٣] يعد قياس الصدق أمرًا هامًا 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 معامل الصدق. من خلال استخدام هذه التقنية، يمكننا قياس مدى ارتباط الاختبار بالبناء المقاس وضمان أن الاختبار يقدم نتائج قابلة للإعتماد وذات مغزى.[٥٤]

$$r = \frac{n(\sum XY) - (\sum X)(\sum Y)}{\sqrt{(n\sum X^2 - (\sum X)^2)(n\sum Y^2 - (\sum Y)^2)}}$$

٢. الثبات

الثبات هو مفهوم يشير إلى مدى موثوقية أو دقة قياس أو اختبار في تقديم نتائج متسقة وثابتة.[٥٥] قياس ثبات فحوصات بنود الاختبار ضروري لأنه يقدم معلومات

[٥٣] Saifuddin Azwar, *Reliabilitas Dan Validitas*, 4th ed. (Yogyakarta: Pustaka Pelajar, 2014).
[٥٤] Arifin, *Evaluasi Pembelajaran*.
[٥٥] Azwar, *Reliabilitas Dan Validitas*.

حول مدى يمكن الاعتماد على بنودات الاختبار في قياس البنية المرغوبة. بعبارة أخرى، يشير ثبات بنودات الاختبار إلى مدى قدرة الاختبار على تقديم نتائج متسقة وقابلة للثقة. يمكن حساب ثبات بنودات الاختبار عن طريق حساب معامل الاتساق الداخلي. واحدة من الصيغ الأكثر شيوعًا لحساب معامل الاتساق الداخلي لاختبار هي صيغة ألفا كرونباخ.[٥٦] هذا المعامل يوفر إشارة حول مدى ترابط بنود الاختبار ويعكس مستوى الاتساق الداخلي للفحص.

$$\alpha = \frac{N}{N-1}\left(1 - \frac{\sum Vi}{Vt}\right)$$

٣. مستوى الصعوبة

في تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية، يُعتبر مستوى الصعوبة أحد المعايير المستخدمة لتقييم جودة بنود الأسئلة. يُقيس مستوى الصعوبة مدى صعوبة أو سهولة بنود الأسئلة على الطلاب المشاركين في الاختبار. يُعتمد قياس مستوى الصعوبة على نسبة الطلاب الذين يجيبون بشكل صحيح. إذا كان هناك الكثير من الطلاب الذين يمكنهم الإجابة بشكل صحيح، فإن بند السؤال يُصنف كسهل. وعلى العكس، إذا كان هناك عدد قليل من الطلاب الذين يمكنهم الإجابة بشكل صحيح، فإن بند السؤال يُعتبر صعبًا. يعد قياس مستوى الصعوبة أمرًا هامًا لضمان أن نطاق صعوبة بنود الأسئلة يتناسب مع احتياجات الاختبار ويميز بين الطلاب ذوي مستويات قدرات مختلفة. تُرمز درجة صعوبة بند السؤال بـ "p_i". إذا كانت قيمة "p_i" تقترب من الصفر، فإن السؤال يُعتبر صعبًا جدًا. بينما إذا كانت قيمة "p_i " تقترب من الواحد، فإن السؤال يُعتبر سهلاً جدًا، وبالتالي يحتاج إلى إلغاء أو مراجعة.[٥٧]

مستوى صعوبة البند، والتي يتم التعبير عنها بنسبة تتراوح بين الصفر والواحد، ويكون القيمة متمركزة حول $p_i = 0.50$، يعتبر لديها معايير متوسطة تفصل بين البنود

[٥٦] Heri Retnawati, "Reliabilitas Instrumen Penelitian" (Lokakarya, Workshop Pembelajaran Berbasis Riset dan Analisis Data, Bukittinggi, 2017).

[٥٧] Heri Retnawati, *Validitas, Reliabilitas, Dan Karakteristik Butir* (Yogyakarta: Parama Publishing, 2016).

السهلة والبنود الصعبة. يُعتبر البنود سهلة إذا كانت قيمة $p_i > 0.50$، ويُعتبر البنود صعبة إذا كانت قيمة $p_i < 0.50$.[٥٨] وهناك وجهات نظر أخرى تشير إلى أن مؤشر صعوبة البنود الجيد يتراوح عادة بين ٠,٣ و ٠,٧. في هذا النطاق، يتم الحصول على معلومات عن قدرات الطلاب بشكل أقصى. إذا كانت قيمة المؤشر $p_i < 0.30$، يعتبر البنود صعبة، إذا كانت تتراوح قيمة p_i بين ٠,٣٠ و ٠,٧٠، يُعتبر مستوى صعوبة البنود متوسطًا، وإذا كانت قيمة $p_i > 0.70$، يعتبر البنود سهلة.[٥٩]

$$p_i = \frac{\sum B}{N} \quad \text{.....}$$

p_i : مستوى الصعوبة

$\sum B$: عدد المشاركين في الاختبار الذين أجابوا على الأسئلة بشكل صحيح

N : عدد المشاركين في الاختبار

٤. درجة التمييز

في تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية، يُعتبر درجة التمييز معيارًا يُستخدم لقياس قدرة بند السؤال على تمييز بين الطلاب في الاختبار الذين لديهم مستوى قدرات عالي والطلاب الذين لديهم مستوى قدرات منخفض. يُعتمد قياس درجة التمييز على اختلاف نسبة الطلاب الذين يجيبون بشكل صحيح بين مجموعة الطلاب ذوي المستوى العالي ومجموعة الطلاب ذوي المستوى المنخفض. إذا كان لدى بند السؤال قدرة تمييزية عالية، فهذا يعني أنه يمكنه تمييز الطلاب بوضوح بين المستويات المختلفة للقدرات.

حساب القوة التمييزية هو قياس لمدى قدرة البنود على تمييز الطلاب الذين يتقنون الكفايات عن الطلاب الذين لا يتقنون الكفايات أو يتقنونها بشكل غير كافٍ، بناءً على معايير محددة. كلما ارتفع معامل القوة التمييزية للبنود ، زادت قدرتها على التمييز

---

[٥٨] Dali S. Naga, *Pengantar Teori Sekor Pada Pengukuran Pendidikan* (Jakarta: Gunadarma, 1992).

[٥٩] Mary J. Allen, *Introduction to Measurement Theory* (Monterey, Calif: Brooks/Cole Pub. Co., 1979).

بين الطلاب الذين يتقنون الكفايات والطلاب الذين يتقنونها بشكل غير كافٍ.[٦٠] يُعتبر مؤشر درجة التمييز جيدًا إذا كان لديه قيمة تساوي أو تفوق ٠,٣. مؤشر درجة التمييز للبنود ذات قيمة صغيرة يؤدي إلى عدم قدرتها على تمييز الطلاب ذوي القدرات العالية عن الطلاب ذوي القدرات المنخفضة.[٦١]

$$\mathrm{DB} = \frac{\sum TB}{\sum T} - \frac{\sum RB}{\sum R}$$

$\sum TB$: عدد المشاركين الذين أجابوا بشكل صحيح في مجموعة المشاركين ذوي الكفاءة العالية

$\sum RB$: عدد المشاركين الذين أجابوا بشكل صحيح في مجموعة المشاركين ذوي الكفاءة المنخفضة

$\sum T$: عدد المشاركين ذوي الكفاءة العالية

$\sum R$: عدد المشاركين ذوي الكفاءة المنخفضة

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

٥. فعالية الخيارات

أسئلة متعددة الاختيارات تحتوي على خيارات يمكن اختيارها من قِبَل المشاركين في الامتحان. يجب أن تكون لكل خيار جودة جيدة حتى لا يُمكن الإجابة عليه بسهولة فقط بالتخمين. لذلك، من المهم قياس فاعلية كل خيار في سؤال معين. الخطوة الأولى هي تحديد عدد المشاركين في الاختبار (N). ثم، حدد عدد العينات

[٦٠] Arifin, *Evaluasi Pembelajaran*.
[٦١] Retnawati, *Validitas, Reliabilitas, Dan Karakteristik Butir*.

(n) للمجموعة العليا والمجموعة السفلى باستخدام الصيغة ٢٧% × N. بعد ذلك، قم بإعداد جدول لاختبار فاعلية الخيارات واحسب عدد الإجابات التي تم اختيارها لكل خيار، سواء في المجموعة العليا أو المجموعة السفلى.

الخيارات	أ/a	ب/b	ج/c	د/d	هـ/e
العليا					
السفلى					

بعد ذلك، يتعين تحديد فعالية الخيارات استنادًا إلى المعايير التالية:[٦٢]

١) بالنسبة للخيار الصحيح للإجابة

- عدد المشاركين في المجموعة العليا والمجموعة السفلى يتراوح بين ٢٥% و ٧٥%. الصيغة المستخدمة لحساب ذلك هي:

$$\frac{\sum PKA + \sum PKB}{n1 + n2} \text{ x } 100\%$$

$\sum PKA$: محدد الخيار من المجموعة العليا

$\sum PKB$: محدد الخيار من المجموعة السفلى

$n1$: عدد العينات من المجموعة العليا (٢٧%)

$n2$: عدد العينات من المجموعة السفلى (٢٧%)

- يجب أن يكون عدد المشاركين في المجموعة العليا أكبر من عدد المشاركين في المجموعة السفلى.

٢) بالنسبة للخيارات الخادعة

- عدد الأشخاص الذين صوتوا في المجموعة العليا والمجموعة السفلى لا يقل عن:

$$25\% \text{ x } \frac{1}{2(\sum d)} \text{ x } (Ka + Kb)$$

[٦٢] Arifin, *Evaluasi Pembelajaran*.

d : عدد الخيارات التشتت

Ka : المجموعة العليا

Kb : المجموعة السفلى

- يجب أن يكون عدد المشاركين في المجموعة السفلى أكبر من عدد المشاركين في المجموعة العليا.

في أسئلة متعدد الاختيار توجد خيارات خادعة. في الأسئلة الجيدة، سيتم اختيار خيارات المشتتات بالتساوي من قبل الممتحنين الذين أجابوا بشكل غير صحيح. على العكس من ذلك، بالنسبة للبنود التي ليست جيدة بما فيه الكفاية، سيتم اختيار الخادعة بشكل غير متساو. يعتبر عامل التشتيت جيدًا إذا كان عدد الطلاب الذين يختارون الخادعة هو نفس العدد المثالي أو قريبًا منه. يتم حساب مؤشر المنتقص بالصيغة:[٦٣]

$$\text{IP} = \frac{\text{P}}{(\text{N}-\text{B})/(\text{n}-1)} \text{ x } 100\%$$

IP : مؤشر التشتيت

P : عدد المشاركين يختارون الإجابة التشتيت

N : عدد مشاركي الإختبار

B : عدد المشاركين يختارون الإجابة الصحيحة

n : عدد الخيارات

المبحث الرابع: نظرية استجابة البنود (Item Response Theory)

أ. تعريف نظرية استجابة البنود (Item Response Theory)

نظرية استجابة البنود هي أحد الأساليب المستخدمة لتقييم ملاءمة بنود الاختبار عن طريق مقارنة أداء البند المتوسط مع مستويات القدرة الجماعية المتوقعة المتنبأة بالنموذج.[٦٤]

---

[٦٣] Arifin.

[٦٤] Wim J. Van Der Linden and Ronald K. Hambleton, eds., *Handbook of Modern Item Response Theory* (New York, NY: Springer New York, 1997), https://doi.org/10.1007/978-1-4757-2691-6.

تعتبر نظرية استجابة البنود تقنية توزيع قوية مع ميزات جذابة مثل ثبات معاملات البند والقدرة. الهدف الأساسي من تطوير نظرية استجابة البنود هو التعامل مع الاعتمادية على مجموعة محددة من الممتحنين والبنود المختبرة التي توجد في نظرية الاختبار التقليدية.[٦٥]

نظرية استجابة البنود، أو نظرية الاختبار الحديثة، تم تطويرها استنادًا إلى مبدأين رئيسيين. أولاً، يمكن التنبؤ بقدرة الفرد من خلال نتائج اختباره. ثانيًا، يمكن التعبير عن العلاقة بين نتائج الاختبار والقدرة من خلال وظيفة تُعرف بمنحنى خصائص البند.[٦٦] من حيث صعوبة البند، يتضمن خاصية الثبات في هذه النظرية أن مستوى صعوبة البند يظل ثابتًا بغض النظر عن مستوى الكفاءة لدى الشخص الذي يجري الاختبار. تعتبر هذه الخاصية للثبات واحدة من مزايا نظرية استجابة البند التي لا تنطبق على الاختبارات التقليدية. علاوة على ذلك، تتطلب نظرية الاختبار الحديثة متطلبات أكثر صرامة فيما يتعلق بالافتراضات وحجم العينة المطلوبة لتحليلها.[٦٧]

يشكل أحد المكونات الحاسمة لنظرية استجابة البند احتمالية تقديم الممتحنين لإجابات صحيحة، حيث يتم ربط معاملات البند ومعاملات الممتحن من خلال وظيفة رياضية أو صيغة نموذج رياضي. في هذه الصيغة، يتم فهم احتمالية ممتحن الإجابة بشكل صحيح على البند على أنها وظيفة لوجستية تعتمد على الفرق بين المعاملات المدرجة في النموذج. نظرية استجابة البنود هي منهج بديل يمكن استخدامه في تحليل الاختبارات. يعتمد هذا النهج على استخدام نماذج احتمالية، مما يعني أن احتمالية إجابة الفرد على البند بشكل صحيح تعتمد على قدراته وخصائص البند نفسه. بعبارة أخرى، المشاركين في الاختبار

---

[٦٥] Tie Liang, Craig S. Wells, and Ronald K. Hambleton, "An Assessment of the Nonparametric Approach for Evaluating the Fit of Item Response Models," *Journal of Educational Measurement* 51, no. 1 (2014): 1–17.

[٦٦] Ronald K. Hambleton, Hariharan Swaminathan, and H. Jane Rogers, *Fundamentals of Item Response Theory*, Fundamentals of Item Response Theory (Thousand Oaks, CA, US: Sage Publications, Inc, 1991).

[٦٧] Muh Syahrul Sarea and Rosnia Ruslan, "KARAKTERISTIK BUTIR SOAL: CLASSICAL TEST THEORY VS ITEM RESPONSE THEORY?," *DIDAKTIKA : Jurnal Kependidikan* 13, no. 1 (August 15, 2019): 1–16, https://doi.org/10.30863/didaktika.v13i1.296.

الذين لديهم قدرات أعلى لديهم احتمالية أكبر للإجابة بشكل صحيح مقارنةً بالذين لديهم قدرات أقل.[٦٨]

هناك العديد من المزايا لنظرية استجابة البنود كما يلي. أولاً، النتائج المولدة تعكس قدرات المتسابقين في الاختبار ولا تعتمد على مستوى صعوبة الاختبار نفسه. وهذا يتيح تقييمًا أكثر دقة للقدرات الفردية عبر أنواع مختلفة من البنود. ثانيًا، يمكن استخدام نظرية استجابة البنود لتأسيس صلة أكثر تفصيلًا بين بنود الاختبار وقدرات المتسابقين فيه. ونتيجة لذلك، يصبح من الممكن تحديد البنود الأكثر صلة وفعالية لقياس القدرات الفردية. ثالثًا، إحدى مزايا نظرية استجابة البنود هي أنها لا تتطلب اختبارات متوازية لتحديد معاملات الموثوقية. هذا النهج يمكن أن يمكّن قياسًا أكثر مرونة وكفاءة في تقييم موثوقية الاختبار.[٦٩]

ب. تحليل بنود الأسئلة باستخدام نظرية استجابة البنود

١. تقدير الصدق

تُجرى تقدير الصدق في نظرية استجابة البنود من خلال حساب الصدق البنائي، والذي يمكن الحصول عليه من خلال طريقة تحليل العوامل الاستكشافية (EFA). تحليل العوامل هو تقنية إحصائية تُستخدم لتحديد أنماط العلاقات بين المتغيرات في أداة الاختبار. يساعد تحليل العوامل الاستكشافية في كشف البنية العامة لاستجابة البنود في أداة الاختبار، من خلال تحديد العوامل البارزة التي قد تكون موجودة.[٧٠]

تحليل العوامل له عدة افتراضات وشروط يجب تحقيقها في تطبيقه. أولاً، يمكن معرفة ما إذا كانت العينة كافية من خلال اختبار بارتليت وقياس كايزر ماير أولكين

---

[٦٨] Ronald K. Hambleton and H. Swaminathan, *Item Response Theory: Principles and Applications* (Springer Science & Business Media, 1985).

[٦٩] Asri Andayani, Purwanto, and Taufik Ramlan Ramalis, "Kajian Implementasi Teori Respon Butir Dalam Menganalisis Instrumen Tes Materi Fisika," *Prosiding Seminar Nasional Fisika (SiNaFi) 5.0* 1, no. 1 (2019).

[٧٠] Konsultan Data Penelitian & ArcGIS, "Mengenal Lebih Jauh Tentang Analisis Faktor Eksploratori," *Konsultan Analis Data Penelitian dan Peta Digital* (blog), August 24, 2020, https://patrastatistika.com/analisis-faktor-eksploratori/; Lu'luatul Chizanah and M. Noor Rochman Hadjam, "Validitas Konstruk Ikhlas: Analisis Faktor Eksploratori Terhadap Instrumen Skala Ikhlas," *Jurnal Psikologi UGM* 38, no. 2 (2011): 129023.

لملائمة العينة (KMO-MSA). يُستخدم اختبار بارتليت لاختبار فرضية عدم وجود ترابط بين المتغيرات في العينة. بينما يُستخدم KMO-MSA لقياس مدى توافق البيانات مع النموذج العاملي المقدر. بالإضافة إلى ذلك، يتم استخدام القيم الخاصة (Eigen values) ورسم الانحدار لتحديد عدد العوامل البارزة التي تتشكل في أداة الاختبار. تشير القيم الخاصة إلى مدى تفسير كل عامل للتباين في البيانات، في حين يساعد رسم الانحدار في تحديد نقطة انحدار ملحوظة في القيم الخاصة. علاوة على ذلك، تحتوي المصفوفة المكونة على عوامل التحميل (loading factor) التي توضح مدى تساهم المتغيرات في تشكيل تلك العوامل.[٧١]

٢. الافتراضات في نظرية استجابة البنود

هناك ثلاثة افتراضات أساسية في نظرية استجابة البنود: الأحادية البُعد، وعدم الترابط المحلي، وثبات المعاملة.[٧٢]

١) يشير افتراض الأحادية البُعد إلى أن كل بند اختبار يقيس سمة خفية واحدة فقط، مثل الاجتهاد في مادة الأخلاق الإسلامية في اختبار لقياس مهارة اللغة العربية. ومع ذلك، في الممارسة، يُعتبر هذا الافتراض تحديًا نظرًا للعوامل مثل القدرات الإدراكية وصفات الشخصية والتأثيرات البيئية والقلق. يمكن استخدام طرق تحليل العوامل لاختبار كفاية افتراض الأحادية البُعد.

٢) يفترض عدم الترابط المحلي عدم وجود علاقة بين استجابات الفرد لأسئلة الاختبار المختلفة. وبعبارة أخرى، لا يوجد ترابط بين استجابات الفرد لأسئلة الاختبار المختلفة. وهذا يشير إلى أن القدرة المشتملة في النموذج هي العامل الوحيد الذي يؤثر في استجابات الفرد لأسئلة الاختبار.

---

[٧١] Danni, Wahyuni, and Tauratiya, "ITEM RESPONSE THEORY APPROACH."

[٧٢] Hambleton, Swaminathan, and Rogers, *Fundamentals of Item Response Theory*; Heri Retnawati, *Teori Respons Butir Dan Penerapannya: Untuk Peneliti, Praktisi Pengukuran Dan Pengujian, Mahasiswa Pascasarjana* (Yogyakarta: Nuha Medika, 2014); R. Michael Furr, *Psychometrics: An Introduction*, 4th ed. (Thousand Oaks, California: SAGE Publications, 2021).

٣) يفترض ثبات المعاملة أن خصائص البند مستقلة عن توزيع معاملات الأفراد، ومعاملة الأفراد مستقلة عن خصائص البند. وبالتالي، فإن قدرة الفرد لن تتغير فقط بسبب إجراء اختبار بمستوى صعوبة مختلف، ولن تتغير خصائص البند ببساطة بوجود مجموعة من الأفراد ذوي مستويات مختلفة من القدرة.

٣. نماذج المعاملات اللوجستية في نظرية استجابة البنود

في نظرية استجابة البنود، ينطوي تحديد المعاملات لكل البند على استخدام نماذج لوجستية مناسبة. هناك ثلاث نماذج للمعاملات اللوجستية المختلفة: النموذج اللوجستي ذو المعامل الواحد لتقدير مستوى الصعوبة، والنموذج اللوجستي ذو المعاملين لتقدير مستوى الصعوبة والتمييز، والنموذج اللوجستي ذو المعاملات الثلاثة لتقدير مستوى الصعوبة والتمييز والتخمين الزائف. اختيار هذه النماذج ليس عشوائيًا ولكنه يتم من خلال اختبار جودة الملاءمة للبنود باستخدام النموذج اللوجستي المطلوب. تُقوم تحديد نموذج معاملات التوزيع اللوجستيكي على أساس توافق البنود مع النموذج. اختيار النموذج اللوجستي المناسب أمر بالغ الأهمية لتقدير المعاملات بدقة، حيث يعد تلاؤم العنصر أمرًا ضروريًا في التطبيق الصحيح لنظرية استجابة البنود.[٧٣]

نماذج	الدالة استجابة البند
1PL	$P(X_{is}=1\|\theta_s,\ \beta_i)=\frac{e^{(\theta_s-\beta_i)}}{1+e^{(\theta_s-\beta_i)}}$
2PL	$P(X_{is}=1\|\theta_s,\ \beta_i,\ \alpha_i)=\frac{e^{(\alpha_i(\theta_s-\beta_i))}}{1+e^{(\alpha_i(\theta_s-\beta_i))}}$
3PL	$P(X_{is}=1\|\theta_s,\ \beta_i,\ \alpha_i,\ c_i)=c_i+(1-c_i)\frac{e^{(\alpha_i(\theta_s-\beta_i))}}{1+e^{(\alpha_i(\theta_s-\beta_i))}}$

٤. المعاملات في نظرية استجابة البنود

١) مستوى الصعوبة

مستوى صعوبة البند هو العامل الذي يؤثر على احتمالية شخص ما الإجابة على البند بشكل صحيح. على سبيل المثال، البند الرياضي الذي يحمل مستوى

[٧٣] Furr, *Psychometrics*.

صعوبة عالٍ سيكون أقل احتمالاً للإجابة عليه بشكل صحيح مقارنة بالبند الرياضي الذي يحمل مستوى صعوبة منخفض (أي البند السهل). على سبيل المثال، البند "ما هو الجذر التربيعي للعدد ١٠،٠٠٠؟" سيكون أقل احتمالاً للإجابة عليه بشكل صحيح مقارنة بالبند "ما هو ٢ + ٢؟".[٧٤]

على الرغم من أن مستوى السمة وصعوبة البند يتم تحليلهما بشكل منفصل في تحليل الاستجابة للبنود، إلا أنهما مترابطان بشكل أساسي. يتم فهم صعوبة البند في سياق مستوى السمة المتعلقة. بشكل خاص، البنود الصعبة تتطلب مستوى سمة مرتفع نسبيًا للإجابة عليها بشكل صحيح، بينما البنود السهلة تتطلب مستوى سمة منخفض. على سبيل المثال، للإجابة على البند المتعلق بجذر تربيعي بشكل صحيح، قد يحتاج الطلاب إلى أن يكون لديهم قدرة رياضية معادلة لطلاب الصف التاسع. ومن ناحية أخرى، قد يحتاجون فقط إلى قدرة رياضية معادلة لطلاب الصف الثاني للإجابة على البند المتعلق بالجمع بشكل صحيح.

مستوى صعوبة البند في تحليل نظرية استجابة البنود يتم تعريفه على أنه المستوى الذي يلزم من الميزة الفردية للفرد ليكون لديه احتمال ٥٠:٥٠ للإجابة على البند بشكل صحيح. إذا كان مستوى صعوبة البند هو ٠، فإن الفرد ذو الميزة الفردية المتوسطة يكون لديه احتمال ٥٠:٥٠ للإجابة على البند بشكل صحيح. ومع ذلك، الأفراد ذوو الميزة الفردية الأعلى من ٠ سيكون لديهم احتمالية أعلى للإجابة على البند بشكل صحيح، بينما الأفراد ذوو الميزة الفردية الأقل من ٠ سيكون لديهم احتمالية أقل. بشكل عام، يشير مستوى صعوبة أعلى إلى ضرورة وجود مستوى ميزة فردية أعلى لدي الأفراد لديهم احتمال ٥٠:٥٠ للإجابة على البند بشكل صحيح. على سبيل المثال، إذا كان مستوى صعوبة البند هو ١,٥، فإن الفرد ذو الميزة الفردية ١,٥ (الذي يبعد ١,٥ انحراف معياري فوق المتوسط) لديه احتمال ٥٠:٥٠ للإجابة على البند بشكل صحيح. وعلى العكس من ذلك،

---

[٧٤] Furr.

يشير مستوى صعوبة أدنى إلى أن الأفراد ذوو الميزة الفردية النسبية المنخفضة يكونون قادرين بالفعل على الحصول على احتمال ٥٠:٥٠ في الإجابة على البند بشكل صحيح. في نماذج الرياضيات لتحليل الاستجابة للبنود التي سنناقشها، يُعبَر عن مستوى صعوبة البند بقيمة بيتا (βi).

٢) درجة التمييز

بعض البنود في الاختبار قد تختلف في درجة صعوبتها (بعض البنود أصعب من الأخرى)، ويمكن أن تختلف أيضًا في مدى قدرتها على تمييز الأفراد الذين يمتلكون مستوى مرتفع من الخاصية الفردية عن الأفراد الذين يمتلكون مستوى منخفض من الخاصية الفردية. يُطلق على هذه الخاصية في البنود اسم "تمييز البند"، وتشبه مرتبط البند الكلي في نظرية الاختبار التقليدية.[٧٥]

تشير درجة التمييز في نظرية استجابة البنود إلى مدى قدرة بنود في الاختبار على تمييز الأفراد الذين يمتلكون مستوى عالٍ من السمة الفردية عن الأفراد الذين يمتلكون مستوى منخفض من السمة الفردية. البنود ذات التمييز الجيد ترتبط بشكل قوي بالسمة المقاسة بواسطة الاختبار، بينما البنود ذات التمييز الضعيف ليست مرتبطة إلى حد كبير بهذه السمة. قيمة التمييز الإيجابية تشير إلى التسليم بين البند والسمة المرتبطة به، بينما القيمة السلبية للتمييز تشير إلى العلاقة المعكوسة بين البند والسمة المرتبطة به. في سياق اختبار الرياضيات، البنود ذات التمييز الجيد ستكون أكثر قدرة على تمييز الأفراد ذوي القدرة الرياضية العالية عن الأفراد ذوي القدرة الرياضية المنخفضة.

في تحليل نظرية استجابة البنود، يعد مستوى التمييز عاملًا مهمًا في تقييم جودة البنود في الاختبار. من خلال فهم مستوى التمييز، يمكننا معرفة إلى أي مدى يمكن للبنود في الاختبار أن تميز بين الأفراد بناءً على مستوى الصفات الفردية

[٧٥] Susan E. Embretson and Steven P. Reise, *Item Response Theory for Psychologists*, Item Response Theory for Psychologists (Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, 2000).

المقاسة. تعد البنود ذات مستوى التمييز العالي أفضل في تمييز الأفراد بمستويات الصفات المختلفة، بينما تكون البنود ذات مستوى التمييز المنخفض أقل فعالية في تمييز الأفراد. من خلال استخدام قيم التمييز، يمكننا تحديد البنود التي تحتاج إلى تحسين أو استبدالها لتحسين دقة وصحة الاختبار. بالتالي، يعد مستوى التمييز جانبًا هامًا في تطوير اختبارات أفضل وفهم إلى أي مدى يمكن للبنود في الاختبار أن تميز بين الأفراد بناءً على مستوى الصفات الفردية المختلفة. يُعبَر عن درجة التمييز بقيمة ألفا (αi).

٣) التخمين الزائف

التخمين الزائف مرتبط بخاصية العنصر الثالث التي قد تؤثر على استجابة المشاركين لبعض أنواع أسئلة الاختبار. في اختبارات مثل اختبارات الاختيار من متعدد أو اختبارات صحيح/خطأ في المعرفة أو القدرة، قد يلجأ بعض المشتركين في الاختبار إلى التخمين إذا كانوا لا يعرفون الإجابة الصحيحة لسؤال ما. عند إجراء مثل هذه الاختبارات، قد يجيب المشتركون على بعض الأسئلة بشكل صحيح فقط عن طريق التخمين والحظ.

لذا، يمكن لنموذج استجابة البنود أن يشمل مكون التخمين الزائف لمراعاة هذا الاحتمال، ويتعلق بالاحتمالية التي يجيب بها المشاركون على سؤال بشكل صحيح بناءً على الصدفة. على سبيل المثال، يأخذ مكون التخمين الزائف في الاعتبار حقيقة أن التخمين سينتج إجابة صحيحة بنسبة ٥٠٪ من الوقت في حالة الأسئلة الصحيح/خطأ. بالمثل، يعتمد مكون التخمين الزائف في حالة الأسئلة ذات الاختيارات المتعددة على عدد الخيارات المتاحة – إذا كان هناك أربعة خيارات، فإن مكون التخمين الزائف يأخذ في الاعتبار حقيقة أن التخمين سينتج إجابة صحيحة بنسبة ٢٥٪ من الوقت. في نماذج استجابة البنود الرياضية التي سنناقشها، يتم توضيح مكون التخمين الزائف لبنود ما بواسطة "ci".[٧٦]

[٧٦] Furr, *Psychometrics*.

الفصل الثالث
منهجية البحث

أ. مدخل البحث ومنهجه

تستخدم هذه الدراسة المنهج الكمّي الوصفي وتهدف إلى مقارنة تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية ونظرية استجابة البنود لدرس اللغة العربية بالمدرسة المتوسطة الإسلامية الحكومية باتو. يستخدم المنهج الكمي في هذه الدراسة نظراً لأن البيانات التي تم جمعها على شكل أرقام وتم تحليلها بأساليب الإحصائي. والنوع الذي يستخدم في هذه الدراسة هو الوصفي التحليلي. يستخدم النوع الوصفي التحليلي في تحليل البيانات بهدف الحصول على وصف دقيق وتفصيلي للظاهرة التي يريد دراستها، وهنا في هذه الدراسة يتعلق الأمر بجودة وخصائص أسئلة الاختبار النهائي لدرس اللغة العربية في المدرسة المتوسطة الإسلامية الحكومية باتو.[٧٧] باستخدام هذا النوع، يمكن للباحث تقديم وصف تفصيلي لجودة وخصائص أسئلة الاختبار النهائي لدرس اللغة العربية في المدرسة المتوسطة الإسلامية الحكومية باتو.

ب. مجتمع البحث وعينته

نظرًا لأن هذا البحث يهدف إلى تحليل نبود أسئلة، فإن مجتمع البحث هو الطلاب في المدرسة المتوسطة الإسلامية الحكومية باتو. والعينة البحثية هي طلاب الصف التاسع في المدرسة المتوسطة الإسلامية الحكومية باتو. في سعينا لتصوير السكان بشكل أفضل، يتم أخذ عينة بأكبر قدر ممكن. بالإضافة إلى ذلك، عند إجراء تحليل بنود أسئلة الاختبار باستخدام نظرية استجابة البنود، يتطلب ذلك عددًا محددًا من العينات يمكن تحديده فقط من خلال اختبار كفاية العينة. لذلك، يتم اتخاذ أكبر قدر ممكن من العينات كإجراء وقائي.

---

[٧٧] Sugiyono Sugiyono, *Metode Penelitian Pendidikan; Pendekatan Kuantitatif, Kualitatif, Dan R&D*, Cet. 23 (Bandung: Alfabeta, 2016).

ج. البيانات ومصادرها

مصدر بيانات البحث تتكون من المصدر الرئيسي والمصدر الفرعي وذلك مايلى[٧٨]:

١. المصدر الرئيسي: أسئلة الإختبار لدرس اللغة العربية ونتائج الطلاب في إجابة الأسئلة ومدرس اللغة العربية في المدرسة.

٢. المصدر الفرعي: الوثائق الداعمة الأخرى المتعلقة بالأسئلة ومواد التقييم مثل المنهاج الدراسي، وخطط تنفيذ التعليم، وشبكة أسئلة الاختبار. وأيضا عدد من الكتب والمقالة العلمية المتعلقة بالموضوع التي تساعد الباحث في عملية البحث.

د. أسلوب جمع البيانات

في هذ البحث، تم جمع البيانات باستخدام أسلوب الوثائق والمقابلة. قام الباحث بجمع البيانات من نتائج الطلاب في الإختبار، وكذلك شبكة السؤال والكفاءة الأساسية وأهداف التعليم. أما المقابلة التي أُجريت مع المدرّس للغة العربية للتحقق من نتائج تحليل الأسئلة التي تم إجراؤها.

هـ. أسلوب تحليل البيانات

١. تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية

تحليل البيانات باستخدام نظرية الاختبار التقليدية في هذا البحث يستخدم تطبيق Microsoft Excel وSPPS مع الخطوات التالية.

أ) ترتيب البيانات وإجراء الترقيم وإدخال البيانات.

ب)اختبار طبيعية البيانات باستخدام برنامج SPPS v.26

ج) حساب الصدق باستخدام الصيغة "المعامل الإنتاجي" والثبات باستخدام الصيغة "ألفا كرونباخ"

$$\alpha = \frac{N}{N-1}\left(1 - \frac{\sum Vi}{Vt}\right) \qquad r = \frac{n(\sum XY) - (\sum X)(\sum Y)}{\sqrt{(n\sum X^2 - (\sum X)^2)(n\sum Y^2 - (\sum Y)^2)}}$$

[٧٨] Sugiyono.

د) حساب مستوى صعوبة كل بنود من الأسئلة باستخدام الصيغة المحددة مسبقاً

$$p_i = \frac{\sum B}{N} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

p_i : مستوى الصعوبة

$\sum B$: عدد المشاركين في الاختبار الذين أجابوا على الأسئلة بشكل صحيح

N : عدد المشاركين في الاختبار

هـ) حساب درجة التمييز باستخدام الصيغة المحددة مسبقاً

$$\mathrm{DB} = \frac{\sum TB}{\sum T} - \frac{\sum RB}{\sum R}$$

$\sum TB$: عدد المشاركين الذين أجابوا صحيحا في المجموعة العالية

$\sum RB$: عدد المشاركين الذين أجابوا بشكل صحيحا في المجموعة المنخفضة

$\sum T$: عدد المشاركين ذوي الكفاءة العالية

$\sum R$: عدد المشاركين ذوي الكفاءة المنخفضة

و) تفسير النتائج باستخدام المعايير التالية.

مستوى الصعوبة	درجة التمييز
نسبة مؤوية الأسئلة السهلة والمتوسطة والصعبة هي ٢٧:٤٦:٢٧ %	درجة التمييز مرتفعة، أي أنها تكون أكبر من أو تساوي ٠,٣

٢. تحليل بنود الأسئلة باستخدام نظرية استجابة البنود

تحليل البيانات باستخدام نظرية استجابة البنود في هذا البحث يستخدم تطبيقات R و RStudio مع حزمة "irtawsi" و"ltm" وفيما يلي الخطوات.

أ) ترتيب البيانات وإجراء الترقيم باستخدام Microsoft Excel

ب)حساب الصدق باستخدام تحليل العوامل التفسيرية في SPSS v.26

ج) تحويل تنسيق البيانات إلى تنسيق csv أو xlsx أوtxt

د) تشغيل حزمة "irtawsi" و"ltm" علىRstudio

هـ) إدخال البيانات إلى الحزمة وتشغيل الكشف المبدئي

و) اختبار مطابقة النموذج واختيار النموذج الموصى به

$$\text{1PL} \quad : \quad P(X_{is}=1\big|\theta_s,\ \beta_i)=\frac{e^{(\theta_s-\beta_i)}}{1+e^{(\theta_s-\beta_i)}}$$

$$\text{2PL} \quad : \quad P(X_{is}=1\big|\theta_s,\ \beta_i,\ \alpha_i)=\frac{e^{(\alpha_i(\theta_s-\beta_i))}}{1+e^{(\alpha_i(\theta_s-\beta_i))}}$$

$$\text{3PL} \quad : \quad P(X_{is}=1\big|\theta_s,\ \beta_i,\ \alpha_i,\ c_i)=c_i+(1-c_i)\frac{e^{(\alpha_i(\theta_s-\beta_i))}}{1+e^{(\alpha_i(\theta_s-\beta_i))}}$$

ز) اختبار افتراضات نظرية استجابة البنود على البيانات. إذا كانت هناك بنود لا تفي بواحدة أو أكثر من افتراضات نظرية استجابة البنود، فيجب إلغاء هذه البنود، ثم يتم بدء عملية التحليل من جديد من مرحلة إدخال البيانات

ح) بعد ترتيب بيانات بنود الأسئلة مرة أخرى، يتم اختبار مطابقة النموذج مرة أخرى واختيار النموذج الموصى به

ط) اختبار افتراضات نظرية استجابة البنود على البيانات مرة أخرى

ي) بعد تحقيق كافة الافتراضات، يتم عرض جدول معاملات كل معامل لبنود السؤال. يتم عرض عدد معاملات بنود الأسئلة بناءً على النموذج المتوافق مع البيانات.

ك) تفسير النتائج باستخدام المعايير التالية.

نموذج	معايير		
	جيد جدا	جيد	ناقص
1PL	$-٢ \le b \le ٢$	-	$b < -٢ \,/\, b > ٢$
2PL	$٠ \le a \le ٢$ $-٢ \le b \le ٢$	لم يتم تحقيق أحد من المعايير	لم يتم تحقيق أكثر من واحد من المعايير
3PL	$٠ \le a \le ٢$ $-٢ \le b \le ٢$ $c \le ٠,٢٥$	لم يتم تحقيق أحد من المعايير	لم يتم تحقيق أكثر من واحد من المعايير

الفصل الرابع
عرض البيانات وتحليلها

الفصل الرابع سيقدم نتائج البحث مع تحليلها، والذي يتكون من مناقشتين رئيسيتين. المناقشة الأولى تتعلق بمقارنة تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية ونظرية استجابة البند في امتحان مادة اللغة العربية. أما المناقشة الثانية، تفسير نتائج تحليل بنود الأسئلة باستخدام النظريتين في درس اللغة العربية بالمدرسة المتوسطة الإسلامية الحكومية باتو، بالإضافة إلى تصنيفها ومعالجة عيوبها. ومع ذلك، سيتم تقديم نظرة عامة أولاً حول المدرسة المتوسط الإسلامية الحكومية باتو وتنفيذ امتحان المدرسة لدرس اللغة العربية في هذه المدرسة.

المدرسة المتوسطة الإسلامية الحكومية في مدينة باتو، جاوا الشرقية، هي المدرسة المتوسطة الإسلامية الحكومية الوحيدة التي توجد في المدينة. رؤية المدرسة هي تحقيق مدرسة البحث الدينية الذكية المتميزة والتنافسية وذات الرؤية البيئية، بشعار "المدرسة البحثية الدينية الذكية".[٧٩] وتتمثل مساهمتها الرئيسية في دعم تطوير الطلاب والمدرسين بناءً على البحث والتكنولوجيا والوعي البيئي. بالإضافة إلى تقديم تدريب منتظم للمدرسين لزيادة كفاءتهم في التدريس في الفصول الدراسية، تدعم المدرسة أيضًا أنشطة متنوعة للطلاب سواء كانت أكاديمية أم غير أكاديمية وفقًا لرؤيتها. واحدة من أشكال هذا الدعم هي برنامج البحث للطلاب الذين نجحوا في تصميم أبحاث بيئية متعددة. بالإضافة إلى ذلك، كمؤسسة متميزة، تفتح مدرسة المتوسط الإسلامية الحكومية باتو أبوابها للتعاون مع مختلف الأطراف الخارجية وترحب دائمًا بالطلاب الجامعيين الذين يرغبون في إجراء أبحاث أو تنفيذ أعمال عملية في المدرسة.

تُنفَّذ تعليم اللغة العربية في المدرسة المتوسطة الإسلامية الحكومية باتو بشكل متميز وبدعم قوي من قبل موارد المدرسين والمرافق المتاحة بشكل جيد. يوجد أربعة مدرسين لمادة

---

[٧٩] "Visi Dan Misi MTs. N. Kota Batu," accessed December 28, 2023, https://mtsnkotabatu.sch.id/halaman/detail/visi-dan-misi.

اللغة العربية في هذه المدرسة، جميعهم يتمتعون بمؤهلات تعليمية كافية، وخبرة، ومهارات تدريس استثنائية. يعملون بشكل تعاوني لرفع مستوى تعلم اللغة العربية في المدرسة. تم تطبيق استخدام مواد تعليمية سواء التقليدية أو المبنية على التكنولوجيا في الفصول الدراسية، وذلك بدعم من مرافق عصرية تتضمن جهاز العرض.[٨٠] بالإضافة إلى ذلك، تم توفير الكتب الأساسية اللازمة للطلاب.

فيما يتعلق بعملية التقويم، تمكنت المدرسة المتوسطة الإسلامية الحكومية باتو من الاستفادة من التكنولوجيا الحاسوبية في عمليات التقويم، بما في ذلك اختبارات اللغة العربية. تم تطوير أدوات التقويم من قبل المدرسين بناءً على مقاييس التقييم المحددة مسبقًا، مع الالتزام بالمنهاج وأهداف التعلم المحددة. هذا النهج يوفر مرونة كبيرة في تنفيذ عمليات التقويم ويضمن أن تتم وفقًا للمعايير المحددة. وبالتالي، تعكس عملية تعلم اللغة العربية في هذه المدرسة التزامها بتقديم تعليم عالي الجودة لطلابها.

بعد تقديم ملخص لملف المدرسة وتنفيذ عملية تعليم درس اللغة العربية في المدرسة بشكل عام، يتم الانتقال إلى جزء النتائج والمناقشة في هذا البحث وهي إجراءات تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية ونظرية استجابة البنود، والمقارنة بينهما، لاختبار درس اللغة العربية في المدرسة المتوسطة الإسلامية الحكومية باتو. سيتم تقسيم هذا الفصل إلى ثلاثة أجزاء، وهي:

أ. تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية (CTT)

تم تنفيذ تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية باستخدام تطبيق مايكروسوفت إكسل ٢٠١٣ (Microsoft Excel 2013). استناداً إلى بساطة حسابات الإحصاء في تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية، والتي يمكن أداؤها بسهولة نسبية باستخدام مايكروسوفت إكسل. إذا لزم الأمر، سيتم أيضاً النظر في

[٨٠] Ibadurrahman Ali et al., "Implementasi KMA 183 dalam Penyajian Materi Bahasa Arab Berbasis E-Learning di MAN 1 Kabupaten Gorontalo," *PIWULANG Jurnal Pendidikan Agama Islam* 4, no. 2 (2022).

استخدام تطبيقات إضافية لتيسير عملية التحليل في المستقبل. ومع ذلك، سيتم إجراء معظم التحليلات الأساسية باستخدام مايكروسوفت إكسل ٢٠١٣ باستخدام ميزات تحليل البيانات، بالإضافة إلى استخدام الصيغ اليدوية التي يمكن تشغيلها يدوياً.

بالإضافة إلى دورها كأداة لتحليل البيانات، تلعب مايكروسوفت إكسل أيضًا دورًا هامًا في تخزين بيانات هذا البحث. تم تنسيق البيانات المستخدمة في هذا البحث باستخدام مايكروسوفت إكسل، لذا يصبح استخدام هذا التطبيق أمرًا بالغ الأهمية. فإن استخدامه ضروري جدًا في عملية التحليل، سواء كان ذلك يتضمن نظرية الاختبار التقليدي أو نظرية استجابة البند. وبالتالي، يتم استخدام مايكروسوفت إكسل ليس فقط كأداة لتحليل البيانات، ولكن أيضًا كمنصة لتخزين البيانات التي تسهل إدارتها بكفاءة في البحث.

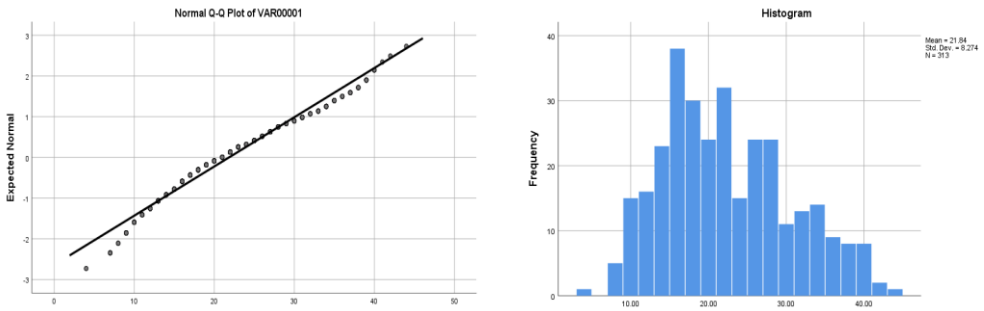
بالإضافة إلى استخدام برنامج مايكروسوفت إكسل ٢٠١٣، استفاد الباحث أيضًا من تطبيق SPSS في عملية البحث لأداء حسابات خاصة والتي لا يمكن تنفيذها. SPSS هو تطبيق حاسوبي يُستخدم عادة لتحليل البيانات الإحصائية والبحث. الإصدار المستخدم من تطبيق SPSS هو SPSS v.26.

عملية تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية تبدأ بإجراء اختبارات الصدق والثبات لتقييم جودة بنود الأسئلة. بعد ذلك، يتم إجراء اختبارات على معاملات بنود الأسئلة، بما في ذلك مستوى الصعوبة ودرجة التمييز. تستمر العملية بإجراء تفسير وتوصيف لكل بند يتم تحليله.

١. طبيعية البيانات

التحليل بناء على نظرية الاختبار التقليدية يشمل حسابات رياضية بسيطة ولا يتطلب افتراضات معقدة. ومع ذلك، يتوجب إجراء اختبار طبيعية البيانات قبل القيام بالتحليل. اختبار طبيعية البيانات هو خطوة تُستخدم للتحقق مما إذا كانت البيانات تتبع التوزيع الطبيعي. الهدف من ذلك هو ضمان أن البيانات المستخدمة في التحليل تتوافق مع سمة التوزيع الطبيعي للبيانات، مما يضمن استيفاء الافتراضات الأساسية لنظرية الاختبار التقليدية بشكل كافٍ.

هناك عدة طرق متاحة لاختبار الطبيعية للبيانات. من بين هذه الطرق تقنيات بصرية، مثل استخدام الهيستوغرامات (Histogram) ورسوم الاحتمال الكمي (Q-Q Plots) ورسوم صندوق الإحصائيات (Box Plots). بالإضافة إلى ذلك، هناك اختبارات إحصائية مثل اختبار كولموغوروف–سميرنوف واختبار شابيرو–ويلك. ومع ذلك، في هذه الدراسة، استخدام الاختبارات الإحصائية أقل ملائمة لأنها تصلح بشكل أفضل للعينات ذات حجم معتدل، عادة حوالي ١٠٠. كل من اختبار كولموغوروف–سميرنوف واختبار شابيرو–ويلك (Kolmogrof-Smirnof & Shapiro-Wilk) حساسين للغاية لأحجام العينات التي تكون إما كبيرة جدًا أو صغيرة جدًا، وبالتالي فهما لا يصلحان لاختبار الطبيعية للبيانات في هذا البحث، حيث حجم العينة هو ٣١٣. وبالعكس، تقنيات بصرية تكون مناسبة للغاية لتقييم الطبيعية للبيانات الكبيرة ذات حجم عينة أكبر من ٢٠٠. وبالتالي، تعتبر تقنيات بصرية مناسبة لتقييم الطبيعية للبيانات في هذا البحث.



الرسم البياني ٤,١ : الهيستوغرامات ورسوم الاحتمال الكمي البيانات

في الهيستوغرام، يُعتبر البيانات موزعة بشكل طبيعي إذا كان الهيستوغرام الناتج يُظهر نمطًا يشبه الجرس المقلوب ومتماثلاً، ولا يكون مفرطًا في العرض أو رقة. وهذا يعني أن البيانات التي تقترب من القيمة المتوسطة أكثر عددًا مما تبتعد عنها. استنادًا إلى الهيستوغرام الناتج، يُمكن استنتاج أن البيانات موزعة بشكل طبيعي، نظرًا لأن شكل الهيستوغرام يشبه الجرس الطبيعي. بالإضافة إلى ذلك، في ورسوم الاحتمال الكمي، يُعتبر البيانات موزعة بشكل طبيعي إذا اقتربت النقاط في الرسم من الخط

الذي يمثل التوزيع الطبيعي المتوقع. عند النظر رسوم الاحتمال الكمي في هذه الدراسة، وجد أن جميع النقاط قريبة من الخط المرجعي، مما يُمكن استنتاج أن البيانات في هذه الدراسة موزعة بشكل طبيعي.

٢. الصدق

الصدق هو مقياس لمدى قدرة الاختبار أو أداة القياس على قياس ما يجب أن يتم قياسه.[٨١] هناك عدد طرق لقياس درجة الصدق للاختبار منها الصدق الحمتوى وصدق التوقع وصدق البناء وصدق القياس وغير ذلك. يعد قياس درجة الصدق أمرًا هامًا 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.[٨٢]

تم إجراء اختبار الصدق باستخدام صيغة نسبة الترابط لبيرسون (Pearson's Product Moment) باستخدام تطبيق مايكروسوفت إكسل. اختبار الصدق الذي تم إجراؤه باستخدام الطريقة المذكورة سابقا، أسفر الحساب قيمة "r" الحساب أو قيمة الارتباط لكل بند. يتم مقارنة هذه القيمة مع القيمة "r" الجدول التي تمثل معيار اختبار الملائمة. إذا كانت القيمة "r" الحساب أكبر من القيمة "r" الجدول، فإن البند يعتبر صادقا. على العكس، إذا كانت القيمة "r" الحساب أصغر من القيمة "r" الجدول، فإن البند يُعتبر غير صادقا.[٨٣] وبعد قام الباحث بتحليل ورقة أجوبة الطلبة، حصل البيانات كما في الجدول الآتي:

---

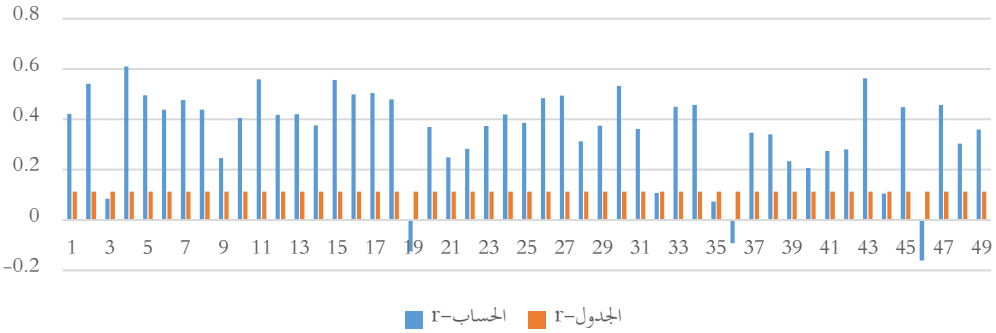
[٨١] Saifuddin Azwar, *Reliabilitas Dan Validitas*, 4th ed. (Yogyakarta: Pustaka Pelajar, 2014).
[٨٢] Arifin, *Evaluasi Pembelajaran*.
[٨٣] Sugiyono, *Metode Penelitian Pendidikan; Pendekatan Kuantitatif, Kualitatif, Dan R&D*.

الجدول ٤,١: الصدق لكل الأسئلة في التحليل بنظرية الاختبار التقليدية

رقم	بند	"r" الحساب	"r" الجدول	حالة	حاصل
١	X1	٠,٤٢١	٠,١١٣	أكبر	صادق
٢	X2	٠,٥٤١	٠,١١٣	أكبر	صادق
٣	X3	٠,٠٨٤	٠,١١٣	أصغر	غير صادق
٤	X4	٠,٦٠٩	٠,١١٣	أكبر	صادق
٥	X5	٠,٤٩٥	٠,١١٣	أكبر	صادق
٦	X6	٠,٤٣٨	٠,١١٣	أكبر	صادق
٧	X7	٠,٤٧٧	٠,١١٣	أكبر	صادق
٨	X8	٠,٤٨٣	٠,١١٣	أكبر	صادق
٩	X9	٠,٢٤٦	٠,١١٣	أكبر	صادق
١٠	X10	٠,٤٠٥	٠,١١٣	أكبر	صادق
١١	X11	٠,٥٥٨	٠,١١٣	أكبر	صادق
١٢	X12	٠,٤١٨	٠,١١٣	أكبر	صادق
١٣	X13	٠,٤٢١	٠,١١٣	أكبر	صادق
١٤	X14	٠,٣٧٦	٠,١١٣	أكبر	صادق
١٥	X15	٠,٥٥٧	٠,١١٣	أكبر	صادق
١٦	X16	٠,٤٩٩	٠,١١٣	أكبر	صادق
١٧	X17	٠,٥٠٤	٠,١١٣	أكبر	صادق
١٨	X18	٠,٤٧٩	٠,١١٣	أكبر	صادق
١٩	X19	-٠,١٢	٠,١١٣	أصغر	غير صادق
٢٠	X20	٠,٣٧	٠,١١٣	أكبر	صادق
٢١	X21	٠,٢٤٩	٠,١١٣	أكبر	صادق
٢٢	X22	٠,٢٨٣	٠,١١٣	أكبر	صادق
٢٣	X23	٠,٠٠٣	٠,١١٣	أصغر	غير صادق
٢٤	X24	٠,٣٧٤	٠,١١٣	أكبر	صادق

٢٥	X25	٠,٤٢	٠,١١٣	أكبر	صادق
٢٦	X26	٠,٣٨٦	٠,١١٣	أكبر	صادق
٢٧	X27	٠,٤٨٤	٠,١١٣	أكبر	صادق
٢٨	X28	٠,٤٩٤	٠,١١٣	أكبر	صادق
٢٩	X29	٠,٣١٢	٠,١١٣	أكبر	صادق
٣٠	X30	٠,٣٧٤	٠,١١٣	أكبر	صادق
٣١	X31	٠,٥٣٣	٠,١١٣	أكبر	صادق
٣٢	X32	٠,٣٦٢	٠,١١٣	أكبر	صادق
٣٣	X33	٠,١٠٧	٠,١١٣	أصغر	غير صادق
٣٤	X34	٠,٤٥	٠,١١٣	أكبر	صادق
٣٥	X35	٠,٤٥٧	٠,١١٣	أكبر	صادق
٣٦	X36	٠,٠٧٣	٠,١١٣	أصغر	غير صادق
٣٧	X37	-٠,٠٩	٠,١١٣	أصغر	غير صادق
٣٨	X38	٠,٣٤٧	٠,١١٣	أكبر	صادق
٣٩	X39	٠,٣٤	٠,١١٣	أكبر	صادق
٤٠	X40	٠,٢٣٤	٠,١١٣	أكبر	صادق
٤١	X41	٠,٢٠٦	٠,١١٣	أكبر	صادق
٤٢	X42	٠,٢٧٥	٠,١١٣	أكبر	صادق
٤٣	X43	٠,٢٨	٠,١١٣	أكبر	صادق
٤٤	X44	٠,٥٦٣	٠,١١٣	أكبر	صادق
٤٥	X45	٠,١٠٥	٠,١١٣	أصغر	غير صادق
٤٦	X46	٠,٤٤٨	٠,١١٣	أكبر	صادق
٤٧	X47	-٠,١٦	٠,١١٣	أصغر	غير صادق
٤٨	X48	٠,٤٥٧	٠,١١٣	أكبر	صادق
٤٩	X49	٠,٣٠٤	٠,١١٣	أكبر	صادق
٥٠	X50	٠,٣٦	٠,١١٣	أكبر	صادق

يمكن ملاحظة أنه من ٥٠ سؤالا، هناك سؤال واحد من نوع الصح أم خطأ والذي يحتوي فقط على خيارين، وهو السؤال رقم ٢٣. وهذا يختلف عن سائر الأسئلة التي تحتوي جميعها على أربعة خيارات. أظهرت نتائج اختبار الصدق أنه من إجمالي ٥٠ سؤالًا تم تحليلها، ثبت صدق ٤٢ سؤالًا، بينما أظهرت ٨ أسئلة أخرى نتائج غير صادق.



الرسم البياني ٤,٢ : مقارنة قيمة r-الحساب وقيمة r-الجدول لكل بنود الأسئلة

من الرسم البياني المقدم أعلاه، يمكن استنتاج أن ليس جميع أسئلة الاختبار تفي بشروط الصدق. هناك بعض أسئلة الاختبار التي تم اعتبارها غير صادقة لأن قيمة r المحساب أقل من قيمة r الجدول والتي تم تحديدها كمعيار في اختبار الصدق. بل وجد أيضًا أن بعض أسئلة الاختبار لديها قيم سلبية للقيمة الزوجية المحسوبة. لتحديد أسباب عدم الصدق في هذه الأسئلة، يتعين إجراء تحليل مفصل لفهم العوامل التي يمكن أن تُسهم في عدم الصدق. يهدف هذا التحليل إلى تحديد الجوانب التي يجب تحسينها في تصميم أسئلة الاختبار لزيادة صدقها.

٣. الثبات

أما الثبات هو مفهوم يشير إلى مدى موثوقية أو دقة قياس أو اختبار في تقديم نتائج متسقة وثابتة.[٨٤] قياس ثبات فحوصات بنود الاختبار ضروري لأنه يقدم معلومات حول مدى يمكن الاعتماد على بنودات الاختبار في قياس البنية المرغوبة. بعبارة أخرى، يشير ثبات بنودات الاختبار إلى مدى قدرة الاختبار على تقديم نتائج

[٨٤] Azwar, *Reliabilitas Dan Validitas*.

متسقة وقابلة للثقة. في نظرية الاختبار التقليدية، يُحتسب الثبات على مستوى الأداة ويُفترض أن يكون خطأ القياس المعياري واحدًا لجميع المشتركين في الاختبار. يمكن حساب ثبات بنودات الاختبار عن طريق حساب معامل الاتساق الداخلي. واحدة من الصيغ الأكثر شيوعًا لحساب معامل الاتساق الداخلي لاختبار هي صيغة ألفا كرونباخ.[٨٥] هذا المعامل يوفر إشارة حول مدى ترابط بنود الاختبار ويعكس مستوى الاتساق الداخلي للفحص.

من الناحية النظرية، يتراوح معامل الثبات بين القيمة ٠ و ١، ولكن معامل الثبات بقيمة ١ لا يمكن أبدًا العثور عليه في ممارسة القياس العملي. وجود التناسق المثالي الذي يظهره معامل الثبات بالقيمة ١ من هذا النوع يكاد يكون مستحيلاً في قياس السمات النفسية أو الاجتماعية التي تشمل البشر كموضوع للقياس. يعود ذلك إلى وجود مصادر متعددة للأخطاء التي يمكن أن تؤثر على نتائج القياس، سواء من ناحية الأداة المستخدمة أو من ناحية خصائص الفرد كموضوع يتم قياسه، أو حتى تنفيذ اختبار القياس نفسه. تلك الأخطاء يمكن أن تظهر على حدة أو معًا، ولها تأثير سلبي على دقة القياس الذي يتم إجراؤه.[٨٦]

نتيجة اختبار الثبات باستخدام الصيغة ألفا كرونباخ أظهرت قيمة قدرها ٠,٨٦، مما يشير إلى أن هذا الأداة تتمتع بمستوى عال من الثبات. المعامل يقترب من الواحد، وكلما اقترب من الواحد زادت الثبات. كما تم شرحه سابقًا، يُعتبر نطاق معامل الثبات بين ٠,٦٠ و ٠,٨٠ مرتفعًا بما يكفي، بينما يُعتبر القيم التي تزيد عن ٠,٨٠ مرتفعة جدًا. لذا، بناءً على معامل الثبات الذي تم الحصول عليه، يمكن الاستنتاج بأن أداة الاختبار هذه تتمتع بمستوى عالٍ من الثبات في قياس قدرة طلاب اللغة العربية.

٤. مستوى الصعوبة

---

[٨٥] Retnawati, "Reliabilitas Instrumen Penelitian."
[٨٦] Azwar, *Reliabilitas Dan Validitas*.

يُعتبر مستوى الصعوبة أحد المعايير المستخدمة لتقييم جودة بنود الأسئلة. يُقيس مستوى الصعوبة مدى صعوبة أو سهولة بنود الأسئلة على الطلاب المشاركين في الاختبار. يُعتمد قياس مستوى الصعوبة على نسبة الطلاب الذين يجيبون بشكل صحيح. إذا كان هناك الكثير من الطلاب الذين يمكنهم الإجابة بشكل صحيح، فإن بند السؤال يُصنف كسهل. وعلى العكس، إذا كان هناك عدد قليل من الطلاب الذين يمكنهم الإجابة بشكل صحيح، فإن بند السؤال يُعتبر صعبًا. يعد قياس مستوى الصعوبة أمرًا هامًا لضمان أن نطاق صعوبة بنود الأسئلة يتناسب مع احتياجات الاختبار ويميز بين الطلاب ذوي مستويات قدرات مختلفة.

في تقدير مستوى الصعوبة في نظرية الاختبار التقليدية، يُستخدم أحد الأساليب بالمقارنة بين عدد الطلاب الذين أجابوا بشكل صحيح وإجمالي عدد الطلاب الذين أجابوا على سؤال معين. لأجراء ذلك، يجب تقديم معلومات مفصلة حول عدد الإجابات الصحيحة وعدد الإجابات الخاطئة لكل سؤال. يتم تقديم البيانات حول عدد الطلاب الذين أجابوا بشكل صحيح وعدد الطلاب الذين أجابوا بشكل خاطئ على كل سؤال في الجدول المرفق.

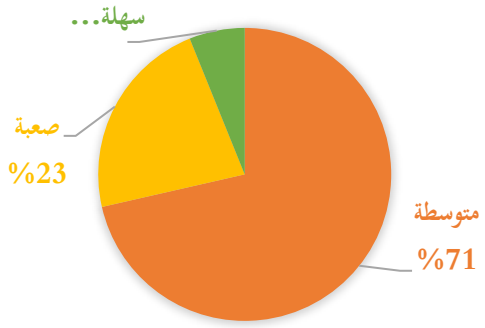
الجدول ٤,٢: مقارنة عدد الإجابات الصحيحة والخاطئة لكل الأسئلة

رقم	بند	صحيح	خطأ	رقم	بند	صحيح	خطأ
١	X1	١٦٥	١٤٨	٢٦	X26	١٢٦	١٨٧
٢	X2	١٩٣	١٢٠	٢٧	X27	٢٠٥	١٠٨
٣	X3	٨٦	٢٢٧	٢٨	X28	١٢٠	١٩٣
٤	X4	١٢٤	١٨٩	٢٩	X29	١١٢	٢٠١
٥	X5	١٨٣	١٣٠	٣٠	X30	٨٦	٢٢٧
٦	X6	١٨٠	١٣٣	٣١	X31	١٧٢	١٤١
٧	X7	٢٣٤	٧٩	٣٢	X32	١١٤	١٩٩
٨	X8	٢٧٢	٤١	٣٣	X33	١٢٦	١٨٧
٩	X9	١١٩	١٩٤	٣٤	X34	١٩٩	١١٤

١٠	X10	١١١	٢٠٢	٣٥	X35	٧٤	٢٣٩
١١	X11	١٥٦	١٥٧	٣٦	X36	٤٣	٢٧٠
١٢	X12	١٧٠	١٤٣	٣٧	X37	٦٩	٢٤٤
١٣	X13	١٤٠	١٧٣	٣٨	X38	١٨٢	١٣١
١٤	X14	٩١	٢٢٢	٣٩	X39	١٤١	١٧٢
١٥	X15	١٩٧	١١٦	٤٠	X40	٨٩	٢٢٤
١٦	X16	١٩١	١٢٢	٤١	X41	١٠٦	٢٠٧
١٧	X17	١١٢	٢٠١	٤٢	X42	١١٢	٢٠١
١٨	X18	١٤١	١٧٢	٤٣	X43	١٠٦	٢٠٧
١٩	X19	٢٥	٢٨٨	٤٤	X44	١٨٤	١٢٩
٢٠	X20	١٦٦	١٤٧	٤٥	X45	٥٣	١٦٠
٢١	X21	٧٧	٢٣٦	٤٦	X46	٢٣٣	٨٠
٢٢	X22	١٣٠	١٨٣	٤٧	X47	٣٥	٢٧٨
٢٣	X23	٢٠٧	١٠٦	٤٨	X48	١٠٩	٢٠٤
٢٤	X24	١٣٤	١٧٩	٤٩	X49	١٤٢	١٧١
٢٥	X25	١٥٤	١٥٩	٥٠	X50	١٤٠	١٧٣

تُرمز درجة صعوبة بند السؤال بـ "p_i". إذا كانت قيمة "p_i" تقترب من الصفر، فإن السؤال يُعتبر صعبًا جدًا. بينما إذا كانت قيمة "p_i " تقترب من الواحد، فإن السؤال يُعتبر سهلاً جدًا، وبالتالي يحتاج إلى إلغاء أو مراجعة.[٨٧] مؤشر صعوبة البنود الجيد يتراوح عادة بين ٠,٣ و ٠,٧. في هذا النطاق، يتم الحصول على معلومات عن قدرات الطلاب بشكل أقصى. إذا كانت قيمة المؤشر $p_i < 0.30$، يعتبر البنود صعبة، إذا كانت تتراوح قيمة p_i بين ٠,٣٠ و ٠,٧٠، يُعتبر مستوى صعوبة البنود متوسطًا، وإذا كانت قيمة $p_i > 0.70$، يعتبر البنود سهلة.[٨٨]

[٨٧] Retnawati, *Validitas, Reliabilitas, Dan Karakteristik Butir*.
[٨٨] Allen, *Introduction to Measurement Theory*.



الرسم البياني ٤,٣ : نسبة مستوى صعوبة الأسئلة في التحليل بنظرية الاختبار التقليدية

بعد إجراء التقدير، يتضح النتائج كما هو موضح في الرسم أعلاه. في المخطط الدائري، يمكن ملاحظة أن نسبة مستوى الصعوبة تهيمن عليها الأسئلة ذات مستوى صعوبة متوسط، حيث بلغ عددها ٣٥ سؤالًا أو ما يعادل ٧١٪ من الإجمالي. بينما تبلغ نسبة الأسئلة ذات مستوى صعوبة عالي ٢٣٪، مما يعني وجود ١١ سؤالًا بهذا المستوى. أما الأسئلة ذات مستوى صعوبة منخفض فتسهم بنسبة ٦٪ فقط من الإجمالي، وهي تعادل ٣ أسئلة. نتائج تقدير معاملات صعوبة البنود لكل سؤال هي كما يلي.

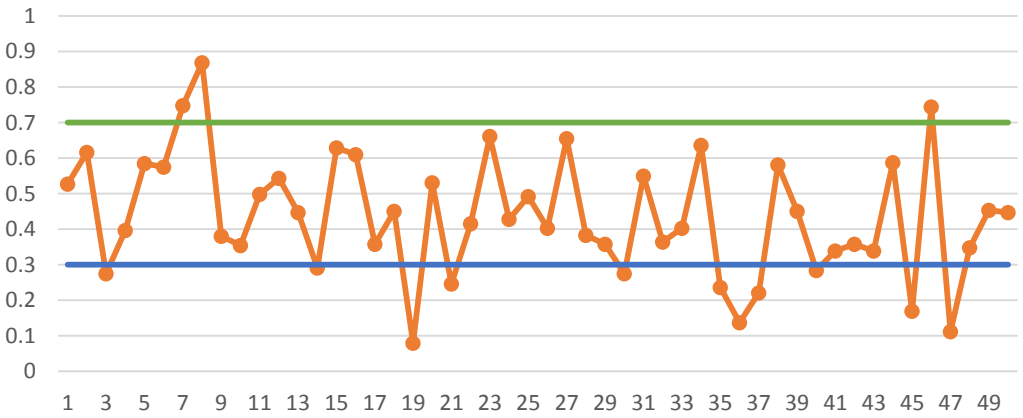
الجدول ٤,٣ : قيمة مستوى الصعوبة لكل الأسئلة في التحليل بنظرية الاختبار التقليدية

رقم	بند	قيمة	حاصل
١	X1	٠,٥٢٧	متوسطة
٢	X2	٠,٦١٧	متوسطة
٣	X3	٠,٢٧٥	صعبة
٤	X4	٠,٣٩٦	متوسطة
٥	X5	٠,٥٨٤	متوسطة
٦	X6	٠,٥٧٥	متوسطة
٧	X7	٠,٧٤٧	سهلة
٨	X8	٠,٨٦٩	سهلة
٩	X9	٠,٣٨٠	متوسطة

رقم	بند	قيمة	حاصل
٢٦	X26	٠,٤٠٢	متوسطة
٢٧	X27	٠,٦٥٥	متوسطة
٢٨	X28	٠,٣٨٣	متوسطة
٢٩	X29	٠,٣٥٨	متوسطة
٣٠	X30	٠,٢٧٥	صعبة
٣١	X31	٠,٥٤٩	متوسطة
٣٢	X32	٠,٣٦٤	متوسطة
٣٣	X33	٠,٤٠٢	متوسطة
٣٤	X34	٠,٦٣٦	متوسطة

١٠	X10	٠,٣٥٤	متوسطة	٣٥	X35	٠,٢٣٦	صعبة
١١	X11	٠,٤٩٨	متوسطة	٣٦	X36	٠,١٣٧	صعبة
١٢	X12	٠,٥٤٣	متوسطة	٣٧	X37	٠,٢٢٠	صعبة
١٣	X13	٠,٤٤٧	متوسطة	٣٨	X38	٠,٥٨١	متوسطة
١٤	X14	٠,٢٩١	صعبة	٣٩	X39	٠,٤٥٠	متوسطة
١٥	X15	٠,٦٢٩	متوسطة	٤٠	X40	٠,٢٨٤	صعبة
١٦	X16	٠,٦١٠	متوسطة	٤١	X41	٠,٣٣٨	متوسطة
١٧	X17	٠,٣٥٨	متوسطة	٤٢	X42	٠,٣٥٨	متوسطة
١٨	X18	٠,٤٥٠	متوسطة	٤٣	X43	٠,٣٣٨	متوسطة
١٩	X19	٠,٠٨٠	صعبة	٤٤	X44	٠,٥٨٨	متوسطة
٢٠	X20	٠,٥٣٠	متوسطة	٤٥	X45	٠,١٦٩	صعبة
٢١	X21	٠,٢٤٦	صعبة	٤٦	X46	٠,٧٤٤	سهلة
٢٢	X22	٠,٤١٥	متوسطة	٤٧	X47	٠,١١١	صعبة
٢٣	X23	٠,٦٦١	متوسطة	٤٨	X48	٠,٣٤٨	متوسطة
٢٤	X24	٠,٤٢٨	متوسطة	٤٩	X49	٠,٤٥٣	متوسطة
٢٥	X25	٠,٤٩٢	متوسطة	٥٠	X50	٠,٤٤٧	متوسطة

تم تحويل البيانات من الجدول أعلاه إلى الرسم البياني التالي.

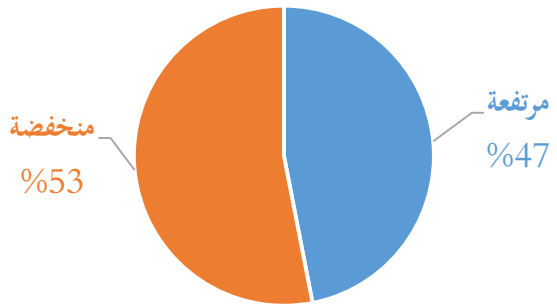


الرسم البياني ٤,٤ : قيمة مستوى صعوبة كل الأسئلة في التحليل بنظرية الاختبار التقليدية

تصنيف مستوى الصعوبة في نظرية الاختبار التقليدية يتم تقسيمه إلى ثلاث فئات، وهي السهلة، والمتوسطة، والصعبة. كلما كانت قيمة مستوى الصعوبة أصغر، زادت احتمالية الإجابة الصحيحة أقل، وزادت صعوبة البند. يُعتبر بندٌ لديه مستوى صعوبة عالٍ إذا كانت قيمته أقل من ٠,٣، ويُعتبر بندٌ لديه مستوى صعوبة منخفضٍ إذا كانت قيمته أكبر من ٠,٧. الخط الأخضر في الرسم البياني هو الحد بين الأسئلة التي تمتلك مستوى صعوبة منخفض وتلك التي تمتلك مستوى صعوبة متوسط. بينما الخط الأزرق هو الحد بين الأسئلة التي تمتلك مستوى صعوبة متوسط وتلك التي تمتلك مستوى صعوبة عال. لذلك، النقاط على الخط البرتقالي التي تقع بين الخطين الأزرق والأخضر تشير إلى أن الأسئلة تمتلك مستوى صعوبة متوسط. بينما النقاط التي تقع تحت الخط الأزرق تشير إلى أن الأسئلة تمتلك مستوى صعوبة عال، والنقاط التي تقع فوق الخط الأخضر تشير إلى أن الأسئلة تمتلك مستوى صعوبة منخفض.

٥. درجة التمييز

أما حساب درجة التمييز هو قياس لمدى قدرة البنود على تمييز الطلاب الذين يتقنون الكفايات عن الطلاب الذين لا يتقنون الكفايات أو يتقنونها بشكل غير كافٍ، بناءً على معايير محددة. كلما ارتفع معامل القوة التمييزية للبنود، زادت قدرتها على التمييز بين الطلاب الذين يتقنون الكفايات والطلاب الذين يتقنونها بشكل غير كافٍ.[٨٩]



الرسم البياني ٤,٥ : نسبة درجة تمييز الأسئلة في التحليل بنظرية الاختبار التقليدية

[٨٩] Arifin, *Evaluasi Pembelajaran*.

درجة التمييز تتم حسابها بطرح احتمالية الإجابة الصحيحة للمجموعة العليا من المشاركين في الاختبار من احتمالية الإجابة الصحيحة للمجموعة السفلى من المشاركين في الاختبار.

يُعتبر مؤشر درجة التمييز جيدًا إذا كان لديه قيمة تساوي أو تفوق ٠,٣. مؤشر درجة التمييز للبنود ذات قيمة صغيرة يؤدي إلى عدم قدرتها على تمييز الطلاب ذوي القدرات العالية عن الطلاب ذوي القدرات المنخفضة.[٩٠] من الرسم البياني أعلاه، نعلم أن هناك ٤٧٪ من الأسئلة لديها قدرة تمييز عالية و٥٣٪ من الأسئلة لديها قدرة تمييز منخفضة. نتائج تقدير معاملات صعوبة البنود لكل سؤال هي كما يلي.

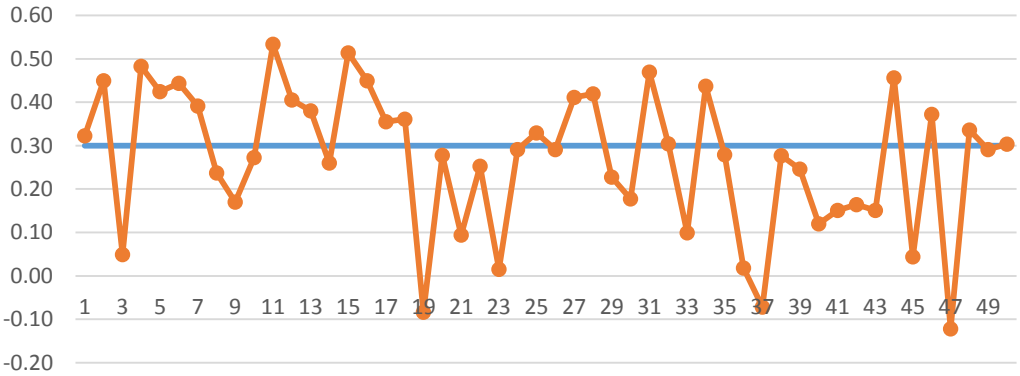
الجدول ٤,٤: قيمة درجة التمييز لكل الأسئلة في التحليل بنظرية الاختبار التقليدية

رقم	بند	قيمة	حاصل	رقم	بند	قيمة	حاصل
١	X1	٠,٣٢٣	مرتفعة	٢٦	X26	٠,٢٩١	منخفضة
٢	X2	٠,٤٥٠	مرتفعة	٢٧	X27	٠,٤١١	مرتفعة
٣	X3	٠,٠٤٩	منخفضة	٢٨	X28	٠,٤٩١	مرتفعة
٤	X4	٠,٤٨٣	مرتفعة	٢٩	X29	٠,٢٢٨	منخفضة
٥	X5	٠,٤٢٤	مرتفعة	٣٠	X30	٠,١٧٧	منخفضة
٦	X6	٠,٤٤٤	مرتفعة	٣١	X31	٠,٤٦٩	مرتفعة
٧	X7	٠,٣٩١	مرتفعة	٣٢	X32	٠,٣٠٤	مرتفعة
٨	X8	٠,٢٣٧	منخفضة	٣٣	X33	٠,١٠٠	منخفضة
٩	X9	٠,١٧٠	منخفضة	٣٤	X34	٠,٤٣٧	مرتفعة
١٠	X10	٠,٢٧٢	منخفضة	٣٥	X35	٠,٢٨٠	منخفضة
١١	X11	٠,٥٣٤	مرتفعة	٣٦	X36	٠,٠١٨	منخفضة
١٢	X12	٠,٤٠٥	مرتفعة	٣٧	X37	-٠,٠٧٢	منخفضة
١٣	X13	٠,٣٨١	مرتفعة	٣٨	X38	٠,٢٧٧	منخفضة
١٤	X14	٠,٢٦٠	منخفضة	٣٩	X39	٠,٢٤٦	منخفضة

[٩٠] Retnawati, *Validitas, Reliabilitas, Dan Karakteristik Butir*.

١٥	X15	٠,٥١٤	مرتفعة	٤٠	X40	٠,١٢٠	منخفضة
١٦	X16	٠,٤٥٠	مرتفعة	٤١	X41	٠,١٥١	منخفضة
١٧	X17	٠,٣٥٦	مرتفعة	٤٢	X42	٠,١٦٤	منخفضة
١٨	X18	٠,٣٦١	مرتفعة	٤٣	X43	٠,١٥١	منخفضة
١٩	X19	-٠,٠٨٤	منخفضة	٤٤	X44	٠,٤٥٦	مرتفعة
٢٠	X20	٠,٢٧٨	منخفضة	٤٥	X45	٠,٠٤٤	منخفضة
٢١	X21	٠,٠٩٤	منخفضة	٤٦	X46	٠,٣٢٧	مرتفعة
٢٢	X22	٠,٢٥٣	منخفضة	٤٧	X47	-٠,١٢٢	منخفضة
٢٣	X23	٠,٠١٥	منخفضة	٤٨	X48	٠,٣٣٦	مرتفعة
٢٤	X24	٠,٢٩١	منخفضة	٤٩	X49	٠,٢٩١	منخفضة
٢٥	X25	٠,٣٢٩	مرتفعة	٥٠	X50	٠,٣٠٤	مرتفعة

تم تحويل البيانات من الجدول أعلاه إلى الرسم البياني التالي.



الرسم البياني ٤,٦: قيمة درجة التمييز لكل الأسئلة في التحليل بنظرية الاختبار التقليدية

تصنيف درجة التمييز في نظرية الاختبار التقليدية يتم تقسيمه إلى فئتين، وهما درجة التمييز المنخفضة ودرجة التمييز العالية. درجة التمييز تعتبر عالية إذا كانت قيمتها أكبر من ٠,٣. الخط الأزرق في الرسم البياني هو الحد بين الأسئلة التي تمتلك درجة التمييز العالية وتلك التي تمتلك درجة التمييز المنخفضة، حيث يكون هذا الحد عند الرقم ٠,٣. لذا، تشير النقاط على الخط البرتقالي التي تقع فوق الخط الأزرق إلى أن الأسئلة

لها درجة التمييز العالية. بينما تشير النقاط التي تقع تحت الخط الأزرق إلى أن الأسئلة لها درجة التمييز المنخفضة.

٦. تفسير نتائج التحليل

تم تقديم تفسير لنتائج تحليل البنود في نظرية الاختبار التقليدية بناءً على معايير محددة تم تطبيقها من قبل الخبراء. المعايير المستخدمة لتفسير نتائج التحليل في نظرية الاختبار التقليدية هي كما يلي:

الجدول ٤,٥: معايير نظرية الاختبار التقليدية

مستوى الصعوبة	درجة التمييز
نسبة مؤوية الأسئلة السهلة والمتوسطة والصعبة هي ٢٧:٤٦:٢٧ %	درجة التمييز مرتفعة، أي أنها تكون أكبر من أو تساوي ٠,٣

قبل تفسير نتائج تقدير معاملات بنود الاختبار، يتعين أولاً تفسير صحة بنود الاختبار. الصحة المناقشة في هذا السياق هي الصحة التي يتم قياسها رياضيًا بناءً على نتائج اختبار الأداء على الأفراد، دون مراعاة محتوى بنود الاختبار ذاته. يتم هذا الحساب لاختبار العلاقة بين درجة بند الاختبار والدرجة الإجمالية للأداء. إذا تبين لاحقًا أن بعض بنود الاختبار لا تلبي معايير الصحة في هذا الحساب، سيتعين إعادة تقييمها من حيث مضمونها أو محتواها. سيتم اتخاذ القرار بشأن ما إذا كان يجب الاحتفاظ ببند معين، أو تحسينه، أو حتى إلغاؤه بالكامل.

تم حساب صدق كل بند اختبار باستخدام صيغة منتج بيرسون، وأظهرت النتائج أن هناك بعض بنود الاختبار التي لا تستوفي معايير الصدق، كما هو موضح في الجدول أدناه.

الجدول ٤,٦: البنود غير صادقة في نظرية الاختبار التقليدية

رقم	بند	"r" الحساب	"r" الجدول	حالة	حاصل
٣	X3	٠,٠٨٤	٠,١١٣	أصغر	غير صادق
١٩	X19	-٠,١٢	٠,١١٣	أصغر	غير صادق

٢٣	X23	٠,٠٠٣	٠,١١٣	أصغر	غير صادق
٣٣	X33	٠,١٠٧	٠,١١٣	أصغر	غير صادق
٣٦	X36	٠,٠٧٣	٠,١١٣	أصغر	غير صادق
٣٧	X37	-٠,٠٩	٠,١١٣	أصغر	غير صادق
٤٥	X45	٠,١٠٥	٠,١١٣	أصغر	غير صادق
٤٧	X47	-٠,١٦	٠,١١٣	أصغر	غير صادق

بعد معرفة الأسئلة غير الصادقة، تم إلغاء هذه الأسئلة مؤقتًا من إجراء التحليل، مما يترك الأسئلة الصادقة فقط لتقدير المعاملات الخاصة بها.

الجدول ٤,٧: نتائج تحليل نظرية الاختبار التقليدية

نظرية الاختبار التقليدية					
رقم	بنود	مستوى الصعوبة		درجة التمييز	
		قيمة	حاصل	قيمة	حاصل
١	X1	٠,٥٢٧	متوسطة	٠,٣٢٣	مرتفعة
٢	X2	٠,٦١٧	متوسطة	٠,٤٥٠	مرتفعة
٣	X3				
٤	X4	٠,٣٩٦	متوسطة	٠,٤٨٣	مرتفعة
٥	X5	٠,٥٨٤	متوسطة	٠,٤٢٤	مرتفعة
٦	X6	٠,٥٧٥	متوسطة	٠,٤٤٤	مرتفعة
٧	X7	٠,٧٤٧	سهلة	٠,٣٩١	مرتفعة
٨	X8	٠,٨٦٩	سهلة	٠,٢٣٧	منخفضة
٩	X9	٠,٣٨٠	متوسطة	٠,١٧٠	منخفضة
١٠	X10	٠,٣٥٤	متوسطة	٠,٢٧٢	منخفضة
١١	X11	٠,٤٩٨	متوسطة	٠,٥٣٤	مرتفعة
١٢	X12	٠,٥٤٣	متوسطة	٠,٤٠٥	مرتفعة
١٣	X13	٠,٤٤٧	متوسطة	٠,٣٨١	مرتفعة
١٤	X14	٠,٢٩١	صعبة	٠,٢٦٠	منخفضة

١٥	X15	٠,٦٢٩	متوسطة	٠,٥١٤	مرتفعة
١٦	X16	٠,٦١٠	متوسطة	٠,٤٥٠	مرتفعة
١٧	X17	٠,٣٥٨	متوسطة	٠,٣٥٦	مرتفعة
١٨	X18	٠,٤٥٠	متوسطة	٠,٣٦١	مرتفعة
١٩	X19				
٢٠	X20	٠,٥٣٠	متوسطة	٠,٢٧٨	منخفضة
٢١	X21	٠,٢٤٦	صعبة	٠,٠٩٤	منخفضة
٢٢	X22	٠,٤١٥	متوسطة	٠,٢٥٣	منخفضة
٢٣	X23				
٢٤	X24	٠,٤٢٨	متوسطة	٠,٢٩١	منخفضة
٢٥	X25	٠,٤٩٢	متوسطة	٠,٣٢٩	مرتفعة
٢٦	X26	٠,٤٠٢	متوسطة	٠,٢٩١	منخفضة
٢٧	X27	٠,٦٥٥	متوسطة	٠,٤١١	مرتفعة
٢٨	X28	٠,٣٨٣	متوسطة	٠,٤٩١	مرتفعة
٢٩	X29	٠,٣٥٨	متوسطة	٠,٢٢٨	منخفضة
٣٠	X30	٠,٢٧٥	صعبة	٠,١٧٧	منخفضة
٣١	X31	٠,٥٤٩	متوسطة	٠,٤٦٩	مرتفعة
٣٢	X32	٠,٣٦٤	متوسطة	٠,٣٠٤	مرتفعة
٣٣	X33				
٣٤	X34	٠,٦٣٦	متوسطة	٠,٤٣٧	مرتفعة
٣٥	X35	٠,٢٣٦	صعبة	٠,٢٨٠	منخفضة
٣٦	X36				
٣٧	X37				
٣٨	X38	٠,٥٨١	متوسطة	٠,٢٧٧	منخفضة
٣٩	X39	٠,٤٥٠	متوسطة	٠,٢٤٦	منخفضة
٤٠	X40	٠,٢٨٤	صعبة	٠,١٢٠	منخفضة

٤١	X41	٠,٣٣٨	متوسطة	٠,١٥١	منخفضة
٤٢	X42	٠,٣٥٨	متوسطة	٠,١٦٤	منخفضة
٤٣	X43	٠,٣٣٨	متوسطة	٠,١٥١	منخفضة
٤٤	X44	٠,٥٨٨	متوسطة	٠,٤٥٦	مرتفعة
٤٥	X45				
٤٦	X46	٠,٧٤٤	سهلة	٠,٣٢٧	مرتفعة
٤٧	X47				
٤٨	X48	٠,٣٤٨	متوسطة	٠,٣٣٦	مرتفعة
٤٩	X49	٠,٤٥٣	متوسطة	٠,٢٩١	منخفضة
٥٠	X50	٠,٤٤٧	متوسطة	٠,٣٠٤	مرتفعة

في مستوى الصعوبة، يمكن ملاحظة أن نسبة مستوى الصعوبة تهيمن عليها الأسئلة ذات مستوى صعوبة متوسط، حيث بلغ عددها ٣٥ سؤالًا أو ما يعادل ٧١٪ من الإجمالي. بينما تبلغ نسبة الأسئلة ذات مستوى صعوبة عالي ٢٣٪، مما يعني وجود ١١ سؤالًا بهذا المستوى. أما الأسئلة ذات مستوى صعوبة منخفض فتسهم بنسبة ٦٪ فقط من الإجمالي، وهي تعادل ٣ أسئلة.

نتائج الاختبارات المذكورة أعلاه، إذا مقارنتها بالمعايير، تشير إلى أن أداة الاختبار هذه لا تفي بمعايير جيدة فيما يتعلق بمستوى الصعوبة وفقًا لنظرية الاختبار التقليدية. نسب الأسئلة السهلة والمتوسطة والصعبة هي ٦٪ و٧١٪ و٢٣٪ على التوالي، وهذا يعني أن التوزيع لا يزال غير مثالي. هناك عدد قليل جدًا من الأسئلة التي تصنف كسهلة، بينما هناك الكثير من الأسئلة ذات مستوى صعوبة متوسط. من ناحية أخرى، فإن عدد الأسئلة الصعبة قد اقترب من التوزيع المثالي.

أما في درجة التمييز، يُعتبر مؤشر درجة التمييز جيدًا إذا كان لديه قيمة تساوي أو تفوق ٠,٣. مؤشر درجة التمييز للبنود ذات قيمة صغيرة يؤدي إلى عدم قدرتها على

تمييز الطلاب ذوي القدرات العالية عن الطلاب ذوي القدرات المنخفضة.[٩١] من الدائرة الدائرية أعلاه، نعلم أن هناك ٤٧٪ من الأسئلة لديها قدرة تمييز عالية و٥٣٪ من الأسئلة لديها قدرة تمييز منخفضة.

ب. تحليل بنود الأسئلة باستخدام نظرية استجابة البنود (IRT)

تم إجراء تحليل بنود الأسئلة باستخدام نظرية استجابة البنود في مرحلة تحليل البيانات، باستخدام برنامج RStudio بدعم من حزمة irtawsi. إن irtawsi هي حزمة تم تطويرها بهدف تبسيط عملية تحليل بنود الأسئلة في تطبيق RStudio. إنها حزمة مدمجة في RStudio تجمع بين مجموعة متنوعة من حزم البرمجيات الموجودة بالفعل، وتقدم واجهة سهلة الاستخدام للمستخدمين. وبالتالي، يتاح لمستخدمي البرنامج الاستفادة من سهولة الاستخدام هذه دون الحاجة إلى الاعتماد على سكربتات أو رموز برمجية لتنفيذ تحليل بنود الأسئلة.[٩٢]

بالإضافة إلى ذلك، يتضمن حزمة irtawsi شروحًا شاملة في كل مرحلة من مراحل التحليل. هذه الشروح لا تقتصر على شرح تفصيلي للخطوات التي يجب اتخاذها في عملية تحليل بنود الأسئلة، بل تشتمل أيضًا على ذكر مصادر أدبية موثوقة. هذه المصادر الأدبية تأتي من مقالات وكتب مؤلفة من قبل خبراء في ميدان تحليل بنود الأسئلة، مما يمكن المستخدمين من الرجوع إلى مصادر ذات جودة وموثوقية لتطوير فهمهم لتحليل بنود الأسئلة.

بناءً على ذلك، لا تقتصر حزمة irtawsi على تبسيط عملية تحليل بنود الأسئلة فقط، بل تضمن أيضًا أن المستخدمين يحصلون على الوصول إلى المعرفة العميقة والمحدثة في هذا الميدان. وهذه خطوة إيجابية بالغة في دعم البحث والتطوير في مجال تحليل بنود الأسئلة، وزيادة جودة وصحة نتائج تحليل بنود الأسئلة التي تتم من خلال تطبيق RStudio.

---

[٩١] Retnawati.

[٩٢] Hari Purnomo Susanto et al., "Irtawsi: Items Response Theory Analysis with Steps and Interpretation" (R, March 2023).

بالإضافة إلى حزمة irtawsi المذكورة أعلاه، تم استخدام حزمة أخرى في تحليل أسئلة الامتحان باستخدام نظرية الاستجابة للبند في هذا البحث، وهي حزمة ltm. تُستخدم حزمة ltm للحصول على معاملات معامل كل معامل مُقدر. بالإضافة إلى ذلك، يُستخدم الحزمة ltm أيضًا لإنتاج مجموعة متنوعة من منحنيات نتائج التحليل المفيدة في تفسير وتوصيف أسئلة الاختبار.[٩٣]

١. مطابقة النموذج

يمكن استخدام نهج نظرية استجابة البنود عندما تتوافق نماذج اللوجستيات مع البيانات الخاصة بالاختبار المراد دراسته. في هذا النهج، هناك ثلاثة أنواع من نماذج اللوجستيات المتاحة، وهي النموذج اللوجستي أحادي المعاملة الذي يحتوي فقط على معاملة مستوى الصعوبة، والنموذج اللوجستي ثنائي المعاملة الذي يحتوي على معاملة مستوى الصعوبة ومعاملة درجة التمييز، والنموذج اللوجستي ثلاثي المعاملات الذي يشمل معاملة مستوى الصعوبة ومعاملة درجة التمييز ومعاملة التخمين الزائف. كلما زاد عدد المعاملات المستخدمة في النموذج، زادت التفاصيل والمعلومات التي يمكن أن يقدمها هذا النموذج اللوجستي حول قدرة المشتركين في الاختبار.[٩٤]

في عملية تحليل بنود الأسئلة باستخدام نظرية استجابة البنود، يعد اختبار ملاءمة النموذج أحد الخطوات الحرجة في مرحلة البداية لهذا التحليل. وفي سياق هذا البحث، يمكن استخدام عدة طرق لاختبار ملاءمة النموذج، منها مؤشرات مثل SABIC (Sample-Adjusted Bayesian Information Criterion)، و HQ (Hannan-Quinn Criterion)، و BIC (Bayesian Information Criterion)، و AIC (Akaike Information Criterion). ومع ذلك، في سياق هذا البحث، تم اعتماد طريقة AIC. الهدف الرئيسي من اختبار ملاءمة النموذج باستخدام AIC هو تحديد النموذج الذي

[٩٣] Dimitris Rizopoulos, "Ltm: An R Package for Latent Variable Modeling and Item Response Analysis," *Journal of Statistical Software* 17 (2007): 1–25, https://doi.org/10.18637/jss.v017.i05.

[٩٤] Hambleton, Swaminathan, and Rogers, *Fundamentals of Item Response Theory*; Danni, Wahyuni, and Tauratiya, "ITEM RESPONSE THEORY APPROACH."

يتناسب بشكل أفضل مع البيانات المتاحة. يتم ذلك من خلال مقارنة قيم AIC بين ثلاثة نماذج للمعاملة اللوجستية المختلفة المتاحة. كلما كانت قيم AIC أقل، كان النموذج أكثر ملاءمة للبيانات، وكانت هناك مزيد من البنود التي تتناسب مع خصائص البيانات.

الجدول ٤,٨: مطابقة النموذج مع البيانات

نموذج	قيمة AIC	مناسب	غير مناسب
Rasch	١٧٦٠١,٥٩	٣٩ بندا	٩ بندا
2PL	١٧١٥٧,١٥	٤٦ بندا	٢ بندا
3PL	١٧٠٩٤,٨١	٤٨ بندا	٠ بند

قبل تحليل البيانات، يتم إزالة البنود التي لا تستوفي الشروط. هناك بندين لا يصلحان للتحليل. البند الأول هو البند رقم ٢٣ والذي يتضمن سؤالًا بخيارين فقط، صح أم خطأ. لذلك تم استبعاد هذا البند من التحليل لأنه يختلف عن باقي البنود التي تحتوي على أربعة خيارات وهي A و B و C وD. البند الثاني هو البند رقم ٣، والذي لا يستوفي شرط استقلال الموقف المحلي، وهذا البند له ارتباط مع أحد البنود الأخرى. لذا، يوصي التطبيق بإزالة هذا البند من التحليل.

أظهرت نتائج الاختبار أن النموذج اللوجستي ثلاثي المعاملات هو الأكثر ملائمة لهذه البيانات. يتميز هذا النموذج بقيمة AIC تبلغ ١٧٨٧٣,٦١، وهي القيمة الأدنى بين النماذج المتاحة. أظهرت نتائج هذا النموذج أن جميع البنود البالغ عددها ٤٨ بندًا مناسبة. على الجانب الآخر، النموذج ذو الاثنين المعاملتين مناسب لمجرد ٤٦ بندًا، والنموذج راش مناسب لمجرد ٣٩ بندًا. استنادًا إلى مقارنة قيم AIC، تم اتخاذ قرار باعتماد النموذج اللوجستي ثلاثي المعاملات لأنه يقدم قيمة AIC الأدنى ومناسب لمعظم بنود الاختبار. لذلك، يعتبر هذا النموذج هو الأكثر ملائمة لتحليل البنود في هذا البحث.

٢. الصدق

تم تحقق الصدق من خلال إثبات صدق البنية باستخدام تحليل العامل التفسيري. تحليل العامل التفسيري هو تقنية إحصائية تُستخدم لتحديد العلاقات بين المتغيرات في مجموعة بيانات معقدة.[٩٥] ومن أجل القيام بتحليل العامل التفسيري بنجاح، يجب أن يتوفر عدد كافٍ من العينات. لتقييم ما إذا كان عدد العينات كافيًا، يمكن المراعة إلى قيمة مؤشر كايزر–ماير–أولكين لقياس كفاية العينات (Kaiser-Meyer-Olkin Measure of Sampling Adequacy/KMO-MSA).

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		٠,٨٥١
Bartlett's Test of Sphericity	Approx. Chi-Square	٢٤٣٧,٦٧٧
	Df	٩٠٣
	Sig.	٠,٠٠٠

وصلت قيمة كايزر–ماير–أولكين لقياس كفاية العينات إلى ٠,٩٩٢، وهذا يتجاوز الحد الأدنى المقدر بحوالى ٠,٥ وقيمة *p* ٠،٠٠٠ حيث أصغر من ٠،٠٥. تشير هذه النتيجة إلى أن عدد العينات المستخدم في البحث وهي ٣١٣ كان كافيًا لدعم تنفيذ تحليل العامل التفسيري بنجاح.[٩٦] وبعبارة أخرى، القيمة التي تتجاوز ٠,٥ تشير إلى أن عدد العينات كان كافيًا لتنفيذ تحليل العامل التفسيري.

بناءً على قيمة مقياس كايزر–ماير–أولكين لقياس كفاية العينات الناتجة، يصبح بإمكان القيام بتحليل العوامل. الخطوة التالية هي النظر في جدول مصفوفة الارتباطات (matrix corelation) ضمن جدول الصور العكسية (anti-image) في نتائج برنامج SPSS. يجب أن تتجاوز قيم الارتباط في هذه المصفوفة الرقم ٠,٥٠. إذا وُجِدت بنود

[٩٥] Purwanto, "ANALISIS FAKTOR: KONSEP, PROSEDUR UJI DAN INTERPRETASI," *Jurnal Teknodik* 8, no. 15 (2018), https://doi.org/10.32550/teknodik.v4i15.388.

[٩٦] Citra Nurmalita, "Karakteristik Perangkat Ujian Kompetensi Teori Teknik Perkayuan Tahun 2014 SMK Di Jawa Timur Menggunakan Model IRT," *Ilmu Pendidikan: Jurnal Kajian Teori Dan Praktik Kependidikan* 3, no. 2 (December 29, 2018): 164–74, https://doi.org/10.17977/um027v3i22018p164.

ذات قيم ارتباط أقل من ٠,٥٠ في هذه المصفوفة، يجب إلغاء هذه البنود وإعادة الحساب للبنود المتبقية. تُعتبر البنود التي تحصل على قيم ارتباط أقل من ٠,٥٠ في هذه المصفوفة غير صحيحة تلقائيًا. فيما يلي جدول قيم الارتباط العكسية للبنود التي يمكن استخدامها، حيث بلغت ٤٣ بندا.

جدول ٤,٩: قيمة الارتباط من الصور العكسية

رقم	بنود	قيمة	رقم	بنود	قيمة
١	X1	–	٢٦	X26	٠,٨١٥
٢	X2	٠,٨٨٧	٢٧	X27	–
٣	X3	–	٢٨	X28	٠,٩٠٠
٤	X4	٠,٩٢١	٢٩	X29	٠,٧٩١
٥	X5	٠,٩١٠	٣٠	X30	٠,٨٦٠
٦	X6	٠,٨٨٧	٣١	X31	٠,٩١٠
٧	X7	٠,٨٩٦	٣٢	X32	٠,٨٢٩
٨	X8	–	٣٣	X33	٠,٥٢١
٩	X9	٠,٦٨٣	٣٤	X34	٠,٨٩٢
١٠	X10	٠,٨٧٠	٣٥	X35	٠,٨٨١
١١	X11	٠,٩١٥	٣٦	X36	–
١٢	X12	٠,٨٨١	٣٧	X37	٠,٥٦٦
١٣	X13	–	٣٨	X38	٠,٨٥٢
١٤	X14	٠,٨٣٧	٣٩	X39	٠,٨١١
١٥	X15	٠,٩٠٥	٤٠	X40	٠,٥٩٨
١٦	X16	٠,٨٧٤	٤١	X41	٠,٦٤٩
١٧	X17	٠,٨٦٥	٤٢	X42	٠,٧٨٨
١٨	X18	٠,٨٥٨	٤٣	X43	٠,٧٥٣
١٩	X19	٠,٥٩٦	٤٤	X44	٠,٩٢٣
٢٠	X20	٠,٨٣٢	٤٥	X45	٠,٥٩٢

٢١	X21	٠,٧٤٦	٤٦	X46	٠,٨٦٧
٢٢	X22	٠,٧٧٦	٤٧	X47	٠,٧١٥
٢٣	X23	–	٤٨	X48	٠,٨٧٣
٢٤	X24	٠,٨١٤	٤٩	X49	٠,٦٩٧
٢٥	X25	٠,٩٠٠	٥٠	X50	٠,٨٣٦

بعد ذلك، بهدف استعراض العوامل التي تكونت، يمكن تحديدها من خلال جدول التفسير الإجمالي الذي يظهر في إخراج SPSS. يمكن تحديد العوامل التي تظهر في هذا الأداة من خلال القيم الذاتية التي أكبر من ١. نتائج التحليل تشير إلى وجود ١٥ مكونًا يتخطى قيمة القيمة الذاتية ١. وهذا يعني وجود ١٥ عاملًا في هذا الاختبار.

الجدول ٤,١٠: القيمة الذاتية/أيجن لكل عوامل من جدول Total Variance Explained

عوامل	Eigen Value	% of Variance	Cumulative %
١	٧,٠٤٠	١٦,٣٧٢	١٦,٣٧٢
٢	١,٩٠٧	٤,٤٣٥	٢٠,٨٠٧
٣	١,٦٢٢	٣,٧٧٢	٢٤,٥٧٨
٤	١,٤٣٤	٣,٣٣٥	٢٧,٩١٤
٥	١,٤٠٢	٣,٢٦١	٣١,١٧٥
٦	١,٣٤٨	٣,١٣٦	٣٤,٣١٠
٧	١,٣٠٠	٣,٠٢٢	٣٧,٣٣٣
٨	١,٢٥٧	٢,٩٢٣	٤٠,٢٥٦
٩	١,٢٠٩	٢,٨١١	٤٣,٠٦٧
١٠	١,١٦٣	٢,٧٠٤	٤٥,٧٧١
١١	١,١٤١	٢,٦٥٤	٤٨,٤٢٥
١٢	١,٠٨٤	٢,٥٢٠	٥٠,٩٤٥

١٣	١,٠٤٧	٢,٣٤٣	٥٣,٣٧٩
١٤	١,٠٤٢	٢,٤٢٣	٥٥,٨٠٢
١٥	١,٠٠٢	٢,٣٣١	٥٨,١٣٣

الخطوة التالية تتمثل في مراقبة عامل التحميل (loading factor) لكل بند في جدول مصفوفة المكونات في برنامج SPSS. في هذا السياق، ومع وجود عدد بيانات أو عينات يبلغ ٣١٣، البنود التي تمتلك عوامل تحميل أكبر من ٠,٣٥ تشير إلى أن هذه البنود تقيس بشكل فعّال البعد. ومع ذلك، يجب التذكر أن معايير قيم عوامل التحميل يجب أن تتكيف مع عدد البيانات المتاحة. وذلك بسبب تأثير الاختلاف في عدد البيانات الذي يمكن أن يؤثر على الحد الأدنى للقيمة المعتبرة مهمة من عامل التحميل.[٩٧] هناك ٣٨ بندًا يظهر لديها معامل التحميل أكبر من ٠,٣٥وهناك ٥ بنود تظهر لديها معامل تحميل أقل.

الجدول ٤,١١: الصدق لكل الأسئلة في التحليل بنظرية استجابة البنود

رقم	بند	عامل	قيمة	حد	حالة	حاصل
١	X1	–	–	–	–	غير صادق
٢	X2	١	٠,٥٥١	٠,٣٥	أكبر	صادق
٣	X3	–	–	–	–	غير صادق
٤	X4	١	٠,٦٢٩	٠,٣٥	أكبر	صادق
٥	X5	١	٠,٥٠٧	٠,٣٥	أكبر	صادق
٦	X6	١	٠,٤٣٧	٠,٣٥	أكبر	صادق
٧	X7	١	٠,٤٩٠	٠,٣٥	أكبر	صادق
٨	X8	–	–	–	–	غير صادق
٩	X9	٤	٠,٤٠٦	٠,٣٥	أكبر	صادق

[٩٧] Joseph F. Hair et al., *Multivariate Data Analysis*, 7th ed. (Upper Saddle River, New Jersey: Prentice-Hall, Inc., 2010); Purwanto, "ANALISIS FAKTOR: KONSEP, PROSEDUR UJI DAN INTERPRETASI."

١٠	X10	١	٠,٤٠٢	٠,٣٥	أكبر	صادق
١١	X11	١	٠,٥٩٥	٠,٣٥	أكبر	صادق
١٢	X12	١	٠,٤٣٣	٠,٣٥	أكبر	صادق
١٣	X13	–	–	–	–	غير صادق
١٤	X14	١	٠,٣٨٤	٠,٣٥	أكبر	صادق
١٥	X15	١	٠,٥٨٦	٠,٣٥	أكبر	صادق
١٦	X16	١	٠,٥٠٨	٠,٣٥	أكبر	صادق
١٧	X17	١	٠,٥١٩	٠,٣٥	أكبر	صادق
١٨	X18	١	٠,٤٨٤	٠,٣٥	أكبر	صادق
١٩	X19	٦	٠,٥٧٧	٠,٣٥	أكبر	صادق
٢٠	X20	١	٠,٣٧٤	٠,٣٥	أكبر	صادق
٢١	X21	٢	٠,٤٩٩	٠,٣٥	أكبر	صادق
٢٢	X22	١	٠,٢٨٧	٠,٣٥	أصغر	غير صادق
٢٣	X23	–	–	–	–	غير صادق
٢٤	X24	١	٠,٣٧٧	٠,٣٥	أكبر	صادق
٢٥	X25	١	٠,٤٣١	٠,٣٥	أكبر	صادق
٢٦	X26	١	٠,٣٨٢	٠,٣٥	أكبر	صادق
٢٧	X27	–	–	–	–	غير صادق
٢٨	X28	١	٠,٥٢١	٠,٣٥	أكبر	صادق
٢٩	X29	٨	٠,٤٦٢	٠,٣٥	أكبر	صادق
٣٠	X30	١	٠,٣٥٢	٠,٣٥	أكبر	صادق
٣١	X31	١	٠,٥٦٠	٠,٣٥	أكبر	صادق
٣٢	X32	١	٠,٣٦٦	٠,٣٥	أكبر	صادق
٣٣	X33	–	–	–	–	غير صادق
٣٤	X34	١	٠,٤٧٥	٠,٣٥	أكبر	صادق
٣٥	X35	١	٠,٤٨١	٠,٣٥	أكبر	صادق

٣٦	X36	–	–	–	–	غير صادق
٣٧	X37	٥	٠,٥٧٨	٠,٣٥	أكبر	صادق
٣٨	X38	١	٠,٣٥٨	٠,٣٥	أكبر	صادق
٣٩	X39	١	٠,٣٤٩	٠,٣٥	أصغر	غير صادق
٤٠	X40	٦	٠,٣٩١	٠,٣٥	أكبر	صادق
٤١	X41	٢	٠,٣٣٠	٠,٣٥	أصغر	غير صادق
٤٢	X42	١٠	٠,٤٣٣	٠,٣٥	أكبر	صادق
٤٣	X43	١٢	٠,٣٨٧	٠,٣٥	أكبر	صادق
٤٤	X44	١	٠,٥٧٨	٠,٣٥	أكبر	صادق
٤٥	X45	٣	٠,٣١٧	٠,٣٥	أصغر	غير صادق
٤٦	X46	١	٠,٤٦٦	٠,٣٥	أكبر	صادق
٤٧	X47	٨	٠,٣٩٠	٠,٣٥	أكبر	صادق
٤٨	X48	١	٠,٤٥٣	٠,٣٥	أكبر	صادق
٤٩	X49	٦	٠,٥٠٥	٠,٣٥	أكبر	صادق
٥٠	X50	١	٠,٣٥٥	٠,٣٥	أكبر	صادق

يُعَتَبَرُ البنود التي لديها معامل التحميل أكبر من ٠,٣٥ بأنها بنود صادقة، وعددها ٣٨ بنداً. ومن بين البنود الصادقة، يقيس ٢٨ بنداً العامل الأوَّل، أي حوالي ٧٤٪. في حين أن ١٠ البنود المتبقية، أي حوالي ٢٦٪، تُوزَع بين العَوَامِل الأُخرَى. يتم تقديم جدول يوضح توزيع معاملات التحميل لكل بند كما يلي. يُقدَّمُ فيما يلي جدول تَوزيع البُنودِ صادقة والبُنودِ غير صادقة.

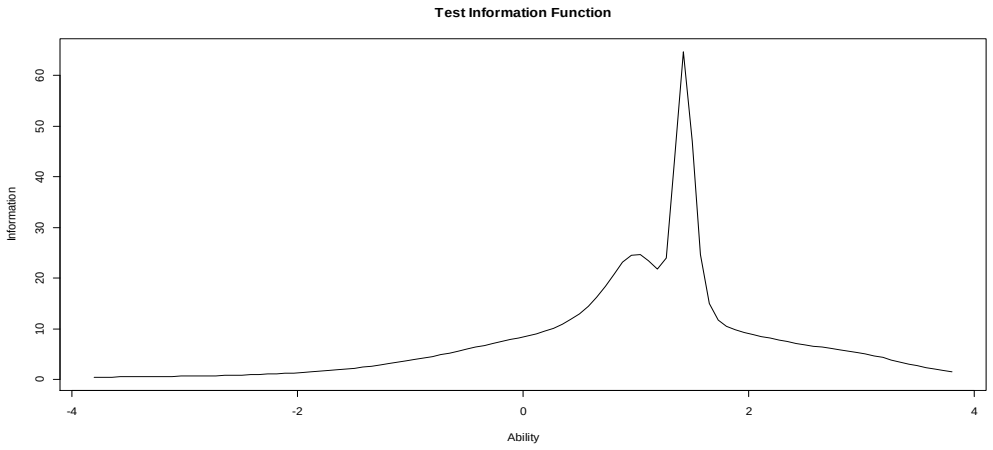
الجدول ٤,١٢: تصنيف الأسئلة الصادقة وغير الصادقة في التحليل بنظرية استجابة البنود

معيار	عدد	نسبة	بنود
صادق	٣٨	%٧٦	٢، ٤، ٥، ٦، ٧، ٩، ١٠، ١١، ١٢، ١٤، ١٥، ١٦، ١٧، ١٨، ١٩، ٢٠، ٢١، ٢٤، ٢٥، ٢٦، ٢٨، ٢٩، ٣٠، ٣١، ٣٢، ٣٤، ٣٥، ٣٧، ٣٨، ٤٠، ٤٢، ٤٣، ٤٤، ٤٦، ٤٧، ٤٨، ٤٩، ٥٠

غير صادق	١٢	٢٤%	٢٢، ٣٣، ٣٩، ٤١، ٤٥، ١ ،٣، ٨، ١٣، ٢٣، ٢٧، ٣٦ ،

٣. الثبات

الثبات في نظرية استجابة البند يمكن مراقبتها من خلال منحنى معلومات الاختبار الذي يتم إنتاجه. من خلال هذا المنحنى، يمكن مشاهدة مستوى دقة الأداة في إجراء القياس على مستويات مختلفة من القدرات. وبعبارة أخرى، تقدم هذه الأداة معلومات مختلفة للمشتركين في الاختبار على مستويات مهارات مختلفة. إذا كان قمة المنحنى مرتفعة، فهذا يشير إلى أن الأداة تقدم الكثير من المعلومات وقادرة على قياس مهارات المشتركين في الاختبار بشكل ممتاز. على النقيض، إذا كانت قمة المنحنى منخفضة، فهذا يشير إلى أن الأداة غير دقيقة بشكل كافي في قياس المهارات.



الرسم البياني ٤,٧: منحنى معلومات الاختبار

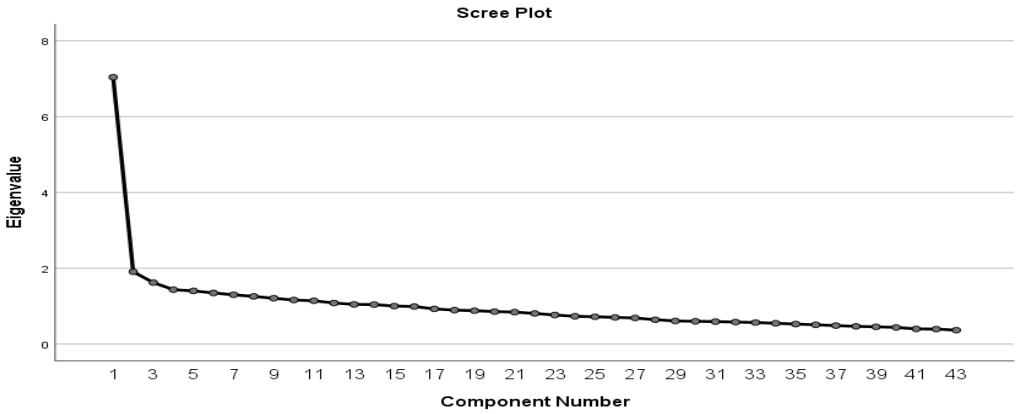
غالبًا ما يتم تقديم منحنى معلومات الاختبار مع خطأ القياس المعياري (SEM). في الاستخدام الشائع، يتم استخدام خطأ القياس المعياري لتحديد مدى قدرة الأداة على قياس نطاق المهارات بشكل جيد. كلما كان خطأ القياس المعياري أقل، زادت دقة الأداة في قياس نطاق المهارات. من المهم أن نلاحظ أنه في حزمة ltm المستخدمة في هذا البحث، لا يمكن عرض الانحراف المعياري مع منحنى معلومات الاختبار. ومع

ذلك، المعلومات المستمدة من منحنى معلومات الاختبار تزودنا بمعلومات حول الثبات للأداة في قياس قدرات الطلاب على مستويات متنوعة.

٤. الافتراضات

الأحادية البُعد

يمكن تحقيق افتراض الأحادية البعد عن طريق اتباع نهجين، أولهما من خلال فحص القيم الذاتية/إيجن (Eigen) والثاني من خلال مراقبة منحنى الانحدار (Scree plot).[٩٨] واحدة من المؤشرات الرئيسية في تقييم تحقيق هذا الافتراض هي عند مقارنة قيمة الذاتية للبند الأول مع البند الثاني، وأيضاً عند مقارنة قيمة الذاتية للبند الثاني مع البند الثالث وما بعده. إذا كانت قيمة الذاتية للبند الأول تظهر فارقًا ملحوظًا مقارنة بالبند الثاني، وإذا كانت قيمة الذاتية للبند الثاني تظهر فارقًا ضئيلاً مقارنة بالبند الثالث وما بعده، فيمكن الاستنتاج بأن افتراض الأحادية البعد قد تحقق.[٩٩]



الرسم البياني ٤,٨ : منحنى الانحدار

يمكن استخدام منحنى الانحدار(Scree Plot) لشرح نتائج الاختبار المذكور أعلاه. في منظر هذا المنحنى، ستلاحظ أن منحنى الانحدار الناتج عن قيم الذاتية بين البند الأول والبند الثاني يظهر فارقًا كبيرًا، مما يشير إلى وجود بعد رئيسي مرتبط بالبند الأول والثاني. وأما القيم الذاتية/إيجن الأولى تبلغ ٧,٠٤٠، حيث يفارق مع سائر القيم

[٩٨] Retnawati, *Validitas, Reliabilitas, Dan Karakteristik Butir*.

[٩٩] Budi Susetyo, *Prosedur Penyusunan Dan Analisis Tes Untuk Penilaian Hasil Belajar Kognitif* (Bandung: Refika Aditama, 2015).

الذاتية/إيجن. من ناحية أخرى، ستظهر الانتقال من البند الثاني إلى البند الثالث في منحنى الانحدار كأنه أكثر انحدارًا وانحسارًا. وهذا يشير إلى عدم وجود بعد إضافي مهم بعد البند الثاني.تعكس كمية التغيير الملحوظة في منحنى الانحدار هذا عدد الأبعاد الرئيسية في البيانات. بالإضافة إلى ذلك، الشكل الانحساري لمنحنى الانحدار يشير إلى عدم وجود بعد إضافي مهم. وبالتالي، يمكن استنتاج أن افتراض الأحادية البعد قد تم تحقيقه في هذا السياق استنادًا إلى تحليل منحنى الانحدار.

عدم الترابط المحلي

بالإضافة إلى افتراض الأحادية البعد، هناك أيضًا افتراض آخر يجب أن يتم تحقيقه في نهج نظرية استجابة البنود، وهو عدم الترابط المحلي. عدم الترابط المحلي هو افتراض يتطلب أن درجات بند الاختبار لا تعتمد على بنود أخرى. بمعنى آخر، إجابة على بند معين لا يجب أن تكون مرتبطة أو تؤثر على إجابات البنود الأخرى.[١٠٠]

أظهرت نتائج التحليل أن هناك بندين لا يستوفيان افتراض عدم الترابط المحلي، وهما البند رقم ٩ والبند رقم ٢٤. ومع ذلك، بعد إجراء فحص إضافي، تبين أن محتوى هذين السؤالين لا يترتبطان بشكل كبير. في هذا السياق، يمكن تجاهل نتائج اختبار عدم الترابط المحلي. إذا كان هناك اثنين أو أكثر من البنود التي لا تستوفي افتراض عدم الترابط المحلي ولكنها من طابع غير مترابط من الناحية المضمونية، فإنه يمكن استخدام تلك البنود في التحليل دون مشكلة.[١٠١]

بالإضافة إلى ذلك، سيتم تحقيق افتراض عدم الترابط المحلي أيضًا إذا تمكن الاختبار من إثبات أنه ذو طابع أحادي البعد. في هذا السياق، إذا تحقق افتراض الأحادية البعد، سيتم تحقيق افتراض عدم الترابط المحلي تلقائيًا.[١٠٢] وبالتالي، يمكن اعتبار هذا الأداة قياسية مستقلة بما أنها قد ثبت أنها تتمتع بخصائص أحادية البعد.

---

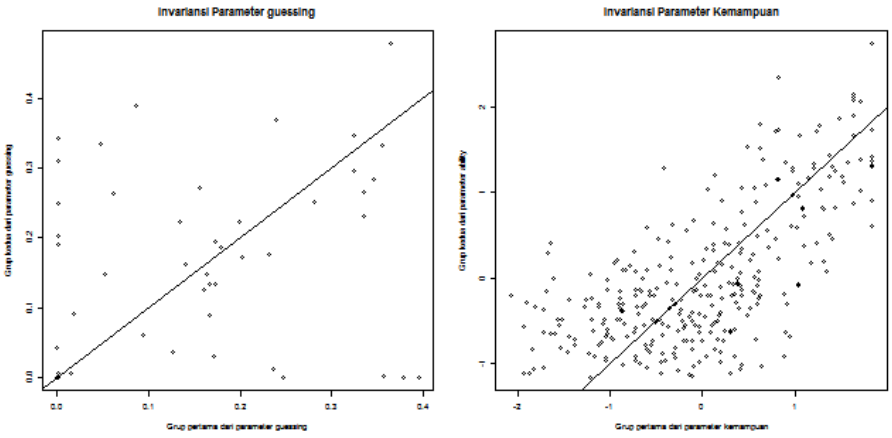
[١٠٠] Naga, *Pengantar Teori Sekor Pada Pengukuran Pendidikan*.

[١٠١] Tam H. Nguyen et al., "An Introduction to Item Response Theory for Patient-Reported Outcome Measurement," *The Patient* 7, no. 1 (2014): 23–35, https://doi.org/10.1007/s40271-013-0041-0.

[١٠٢] Retnawati, *Teori Respons Butir Dan Penerapannya: Untuk Peneliti, Praktisi Pengukuran Dan Pengujian, Mahasiswa Pascasarjana*.

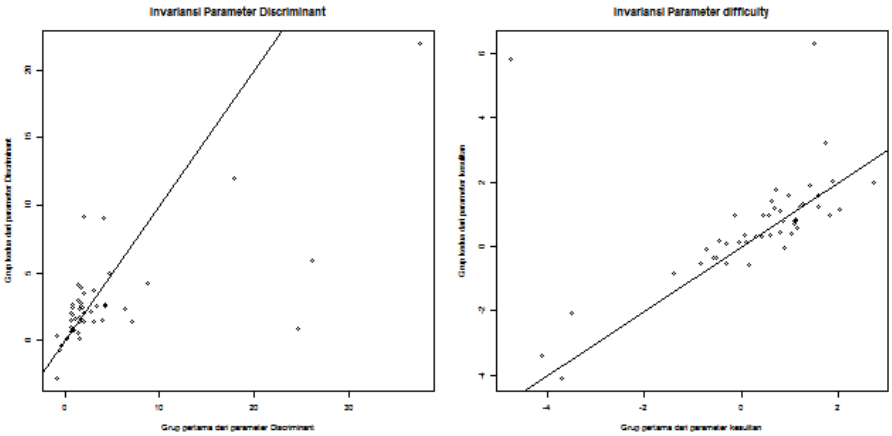
ثبات المعاملة

فرضية ثبات المعامل يمكن تقسيمها إلى فئتين، وهما فرضية ثبات معامل القدرة وفرضية ثبات معامل البند. معامل البند يمكن تقسيمها بدورها إلى ثلاث فئات، وهي معامل التمييز، ومعامل مستوى الصعوبة، ومعامل التخمين الزائف. عدد معاملات البند يعتمد على عدد نماذج تحليل بنود الأسئلة المنطبقة. في هذا السياق، يكون عدد معاملات البند ثلاثة لأن النموذج المستخدم هو نموذج اللوجستية ثلاث المعاملات. نتائج الاختبار تشير إلى أن فرضية ثبات المعامل متوفر في جميع المعاملات، باستثناء معامل التخمين الزائف. يمكن ملاحظة في الرسم البياني أدناه أن معامل التخمين الزائف لها توزيع واسع وتتجاوز بشكل ملحوظ الخط القطري. بينما كلما اقتربت أكثر من الخط القطري، زادت معامل التخمين هذه في الاستقرار.[١٠٣] عدم توفر هذه الفرضية في معامل التخمين الزائف يؤثر على نتائج التحليل. لا يمكن متابعة التحليل إذا كانت هناك أي فرضيات غير متوفرة.



الرسم البياني ٤,٨: مخطط النقاط لمعاملات البند ومعاملات القدرة

[١٠٣] Retnawati.



الرسم البياني ٤,٨: مخطط النقاط لمعاملات البند ومعاملات القدرة

أحد الأساليب التي يمكن استخدامها للتعامل مع عدم توفر فرضية ثبات المعامل هذه هو إزالة بعض الأسئلة التي تسبب عدم توفر هذه الفرضية في معامل التخمين الزائف.[١٠٤] الأسئلة التي تظهر فروقًا كبيرة في قيم معامل التخمين الزائف بين مجموعتي البيانات هي الأسئلة الأولى والسابعة والثانية عشرة والخامسة والعشرون. من خلال إزالة هذه الأسئلة، نأمل أن تتوفر فرضية ثبات المعامل وأن تصبح نتائج التحليل أكثر دقة.

٥. مستوى الصعوبة

مستوى صعوبة السؤال هو عامل رئيسي يؤثر على احتمالية شخص ما في الإجابة على السؤال بشكل صحيح. سيكون لدى السؤال الذي يحمل مستوى صعوبة عالي احتمالية أقل للإجابة عليه بشكل صحيح مقارنة بالسؤال الذي يحمل مستوى صعوبة منخفض. مستوى صعوبة السؤال، والذي يُقاس غالبًا في شكل قيمة بيتا (βi)، مرتبط بشكل كبير بالمستوى المطلوب من القدرة. الأسئلة الصعبة ستتطلب مستوى قدرة أعلى لدى الفرد حتى يكون لديه فرصة ٥٠:٥٠ للإجابة عليها بشكل صحيح، بينما ستمنح الأسئلة السهلة الفرصة نفسها للأفراد ذوي المستوى المنخفض من القدرة. هذا هو مفهوم مهم في تحليل نظرية استجابة البنود ويساعد في فهم كيفية تفاعل مستوى

[١٠٤] Nigel Guenole and Anna Brown, "The Consequences of Ignoring Measurement Invariance for Path Coefficients in Structural Equation Models," *Frontiers in Psychology* 5 (September 17, 2014), https://doi.org/10.3389/fpsyg.2014.00980.

الصعوبة ومستوى القدرة للأفراد في قياس الأسئلة.[١٠٥] ومع ذلك، ذلك لا يعني أن تحديد مستوى الصعوبة في نظرية استجابة البنود يتأثر بموضوع الاختبار بنفس الطريقة التي يحدث في نظرية الاختبار التقليدية لأن أساليب التقدير مختلفة. القيمة المناسبة لمستوى الصعوبة تتراوح بين −٢ و ٢.[١٠٦]

الجدول ٤,١٣ : مستوى صعوبة الأسئلة في التحليل بنظرية استجابة البنود

رقم	بند	قيمة	حاصل	رقم	بند	قيمة	حاصل
١	X1	–	–	٢٦	X26	٠,٦٤٩	جيد
٢	X2	−٠,٤١٨	جيد	٢٧	X27	–	–
٣	X3	–	–	٢٨	X28	٠,٩٠٣	جيد
٤	X4	٠,٧١١	جيد	٢٩	X29	١,٥٣٧	جيد
٥	X5	−٠,٣٣١	جيد	٣٠	X30	١,٤١٠	جيد
٦	X6	−٠,١٦٥	جيد	٣١	X31	٠,٢٤٦	جيد
٧	X7	−٠,٢٥٨	جيد	٣٢	X32	١,٢٦٩	جيد
٨	X8	–	–	٣٣	X33	٣,٥٧٦	ناقص
٩	X9	١,٧٥٠	جيد	٣٤	X34	−٠,٢٧٨	جيد
١٠	X10	٠,٨٧٢	جيد	٣٥	X35	١,٣٥٢	جيد
١١	X11	٠,٣٨١	جيد	٣٦	X36	٢,٩٦٥	ناقص
١٢	X12	−٠,١٣٤	جيد	٣٧	X37	−٣,٦٥٢	ناقص
١٣	X13	–	–	٣٨	X38	٠,٩٠٧	جيد
١٤	X14	١,٤٧٠	جيد	٣٩	X39	١,١٩٩	جيد
١٥	X15	−٠,٤٢٣	جيد	٤٠	X40	٢,٥٣٠	ناقص
١٦	X16	٠,٢٢٦	جيد	٤١	X41	٢,٦٥٢	ناقص
١٧	X17	٠,٩٩٧	جيد	٤٢	X42	١,٨٧٨	جيد
١٨	X18	٠,٩٦٣	جيد	٤٣	X43	١,٧٤٩	جيد

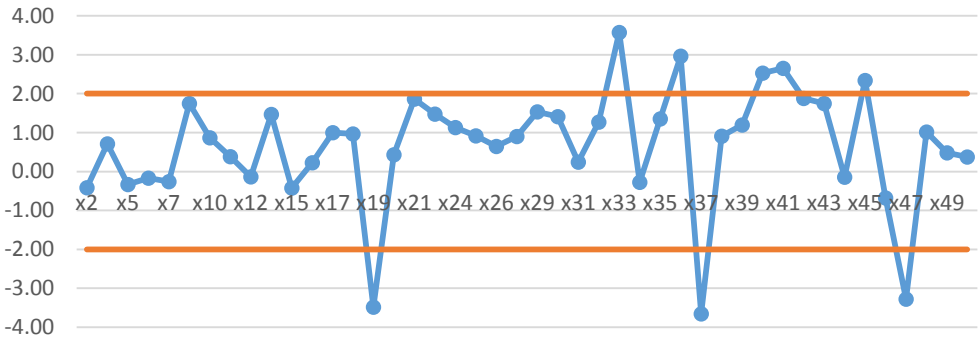
[١٠٥] Furr, *Psychometrics*.

[١٠٦] Retnawati, *Teori Respons Butir Dan Penerapannya: Untuk Peneliti, Praktisi Pengukuran Dan Pengujian, Mahasiswa Pascasarjana*; Hambleton and Swaminathan, *Item Response Theory*.

١٩	X19	-٣,٤٨٥	ناقص	٤٤	X44	-٠,١٤٦	جيد
٢٠	X20	٠,٤٣٥	جيد	٤٥	X45	٢,٣٤٣	ناقص
٢١	X21	١,٨٦٠	جيد	٤٦	X46	-٠,٦٧٦	جيد
٢٢	X22	١,٤٧٧	جيد	٤٧	X47	-٣,٢٧٦	ناقص
٢٣	X23	-	-	٤٨	X48	١,٠١٧	جيد
٢٤	X24	١,١٢٨	جيد	٤٩	X49	٠,٤٨٠	جيد
٢٥	X25	٠,٩١٩	جيد	٥٠	X50	٠,٣٧٠	جيد

قيمة مستوى الصعوبة في نظرية استجابة البند هي النقطة التي تكون فيها احتمالية الإجابة بشكل صحيح مساوية لـ ٠,٥ على محور الصعود (المحور الرأسي). لتحديد قيمة مستوى الصعوبة هذه، يتم رسم خط مستقيم يبدأ من النقطة ٠,٥ على محور الصعود ويتقاطع مع منحنى سمات البند على المحور الأفقي (المحور السيني). على سبيل المثال، إذا كانت القيمة على المحور السيني تساوي ١، فذلك يعني أنه يجب أن يكون لدى الشخص القدرة على مستوى ١ من الكفاءة للإجابة بشكل صحيح على البند. يقدم الجدول أعلاه قيم مستوى الصعوبة لكل بند. تعكس هذه القيم مستوى الكفاءة الذي يحتاجه الشخص للإجابة بشكل صحيح على البند. القيمة المناسبة لمستوى الصعوبة تتراوح بين -٢ و ٢.[١٠٧]

تم تحويل البيانات من الجدول أعلاه إلى الرسم البياني التالي.



الرسم البياني ٤,٩ : مستوى صعوبة الأسئلة في التحليل بنظرية استجابة البنود

[١٠٧] Retnawati, *Teori Respons Butir Dan Penerapannya: Untuk Peneliti, Praktisi Pengukuran Dan Pengujian, Mahasiswa Pascasarjana*; Hambleton and Swaminathan, *Item Response Theory*.

في الرسم البياني أعلاه، الخطان البرتقاليان هما حد لنطاق مستوى الصعوبة الذي يُعتبر جيدًا في نظرية استجابة البنود. لذلك، إذا وجدنا نقاطًا على الخط الأزرق تحت أو فوق الخطين البرتقاليين، فهذا يُشير إلى أن مستوى صعوبة البند غير جيد. إذا وجدت أعلى الخطين البرتقاليين، فإن ذلك يشير إلى أن مستوى صعوبته مرتفع جدًا، بينما إذا وجدت أسفل الخطين البرتقاليين، فإن ذلك يشير إلى أن مستوى صعوبته منخفض جدًا.

٦. درجة التمييز

في تحليل نظرية استجابة البنود، تعتبر درجة التمييز واحدة من العوامل المهمة في تقييم جودة البنود في الاختبار. البنود التي تمتلك قدرة تمييزية عالية تكون أكثر فعالية في تمييز الأفراد الذين يمتلكون مستويات متفاوتة من الكفاءة. باستخدام قيم درجة التمييز، يمكن التعرف على البنود التي تحتاج إلى تعديل أو استبدالها لزيادة دقة وصدق الاختبار.[١٠٨] تتجلى درجة التمييز في نظرية استجابة البنود من خلال انحدار منحنى خصائص البند. البنود ذات درجة التمييز العالية ستكون لها منحنى منحنٍ جدًا، بينما البنود ذات درجة التمييز الضعيفة ستكون لها منحنى منحنٍ جدًا. لذلك، درجة التمييز المثلى تتجلى في منحنى لا يكون شديد الانحدار ولا يكون متسطحًا للغاية. نطاق قيم درجة التمييز المثلى يتراوح بين ٠ و ٢. تمثل درجة التمييز في نظرية استجابة البنود بواسطة القيمة الألفا (αi).[١٠٩]

الجدول ٤,١٤: درجة تمييز الأسئلة في التحليل بنظرية استجابة البنود

رقم	بند	قيمة	حاصل	رقم	بند	قيمة	حاصل
١	X1	–	–	٢٦	X26	٠,٧٥٨	جيد
٢	X2	١,٤٩٩	جيد	٢٧	X27	–	–
٣	X3	–	–	٢٨	X28	٣,٧١١	ناقص
٤	X4	٤,٠٠٤	ناقص	٢٩	X29	٠,٩٧١	جيد

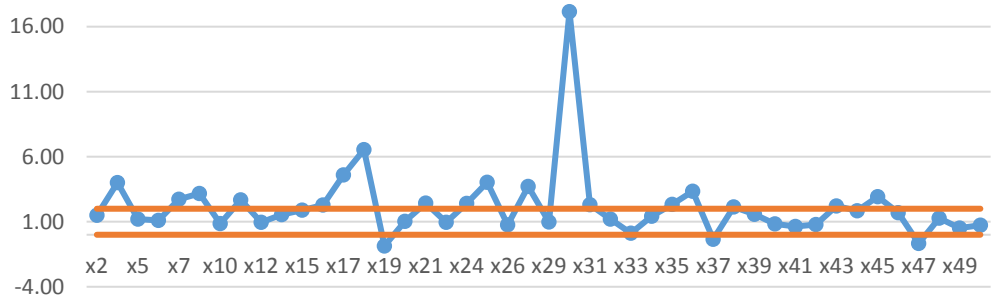
[١٠٨] Embretson and Reise, *Item Response Theory for Psychologists*.

[١٠٩] Furr, *Psychometrics*; Retnawati, *Teori Respons Butir Dan Penerapannya: Untuk Peneliti, Praktisi Pengukuran Dan Pengujian, Mahasiswa Pascasarjana*.

٥	X5	١,١٩٦	جيد	٣٠	X30	١٧,١٥٧	ناقص
٦	X6	١,١١٣	جيد	٣١	X31	٢,٢٩٧	ناقص
٧	X7	٢,٧٣٣	ناقص	٣٢	X32	١,٢٠٣	جيد
٨	X8			٣٣	X33	٠,١١٣	جيد
٩	X9	٣,١٦٣	ناقص	٣٤	X34	١,٤٢٥	جيد
١٠	X10	٠,٨٥٤	جيد	٣٥	X35	٢,٣٣٨	ناقص
١١	X11	٢,٦٩٠	ناقص	٣٦	X36	٣,٣٤١	ناقص
١٢	X12	٠,٩٦٦	جيد	٣٧	X37	-٠,٣٥٢	ناقص
١٣	X13	-	-	٣٨	X38	٢,١٢٦	ناقص
١٤	X14	١,٥٤٥	جيد	٣٩	X39	١,٥٧٨	جيد
١٥	X15	١,٨٧٨	جيد	٤٠	X40	٠,٨٣٧	جيد
١٦	X16	٢,٢٧٧	ناقص	٤١	X41	٠,٦٣٨	جيد
١٧	X17	٤,٥٨٦	ناقص	٤٢	X42	٠,٧٨٠	جيد
١٨	X18	٦,٥٣٢	ناقص	٤٣	X43	٢,٢١٢	ناقص
١٩	X19	-٠,٨٦٢	ناقص	٤٤	X44	١,٨٢٩	جيد
٢٠	X20	١,٠٣٧	جيد	٤٥	X45	٢,٩٢٥	ناقص
٢١	X21	٢,٤٣٨	ناقص	٤٦	X46	١,٦٩٤	جيد
٢٢	X22	٠,٩٥٦	جيد	٤٧	X47	-٠,٦٨٠	ناقص
٢٣	X23	-	-	٤٨	X48	١,٢٨١	جيد
٢٤	X24	٢,٤١٨	ناقص	٤٩	X49	٠,٥٠٩	جيد
٢٥	X25	٤,٠٢٧	ناقص	٥٠	X50	٠,٧٣٧	جيد

من الجدول أعلاه، يمكن ملاحظة وجود عدد من بنود الاختبار التي تحمل قيم قدرة تمييزية غير كافية. يمكن تحديد هذه البنود استنادًا إلى القيم التي تقل قدرة تمييزها عن ٠ أو تزيد عن ٢. القيم التي تقل قدرة تمييزها عن ٠ تشير إلى أن تلك البنود تمتلك قدرة تمييزية ضعيفة للغاية، وهذا يعني أن تلك البنود لا تكون فعالة في تمييز

الأشخاص الذين يمتلكون مهارات عالية من الأشخاص الذين يمتلكون مهارات منخفضة. من ناحية أخرى، القيم التي تزيد قدرة تمييزها عن ٢ تشير إلى أن قدرة تمييز تلك البنود مرتفعة للغاية، وهذا أيضًا غير مرغوب فيه. لذلك، البنود ذات درجة التمييز العالية أو القليلة تحتاج إلى التحسين أو الاستبدال لزيادة جودة الاختبار. تم تحويل البيانات من الجدول أعلاه إلى الرسم البياني التالي.



الرسم البياني ٤,١٠: درجة تمييز الأسئلة في التحليل بنظرية استجابة البنود

في الرسم البياني أعلاه، الخطين البرتقاليين هما الحدان لنطاق درجة التمييز الذي يُعتبر جيدًا في نظرية استجابة البنود. لذلك، إذا وجدنا نقاطًا على الخط الأزرق تحت أو فوق الخطين البرتقاليين، فهذا يشير إلى أن درجة التمييز للبند غير جيدة. إذا وجدت أعلى الخطين البرتقاليين، فإن ذلك يشير إلى أن درجة التمييز مرتفعة جدًا، بينما إذا وجدت أسفل الخطين البرتقاليين، فإن ذلك يشير إلى أن درجة التمييز منخفضة جدًا.

٧. التخمين الزائف

في نموذج اللوجستيات ثلاثي البارامترات، يُعد معامل التخمين الزائف هو المعامل الثالث الذي يمكن أن يؤثر على استجابة المشاركين لسؤال الاختبار. في اختبار الاختيار من متعدد أو الاختبار الصحيح/الخطأ، قد يقوم بعض المشاركين بتخمين الإجابة إذا كانوا لا يعرفون الإجابة الصحيحة. وهذا يتيح للمشاركين الإجابة على السؤال بشكل صحيح عن طريق الصدفة من خلال التخمين. لذلك، يتضمن نموذج اللوجستيات ثلاثي البارامترات معامل التخمين الزائف لمراعاة هذا الاحتمال. يأخذ معامل التخمين الزائف بعين الاعتبار عدد الخيارات المتاحة في السؤال.

القيمة المثلى لمعامل التخمين الزائف لسؤال معين لا يجب أن تتجاوز ١/k، حيث k هو عدد الخيارات المتاحة. في هذا السياق، بوجود أربع خيارات، القيمة المثلى لمعامل التخمين الزائف هي أقل من ٠,٢٥. كلما انخفضت القيمة المثلى لمعامل التخمين الزائف، زادت صعوبة تخمين الإجابة على السؤال وزادت جودة السؤال أيضًا. وفي نظرية استجابة البنود، يُمثل معامل التخمين الزائف بالرمز "ci".[١١٠]

الجدول ٤,١٥: التخمين الزائف للأسئلة في التحليل بنظرية استجابة البنود

رقم	بند	قيمة	حاصل	رقم	بند	قيمة	حاصل
١	X1	–	–	٢٦	X26	٠,٠٠٣	جيد
٢	X2	٠,٠٠٠	جيد	٢٧	X27	–	–
٣	X3	–	–	٢٨	X28	٠,١٩٣	جيد
٤	X4	٠,١٥٤	جيد	٢٩	X29	٠,١٦٥	جيد
٥	X5	٠,٠٠٠	جيد	٣٠	X30	٠,٢٠٢	جيد
٦	X6	٠,٠٦٥	جيد	٣١	X31	٠,١٩٨	جيد
٧	X7	٠,٣٧٩	ناقص	٣٢	X32	٠,١٦٠	جيد
٨	X8	–	–	٣٣	X33	٠,٠٠٢	جيد
٩	X9	٠,٣٢٩	ناقص	٣٤	X34	٠,١٣٣	جيد
١٠	X10	٠,٠٠١	جيد	٣٥	X35	٠,٠٩٢	جيد
١١	X11	٠,١٧٢	جيد	٣٦	X36	٠,١٣٢	جيد
١٢	X12	٠,٠١٧	جيد	٣٧	X37	٠,٠٠٠	جيد
١٣	X13	–	–	٣٨	X38	٠,٤٣٣	ناقص
١٤	X14	٠,١٣٨	جيد	٣٩	X39	٠,٢٩٢	ناقص
١٥	X15	٠,٠٠٠	جيد	٤٠	X40	٠,١٦٩	جيد
١٦	X16	٠,٢٩٩	ناقص	٤١	X41	٠,١٩٥	جيد
١٧	X17	٠,١٩١	جيد	٤٢	X42	٠,١٧٦	جيد
١٨	X18	٠,٣٠٧	ناقص	٤٣	X43	٠,٢٦٦	ناقص

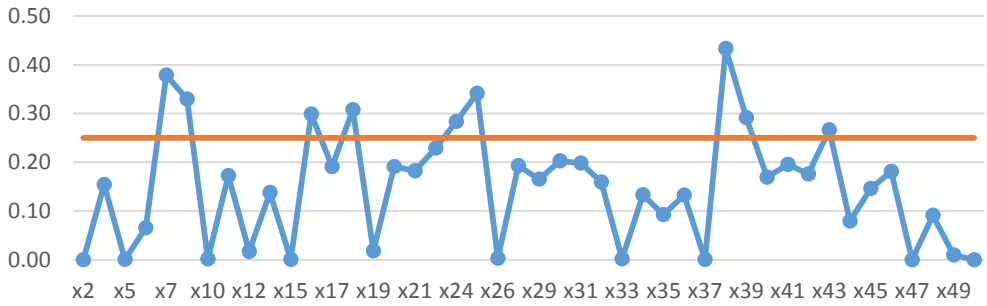
[١١٠] Furr, *Psychometrics*.

١٩	X19	٠,٠١٨	جيد
٢٠	X20	٠,١٩١	جيد
٢١	X21	٠,١٨٣	جيد
٢٢	X22	٠,٢٢٩	جيد
٢٣	X23	–	–
٢٤	X24	٠,٢٨٣	ناقص
٢٥	X25	٠,٣٤١	ناقص

٤٤	X44	٠,٠٧٩	جيد
٤٥	X45	٠,١٤٦	جيد
٤٦	X46	٠,١٨١	جيد
٤٧	X47	٠,٠٠٠	جيد
٤٨	X48	٠,٠٩١	جيد
٤٩	X49	٠,٠١٠	جيد
٥٠	X50	٠,٠٠٠	جيد

تستند النتائج الموجودة في الجدول أعلاه إلى ملاحظة وجود تباين كبير في معامل التخمين الزائف بين أسئلة الاختبار المختلفة. هذا التباين يمتد عبر مجموعة واسعة من القيم، بدءًا من القيمة ٠ وصولاً إلى ٠,٤. القيمة ٠ لمعامل التخمين الزائف تشير إلى أن احتمالية الإجابة على هذا السؤال بالتخمين ضئيلة جدًا أو حتى تكاد تكون معدومة. من ناحية أخرى، الأسئلة التي تحمل قيمة ٠,٤ لمعامل التخمين الزائف تشير إلى أن هذه الأسئلة سهلة جدًا في التخمين على إجاباتها. حتى أن قيمتها تتجاوز الحد الطبيعي المعتاد لأداة الاختبار هذه، والذي يبلغ ٠,٢٥.

تم تحويل البيانات من الجدول أعلاه إلى الرسم البياني التالي.



الرسم البياني ٤,١١: التخمين الزائف للأسئلة في التحليل بنظرية استجابة البنود

في الرسم البياني أعلاه، الخط البرتقالي هو الحد الذي يُعتبر جيدًا للتخمين الزائف في نظرية استجابة البنود. لذا، إذا وجدت نقاطًا على الخط الأزرق توجد أعلى الخط البرتقالي، فهذا يشير إلى أن التخمين الزائف للبند غير جيد. إذا وجدت أنها توجد

أعلى الخط البرتقالي، فإن ذلك يشير إلى أن التخمين الزائف مرتفع للغاية، مما يعني أن هذه الأسئلة يمكن الإجابة عليها بشكل صحيح فقط من خلال التخمين.

٨. تفسير نتائج التحليل

نتائج تحليل أسئلة الاختبار التي تمت باستخدام نظرية استجابة البنود يمكن تفسيرها سواء في سياق البنود بشكل فردي أو في الأداة ككل. ولكن قبل أن يمكن القيام بالتفسير والتوصيف، يجب تحديد معايير القياس المناسبة. يتم الحصول على هذه المعايير استنادًا إلى آراء الخبراء والأبحاث السابقة. فيما يلي جدول يوضح المعايير المعتمدة لبنود الاختبار.

الجدول ٤,١٦: معايير نظرية استجابة البنود

نموذج	معايير		
	جيد جدا	جيد	ناقص
1PL	$-٢ \leq b \leq ٢$	_	$b < -٢ / b > ٢$
2PL	$٠ \leq a \leq ٢$ $-٢ \leq b \leq ٢$	لم يتم تحقيق أحد من المعايير	لم يتم تحقيق أكثر من واحد من المعايير
3PL	$٠ \leq a \leq ٢$ $-٢ \leq b \leq ٢$ $c \leq ٠,٢٥$	لم يتم تحقيق أحد من المعايير	لم يتم تحقيق أكثر من واحد من المعايير

القيمة المناسبة لمستوى الصعوبة تتراوح بين −٢ و ٢. و درجة التمييز المثلى تتجلى في منحنى لا يكون شديد الانحدار ولا يكون متسطحًا للغاية. نطاق قيم درجة التمييز المثلى يتراوح بين ٠ و ٢. القيمة المثلى لمعامل التخمين الزائف لسؤال معين لا يجب أن تتجاوز $1/k$، حيث k هو عدد الخيارات المتاحة. في هذا السياق، بوجود أربع خيارات، القيمة المثلى لمعامل التخمين الزائف هي أقل من ٠,٢٥. كلما انخفضت القيمة المثلى لمعامل التخمين الزائف، زادت صعوبة تخمين الإجابة على السؤال وزادت

جودة السؤال أيضًا.[١١١] فيما يلي توضيح لسمات كل معامل بنود الاختبار استنادًا إلى المعايير المذكورة أعلاه. وبعد أن تم توصيف معاملات كل بنود، يتم متابعة التفسير بناءً على معايير نموذج الثلاث معاملات اللوجستية

الجدول ٤,١٧ : نتائج تحليل نظرية استجابة البنود

رقم	بنود	نظرية استجابة البنود						
		مستوى الصعوبة		درجة التمييز		التخمين الزائف		توصيف
		قيمة	حاصل	قيمة	حاصل	قيمة	حاصل	
١	X1							
٢	X2	-٠,٤١٨	جيد	١,٤٩٩	جيد	٠,٠٠٠	جيد	جيد جدا
٣	X3							
٤	X4	٠,٧١١	جيد	٤,٠٠٤	ناقص	٠,١٥٤	جيد	جيد
٥	X5	-٠,٣٣١	جيد	١,١٩٦	جيد	٠,٠٠٠	جيد	جيد جدا
٦	X6	-٠,١٦٥	جيد	١,١١٣	جيد	٠,٠٦٥	جيد	جيد جدا
٧	X7	-٠,٢٥٨	جيد	٢,٧٣٣	ناقص	٠,٣٧٩	ناقص	ناقص
٨	X8							
٩	X9	١,٧٥٠	جيد	٣,١٦٣	ناقص	٠,٣٢٩	ناقص	ناقص
١٠	X10	٠,٨٧٢	جيد	٠,٨٥٤	جيد	٠,٠٠١	جيد	جيد جدا
١١	X11	٠,٣٨١	جيد	٢,٦٩٠	ناقص	٠,١٧٢	جيد	جيد
١٢	X12	-٠,١٣٤	جيد	٠,٩٦٦	جيد	٠,٠١٧	جيد	جيد جدا
١٣	X13							
١٤	X14	١,٤٧٠	جيد	١,٥٤٥	جيد	٠,١٣٨	جيد	جيد جدا
١٥	X15	-٠,٤٢٣	جيد	١,٨٧٨	جيد	٠,٠٠٠	جيد	جيد جدا
١٦	X16	٠,٢٢٦	جيد	٢,٢٧٧	ناقص	٠,٢٩٩	ناقص	ناقص
١٧	X17	٠,٩٩٧	جيد	٤,٥٨٦	ناقص	٠,١٩١	جيد	جيد

[١١١] Retnawati, *Teori Respons Butir Dan Penerapannya: Untuk Peneliti, Praktisi Pengukuran Dan Pengujian, Mahasiswa Pascasarjana*; Hambleton and Swaminathan, *Item Response Theory*.

١٨	X18	٠,٩٦٣	جيد	٦,٥٣٢	ناقص	٠,٣٠٧	ناقص	ناقص
١٩	X19	-٣,٤٨٥	ناقص	-٠,٨٦٢	ناقص	٠,٠١٨	جيد	ناقص
٢٠	X20	٠,٤٣٥	جيد	١,٠٣٧	جيد	٠,١٩١	جيد	جيد جدا
٢١	X21	١,٨٦٠	جيد	٢,٤٣٨	ناقص	٠,١٨٣	جيد	جيد
٢٢	X22	١,٤٧٧	جيد	٠,٩٥٦	جيد	٠,٢٢٩	جيد	جيد جدا
٢٣	X23							
٢٤	X24	١,١٢٨	جيد	٢,٤١٨	ناقص	٠,٢٨٣	ناقص	ناقص
٢٥	X25	٠,٩١٩	جيد	٤,٠٢٧	ناقص	٠,٣٤١	ناقص	ناقص
٢٦	X26	٠,٦٤٩	جيد	٠,٧٥٨	جيد	٠,٠٠٣	جيد	جيد جدا
٢٧	X27							
٢٨	X28	٠,٩٠٣	جيد	٣,٧١١	ناقص	٠,١٩٣	جيد	جيد
٢٩	X29	١,٥٣٧	جيد	٠,٩٧١	جيد	٠,١٦٥	جيد	جيد جدا
٣٠	X30	١,٤١٠	جيد	١٧,١٥٧	ناقص	٠,٢٠٢	جيد	جيد
٣١	X31	٠,٢٤٦	جيد	٢,٢٩٧	ناقص	٠,١٩٨	جيد	جيد
٣٢	X32	١,٢٦٩	جيد	١,٢٠٣	جيد	٠,١٦٠	جيد	جيد جدا
٣٣	X33	٣,٥٧٦	ناقص	٠,١١٣	جيد	٠,٠٠٢	جيد	جيد
٣٤	X34	-٠,٢٧٨	جيد	١,٤٢٥	جيد	٠,١٣٣	جيد	جيد جدا
٣٥	X35	١,٣٥٢	جيد	٢,٣٣٨	ناقص	٠,٠٩٢	جيد	جيد
٣٦	X36	٢,٩٦٥	ناقص	٣,٣٤١	ناقص	٠,١٣٢	جيد	ناقص
٣٧	X37	-٣,٦٥٢	ناقص	-٠,٣٥٢	ناقص	٠,٠٠٠	جيد	ناقص
٣٨	X38	٠,٩٠٧	جيد	٢,١٢٦	ناقص	٠,٤٣٣	ناقص	ناقص
٣٩	X39	١,١٩٩	جيد	١,٥٧٨	جيد	٠,٢٩٢	ناقص	جيد
٤٠	X40	٢,٥٣٠	ناقص	٠,٨٣٧	جيد	٠,١٦٩	جيد	جيد
٤١	X41	٢,٦٥٢	ناقص	٠,٦٣٨	جيد	٠,١٩٥	جيد	جيد
٤٢	X42	١,٨٧٨	جيد	٠,٧٨٠	جيد	٠,١٧٦	جيد	جيد جدا
٤٣	X43	١,٧٤٩	جيد	٢,٢١٢	ناقص	٠,٢٦٦	ناقص	ناقص

٤٤	X44	-٠,١٤٦	جيد	١,٨٢٩	جيد	٠,٠٧٩	جيد	جيد جدا
٤٥	X45	٢,٣٤٣	ناقص	٢,٩٢٥	ناقص	٠,١٤٦	جيد	ناقص
٤٦	X46	-٠,٦٧٦	جيد	١,٦٩٤	جيد	٠,١٨١	جيد	جيد جدا
٤٧	X47	-٣,٢٧٦	ناقص	-٠,٦٨٠	ناقص	٠,٠٠٠	جيد	ناقص
٤٨	X48	١,٠١٧	جيد	١,٢٨١	جيد	٠,٠٩١	جيد	جيد جدا
٤٩	X49	٠,٤٨٠	جيد	٠,٥٠٩	جيد	٠,٠١٠	جيد	جيد جدا
٥٠	X50	٠,٣٧٠	جيد	٠,٧٣٧	جيد	٠,٠٠٠	جيد	جيد جدا

واستنادًا إلى جدول التوصيف أعلاه، يمكن استنتاج أن التوصيف لكل بند اختبار متنوع للغاية. هناك بنود اختبار تستوفي معايير "جيد جدًا"، مما يعني أن هذه البنود ممتازة جدًا من حيث مستوى الصعوبة ودرجة التمييز والتخمين الزائف. بالإضافة إلى ذلك، هناك بنود اختبار تستوفي معايير "جيد"، مما يعني أن هذه البنود تحقق معيار واحد فقط غير متوافق معه، سواء كان ذلك متعلقًا بمستوى الصعوبة أو درجة التمييز أو التخمين الزائف. من ناحية أخرى، هناك أيضًا بنود اختبار لا تستوفي معايير البنود الجيدة وفقًا لتحليل بنود الاختبار بنظرية استجابة البنود، أو يمكن تصنيفها على أنها "ناقص"، حيث تحتوي هذه البنود على أكثر من معيار واحد غير متوافق معه. يتم تقديم ملخص للتوصيف مع البنود ذات الصلة في الجدول أدناه.

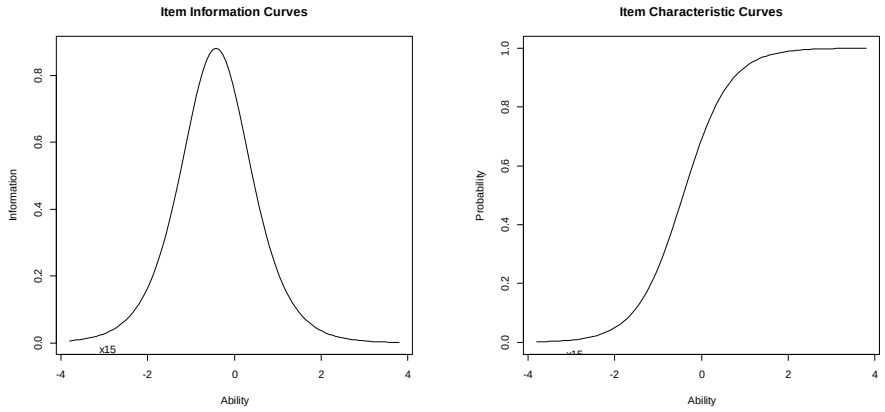
الجدول ٤,١٨: تصنيف البنود إلى المعايير الثلاثة

رقم	معايير	عدد الأسئلة	نسبة مئوية	بنود
١	جيد جدا	١٩	%٤٣,١٨	٢, ٥, ٦, ١٠, ١٢, ١٤, ١٥, ٢٠, ٢٢, ٢٦, ٢٩, ٣٢, ٣٤, ٤٢, ٤٤, ٤٦, ٤٨, ٤٩, ٥٠
٢	جيد	١٢	%٢٧,٢٧	٤, ١١, ١٧, ٢١, ٢٨, ٣٠, ٣١, ٣٣, ٣٥, ٣٩, ٤٠, ٤١
٣	ناقص	١٣	%٢٩,٥٤	٧, ٩, ١٦, ١٨, ١٩, ٢٤, ٢٥, ٣٦, ٣٧, ٣٨, ٤٣, ٤٥, ٤٧

بناءً على جداول توصيف البنود التي تم شرحها أعلاه، سيتم مناقشة بعض أمثلة على البنود الاختبارية من كل معيار مع النظر في منحنى الخصائص للبند (ICC) ومنحنى معلومات البند (IIC). يُمثل منحنى خصائص البند العلاقة بين احتمالية الإجابة الصحيحة ومستوى قدرة الجماهير. بينما يُمثل منحنى معلومات البند القدرة على مدى كفاءة بند الاختبار في قياس القدرة على مستوى معين ضمن نطاق القدرة.

أ) البنود من معيار "جيد جدا"

بنود الامتحان تنطبق عليها معايير "جيد جدا" هي التي تتضمن جميع المعاملات في حالة جيدة. فيما يلي شرح مفصل حول أمثلة على البنود التي تندرج ضمن معايير "جيد جدا"، وهي البنود رقم ١٥ و ٢٦ و ٤٤.

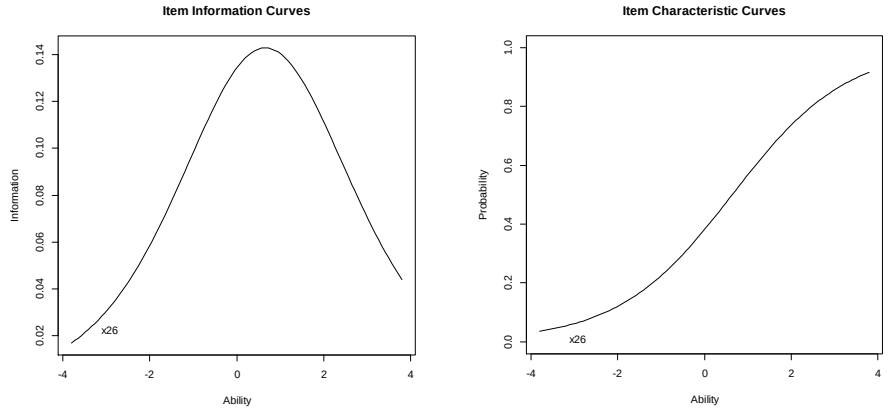


مثال أول من المجموعة التي تحمل معيار "جيد جدا" هو البند رقم ١٥. البند رقم ١٥ يتمتع بمستوى صعوبة جيد، حيث يبلغ قيمته -٠,٤٢٣. تقع هذه القيمة ضمن النطاق الممتد بين -٢ و٢، مما يشير إلى أن مستوى صعوبة هذا البند يتوسط النطاق. كلما تقدمنا يمينًا على منحنى الخصائص الخاص بالبند، زاد مستوى صعوبته. تشير القيمة -٠,٤٢٣ إلى أن القدرة الدنيا المطلوبة للإجابة على البند رقم ١٥ بشكل صحيح هي -٠,٤٢٣.

هذا البند يتميز أيضًا بقدرته التمييزية الجيدة، حيث تبلغ قيمته ١,٨٧٨. النطاق الممتاز لقيم درجة التمييز يمتد بين ٠ و٢. تشير القيم ضمن هذا النطاق إلى أن منحنى الخصائص الخاص بالبند يشكل نمطًا يشبه الحرف S بشكل مثالي

مع منحنى المعلومات الطبيعي. وهذا يشير إلى أن البند رقم ١٥ يتمتع بقدرة تمييزية جيدة.

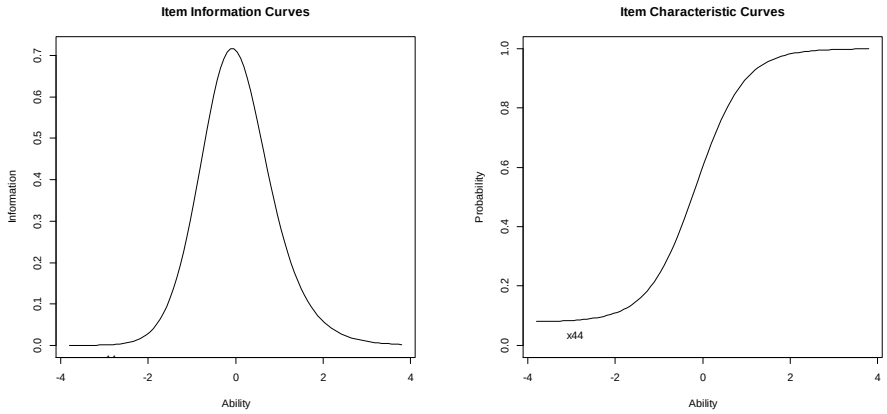
بالإضافة إلى ذلك، يتمتع هذا البند بقيمة تخمين زائف جيدة، حيث تبلغ قيمتها ٠. النطاق الممتاز لقيم التخمين الزائف يمتد بين ٠ و٠,٢٥. كلما كانت القيمة التخمينية الزائفة أقل، كلما كانت أقل احتمالية أن يتم الإجابة على هذا البند بشكل صحيح عن طريق التخمين فقط. وبالتالي، يتطلب هذا البند قدرة حقيقية للإجابة عليه بشكل صحيح، وليس مجرد تخمين.



مثال ثاني من المجموعة التي تحمل معيار "جيد جدا" هو البند رقم ٢٦. البند رقم ٢٦ يتمتع بمستوى صعوبة جيد، حيث يبلغ قيمته ٠,٦٤٩. تقع هذه القيمة ضمن النطاق الممتد بين −٢ و٢، مما يشير إلى أن مستوى صعوبة هذا البند يتوسط النطاق. كلما تقدمنا يمينًا على منحنى الخصائص الخاص بالبند، زاد مستوى صعوبته. تشير القيمة ٠,٦٤٩ إلى أن القدرة الأدنى المطلوبة للإجابة على البند رقم ٢٦ بشكل صحيح هي ٠,٦٤٩.

هذا البند يتميز أيضًا بقدرته التمييزية الجيدة، حيث تبلغ قيمته ٠,٧٥٨. النطاق الممتاز لقيم درجة التمييز يمتد بين ٠ و٢. تشير القيم ضمن هذا النطاق إلى أن منحنى الخصائص الخاص بالبند يشكل نمطًا يشبه الحرف S بشكل مثالي مع منحنى المعلومات الطبيعي. وهذا يشير إلى أن البند رقم ٢٦ يتمتع بقدرة تمييزية جيدة.

بالإضافة إلى ذلك، يتمتع هذا البند بقيمة تخمين زائف جيدة، حيث تبلغ قيمتها ٠,٠٠٣. النطاق الممتاز لقيم التخمين الزائف يمتد بين ٠ و٠,٢٥. كلما كانت القيمة التخمينية الزائفة أقل، كلما كانت أقل احتمالية أن يتم الإجابة على هذا البند بشكل صحيح عن طريق التخمين فقط. وبالتالي، يتطلب هذا البند قدرة حقيقية للإجابة عليه بشكل صحيح، وليس مجرد تخمين.



مثال ثالث من المجموعة التي تحمل معيار "جيد جدا" هو البند رقم ٤٤. البند رقم ٤٤ يتمتع بمستوى صعوبة جيد، حيث يبلغ قيمته -٠,١٤٦. تقع هذه القيمة ضمن النطاق الممتد بين -٢ و٢، مما يشير إلى أن مستوى صعوبة هذا البند يتوسط النطاق. كلما تقدمنا يمينًا على منحنى الخصائص الخاص بالبند، زاد مستوى صعوبته. تشير القيمة -٠,١٤٦ إلى أن القدرة الأدنى المطلوبة للإجابة على البند رقم ٤٤ بشكل صحيح هي -٠,١٤٦.

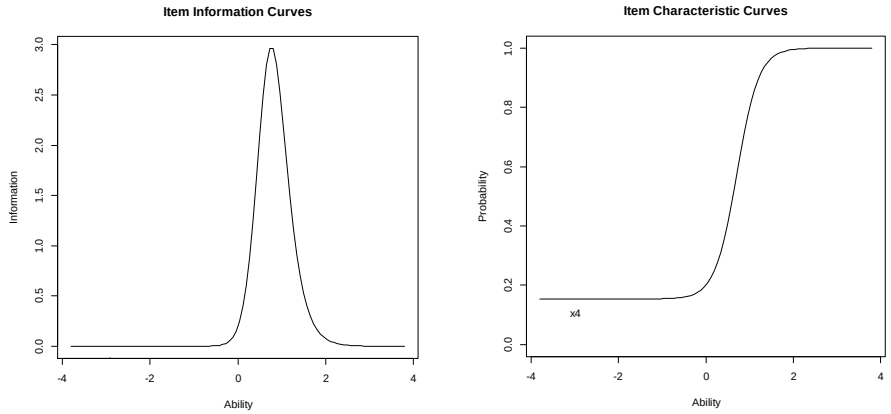
هذا البند يتميز أيضًا بقدرته التمييزية الجيدة، حيث تبلغ قيمته ١,٨٢٩. النطاق الممتاز لقيم درجة التمييز يمتد بين ٠ و٢. تشير القيم ضمن هذا النطاق إلى أن منحنى الخصائص الخاص بالبند يشكل نمطًا يشبه الحرف S بشكل مثالي مع منحنى المعلومات الطبيعي. وهذا يشير إلى أن البند رقم ٤٤ يتمتع بقدرة تمييزية جيدة.

بالإضافة إلى ذلك، يتمتع هذا البند بقيمة تخمين زائف جيدة، حيث تبلغ قيمتها ٠,٠٧٩. النطاق الممتاز لقيم التخمين الزائف يمتد بين ٠ و٠,٢٥. كلما

كانت القيمة التخمينية الزائفة أقل، كلما كانت أقل احتمالية أن يتم الإجابة على هذا البند بشكل صحيح عن طريق التخمين فقط. وبالتالي، يتطلب هذا البند قدرة حقيقية للإجابة عليه بشكل صحيح، وليس مجرد تخمين.

٢) البنود من معيار "جيد"

بنود الامتحان التي تنطبق عليها معايير "جيد" هي تلك التي تحتوي على معامل واحد غير جيد. فيما يلي سيتم شرح تفصيلي لأمثلة على البنود التي تندرج ضمن معايير "جيد"، وهي البنود رقم ٤، ٣٣، و٣٩.

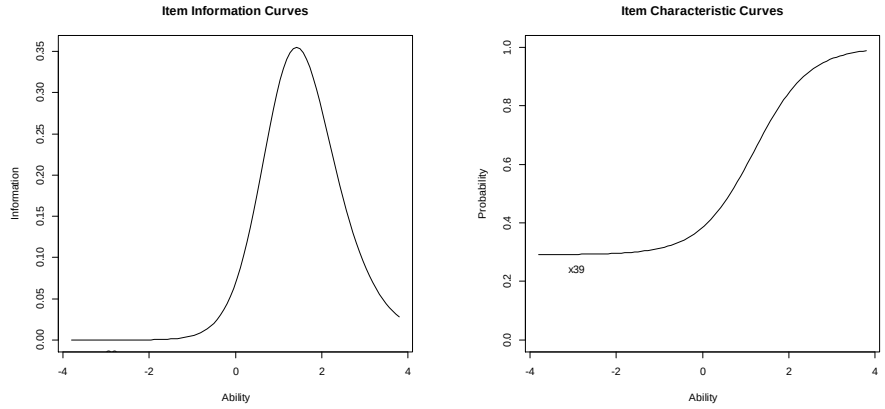


مثال أول من المجموعة التي تحمل معيار "جيد" هو البند رقم ٤. البند رقم ٤ يتمتع بمستوى صعوبة جيد، حيث يبلغ قيمته ٠,٧١١. تقع هذه القيمة ضمن النطاق الممتد بين −٢ و٢، مما يشير إلى أن مستوى صعوبة هذا البند يتوسط النطاق. كلما تقدمنا يمينًا على منحنى الخصائص الخاص بالبند، زاد مستوى صعوبته. تشير القيمة ٠,٧١١ إلى أن القدرة الأدنى المطلوبة للإجابة على البند رقم ٤ بشكل صحيح هي ٠,٧١١.

هذا البند يتميز بقدرة تمييزية غير جيدة، حيث تبلغ قيمته ٤,٠٠٤. النطاق الممتاز لقيم درجة التمييز يمتد بين ٠ و٢. القيم ضمن هذا النطاق تشير إلى أن منحنى الخصائص الخاص بالبند يشكل نمطًا يشبه الحرف S بشكل مثالي مع منحنى المعلومات الطبيعي. القيمة العالية لدرجة التمييز تشير إلى أن منحنى البند

مفرطًا في الانحراف أو يظهر منحنى حادًا بشكل زائد. هذا يشير إلى أن البند رقم ٤ يمتلك قدرة تمييزية غير جيدة.

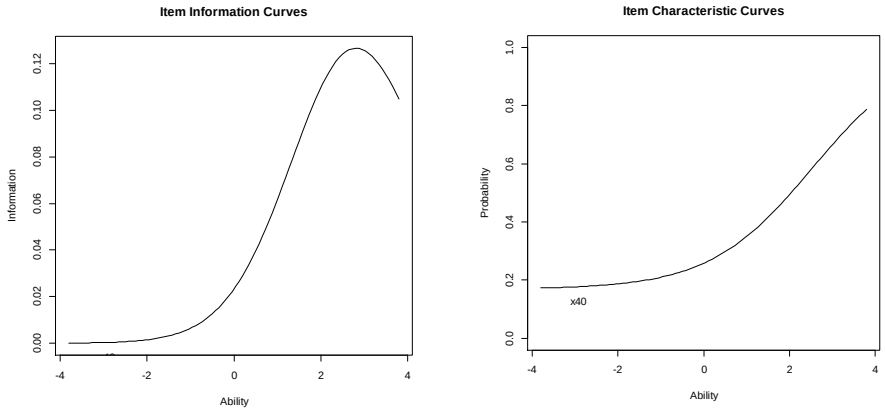
بالإضافة إلى ذلك، يتمتع هذا البند بقيمة تخمين زائف جيدة، حيث تبلغ قيمتها ٠,١٥٤. النطاق الممتاز لقيم التخمين الزائف يمتد بين ٠ و٠,٢٥. كلما كانت القيمة التخمينية الزائفة أقل، كلما كانت أقل احتمالية أن يتم الإجابة على هذا البند بشكل صحيح عن طريق التخمين فقط. وبالتالي، يتطلب هذا البند قدرة حقيقية للإجابة عليه بشكل صحيح، وليس مجرد تخمين.



مثال ثالث من المجموعة التي تحمل معيار "جيد جدا" هو البند رقم ٣٩. البند رقم ٣٩ يتمتع بمستوى صعوبة جيد، حيث يبلغ قيمته ١,١٩٩. تقع هذه القيمة ضمن النطاق الممتد بين −٢ و٢، مما يشير إلى أن مستوى صعوبة هذا البند يتوسط النطاق. كلما تقدمنا يمينًا على منحنى الخصائص الخاص بالبند، زاد مستوى صعوبته. تشير القيمة ١,١٩٩ إلى أن القدرة الدنيا المطلوبة للإجابة على البند رقم ١٥ بشكل صحيح هي ١,١٩٩.

هذا البند يتميز أيضًا بقدرته التمييزية الجيدة، حيث تبلغ قيمته ١,٥٧٨. النطاق الممتاز لقيم درجة التمييز يمتد بين ٠ و٢. تشير القيم ضمن هذا النطاق إلى أن منحنى الخصائص الخاص بالبند يشكل نمطًا يشبه الحرف S بشكل مثالي مع منحنى المعلومات الطبيعي. وهذا يشير إلى أن البند رقم ٣٩ يتمتع بقدرة تمييزية جيدة.

بالإضافة إلى ذلك، حيث تبلغ قيمته ٠,٢٩٢. النطاق الممتاز لقيم التخمين الزائف يمتد بين ٠ و٠,٢٥. كلما كانت قيمة التخمين الزائف أقل، زادت احتمالات أن يتم الإجابة على هذا البند بشكل صحيح عن طريق التخمين. وبالتالي، هناك احتمالية للمشاركين في الاختبار للإجابة على هذا البند بشكل صحيح فقط من خلال تخمين الخيارات المتاحة.



مثال ثاني من المجموعة التي تحمل معيار "جيد" هو البند رقم ٤٠. البند رقم ٤٠ يتمتع بمستوى صعوبة غير جيدة، حيث يبلغ قيمته ٢,٥٣٠. تقع هذه القيمة ضمن النطاق الممتد بين −٢ و٢. كلما تقدم يمينًا على منحنى الخصائص الخاص بالبند، زاد مستوى صعوبته. تشير القيمة ٢,٥٣٠ إلى أن القدرة الأدنى المطلوبة للإجابة على البند رقم ٤٠ بشكل صحيح هي ٢,٥٣٠. القيمة العالية لمستوى الصعوبة تشير إلى أن البند أكثر صعوبةً بشكل زائد.

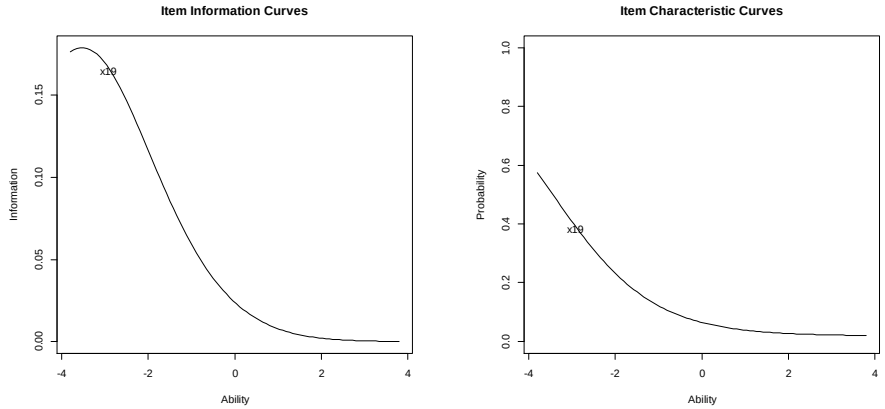
هذا البند يتميز أيضًا بقدرته التمييزية الجيدة، حيث تبلغ قيمته ٠,٨٣٨. النطاق الممتاز لقيم درجة التمييز يمتد بين ٠ و٢. وهذا يشير إلى أن البند رقم ٤٠ يتمتع بقدرة تمييزية جيدة.

بالإضافة إلى ذلك، يتمتع هذا البند بقيمة تخمين زائف جيدة، حيث تبلغ قيمتها ٠,١٦٩. النطاق الممتاز لقيم التخمين الزائف يمتد بين ٠ و٠,٢٥. كلما كانت القيمة التخمينية الزائفة أقل، كلما كانت أقل احتمالية أن يتم الإجابة

على هذا البند بشكل صحيح عن طريق التخمين فقط. وبالتالي، يتطلب هذا البند قدرة حقيقية للإجابة عليه بشكل صحيح، وليس مجرد تخمين.

٣) البنود من معيار "ناقص"

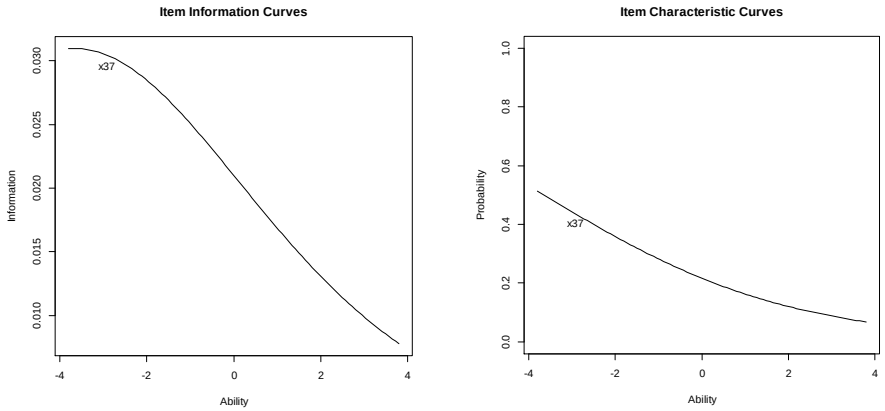
البنود التي تتوافق مع معيار "ناقص" هي تلك التي تحمل أكثر من معامل واحدة غير جيدة. فيما يلي شرح مفصل لأمثلة من البنود التي تتوافق مع معايير "ناقص"، وهي البنود رقم ١٩، ٣٧، و٤٧.



مثال أول من المجموعة التي تحمل معيار "ناقص" هو البند رقم ١٩. البند رقم ١٩ يتمتع بمستوى صعوبة منخفض جدًا، حيث تبلغ قيمته −٣,٤٨٥. تقع هذه القيمة خارج النطاق الذي يمتد من −٢ إلى ٢. كلما تقدمت يسارًا على منحنى الخصائص الخاص بالبند، قلت صعوبة البند. هذا يشير إلى أن القدرة الأدنى المطلوبة للإجابة على البند رقم ١٩ بشكل صحيح هي −٣,٤٨٥.

هذا البند أيضًا يتميز بقدرته التمييزية الضعيفة، حيث تبلغ قيمته −٠,٨٦٢. بينما يمتد النطاق الممتاز لقيم درجة التمييز بين ٠ و٢. هذا يشير إلى أن البند رقم ١٩ يتمتع بقدرة تمييزية ضعيفة. القيمة السالبة لدرجة التمييز تشير إلى أن هناك علاقة عكسية بين احتمالية الإجابة الصحيحة وبين قدرة المشترك. ببساطة، يمتلك المشتركون ذوو القدرات المنخفضة فرصة أعلى للإجابة بشكل صحيح.

على الرغم من ذلك، يتمتع هذا البند بقيمة تخمين زائف جيدة، حيث تبلغ قيمتها ٠,٠١٨. يمتد النطاق الممتاز لقيم التخمين الزائف بين ٠ و٠,٢٥. كلما كانت القيمة التخمينية الزائفة أقل، كلما كانت أقل احتمالية أن يتم الإجابة على هذا البند بشكل صحيح عن طريق التخمين فقط.

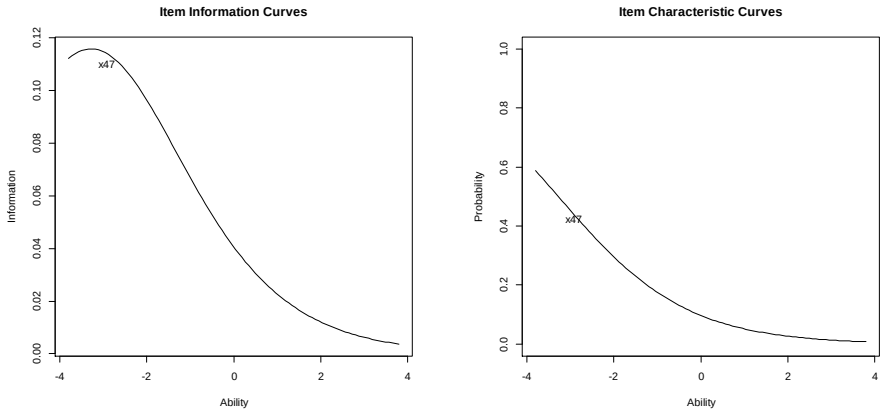


مثال ثاني من المجموعة التي تحمل معيار "ناقص" هو البند رقم ٣٧. البند رقم ٣٧ يتمتع بمستوى صعوبة منخفض جدًا، حيث تبلغ قيمته −٣,٦٥٢. تقع هذه القيمة خارج النطاق الذي يمتد من −٢ إلى ٢. كلما تقدمت يسارًا على منحنى الخصائص الخاص بالبند، قلت صعوبة البند. هذا يشير إلى أن القدرة الأدنى المطلوبة للإجابة على البند رقم ٣٧ بشكل صحيح هي −٣,٦٥٢.

هذا البند أيضًا يتميز بقدرته التمييزية الضعيفة، حيث تبلغ قيمته −٠,٣٥٢. بينما يمتد النطاق الممتاز لقيم درجة التمييز بين ٠ و٢. هذا يشير إلى أن البند رقم ٣٧ يتمتع بقدرة تمييزية ضعيفة. القيمة السالبة لدرجة التمييز تشير إلى أن هناك علاقة عكسية بين احتمالية الإجابة الصحيحة وبين قدرة المشترك. ببساطة، يمتلك المشتركون ذوو القدرات المنخفضة فرصة أعلى للإجابة بشكل صحيح.

على الرغم من ذلك، يتمتع هذا البند بقيمة تخمين زائف جيدة، حيث تبلغ قيمتها ٠,٠٠٠. يمتد النطاق الممتاز لقيم التخمين الزائف بين ٠ و٠,٢٥. كلما

كانت القيمة التخمينية الزائفة أقل، كلما كانت أقل احتمالية أن يتم الإجابة على هذا البند بشكل صحيح عن طريق التخمين فقط.

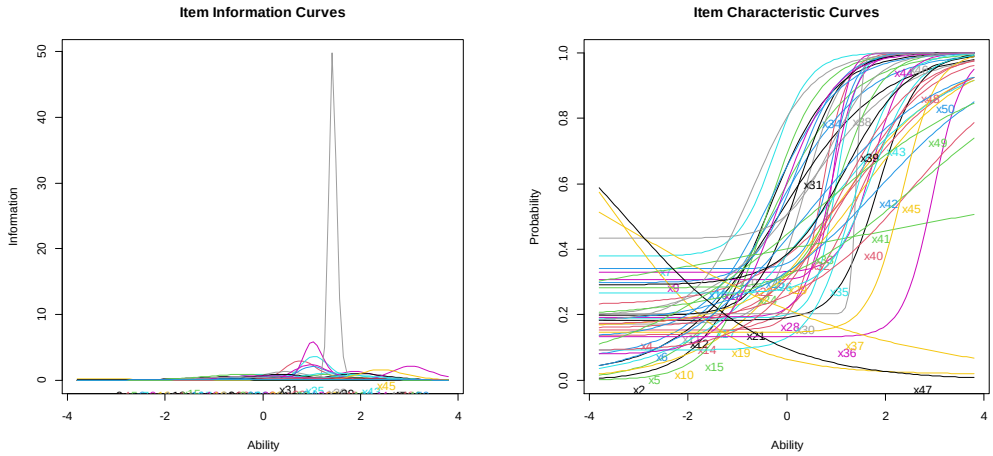


مثال ثالث من المجموعة التي تحمل معيار "ناقص" هو البند رقم ٤٧. البند رقم ٣٧ يتمتع بمستوى صعوبة منخفض جدًا، حيث تبلغ قيمته −٣,٢٧٦. تقع هذه القيمة خارج النطاق الذي يمتد من −٢ إلى ٢. كلما تقدمت يسارًا على منحنى الخصائص الخاص بالبند، قلت صعوبة البند. هذا يشير إلى أن القدرة الأدنى المطلوبة للإجابة على البند رقم ٤٧ بشكل صحيح هي −٣,٢٧٦.

هذا البند أيضًا يتميز بقدرته التمييزية الضعيفة، حيث تبلغ قيمته −٠,٦٨٠. بينما يمتد النطاق الممتاز لقيم درجة التمييز بين ٠ و٢. هذا يشير إلى أن البند رقم ٤٧ يتمتع بقدرة تمييزية ضعيفة. القيمة السالبة لدرجة التمييز تشير إلى أن هناك علاقة عكسية بين احتمالية الإجابة الصحيحة وبين قدرة المشترك. ببساطة، يمتلك المشتركون ذوو القدرات المنخفضة فرصة أعلى للإجابة بشكل صحيح.

على الرغم من ذلك، يتمتع هذا البند بقيمة تخمين زائف جيدة، حيث تبلغ قيمتها ٠,٠٠٠. يمتد النطاق الممتاز لقيم التخمين الزائف بين ٠ و٠,٢٥. كلما كانت القيمة التخمينية الزائفة أقل، كلما كانت أقل احتمالية أن يتم الإجابة على هذا البند بشكل صحيح عن طريق التخمين فقط.

بالإضافة إلى إمكانية إجراء تحليل على مستوى بنود الأسئلة للاختبار، يمكن أيضًا إجراء تحليل باستخدام نظرية استجابة البنود على مستوى الأداة نفسها. يتم تنفيذ التحليل على مستوى الأداة لتقييم جودة الاختبار بشكل عام من خلال مراعاة إلى أي مدى يمكن للأداة تقديم معلومات صادقة وثابتة ومفيدة في قياس القدرات أو الخصائص المرغوبة. الاستنتاجات التي يمكن استخلاصها من التحليل على مستوى الأداة تتعلق بمدى قدرة الاختبار على قياس شيء ما بدقة واعتمادية، وكذلك مدى قدرة الاختبار على تقديم نتائج قابلة للاعتماد لأغراض التقييم المحددة.



الرسم البياني ٤,١٢: منحنيات خصائص البند ومنحنيات معلومات البند

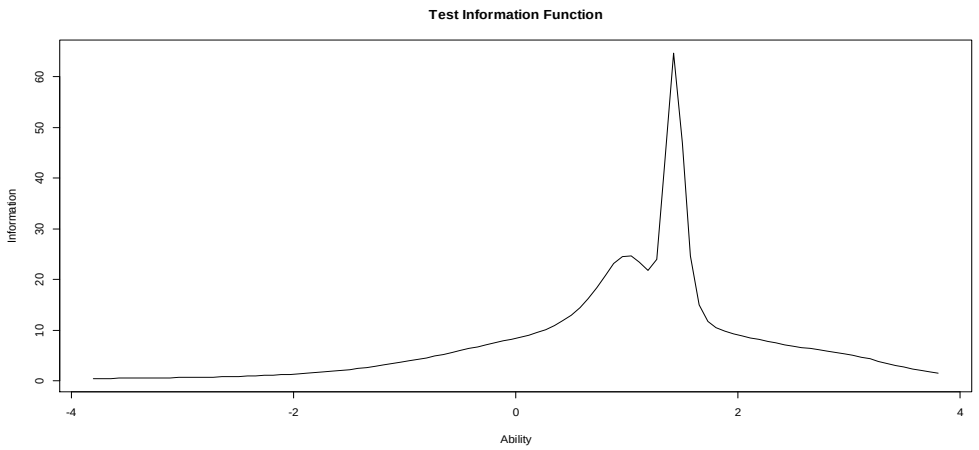
عند تفسير نتائج تحليل بنود الاختبار باستخدام نظرية استجابة البنود، يجب مراعاة منحنى خصائص البند (ICC) ومنحنى معلومات البند (IIC) اللذين تم دمجهما من كل بند متاح. يُمثل منحنى خصائص البند العلاقة بين احتمالية الإجابة الصحيحة ومستوى قدرة الجماهير. بينما يُمثل منحنى معلومات البند القدرة على مدى كفاءة بند الاختبار في قياس القدرة على مستوى معين ضمن نطاق القدرة.

عند دمج منحنيات خصائص البند ومنحنيات معلومات البند، يمكننا متابعة سلوك بند الاختبار عبر مدى القدرة. يقدم منحنى خصائص البند معلومات حول إمكانية الجماهير ذوي مستوى قدرة معين في الإجابة بشكل صحيح أو غير صحيح على

البند. بينما يشير منحنى معلومات البند إلى كمية المعلومات التي يقدمها بند الاختبار عند مستوى معين من القدرة.

عند النظر إلى منحنيات السمة للبنود أعلاه، يمكننا استنتاج أن كل بند ينتج منحنى متنوع. المنحنى المثلى للخصائص يجب أن يكون منحنى احتمال إيجابيًا. ومع ذلك، هناك بعض البنود التي تنتج منحنى احتمال سلبي، مما يشير إلى وجود مشكلة في هذه البنود. هناك أيضًا بنود تظهر أساسًا من الإحتمال يتجاوز ٠,٢٥، مما يشير إلى أنه من الممكن التنبؤ بإجابات هذه البنود بسهولة.

بينما عند مراجعة منحنيات معلومات البنود، يمكن أن نفهم أن كل بند أيضًا ينتج منحنى متنوع. المنحنى المثلى لمعلومات البنود يجب أن يكون بشكل يشبه جسم النطاقة مع شكل متناسب، ليس حادًا جدًا وليس أيضًا عريضًا جدًا. ولكن عند دمج هذه المنحنيات، هناك منحنى واحد يبرز بشكل كبير بذروة عالية جدًا وشكل حاد للغاية. هذا يشير إلى أن البند لديه خصائص مختلفة تمامًا عن البنود الأخرى. ببساطة، البند لديه قدرة فريدة وأعلى نسبيًا من البنود الأخرى، لأن منحناه حاد جدًا ومنحناه ينحدر بشدة.



الرسم البياني ٤,١٣: منحنى معلومات الاختبار

إليك تجميع لمنحنيات معلومات البنود التي تشكل منحنى معلومات الاختبار. منحنى معلومات الاختبار (Test Information Curve/TIC) في نظرية استجابة البنود

هو منحنى يقدم معلومات هامة في سياق أداة الاختبار بشكل شامل. بعض الأمور التي يمكن تفسيرها من منحنى معلومات الاختبار أعلاه تشمل:

نطاق القدرة: نطاق القدرة الذي يمكن قياسه بواسطة أداة الاختبار هذه يتراوح من −٢ إلى ما يقترب من ٤. منحنى معلومات الاختبار يوضح نطاق القدرة الذي يتم قياسه بواسطة الاختبار. يشمل هذا النطاق مستوى القدرة الذي يوفر الاختبار معلومات مفيدة حوله. كلما كان منحنى معلومات الاختبار أوسع، كلما كان نطاق القدرة الذي يمكن قياسه بواسطة الاختبار أكبر.

الحساسية للقدرة: يصف الاختبار بدقة أكبر قدرة المشتركين في الاختبار عند مستوى القدرة من ٠,٥ إلى ٢. هذا هو مستوى القدرة الذي يوفر الاختبار معلومات أكثر بكثير. نقطة ذروة منحنى معلومات الاختبار تشير إلى مستوى القدرة الذي يكون الاختبار فيه الأكثر حساسية. كلما ارتفعت نقطة ذروة منحنى معلومات الاختبار، كلما كان الاختبار يصف قدرة المشتركين في الاختبار بدقة أكبر في هذا النطاق. وعلى العكس، في الجزء العلوي من منحنى معلومات الاختبار الذي يكون فيه نقطة الذروة منخفضة، تقل دقة نتائج القياس ويزيد انحراف القياس القياسي.

ج. مقارنة تحليل بنود الأسئلة باستخدام النظرتين

في مناقشة المقارنة بين تحليل بنود الاختبار في نظرية الاختبار التقليدية ونظرية استجابة البنود، يُقدّم المقارنة في عدة جوانب، وهي:

١. تقدير الصدق

اختبار الصدق في نظرية الاختبار التقليدية ونظرية استجابة البنود له فروقات. تم إجراء اختبار الصدق لنظرية الاختبار التقليدية في هذا البحث باستخدام صيغة منتج اللحظة بيرسون.[١١٢] بينما تم إجراء اختبار الصدق في نظرية استجابة البنود في هذا البحث باستخدام تحليل العوامل التفسيرية الذي تم تنفيذه أيضًا كجزء من اختبار افتراض البعدية. يستخدم اختبار الصدق بمنتج اللحظة بيرسون لفحص العلاقة بين

[١١٢] Arifin, *Evaluasi Pembelajaran*; Azwar, *Reliabilitas Dan Validitas*.

درجة السؤال والدرجة الإجمالية التي يحصل عليها الطالب في الاختبار. بينما يقوم اختبار الصدق باستخدام تحليل العوامل التفسيرية بفحص ما إذا كان أداة الاختبار معيارية واحدة مهيمنة تقاس.[١١٣] من خلال اختبار الصدق باستخدام منتج اللحظة، تم التحقق من عدم صدق بعض الأسئلة. بينما أسفر تحليل العوامل التفسيرية الذي تم تنفيذه عن رسم بياني يوضح وجود عامل واحد مهيمن يتم قياسه في أداة الاختبار هذه.

الجدول ٤,١٩: مقارنة تقدير الصدق بين نظرية الاختبار التقليدية ونظرية استجابة البنود

رقم	جوانب	نظرية الاختبار التقليدية	نظرية استجابة البنود
١	طريقة	منتج اللحظة بيرسون	التحليل العوامل التفسيرية
٢	هدف	تقدير علاقة بين درجة السؤال والدرجة الإجمالية	تقدير عدد العوامل المتضمن في أداة الاختبار
٣	نتيجة	هناك بعض البنود غير صادق	هناك عامل واحد مهيمن في الاختبار

٢. تقدير الثبات

فيما يتعلق بقياس الثبات، تم حساب الثبات في تحليل بنود الاختبار باستخدام نظرية الاختبار التقليدية باستخدام صيغة ألفا كرونباخ. في الوقت نفسه، تم تقييم الثبات في تحليل بنود الاختبار باستخدام نظرية استجابة البنود من خلال منحنى معلومات الاختبار. يتم قياس الثبات في نظرية الاختبار التقليدية على مستوى الأداة باستخدام معيار الخطأ في القياس نفسه لكل مشاركي الاختبار. وهذا يؤدي إلى استنتاج بشأن ثبات الأداة نفسه، والذي عادة ما يتم قياسه وفقًا لمعايير معينة، عادة ما تكون أكبر من ٠,٧، مما يشير إلى مستوى ثبات جيد. من ناحية أخرى، في نظرية استجابة البنود، الثبات سيختلف بالنسبة لكل فرد مشاركي الاختبار، اعتمادًا على

[١١٣] Lan Luo, Cara Arizmendi, and Kathleen M. Gates, "Exploratory Factor Analysis (EFA) Programs in R," *Structural Equation Modeling: A Multidisciplinary Journal* 26, no. 5 (September 3, 2019): 819–26, https://doi.org/10.1080/10705511.2019.1615835.

مستوى قدرته، وهو ما بين في منحنى معلومات الاختبار. وكذلك مع معيار الخطأ في القياس، الذي يُعتقد أنه مختلف لكل فرد. وعلى الرغم من ذلك، هناك أيضًا أبحاث في تحليل أسئلة الاختبار باستخدام نظرية استجابة الفرد تستخدم فيها طريقة قياس الثبات بواسطة معادلة الألفا، تمامًا كما يُستخدم في قياس الصدق في نظرية الاختبار التقليدية.

الجدول ٤,٢٠ : مقارنة تقدير الثبات بين نظرية الاختبار التقليدية ونظرية استجابة البنود

جوانب	نظرية الاختبار التقليدية	نظرية استجابة البنود
طريقة	باستخدام صيغة ألفا كرونباخ	من خلال منحنى معلومات الاختبار
نتيجة	مقَدَّر على مستوى أداة الاختبار	مختلف لكل فرد وفقا لقدرته

٣. المعاملات المقَدّرة

المعاملات المقَدّرة في هذا البحث تختلف بين نظرية الاختبارات التقليدية ونظرية استجابة البنود الحديثة. في تحليل باستخدام نظرية الاختبارات التقليدية، يتم تقدير معامل مستوى الصعوبة ودرجة التمييز. عمومًا، تشمل تحليل البنود باستخدام النهج التقليدي ثلاث معاملات، وهي مستوى الصعوبة ودرجة التمييز، وفعالية خيارات الإجابة. ومع ذلك، في هذا البحث، تم تقدير معاملي مستوى الصعوبة ودرجة التمييز فقط بناءً على البيانات التي حصل عليها الباحث والتي تتضمن درجات ولا تتضمن بيانات الاختيارات الخام لإجابات الطلاب بصيغة A، B، C، أو D.

وذلك بسبب تنفيذ الاختبار باستخدام الحاسوب، مما يعني أن اختيارات الإجابة بين الطلاب تُرتب بشكل عشوائي. وفقط إذا أجاب المشترك بشكل صحيح، ستزيد درجته، دون مراعاة الاختيار الذي قاموا به على الشاشة أثناء الاختبار. من ناحية أخرى، في تحليل باستخدام نظرية استجابة البنود، تُشمل ثلاث معاملات، وهي مستوى الصعوبة ودرجة التمييز، والتخمين الزائف. يُستند استخدام هذه الثلاث معاملات إلى نموذج المعاملات اللوجستية المقَتَرَح بناءً على نتائج اختبار تطابق

النموذج. ونظرًا لأن نموذج المعاملات اللوجستية الثلاثي هو الأكثر تناسبًا مع البيانات، فإن هناك ثلاث معاملات تُقَدَّر.

الجدول ٤,٢١: مقارنة المعاملات المستخدمة في التحليل

المعامل	نظرية الاختبار التقليدية	نظرية استجابة البنود
الصعوبة	✓	✓
التمييز	✓	✓
التخمين	–	✓

٤. مقارنة نتائج التحليل باستخدام النظريتين

بعد إجراء التحليل، سيتم عرض البيانات لتوضيح الاختلافات بين نتائج التحليل باستخدام النظريتين. أدناه يتم تقديم جدول مقارنة بين نتائج تقدير معاملات التي تمت باستخدام نظرية الاختبار التقليدية ونظرية استجابة البنود.

الجدول ٤,٢٢: مقارنة كل معاملات بين نظرية الاختبار التقليدية ونظرية استجابة البنود

رقم	بنود	نظرية الاختبار التقليدية		نظرية استجابة البنود		
		م.صعوبة	ق.تمييزية	م.صعوبة	ق.تمييزية	تخمين زائف
١	X1	٠,٥٢٧	٠,٣٢٣			
٢	X2	٠,٦١٧	٠,٤٥٠	-٠,٤١٨	١,٤٩٩	٠,٠٠٠
٣	X3					
٤	X4	٠,٣٩٦	٠,٤٨٣	٠,٧١١	٤,٠٠٤	٠,١٥٤
٥	X5	٠,٥٨٤	٠,٤٢٤	-٠,٣٣١	١,١٩٦	٠,٠٠٠
٦	X6	٠,٥٧٥	٠,٤٤٤	-٠,١٦٥	١,١١٣	٠,٠٦٥
٧	X7	٠,٧٤٧	٠,٣٩١	-٠,٢٥٨	٢,٧٣٣	٠,٣٧٩
٨	X8	٠,٨٦٩	٠,٢٣٧			
٩	X9	٠,٣٨٠	٠,١٧٠	١,٧٥٠	٣,١٦٣	٠,٣٢٩
١٠	X10	٠,٣٥٤	٠,٢٧٢	٠,٨٧٢	٠,٨٥٤	٠,٠٠١
١١	X11	٠,٤٩٨	٠,٥٣٤	٠,٣٨١	٢,٦٩٠	٠,١٧٢

١٢	X12	٠,٥٤٣	٠,٤٠٥	-٠,١٣٤	٠,٩٦٦	٠,٠١٧
١٣	X13	٠,٤٤٧	٠,٣٨١			
١٤	X14	٠,٢٩١	٠,٢٦٠	١,٤٧٠	١,٥٤٥	٠,١٣٨
١٥	X15	٠,٦٢٩	٠,٥١٤	-٠,٤٢٣	١,٨٧٨	٠,٠٠٠
١٦	X16	٠,٦١٠	٠,٤٥٠	٠,٢٢٦	٢,٢٧٧	٠,٢٩٩
١٧	X17	٠,٣٥٨	٠,٣٥٦	٠,٩٩٧	٤,٥٨٦	٠,١٩١
١٨	X18	٠,٤٥٠	٠,٣٦١	٠,٩٦٣	٦,٥٣٢	٠,٣٠٧
١٩	X19			-٣,٤٨٥	-٠,٨٦٢	٠,٠١٨
٢٠	X20	٠,٥٣٠	٠,٢٧٨	٠,٤٣٥	١,٠٣٧	٠,١٩١
٢١	X21	٠,٢٤٦	٠,٠٩٤	١,٨٦٠	٢,٤٣٨	٠,١٨٣
٢٢	X22	٠,٤١٥	٠,٢٥٣	١,٤٧٧	٠,٩٥٦	٠,٢٢٩
٢٣	X23					
٢٤	X24	٠,٤٢٨	٠,٢٩١	١,١٢٨	٢,٤١٨	٠,٢٨٣
٢٥	X25	٠,٤٩٢	٠,٣٢٩	٠,٩١٩	٤,٠٢٧	٠,٣٤١
٢٦	X26	٠,٤٠٢	٠,٢٩١	٠,٦٤٩	٠,٧٥٨	٠,٠٠٣
٢٧	X27	٠,٦٥٥	٠,٤١١			
٢٨	X28	٠,٣٨٣	٠,٤٩١	٠,٩٠٣	٣,٧١١	٠,١٩٣
٢٩	X29	٠,٣٥٨	٠,٢٢٨	١,٥٣٧	٠,٩٧١	٠,١٦٥
٣٠	X30	٠,٢٧٥	٠,١٧٧	١,٤١٠	١٧,١٥٧	٠,٢٠٢
٣١	X31	٠,٥٤٩	٠,٤٦٩	٠,٢٤٦	٢,٢٩٧	٠,١٩٨
٣٢	X32	٠,٣٦٤	٠,٣٠٤	١,٢٦٩	١,٢٠٣	٠,١٦٠
٣٣	X33			٣,٥٧٦	٠,١١٣	٠,٠٠٢
٣٤	X34	٠,٦٣٦	٠,٤٣٧	-٠,٢٧٨	١,٤٢٥	٠,١٣٣
٣٥	X35	٠,٢٣٦	٠,٢٨٠	١,٣٥٢	٢,٣٣٨	٠,٠٩٢
٣٦	X36			٢,٩٦٥	٣,٣٤١	٠,١٣٢
٣٧	X37			-٣,٦٥٢	-٠,٣٥٢	٠,٠٠٠

٣٨	X38	٠,٥٨١	٠,٢٧٧	٠,٩٠٧	٢,١٢٦	٠,٤٣٣
٣٩	X39	٠,٤٥٠	٠,٢٤٦	١,١٩٩	١,٥٧٨	٠,٢٩٢
٤٠	X40	٠,٢٨٤	٠,١٢٠	٢,٥٣٠	٠,٨٣٧	٠,١٦٩
٤١	X41	٠,٣٣٨	٠,١٥١	٢,٦٥٢	٠,٦٣٨	٠,١٩٥
٤٢	X42	٠,٣٥٨	٠,١٦٤	١,٨٧٨	٠,٧٨٠	٠,١٧٦
٤٣	X43	٠,٣٣٨	٠,١٥١	١,٧٤٩	٢,٢١٢	٠,٢٦٦
٤٤	X44	٠,٥٨٨	٠,٤٥٦	-٠,١٤٦	١,٨٢٩	٠,٠٧٩
٤٥	X45			٢,٣٤٣	٢,٩٢٥	٠,١٤٦
٤٦	X46	٠,٧٤٤	٠,٣٢٧	-٠,٦٧٦	١,٦٩٤	٠,١٨١
٤٧	X47			-٣,٢٧٦	-٠,٦٨٠	٠,٠٠٠
٤٨	X48	٠,٣٤٨	٠,٣٣٦	١,٠١٧	١,٢٨١	٠,٠٩١
٤٩	X49	٠,٤٥٣	٠,٢٩١	٠,٤٨٠	٠,٥٠٩	٠,٠١٠
٥٠	X50	٠,٤٤٧	٠,٣٠٤	٠,٣٧٠	٠,٧٣٧	٠,٠٠٠

كلتا نظريتي نظرية الاختبار التقليدي ونظرية استجابة البنود لديهما مزايا وعيوب في سياق استخدامهما كنهج في تحليل أداة الاختبار. بعد أن تم عرض مقارنة نتائج تحليل المعاملات من النظريتين في المنقاشة السابقة، سيتم في هذا الجزء دمج النتائج من النظريتين لإعطاء تقييم لكل بند اختبار. سيتم مراعاة كل نتيجة تقدير للمعامل لكل بند، وسيتم اتخاذ قرار بشأن ما إذا كان البند ملائمًا أم غير ملائم للاستخدام في الاختبارات المستقبلية.

نتائج التقييم للبنود التي تم دمج نتائج تحليليها من كلا النظريتين مُعرضة في الجدول أدناه. تم استبعاد البنود غير الصادقة في تحليل باستخدام نظرية الاختبار التقليدية والبنود غير المناسبة للنموذج في نظرية استجابة البنود تلقائياً من عملية التقييم. يُستند ذلك إلى افتراض أنه في هذا التحليل ليس هناك اعتبار لجانب جوهر البنود، وإنما تعتمد فقط على الحسابات الإحصائية بشكل تجريبي. إذا تبين فيما بعد أن البنود التي تم استبعادها تتمتع بجودة جوهرية جيدة، فسيكون ذلك خارج نطاق هذا البحث.

الجدول ٤,٢٣: تحديد البنود المستخدمة

بنود	نظرية الاختبار التقليدية		نظرية استجابة البنود			حاصل
	م.صعوبة	ق.تمييزية	م.صعوبة	ق.تمييزية	ت.زائف	
X1	متوسطة	مرتفعة				–
X2	متوسطة	مرتفعة	جيد	جيد	جيد	مستخدم
X3						–
X4	متوسطة	مرتفعة	جيد	ناقص	جيد	مستخدم
X5	متوسطة	مرتفعة	جيد	جيد	جيد	مستخدم
X6	متوسطة	مرتفعة	جيد	جيد	جيد	مستخدم
X7	سهلة	مرتفعة	جيد	ناقص	ناقص	غير مستخدم
X8	سهلة	منخفضة				–
X9	متوسطة	منخفضة	جيد	ناقص	ناقص	غير مستخدم
X10	متوسطة	منخفضة	جيد	جيد	جيد	مستخدم
X11	متوسطة	مرتفعة	جيد	ناقص	جيد	مستخدم
X12	متوسطة	مرتفعة	جيد	جيد	جيد	مستخدم
X13	متوسطة	مرتفعة				–
X14	صعبة	منخفضة	جيد	جيد	جيد	مستخدم
X15	متوسطة	مرتفعة	جيد	جيد	جيد	مستخدم
X16	متوسطة	مرتفعة	جيد	ناقص	ناقص	مستخدم
X17	متوسطة	مرتفعة	جيد	ناقص	جيد	مستخدم
X18	متوسطة	مرتفعة	جيد	ناقص	ناقص	غير مستخدم
X19			ناقص	ناقص	جيد	–
X20	متوسطة	منخفضة	جيد	جيد	جيد	مستخدم
X21	صعبة	منخفضة	جيد	ناقص	جيد	مستخدم
X22	متوسطة	منخفضة	جيد	جيد	جيد	مستخدم
X23						–

X24	متوسطة	منخفضة	جيد	ناقص	ناقص	غير مستخدم
X25	متوسطة	مرتفعة	جيد	ناقص	ناقص	غير مستخدم
X26	متوسطة	منخفضة	جيد	جيد	جيد	مستخدم
X27	متوسطة	مرتفعة				–
X28	متوسطة	مرتفعة	جيد	ناقص	جيد	مستخدم
X29	متوسطة	منخفضة	جيد	جيد	جيد	مستخدم
X30	صعبة	منخفضة	جيد	ناقص	جيد	غير مستخدم
X31	متوسطة	مرتفعة	جيد	ناقص	جيد	مستخدم
X32	متوسطة	مرتفعة	جيد	جيد	جيد	مستخدم
X33			ناقص	جيد	جيد	–
X34	متوسطة	مرتفعة	جيد	جيد	جيد	مستخدم
X35	صعبة	منخفضة	جيد	ناقص	جيد	غير مستخدم
X36			ناقص	ناقص	جيد	–
X37			ناقص	ناقص	جيد	–
X38	متوسطة	منخفضة	جيد	ناقص	ناقص	غير مستخدم
X39	متوسطة	منخفضة	جيد	جيد	ناقص	غير مستخدم
X40	صعبة	منخفضة	ناقص	جيد	جيد	مستخدم
X41	متوسطة	منخفضة	ناقص	جيد	جيد	مستخدم
X42	متوسطة	منخفضة	جيد	جيد	جيد	مستخدم
X43	متوسطة	منخفضة	جيد	ناقص	ناقص	غير مستخدم
X44	متوسطة	مرتفعة	جيد	جيد	جيد	مستخدم
X45			ناقص	ناقص	جيد	–
X46	سهلة	مرتفعة	جيد	جيد	جيد	مستخدم
X47			ناقص	ناقص	جيد	–
X48	متوسطة	مرتفعة	جيد	جيد	جيد	مستخدم
X49	متوسطة	منخفضة	جيد	جيد	جيد	مستخدم

X50	متوسطة	مرتفعة	جيد	جيد	مستخدم

هناك ١٢ سؤالًا تم استبعادها من التقييم، مما يترك ٣٨ سؤالًا. من بين الأسئلة ال ٣٨ المتبقية، هناك ٢٨ سؤالًا يستوفي المعايير للاستخدام، بينما ال ١٠ الأسئلة الباقية لا تستوفي المعايير بسبب وجود قيم معاملات غير جيدة. ومعظم هذه الأسئلة التي لا تستوفي المعايير هي أسئلة تحمل قيم معاملات اختبار عشوائي عالية وعدم تطابق معاملات درجة التمييز بين منهج نظرية الاختبار التقليدي ومنهج نظرية استجابة البنود.

الفصل الخامس

مناقشة نتائج البحث

استنادا من البيانات التي عرضها الباحث في مقارنة تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية ونظرية استجابة البنود لدرس اللغة العربية في المدرسة المتوسطة الإسلامية الحكومية باتو سيعرض الباحث مناقشة نتائج البحث فيما يلي:

أ. مناقشة تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية (CTT)

تعتبر تقنيات بصرية مناسبة لتقييم الطبيعية للبيانات في هذا البحث. في الهيستوغرام، يُعتبر البيانات موزعة بشكل طبيعي إذا كان الهيستوغرام الناتج يُظهر نمطًا يشبه الجرس المقلوب ومتماثلاً، ولا يكون مفرطًا في العرض أو رقة. وهذا يعني أن البيانات التي تقترب من القيمة المتوسطة أكثر عددًا مما تبتعد عنها. استنادًا إلى الهيستوغرام الناتج في هذه الدراسة، يُمكن استنتاج أن البيانات موزعة بشكل طبيعي، نظرًا لأن شكل الهيستوغرام يشبه الجرس الطبيعي.

تم إجراء اختبار الصدق باستخدام صيغة نسبة الترابط لبيرسون (Pearson's Product Moment) باستخدام تطبيق مايكروسوفت إكسل. [١١٤] اختبار الصدق الذي تم إجراؤه باستخدام الطريقة المذكورة سابقا، أسفر الحساب قيمة "r" الحساب أو قيمة الارتباط لكل بند. يتم مقارنة هذه القيمة مع القيمة "r" الجدول التي تمثل معيار اختبار الملائمة. [١١٥] إذا كانت القيمة "r" الحساب أكبر من القيمة "r" الجدول، فإن البند يعتبر صادقا. على العكس، إذا كانت القيمة "r" الحساب أصغر من القيمة "r" الجدول، فإن البند يُعتبر غير صادقا.[١١٦] أظهرت نتائج اختبار الصدق أنه من إجمالي ٥٠ سؤالًا تم تحليلها، ثبت صدق ٤٢ سؤالًا، بينما أظهرت ٨ أسئلة أخرى نتائج غير صادق. يمكن استنتاج أن ليس جميع أسئلة الاختبار تفي بشروط الصدق. هناك بعض أسئلة الاختبار التي تم اعتبارها غير

[١١٤] Arifin, *Evaluasi Pembelajaran*.
[١١٥] Saifuddin Azwar, *Reliabilitas Dan Validitas*, 4th ed. (Yogyakarta: Pustaka Pelajar, 2014).
[١١٦] Sugiyono, *Metode Penelitian Pendidikan; Pendekatan Kuantitatif, Kualitatif, Dan R&D*.

صادقة لأن قيمة r المحسوبة أقل من قيمة r في الجدول والتي تم تحديدها كمعيار في اختبار الصدق. بل وجد أيضًا أن بعض أسئلة الاختبار لديها قيم سلبية للقيمة الزوجية المحسوبة.

أما الثبات، تم قياس الثبات باستخدام صيغة ألفا كرونباخ.[١١٧] نتيجة اختبار الثبات باستخدام الصيغة ألفا كرونباخ أظهرت قيمة قدرها ٠,٨٦، مما يشير إلى أن هذا الأداة تتمتع بمستوى عال من الثبات. المعامل يقترب من الواحد، وكلما اقترب من الواحد زادت الثبات. كما تم شرحه سابقًا، يُعتبر نطاق معامل الثبات بين ٠,٦٠ و ٠,٨٠ مرتفعًا بما يكفي، بينما يُعتبر القيم التي تزيد عن ٠,٨٠ مرتفعة جدًا.[١١٨] لذا، بناءً على معامل الثبات الذي تم الحصول عليه، يمكن الاستنتاج بأن أداة الاختبار هذه تتمتع بمستوى عالٍ من الثبات في قياس قدرة طلاب اللغة العربية.

في مستوى الصعوبة، يمكن ملاحظة أن نسبة مستوى الصعوبة تهيمن عليها الأسئلة ذات مستوى صعوبة متوسط، حيث بلغ عددها ٣٥ سؤالًا أو ما يعادل ٧١٪ من الإجمالي. بينما تبلغ نسبة الأسئلة ذات مستوى صعوبة عالي ٢٣٪، مما يعني وجود ١١ سؤالًا بهذا المستوى. أما الأسئلة ذات مستوى صعوبة منخفض فتسهم بنسبة ٦٪ فقط من الإجمالي، وهي تعادل ٣ أسئلة. نتائج الاختبارات المذكورة، إذا مقارنتها بالمعايير، تشير إلى أن أداة الاختبار هذه لا تفي بمعايير جيدة فيما يتعلق بمستوى الصعوبة وفقًا لنظرية الاختبار التقليدية. نسب الأسئلة السهلة والمتوسطة والصعبة هي ٦٪ و٧١٪ و٢٣٪ على التوالي، وهذا يعني أن التوزيع لا يزال غير مثالي. هناك عدد قليل جدًا من الأسئلة التي تصنف كسهلة، بينما هناك الكثير من الأسئلة ذات مستوى صعوبة متوسط. من ناحية أخرى، فإن عدد الأسئلة الصعبة قد اقترب من التوزيع المثالي.

أما في درجة التمييز، يُعتبر مؤشر درجة التمييز جيدًا إذا كان لديه قيمة تساوي أو تفوق ٠,٣. مؤشر درجة التمييز للبنود ذات قيمة صغيرة يؤدي إلى عدم قدرتها على تمييز الطلاب ذوي القدرات العالية عن الطلاب ذوي القدرات المنخفضة.[١١٩] من الدائرة

---

[١١٧] Retnawati, "Reliabilitas Instrumen Penelitian."
[١١٨] Azwar, *Reliabilitas Dan Validitas*.
[١١٩] Retnawati, *Validitas, Reliabilitas, Dan Karakteristik Butir*.

الدائرية أعلاه، نعلم أن هناك ٤٧٪ من الأسئلة لديها قدرة تمييز عالية و٥٣٪ من الأسئلة لديها قدرة تمييز منخفضة.

ب. مناقشة تحليل بنود الأسئلة باستخدام نظرية استجابة البنود (IRT)

يمكن استخدام نهج نظرية استجابة البنود عندما تتوافق نماذج اللوجستيات مع البيانات الخاصة بالاختبار المراد دراسته.[١٢٠] أظهرت نتائج الاختبار أن النموذج اللوجستي ثلاثي المعاملات هو الأكثر ملائمة لهذه البيانات. يتميز هذا النموذج بقيمة AIC تبلغ ١٧٨٧٣,٦١، وهي القيمة الأدنى بين النماذج المتاحة. أظهرت نتائج هذا النموذج أن جميع البنود البالغ عددها ٤٨ بندًا مناسبة. على الجانب الآخر، النموذج ذو الاثنين المعاملتين مناسب لمجرد ٤٦ بندًا، والنموذج راش مناسب لمجرد ٣٩ بندًا. استنادًا إلى مقارنة قيم AIC، تم اتخاذ قرار باعتماد النموذج اللوجستي ثلاثي المعاملات لأنه يقدم قيمة AIC الأدنى ومناسب لمعظم بنود الاختبار. لذلك، يعتبر هذا النموذج هو الأكثر ملائمة لتحليل البنود في هذا البحث.

تم تحقق الصدق من خلال إثبات صدق البنية باستخدام تحليل العامل التفسيري. تحليل العامل التفسيري هو تقنية إحصائية تُستخدم لتحديد العلاقات بين المتغيرات في مجموعة بيانات معقدة.[١٢١]

الجدول ٥,١: الصدق لكل بنود باعتبار عامل التحميل في التحليل بنظرية استجابة البنود

معيار	عدد	نسبة	بنود
صادق	٣٨	٧٦%	٢، ٤، ٥، ٦، ٧، ٩، ١٠، ١١، ١٢، ١٤، ١٥، ١٦، ١٧، ١٨، ١٩، ٢٠، ٢١، ٢٤، ٢٥، ٢٦، ٢٨، ٢٩، ٣٠، ٣١، ٣٢، ٣٤، ٣٥، ٣٧، ٣٨، ٤٠، ٤٢، ٤٣، ٤٤، ٤٦، ٤٧، ٤٨، ٤٩، ٥٠
غير صادق	١٢	٢٤%	٢٢، ٣٣، ٣٩، ٤١، ٤٥، ١، ٣، ٨، ١٣، ٢٣، ٢٧، ٣٦

[١٢٠] Hambleton, Swaminathan, and Rogers, *Fundamentals of Item Response Theory*; Danni, Wahyuni, and Tauratiya, "ITEM RESPONSE THEORY APPROACH."

[١٢١] Purwanto, "ANALISIS FAKTOR: KONSEP, PROSEDUR UJI DAN INTERPRETASI."

يُعَتَبَرُ البنود التي لديها معامل التحميل أكبر من ٠,٣٥ بأنها بنود صادقة، وعددها ٣٨ بنداً. ومن بين البنود الصادقة، يقيس ٢٨ بنداً العامل الأوَّل، أي حوالي ٧٤٪. في حين أن ١٠ البنود المتبقية، أي حوالي ٢٦٪، تُوزَع بين العَوَامِل الأُخرَى. يتم تقديم جدول يوضح توزيع معاملات التحميل لكل بند كما يلي. يُقدَّمُ فيما يلي جدول تَوزيع البُنودِ صادقة والبُنودِ غير صادقة.

والثبات في نظرية استجابة البنود يمكن مراقبتها من خلال منحنى معلومات الاختبار الذي يتم إنتاجه. من خلال هذا المنحنى، يمكن مشاهدة مستوى دقة الأداة في إجراء القياس على مستويات مختلفة من القدرات. وبعبارة أخرى، تقدم هذه الأداة معلومات مختلفة للمشتركين في الاختبار على مستويات مهارات مختلفة. إذا كان قمة المنحنى مرتفعة، فهذا يشير إلى أن الأداة تقدم الكثير من المعلومات وقادرة على قياس مهارات المشتركين في الاختبار بشكل ممتاز. على النقيض، إذا كانت قمة المنحنى منخفضة، فهذا يشير إلى أن الأداة غير دقيقة بشكل كافي في قياس المهارات. من المهم أن نلاحظ أنه في حزمة ltm المستخدمة في هذا البحث، لا يمكن عرض الانحراف المعياري مع منحنى معلومات الاختبار. ومع ذلك، المعلومات المستمدة من منحنى معلومات الاختبار تزودنا بمعلومات حول الثبات للأداة في قياس قدرات الطلاب على مستويات متنوعة.

أما يتعلق عن الافتراضات، هناك ثلاثة افتراضات أساسية في نظرية استجابة البنود: الأحادية البُعد، وعدم الترابط المحلي، وثبات المعامل.[١٢٢] يمكن تحقيق افتراض الأحادية البعد عن طريق اتباع نهجين، أولهما من خلال فحص القيم الذاتية/إيجن (Eigen) والثاني من خلال مراقبة منحنى الانحدار (Scree plot).[١٢٣] وعدم الترابط المحلي هو افتراض يتطلب أن درجات بند الاختبار لا تعتمد على بنود أخرى. بمعنى آخر، إجابة على بند معين لا يجب أن تكون مرتبطة أو تؤثر على إجابات البنود الأخرى.[١٢٤] عدم توفر فرضية ثبات

---

[١٢٢] Hambleton, Swaminathan, and Rogers, *Fundamentals of Item Response Theory*; Retnawati, *Teori Respons Butir Dan Penerapannya: Untuk Peneliti, Praktisi Pengukuran Dan Pengujian, Mahasiswa Pascasarjana*; Furr, *Psychometrics*.

[١٢٣] Retnawati, *Validitas, Reliabilitas, Dan Karakteristik Butir*.

[١٢٤] Naga, *Pengantar Teori Sekor Pada Pengukuran Pendidikan*.

المعامل في معامل التخمين الزائف يؤثر على نتائج التحليل. يُقترح إزالة بعض الأسئلة التي تسبب عدم توفر هذه الفرضية في معامل التخمين الزائف، مع التركيز على الأسئلة التي تظهر فروقًا كبيرة بين مجموعات البيانات. هذا يهدف إلى تحسين دقة نتائج التحليل.[١٢٥]

مستوى صعوبة السؤال هو عامل رئيسي يؤثر على احتمالية شخص ما في الإجابة على السؤال بشكل صحيح.[١٢٦] القيمة المناسبة لمستوى الصعوبة تتراوح بين −٢ و ٢.[١٢٧] قيمة مستوى الصعوبة في نظرية استجابة البنود هي النقطة التي تكون فيها احتمالية الإجابة بشكل صحيح مساوية لـ ٠,٥ على محور الصعود (المحور الرأسي). من نتائج التحليل، هناك ٨ بنود من الأسئلة لديها مستوى الصعوبة غير جيدة.

درجة التمييز واحدة من العوامل المهمة في تقييم جودة البنود في الاختبار. البنود التي تمتلك درجة التمييز عالية تكون أكثر فعالية في تمييز الأفراد الذين يمتلكون مستويات متفاوتة من الكفاءة.[١٢٨] نطاق قيم درجة التمييز المثلى يتراوح بين ٠ و ٢. تمثل درجة التمييز في نظرية استجابة البنود بواسطة القيمة الألفا (αi).[١٢٩] تحتاج بعض بنود الاختبار إلى تحسين أو استبدال بسبب قيم تمييز غير كافية. القيم التي تقل عن ٠ تشير إلى فعالية ضعيفة، بينما القيم التي تزيد عن ٢ تشير إلى فعالية زائدة. يجب تعديل هذه البنود لتحسين جودة الاختبار. من نتائج التحليل، هناك ٢١ بنود من الأسئلة لديها درجة التمييز غير جيدة.

يُعد معامل التخمين الزائف هو المعامل الثالث الذي يمكن أن يؤثر على استجابة المشاركين لسؤال الاختبار. القيمة المثلى لمعامل التخمين الزائف لسؤال معين لا يجب أن تتجاوز $1/k$، حيث k هو عدد الخيارات المتاحة. في هذا السياق، بوجود أربع خيارات، القيمة المثلى لمعامل التخمين الزائف هي أقل من ٠,٢٥. كلما انخفضت القيمة المثلى

---

[١٢٥] Guenole and Brown, "The Consequences of Ignoring Measurement Invariance for Path Coefficients in Structural Equation Models."

[١٢٦] Furr, *Psychometrics*.

[١٢٧] Retnawati, *Teori Respons Butir Dan Penerapannya: Untuk Peneliti, Praktisi Pengukuran Dan Pengujian, Mahasiswa Pascasarjana*; Hambleton and Swaminathan, *Item Response Theory*.

[١٢٨] Embretson and Reise, *Item Response Theory for Psychologists*.

[١٢٩] Furr, *Psychometrics*; Retnawati, *Teori Respons Butir Dan Penerapannya: Untuk Peneliti, Praktisi Pengukuran Dan Pengujian, Mahasiswa Pascasarjana*.

لمعامل التخمين الزائف، زادت صعوبة تخمين الإجابة على السؤال وزادت جودة السؤال أيضًا. وفي نظرية استجابة البنود، يُمثل معامل التخمين الزائف بالرمز "ci".[١٣٠] من نتائج التحليل، هناك ٩ بنود من الأسئلة لديها التخمين الزائف غير جيد.

يتم تقديم ملخص للتوصيف مع البنود ذات الصلة في الجدول أدناه.

الجدول ٥,٢: تصنيف البنود إلى المعايير الثلاثة

رقم	معايير	عدد الأسئلة	نسبة مئوية	بنود
١	جيد جدا	١٩	%٤٣,١٨	٢, ٥, ٦, ١٠, ١٢, ١٤, ١٥, ٢٠, ٢٢, ٢٦, ٢٩, ٣٢, ٣٤, ٤٢, ٤٤, ٤٦, ٤٨, ٤٩, ٥٠
٢	جيد	١٢	%٢٧,٢٧	٤, ١١, ١٧, ٢١, ٢٨, ٣٠, ٣١, ٣٣, ٣٥, ٣٩, ٤٠, ٤١
٣	ناقص	١٣	%٢٩,٥٤	٧, ٩, ١٦, ١٨, ١٩, ٢٤, ٢٥, ٣٦, ٣٧, ٣٨, ٤٣, ٤٥, ٤٧

ج. مناقشة مقارنة تحليل بنود الأسئلة باستخدام النظريتين

كلتا نظريتي نظرية الاختبار التقليدي ونظرية استجابة البنود لديهما مزايا وعيوب في سياق استخدامهما كنهج في تحليل أداة الاختبار. بعد أن تم عرض مقارنة نتائج تحليل المعاملات من النظريتين في المنقاشة السابقة، سيتم في هذا الجزء دمج النتائج من النظريتين لإعطاء تقييم لكل بند اختبار. سيتم مراعاة كل نتيجة تقدير للمعامل لكل بند، وسيتم اتخاذ قرار بشأن ما إذا كان البند ملائمًا أم غير ملائم للاستخدام في الاختبارات المستقبلية.

نتائج التقييم للبنود التي تم دمج نتائج تحليليها من كلا النظريتين مُعرضة في الجدول أدناه. تم استبعاد البنود غير الصادقة في تحليل باستخدام نظرية الاختبار التقليدية والبنود غير المناسبة للنموذج في نظرية استجابة البنود تلقائياً من عملية التقييم. يُستند ذلك إلى افتراض أنه في هذا التحليل ليس هناك اعتبار لجانب جوهر البنود، وإنما تعتمد فقط على

[١٣٠] Furr, *Psychometrics*.

الحسابات الإحصائية بشكل تجريبي. إذا تبين فيما بعد أن البنود التي تم استبعادها تتمتع بجودة جوهرية جيدة، فسيكون ذلك خارج نطاق هذا البحث.

الجدول ٥,٣: تصنيف بنود مستخدم وغير مستخدم

رقم	معايير	عدد الأسئلة	نسبة مئوية	بنود
١	مستبعد	١٢	٢٤%	١، ٣، ٨، ١٣، ١٩، ٢٣، ٢٧، ٣٣، ٣٦، ٣٧، ٤٥، ٤٧
٢	مستخدم	٢٨	٥٦%	٢، ٤، ٥، ٦، ١٠، ١١، ١٢، ١٤، ١٥، ١٦، ١٧، ٢٠، ٢١، ٢٢، ٢٦، ٢٨، ٢٩، ٣١، ٣٢، ٣٤، ٤٠، ٤١، ٤٢، ٤٤، ٤٦، ٤٨، ٤٩، ٥٠
٣	غير مستخدم	١٠	٢٩,٥٤%	٧، ٩، ١٨، ٢٤، ٢٥، ٣٠، ٣٥، ٣٨، ٣٩، ٤٣،

هناك ١٢ سؤالًا تم استبعادها من التقييم، مما يترك ٣٨ سؤالًا. من بين الأسئلة ٣٨ المتبقية، هناك ٢٨ سؤالًا يستوفي المعايير للاستخدام، بينما ال ١٠ الأسئلة الباقية لا تستوفي المعايير بسبب وجود قيم معاملات غير جيدة. ومعظم هذه الأسئلة التي لا تستوفي المعايير هي أسئلة تحمل قيم معاملات اختبار عشوائي عالية وعدم تطابق معاملات درجة التمييز بين منهج نظرية الاختبار التقليدي ومنهج نظرية استجابة البنود.

الفصل السادس

الخاتمة

أ. ملخص نتائج البحث

بناءً على النتائج التي تم تقديمها بشكل منهجي في الفصل السابق، يقوم الباحث بصياغة استنتاجات من البحث المعنون "مقارنة تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية ونظرية استجابة البنود لدرس اللغة العربية في المدرسة المتوسطة الإسلامية الحكومية باتو" إلى عدة نقاط على النحو التالي.

١. إجراء تحليل البنود لدرس اللغة العربية باستخدام نظرية الاختبار التقليدية ونظرية استجابة البنود يختلفان. وبالمثل، تُظهر نتائج التحليل للأسئلة باستخدام نظرية الاختبار التقليدية ونظرية استجابة البنود اختلافًا، وخاصة فيما يتعلق بالمعاملات التي يتم تقديرها، سواء من حيث القيمة أو التصنيف، نظرًا لاختلاف المعايير المستخدمة في النظريتين. ومع ذلك، فيما يتعلق بكشف الأخطاء في الأسئلة، فإن كل منهما لا يفضي إلى اختلاف نتائج يعتبر ذلك ملحوظًا. يمكن استخدام نظريتي تحليل بنود الاختبار معًا للوصول إلى تحليل شامل لهذا الاختبار.

٢. تم دمج نتائج التفسير من التحليل باستخدام نظرية الاختبار التقليدية ونظرية استجابة البنود للوصول إلى استنتاجات تفسيرية بشأن بنود الاختبار في هذا الاختبار. من بين جميع الأسئلة، هناك ٥٦٪ من الأسئلة تظهر نتائج مرضية. لذا، من الضروري تحسين هذه الأسئلة قبل إعادة استخدامها بناءً على نتائج التحليل. العيوب لهذه أسئلة الاختبار تظهر في بعض الجوانب، وإذا تم تصنيفها من الأكثر أهمية فهي: درجة التمييز، فعالية خيارات الإجابة، صدق الأسئلة، ونسبة مستوى الصعوبة. للتغلب على هذه العيوب، ينبغي أن تؤخذ في الاعتبار نتائج هذا البحث وأن يتم أيضًا تقييم مضمون الأسئلة.

ب. التوصيات والاقتراحات

أما التوصيات والاقتراحات من البحث "مقارنة تحليل بنود الأسئلة باستخدام نظرية الاختبار التقليدية ونظرية استجابة البنود لدرس اللغة العربية في المدرسة المتوسطة الإسلامية الحكومية باتو" على النحو التالي.

١. ينبغي على مطوري أدوات التقييم، خاصة في سياق الاختبارات ذات الاختيارات المتعددة، من المهم جدًا فهم والاهتمام بمعاملات أسئلة الاختبار. هذا ضروري لإنشاء أداة تقييم عالية الجودة والتي يمكنها قياس قدرات الطلاب بدقة أكبر.

٢. يرجى للباحثين الآخرين الذين يهتمون بمجال تحليل أسئلة الاختبار، يُوصى بأن يُجروا أبحاثًا ذات صلة باستخدام نهج نظري مختلف وأساليب حديثة أكثر تقدمًا. وبهذه الطريقة، يمكن للبحوث الإضافية أن تسهم في تطوير وتعميق الفهم في هذا المجال، وتقديم إسهامات في تقدم المعرفة العلمية.

قائمة المراجع

أولا: المصادر والمراجع العربية

القرني, أحمد سالم سعيد. ‘‘أثر طرق تقدير القدرة للأفراد على دقة معادلة الاختبارات في نظرية الاستجابة للمفردة.’’ *المجلة العربية للعلوم التربوية والنفسية* ٦ no. 30 , (October 1, 2022): 1–22. https://doi.org/10.21608/jasep.2022.265536.

دبوس, محمد. ‘‘استخدام نظرية االستجابة للفقرة في بناء فقرات اختبار محكي المرجع في الرياضيات بفقرات ثنائية التدريج ومتعددة التدريج وفق النموذج اللوجستي ثنائي المعلم.’’ *مجلة جامعة النجاح لأبحاث (العلوم الإنسانية)* ٣٠ .(2016) 7 no. , https://journals.najah.edu/media/journals/full_texts/7_B2fCogF.pdf.

علام, رجاء محمود أبو. *تقويم التعلم*. دار المسيرة للطباعة والنشر, ٢٠٠٥.

علي, مصطفى علي خلف. ‘‘استخدام نظرية الاستجابة للمفردة الاختبارية في بناء مقياس الاتجاه نحو مهنة التدريس.’’ *مجلة البحث في التربية وعلم النفس* ٣٤ 1 no. , (January 1, 2019): 1–47. https://doi.org/10.21608/mathj.2019.82386.

عليمات, مهدي محمد. ‘‘بناء اختبار محكي المرجع في الرياضيات وفق النموذج اللوجستي ثلاثي المعلمة.’’ *مجلة جامعة القدس المفتوحة للأبحاث والدراسات التربوية والنفسية* ١٣.004-038-013-1182/10.33977/https://doi.org (April 19, 2022). 38 no. ,

عودة, أحمد. *القياس والتقويم في العملية التدريسية*. شركة الأمل للطباعة والنشر والتوزيع, ٢٠٠٥.

عواضة, هاشم. *تقويم التعلم*. بيروت: دار العلم للملايين, ٢٠١٠

فريدة، شريفة، ومحمد. تحليل بنود الأسئلة في الإمتحان النهائي لمادة اللغة العربية في الفصل العاشر بالمدرسة العالية ‘الإسلام’ كموجا *AL-MU’ARRIB: JOURNAL OF* ”. *ARABIC EDUCATION* 1, no. 2 (November 12, 2021): 1–15. https://doi.org/10.32923/al-muarrib.v1i2.2064.

نورا الشامخ. *التقويم في التعليم*. الألوكة, ٢٠١٨

ثانيا: المصادر والمراجع الأجنبية

Adam, Muhammad Zikran, Muhammad Jundi, Ibadurrahman Ali, Siti Aliyya Laubaha, and Suleman D. Kadir. "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, no. 1 (July 2023).

Alfarisa, Fitri, and Dian Purnama. "ANALISIS BUTIR SOAL ULANGAN AKHIR SEMESTER MATA PELAJARAN EKONOMI SMA MENGGUNAKAN RASCH MODEL." *Jurnal Pendidikan Ekonomi* 11, no. 2 (2019).

Ali, Ibadurrahman, Muhammad Jundi, Muhammad Zikran Adam, Siti Aliyya Laubaha, and Suleman D. Kadir. "IMPLEMENTASI KMA 183 DALAM PENYAJIAN MATERI BAHASA ARAB BERBASIS E-LEARNING DI MAN 1 KABUPATEN GORONTALO." *PIWULANG Jurnal Pendidikan Agama Islam* 4, no. 2 (2022).

Allen, Mary J. *Introduction to Measurement Theory*. Monterey, Calif: Brooks/Cole Pub. Co., 1979.

Amelia, Rizki Nor, and Kriswantoro Kriswantoro. "Implementation of Item Response Theory for Analysis of Test Items Quality and Students' Ability in Chemistry." *JKPK (Jurnal Kimia dan Pendidikan Kimia)* 2, no. 1 (May 1, 2017): 1. https://doi.org/10.20961/jkpk.v2i1.8512.

American Educational Research Association, National Council on Measurement in Educatio, and American Psychological Association. *Standards for Educational and Psychological Testing*. Washington DC: American Educational Research Association, 2014. https://www.testingstandards.net/uploads/7/6/6/4/76643089/9780935302356.pdf.

Andayani, Asri, Purwanto, and Taufik Ramlan Ramalis. "Kajian Implementasi Teori Respon Butir Dalam Menganalisis Instrumen Tes Materi Fisika." *Prosiding Seminar Nasional Fisika (SiNaFi) 5.0* 1, no. 1 (2019).

Andriani, Rina Suci. "LANGUAGE TESTS." *Jurnal Educate* 3, no. 2. Accessed September 1, 2023. https://media.neliti.com/media/publications/241552-tes-kebahasan-9bda9d9f.pdf.

ArcGIS, Konsultan Data Penelitian &. "Mengenal Lebih Jauh Tentang Analisis Faktor Eksploratori." *Konsultan Analis Data Penelitian dan Peta Digital* (blog), August 24, 2020. https://patrastatistika.com/analisis-faktor-eksploratori/.

Arifin, Zainal. *Evaluasi Pembelajaran*. Jakarta: Direktorat Jenderal Pendidikan Islam Kementerian Agama RI, 2012.

Arikunto, Suharsimi. *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara, 2009.

Azwar, Saifuddin. *Reliabilitas Dan Validitas*. 4th ed. Yogyakarta: Pustaka Pelajar, 2014.

Brennan, Robert L. *Educational Measurement*. Praeger Publishers, 2006.

Chizanah, Lu'luatul, and M. Noor Rochman Hadjam. "Validitas Konstruk Ikhlas: Analisis Faktor Eksploratori Terhadap Instrumen Skala Ikhlas." *Jurnal Psikologi UGM* 38, no. 2 (2011): 129023.

Danni, Rahmat, Ajeng Wahyuni, and Tauratiya Tauratiya. "ITEM RESPONSE THEORY APPROACH: KALIBRASI BUTIR SOAL PENILAIAN AKHIR SEMESTER MATA PELAJARAN BAHASA ARAB." *Arabi : Journal of Arabic Studies* 6, no. 1 (June 30, 2021): 93–104. https://doi.org/10.24865/ajas.v6i1.320.

Embretson, Susan E., and Steven P. Reise. *Item Response Theory for Psychologists*. Item Response Theory for Psychologists. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, 2000.

Fahmi, Bahrudin, Syahrul Rizqi, and Nurul Elmira Harmeilinda. "Analisis Butir Soal Bahasa Arab MAS Pondok Pesantren Assalam Kampar Riau." *Ta'lim al-'Arabiyyah: Jurnal Pendidikan Bahasa Arab & Kebahasaaraban* 6, no. 1 (June 1, 2022): 95–105. https://doi.org/10.15575/jpba.v6i1.16193.

Farida, Syarifah, and Muhamad. "تحليل بنود الأسئلة في الإمتحان النهائي لمادة اللغة العربية في الفصل العاشر بالمدرسة العالية 'الإسلام' كموجا." *AL-MU'ARRIB: JOURNAL OF ARABIC EDUCATION* 1, no. 2 (November 12, 2021): 1–15. https://doi.org/10.32923/al-muarrib.v1i2.2064.

Furr, R. Michael. *Psychometrics: An Introduction*. 4th ed. Thousand Oaks, California: SAGE Publications, 2021.

Guenole, Nigel, and Anna Brown. "The Consequences of Ignoring Measurement Invariance for Path Coefficients in Structural Equation Models." *Frontiers in Psychology* 5 (September 17, 2014). https://doi.org/10.3389/fpsyg.2014.00980.

Hair, Joseph F., William C. Black, Barry J. Babin, and Rolph E. Anderson. *Multivariate Data Analysis*. 7th ed. Upper Saddle River, New Jersey: Prentice-Hall, Inc., 2010.

Hambleton, Ronald K., and Russell W. Jones. "An NCME Instructional Module on: Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test Development." *Educational Measurement: Issues and Practice* 12, no. 3 (October 25, 2005): 38–47. https://doi.org/10.1111/j.1745-3992.1993.tb00543.x.

Hambleton, Ronald K., and H. Swaminathan. *Item Response Theory: Principles and Applications*. Springer Science & Business Media, 1985.

Hambleton, Ronald K., Hariharan Swaminathan, and H. Jane Rogers. *Fundamentals of Item Response Theory*. Fundamentals of Item Response Theory. Thousand Oaks, CA, US: Sage Publications, Inc, 1991.

Himelfarb, Igor. "A Primer on Standardized Testing: History, Measurement, Classical Test Theory, Item Response Theory, and Equating." *Journal of Chiropractic Education* 33, no. 2 (October 1, 2019): 151–63. https://doi.org/10.7899/JCE-18-22.

Hula, Ibnu Rawandhy. "EVALUASI DAN TES BAHASA ARAB: Tinjauan Teori." *Language Development Center* 12, no. 7 (December 9, 2021). https://doi.org/10.31219/osf.io/uwt8x.

Huriaty, Dina. “METODE KALIBRASI DAN DESAIN TES BERDASARKAN TEORI RESPONS BUTIR (IRT).” *Math Didactic: Jurnal Pendidikan Matematika* 1, no. 3 (2015).

Jundi, Muhammad. “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 الأسلة في إندونيسيا.” *ATHLA: Journal of Arabic Teaching, Linguistic and Literature* 4, no. 2 (December 2023).

———. “Persepsi Mahasiswa terhadap Pelaksanaan Perkuliahan Naḥwu dengan Sistem Daring.” *Al-Ma‘rifah: Jurnal Budaya, Bahasa, dan Sastra Arab* 17, no. 2 (October 31, 2020): 105–18. https://doi.org/10.21009/almakrifah.17.02.01.

———. “تصميم برنامج أندرويد لتعزيز إتقان المفردات لدى طلاب المدرسة العالية.” *Lahjah Arabiyah: Jurnal Bahasa Arab Dan Pendidikan Bahasa Arab* 4, no. 2 (July 31, 2023): 140–58. https://doi.org/10.35316/lahjah.v4i2.140-158.

Jundi, Muhammad, and Yuslin Kasan. “GAYA DAN MAKNA BAHASA TULISAN: KAJIAN DESKRIPTIF CHAT MAHASISWA KEPADA DOSEN.” *MARDIBASA: Jurnal Pembelajaran Bahasa Dan Sastra Indonesia* 1, no. 2 (September 2021). https://doi.org/10.21274/jpbsi.2021.1.2.290-315.

Jundi, Muhammad, Yuslin Kasan, and Suleman D. Kadir. “Maf’ul Mutlaq Bahasa Arab Dan Cognate Object Bahasa Inggris: Analisis Kontrastif Dan Sintaksis Pada Juz ’Amma.” *Lisanan Arabiya: Jurnal Pendidikan Bahasa Arab* 5, no. 2 (December 31, 2021): 153–78. https://doi.org/10.32699/liar.v5i2.2190.

Jundi, Muhammad, and Nabila Nabila. “Transformational-Generative Theory Perspective in Teaching The Four Language Skills of Arabic.” *Al-Fusha : Arabic Language Education Journal* 5, no. 2 (June 30, 2023): 72–81. https://doi.org/10.36835/alfusha.v5i2.1193.

Jundi, Muhammad, and Najamudin Petta Solong. “Analisis Kesesuaian Indikator Dan Kompetensi Dasar Bahasa Arab KMA 183.” *Tatsqifiy: Jurnal Pendidikan Bahasa Arab* 2, no. 1 (January 28, 2021): 61–70. https://doi.org/10.30997/tjpba.v2i1.3642.

Kementrian Pendidikan dan Kebudayaan. “Panduan Penulisan Soal Pilihan Ganda.” Depdiknas, 2007.

———. “Permendiknas No.16, Tahun 2007, Tentang Standar Kualifikasi Akademik Dan Kompetensi Guru,” 2007.

Komala, Popon, Ida Kaniawati, and Ridwan Efendi. “KARAKTERISASI SOAL TES KETERAMPILAN BERPIKIR KRITIS MENGGUNAKAN ANALISIS ITEM RESPONSE THEORY PADA MATERI FLUIDA STATIS.” *WaPFi (Wahana Pendidikan Fisika)* 5, no. 2 (September 11, 2020): 100–109. https://doi.org/10.17509/wapfi.v5i2.7755.

Kumari, Amrita, and S B Bhattacharya. “Item Analysis of Diagnostic Test in English Language Skills of Secondary School Students,” n.d.

Laubaha, Siti Aliyya Laubaha, Ratni Bt H. Bahri, Suleman D. Kadir, muhammad Zikran Adam, Ibadurrahman Ali, and Muhammad Jundi. “Kesesuaian Materi Berdasarkan Teori Dan Praktek Pada Bahan Ajar Bahasa Arab.”

Naskhi: Jurnal Kajian Pendidikan Dan Bahasa Arab 4, no. 2 (October 24, 2022): 10–21. https://doi.org/10.47435/naskhi.v4i2.1198.

Liang, Tie, Craig S. Wells, and Ronald K. Hambleton. "An Assessment of the Nonparametric Approach for Evaluating the Fit of Item Response Models." *Journal of Educational Measurement* 51, no. 1 (2014): 1–17.

Luo, Lan, Cara Arizmendi, and Kathleen M. Gates. "Exploratory Factor Analysis (EFA) Programs in R." *Structural Equation Modeling: A Multidisciplinary Journal* 26, no. 5 (September 3, 2019): 819–26. https://doi.org/10.1080/10705511.2019.1615835.

Mardapi, Djemari. *Pengukuran Penilaian Dan Evaluasi Pendidikan*. Yogyakarta: Nuha Medika, 2012.

———. *Pengukuran Penilaian Dan Evaluasi Pendidikan*. 2nd ed. Yogyakarta: Nuha Litera, 2016.

Miller, M.D., R.L. Linn, and N.E. Grondlund. *Measurement and Assessment in Teaching*. 10th ed. New Jersey: Pearson Education, 2009.

Naga, Dali S. *Pengantar Teori Sekor Pada Pengukuran Pendidikan*. Jakarta: Gunadarma, 1992.

Nguyen, Tam H., Hae-Ra Han, Miyong T. Kim, and Kitty S. Chan. "An Introduction to Item Response Theory for Patient-Reported Outcome Measurement." *The Patient* 7, no. 1 (2014): 23–35. https://doi.org/10.1007/s40271-013-0041-0.

Nurmalita, Citra. "Karakteristik Perangkat Ujian Kompetensi Teori Teknik Perkayuan Tahun 2014 SMK Di Jawa Timur Menggunakan Model IRT." *Ilmu Pendidikan: Jurnal Kajian Teori Dan Praktik Kependidikan* 3, no. 2 (December 29, 2018): 164–74. https://doi.org/10.17977/um027v3i22018p164.

Pratama, Dinar. "Analisis Kualitas Tes Buatan Guru Melalui Pendekatan Item Response Theory (IRT) Model Rasch." *Tarbawy : Jurnal Pendidikan Islam* 7, no. 1 (April 24, 2020): 61–70. https://doi.org/10.32923/tarbawy.v7i1.1187.

———. "ANALYSIS OF CLASICAL TEST THEORY (CTT) APPROCH ON ACADEMIC ABILITY TEST INSTRUMENT." *JISAE: Journal of Indonesian Student Assessment and Evaluation* 5, no. 2 (2019): 43–54. https://doi.org/10.21009/jisae.v5i2.12701.

Purwanto. "ANALISIS FAKTOR: KONSEP, PROSEDUR UJI DAN INTERPRETASI." *Jurnal Teknodik* 8, no. 15 (2018). https://doi.org/10.32550/teknodik.v4i15.388.

Purwati, Heni, Heri Retnawati, Jailani, and Trie Hartiti Retnowati. "Analisis Karakteristik Butir Soal Ujian Nasional Matematika SMP/MTs Berdasarkan Pendekatan Teori Tes Klasik." *Jurnal Sains Dan Edukasi Sains* 4, no. 2 (Agustus 2021). https://doi.org/10.24246/juses.v4i2p46-51.

Retnawati, Heri. "Reliabilitas Instrumen Penelitian." Lokakarya presented at the Workshop Pembelajaran Berbasis Riset dan Analisis Data, Bukittinggi, 2017.

———. *Teori Respons Butir Dan Penerapannya: Untuk Peneliti, Praktisi Pengukuran Dan Pengujian, Mahasiswa Pascasarjana*. Yogyakarta: Nuha Medika, 2014.

———. *Validitas, Reliabilitas, Dan Karakteristik Butir*. Yogyakarta: Parama Publishing, 2016.

Rizopoulos, Dimitris. "Ltm: An R Package for Latent Variable Modeling and Item Response Analysis." *Journal of Statistical Software* 17 (2007): 1–25. https://doi.org/10.18637/jss.v017.i05.

Sarea, Muh Syahrul, and Rosnia Ruslan. "KARAKTERISTIK BUTIR SOAL: CLASSICAL TEST THEORY VS ITEM RESPONSE THEORY?" *DIDAKTIKA : Jurnal Kependidikan* 13, no. 1 (August 15, 2019): 1–16. https://doi.org/10.30863/didaktika.v13i1.296.

Sugiyono, Sugiyono. *Metode Penelitian Pendidikan; Pendekatan Kuantitatif, Kualitatif, Dan R&D*. Cet. 23. Bandung: Alfabeta, 2016.

Sumintono, Bambang, and Wahyu Widhiarso. *Aplikasi Model Rasch Untuk Penelitian Ilmu-Ilmu Sosial (Edisi Revisi)*, 2014.

Surapranata, Sumarna. *Analisis Validitas, Reliabilitas, Dan Interpretasi Hasil Tes: Implementasi Kurikulum 2004*. 4th ed. Bandung: Remaja Rosdakarya, 2009.

———. *Panduan Penulisan Tes Tertulis (Penilaian Berbasis Kelas)*. Bandung: Remaja Rosdakarya, 2005.

Susanto, Hari Purnomo, Heri Retnawati, Agus Maman Abadi, Haryanto, and Raden Muhammad Ali. "Irtawsi: Items Response Theory Analysis with Steps and Interpretation." R, March 2023.

Susetyo, Budi. *Prosedur Penyusunan Dan Analisis Tes Untuk Penilaian Hasil Belajar Kognitif*. Bandung: Refika Aditama, 2015.

Triantafillou, Evangelos, Elissavet Georgiadou, and Anastasios A. Economides. "The Design and Evaluation of a Computerized Adaptive Test on Mobile Devices." *Computers & Education* 50, no. 4 (May 2008): 1319–30. https://doi.org/10.1016/j.compedu.2006.12.005.

Umar, J., G. N. Masters, and J.P. Keeves. "Item Banking." In *Advances in Measurement in Educational Research and Assessment*, 207–18. Oxford: Elsevier Science Ltd., 1999.

Van Der Linden, Wim J., and Ronald K. Hambleton, eds. *Handbook of Modern Item Response Theory*. New York, NY: Springer New York, 1997. https://doi.org/10.1007/978-1-4757-2691-6.

"Visi Dan Misi MTs. N. Kota Batu." Accessed December 28, 2023. https://mtsnkotabatu.sch.id/halaman/detail/visi-dan-misi.

Wardhani, Dyah Febria, and Aminuddin Prahatama Putra. "Pengembangan Instrumen Tes Standar Kognitif pada Mata Pelajaran IPA Kelas 7 SMP di Kabupaten Banjar." In *Seminar Nasional XIII Pendidikan Biologi FKIP UNS 2016*. Sebelas Maret University, 2016. https://www.neliti.com/publications/175334/.

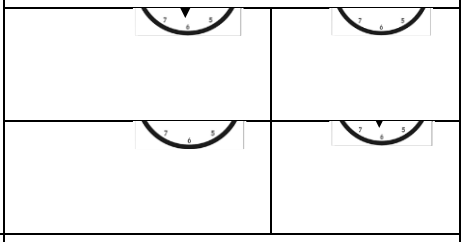
Wells, Craig S., Michael J. Subkoviak, and Ronald C. Serlin. "The Effect of Item Parameter Drift on Examinee Ability Estimates." *Applied Psychological*

Measurement 26 (2002): 77–87. https://doi.org/10.1177/0146621602026001005.
Widoyoko, S. E. P. *Teknik Penyusunan Instrumen Penelitian*. Yogyakarta: Pustaka Pelajar, 2012. https://www.semanticscholar.org/paper/Teknik-Penyusunan-Instrumen-Penelitian-Widoyoko/3732ac3bd6dde089e841c7353f71c2f44bc0a702.
Yen, Wendy, and Anne Fitzpatrick. "Item Response Theory," 2006.
Yusuf, A. Muri. *Asesmen Dan Evaluasi Pendidikan*. Jakarta: Prenada Media, 2017.

LAMPIRAN 1: **SOAL ASESMEN MADRASAH BAHASA ARAB TA.2022/2023 MTs. N. Kota Batu**

No	Kompetensi Dasar	Materi	Indikator Soal	Soal
1.	3.2 Menganalisis gagasan dari teks sederhana yang berkaitan dengan tema: التعارف dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal المبتدأ (ضمير) + الخبر	• Teks Khiwar tema التعارف • Bentuk, makna dan fungsi dari susunan gramatikal المبتدأ (ضمير) + الخبر	Menganalisis penggunaan susunan gramatikal المبتدأ (ضمير) + الخبر	- السَّلامُ عَلَيْكُمْ + وَعَلَيْكُمُ السَّلام وَرَحْمَةُ اللهِ وَبَرَكَاتُهُ اَنَا تِلْمِيْذَةٌ، اسْمِيْ صَا فِيَةٌ، وَ أَنْتِ مَاسْمُكِ؟ - ..... اسْمُهُ أَحْمَدٌ اسْمِيْ صَالِحَةٌ اسْمُكَ صَالِحَةٌ اِسْمِيْ صَافِيَةٌ
2.	3.2 Menganalisis gagasan dari teks sederhana yang berkaitan dengan tema: التعارف dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal المبتدأ (ضمير) + الخبر	• Teks khiwar tema التعارف • Bentuk, makna dan fungsi dari susunan gramatikal المبتدأ (ضمير) + الخبر	Menganalisis penggunaan susunan gramatikal المبتدأ (ضمير) + الخبر	مَنْ أَنْتِ ؟ ..... أَنَا مُدَرِّسَةٌ لَا، أَنَا طَالِبَةٌ نَعَمْ , أَنَا مُدَرِّسَةٌ أَنْتَ مُدَرِّسَةٌ
3.	3.4 Menerapkan kaidah tentang bentuk, makna dan fungsi gramatikal المبتدأ (إشارة) + الخبر (نعت/ظرف المكان)	Bentuk, makna dan fungsi gramatikal المبتدأ (إشارة) + الخبر (نعت/ظرف المكان)	Menerapkan fungsi gramatikal tentang المبتدأ (إشارة) + الخبر (نعت/ظرف المكان)	ذَهَبَ اَحْمَدُ إِلَى الْمَسْجِدِ، ذَالِكَ الْمَسْجِدُ وَاسِعٌ. عَيِّنْ الْجُمْلَةُ الَّتِى تَتَرَكَّبُ مِنَ الْمُبْتَدَاوَالْخَبَرُ .... ذَهَبَ اَحْمَدُ إِلَى الْمَسْجِدِ ذَالِكَ الْمَسْجِدُ وَاسِعٌ اَحْمَدُ إِلَى الْمَسْجِدِ
4.	3.4 Menerapkan kaidah tentang bentuk, makna dan fungsi gramatikal المبتدأ (إشارة) + الخبر (نعت/ظرف المكان)	Bentuk, makna dan fungsi gramatikal المبتدأ (إشارة) + الخبر (نعت/ظرف المكان)	Menerapkan fungsi gramatikal tentang المبتدأ (إشارة) + الخبر (نعت/ظرف المكان)	الْمَاءُ وَالصَّا بُوْنُ فِيْ .... غُرْفَةِ ٱلْجُلُوْسِ غُرْفَةِ النَّوْمِ ٱلْحَمَّامِ غُرْفَةِ ٱلْمُذَاكَرَةِ
5.	3.6 Menganalisis gagasan dari teks naratif sederhana tentang jenis dan warna yang berkaitan dengan tema: الأدوات المدرسية dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatika الضمير المتصل	• Teks naratif tema الأدوات المدرسية Bentuk, makna dan fungsi dari susunan gramatikal المتصل الضمير	Menyimpulkan isi teks naratif tema الأدوات المدرسية	فِي غُرْفَةِ الْجُلُوْسِ مِنْضَدَة وَاحِدَة وَخَمْسَةُ كَرَاسِي مُنَظَّمَةٌ. وَعَلَى الْمِنْضَدَةِ زَهْرِيَّةٌ . وَعَلَىَ الزَّهْرِيَّةِ أَزْهَارٌ جَمِيْلَةٌ مِثْلُ زَهْرَة، وَرْدَة، وَيَسْمِيْن. وفَوْقَ الْمِنْضَدَةِ مِصْبَاحٌ. مَاذَا فِيْ غُرْفَةِ الْجُلُوْسِ؟ فِيْ غُرْفَةِ الْجُلُوْسِ ... مِنْضَدَةٌ وَ مَائِدَةٌ مِنْضَدَةٌ وَ كُرْسِيّ كُرْسِيّ وَ أَزْهَارٌ مِنْضَدَةٌ وَ مِقْلَمَةٌ
6.	3.6 Menganalisis gagasan dari teks naratif sederhana tentang jenis dan warna yang berkaitan dengan tema: الأدوات المدرسية dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatika الضمير المتصل	• Teks naratif tema الأدوات المدرسية Bentuk, makna dan fungsi dari susunan gramatikal • المتصل الضمير	Menyimpulkan isi teks naratif tema الأدوات المدرسية	وَهَذِهِ غُرْفَةُ الْمُذَاكَرَةِ. الغُرْفَةُ مُذَاكَرَةٌ وَرَاءَ غُرْفَةِ الْجُلُوْسِ. تِلْكَ غُرْفَةُ لِيُوْسُفُ، هُوَ اِبْنُ عَمِّيْ الصَّغِيْر. فِيْهَا مِرْسَم، وَ كِتَابٌ، وَمِقْلَمَةٌ، وَقَلَمٌ، وَمَكْتَبٌ، وَ رَفٌّ. وَهَذِهِ غُرْفَةُ ٱلْأَكْلِ . وَهِيَ جَانِبَ الْمَطْبَخِ . فِيْهَا مَائِدَةٌ كَبِيْرَةٌ . وَعَلَى الْمَائِدَةِ رُزٌّ وَمِلْعَقَةٌ وَصَحْنٌ وَشَايٌ وَقَهْوَةٌ. هَلْ الْمِقْلَمَةِ وَ الصَّحْنِ عَلَى الْمَائِدَةِ؟ لَا، الْمِقْلَمَةُ وَ الصَّحْنُ عَلَى الْمِنْضَدَةِ لَا، الْمِقْلَمَةُ وَ الصَّحْنُ عَلَى الْمَكْتَبِ نَعَمْ، الْمِقْلَمَةُ وَ الصَّحْنُ عَلَى الْمَائِدَةِ

				نَعَمْ، الْمِقْلَمَةُ وَ الصَّحْنُ عَلَى الْمِنْضَدَةِ
7.	3.8 Menganalisis gagasan dari teks sederhana yang berkaitan dengan tema: العنوان dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal ١ الأرقام –١٠٠	• Teks tema العنوان Bentuk, makna dan fungsi dari susunan gramatikal ١ الأرقام –١٠٠	Mengidentifikasi bentuk, makna dan fungsi dari susunan gramatikal ١ الأرقام – ١٠٠	رَقْمُ تِلْفُوْنِ بَيْتِ إِبْرَاهِيْمُ ... (خَمْسَةُ أَرْبَعَةُ سَبْعَةُ تِسْعَةُ سِتَّةُ صِفْرُ ثَلَاثَةُ) ٠٣٧٩٦٤٥ ٣٤٧٦٩٠٥ ٣٠٦٩٧٤٥ ٣٥٧٩٦٠٣
8.	3.8 Menganalisis gagasan dari teks sederhana yang berkaitan dengan tema: العنوان dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal ١ الأرقام –١٠٠	• Teks tema العنوان Bentuk, makna dan fungsi dari susunan gramatikal • ١ الأرقام –١٠٠	Mengidentifikasi bentuk, makna dan fungsi dari susunan gramatikal ١ الأرقام – ١٠٠	Terjemahkan ke Bahasa Arab "Nomor teleponku 742015" رَقْمُ تِلِيْفُوْنِيْ: سَبْعَةٌ – أَرْبَعَةٌ – وَاحِدٌ – صِفْرٌ – خَمْسَةٌ – إِثْنَانِ رَقْمُ تِلِيْفُوْنِيْ: سَبْعَةٌ – أَرْبَعَةٌ – إِثْنَانِ – صِفْرٌ – وَاحِدٌ – خَمْسَةٌ رَقْمُ تِلِيْفُوْنِيْ: سَبْعَةٌ – إِثْنَانِ – أَرْبَعَةٌ - صِفْرٌ – وَاحِدٌ – خَمْسَةٌ رَقْمُ تِلِيْفُوْنِيْ: سَبْعَةٌ – صِفْرٌ – أَرْبَعَةٌ – إِثْنَانِ - وَاحِدٌ – خَمْسَةٌ
9.	3.10 Menerapkan kaidah tentang bentuk, makna dan fungsi dari susunan gramatikal الخبر المقدم + المبتدأ المؤخر dalam teks deskriptif sederhana yang berkaitan dengan tema: البيت	Bentuk, makna dan fungsi dari susunan gramatikal الخبر المقدم + المبتدأ المؤخر teks deskriptif sederhana yang berkaitan dengan tema: البيت	Menerapkan bentuk, makna dan fungsi dari susunan gramatikal الخبر المقدم + المبتدأ المؤخر Sesuai dengan tema: البيت	فَوْقَ الْمَكْتَبِ مِصْبَاحٌ. Manakah yang termasuk *Mubtada' Mu'akhor* dari kalimat tersebut ... فَوْقَ الْمَكْتَبِ الْمَكْتَبِ الْمَكْتَبِ مِصْبَاحٌ مِصْبَاحٌ
10.	3.10 Menerapkan kaidah tentang bentuk, makna dan fungsi dari susunan gramatikal الخبر المقدم + المبتدأ المؤخر dalam teks deskriptif sederhana yang berkaitan dengan tema: البيت	Bentuk, makna dan fungsi dari susunan gramatikal الخبر المقدم + المبتدأ المؤخر teks deskriptif sederhana yang berkaitan dengan tema: البيت	Menerapkan bentuk, makna dan fungsi dari susunan gramatikal الخبر المقدم + المبتدأ المؤخر Sesuai dengan tema: البيت	عَلَى الْمِنْضَدَةِ زَهْرِيَّةٌ Bentuk struktur kalimat diatas adalah .... خَبَرٌ مُقَدَّمٌ + مُبْتَدَاءٌ مُؤَخَّرٌ مُبْتَدَاءٌ مُؤَخَّرٌ + خَبَرٌ مُقَدَّمٌ خَبَرٌ مُؤَخَّرٌ + مُبْتَدَاءٌ مُؤَخَّرٌ مُبْتَدَاءٌ مُقَدَّمٌ + خَبَرٌ مُقَدَّمٌ
11.	3.12 Menganalisis gagasan dari teks naratif sederhana yang berkaitan dengan tema: من يوميات الأسرة dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal تصريف المضارع للمفرد	Teks naratif tema من يوميات الأسرة Bentuk, makna dan fungsi dari susunan gramatikal تصريف المضارع للمفرد	• Menentukan ide pokok dalam teks naratif tema من يوميات الأسرة	أَسْتَيْقِظُ مُبَكِرًا فِي الصَّبَاحِ. حَوَلَي السَّاعَةِ الرَّابِعَةِ وَ النِّصْفِ. ثُمَّ أَذْهَبُ إِلَى الحَمَّامِ، وَأَتَوَضَّؤُ لِأُصَلِّي الصُّبْحِ فِيْ الْمَسْجِدِ. الأَنْشِطَةُ مِنَ الْفَقْرَةِ السَّابِقَةِ هِيَ .... يَوْمِيَاتِنَا فِيْ الْمَدْرَسَةِ يَوْمِيَاتِنَا فِيْ الفَصْلِ يَوْمِيَاتِنَا فِيْ الْمَكْتَبَةِ يَوْمِيَاتِنَا فِيْ البَيْتِ

12.	3.12 Menganalisis gagasan dari teks naratif sederhana yang berkaitan dengan tema: من يوميات الأسرة dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal تصريف المضارع للمفرد	Teks naratif tema من يوميات الأسرة Bentuk, makna dan fungsi dari susunan gramatikal تصريف المضارع للمفرد	Membuat pola hubungan antar dhomir pada تصريف الفعل المضارع	أَنَا ... مِنَ النَّوْمِ فِيْ السَّاعَةِ الرَّابِعَةِ صَبَاحًا الكَلِمَةُ الْمُنَاسِبَةُ لِإِكْمَالِ الجُمْلَةِ هِيَ .... أُنَظِّفُ الأَسْنَانِ أَرْتَدِيْ مَلاَبِسَ أَسْتَحِمُّ أَسْتَيْقِظُ
13.	3.2 Menganalisis gagasan dari teks naratif sederhana yang berkaitan dengan tema: الساعة dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal العدد الترتيبي	• Teks naratif tema الساعة Bentuk, makna dan fungsi dari susunan gramatikal العدد الترتيبي	Menyimpulkan kegiatan berdasarkan jam pada teks naratif tentang الساعة	تَنْتَهِي الاسْتِرَاحَةُ فِي السَّاعَةِ العَاشِرَةِ وَ الرُّبْعِ. صُوْرَةُ السَّاعَةِ الْمُنَاسِبَةِ هِيَ ... 
14.	3.2 Menganalisis gagasan dari teks naratif sederhana yang berkaitan dengan tema: الساعة dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal العدد الترتيبي	• Teks naratif tema الساعة Bentuk, makna dan fungsi dari susunan gramatikal العدد الترتيبي	Menyimpulkan kegiatan berdasarkan jam pada teks naratif tentang الساعة	إِقْرَأْ نَصَّ القِرَاءَةِ الآتي! آمِنَةُ تَذْهَبُ إِلَى الْمَكْتَبَةِ فِي السَّاعَةِ الثَّامِنَةِ صَبَاحًا، وَتَقْرَأُ الْمَجَلَّةَ فِيْهَا. تَرْجِعُ مِنَ الْمَكْتَبَةِ فِيْ السَّاعَةِ الْعَاشِرَةِ. كَمْ سَاعَةً تَقْرَأُ آمِنَةُ فِيْ الْمَكْتَبَةِ ؟ سَاعَةً وَاحِدَةً نِصْفَ السَّاعَةِ سَاعَتَيْنِ ثَلاَثُ سَاعَة
15.	3.4 Menganalisis gagasan dari teks naratif sederhana yang erkaitan dengan tema: يومياتنا dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal الجملة الاسمية	• Teks naratif tema يومياتنا Bentuk, makna dan fungsi dari susunan gramatikal الجملة الاسمية	• Membandingkan kegiatan siswa dirumah dan disekolah pada teks tema يومياتنا	النَّصُّ- ١ حُسَيْنُ يُصَلِّى الْمَغْرِبِ فِيْ السَّادِسَةِ ثُمَّ يَتَنَاوَلُ الْعَشَاءِ، بَعْدَ الْعَشَاءِ يَقْرَأُ الْكِتَابِ وَ الْمَجَلَّةِ . حُسَيْنُ يُحِبُّ الْقِرَاءَةَ النَّصُّ- ٢ فِيْ وَقْتِ الإِسْتِرَاحَةِ فَاطِمَةُ تَذْهَبُ إِلَى الْمَكْتَبَةِ وَ تَقْرَأُ الْكِتَابَ وَ الْمَجَلَّةَ. أَيْنَ يَقْرَأُ حُسَيْنُ وَ فَاطِمَة الْكِتَابَ ؟ حُسَيْن يَقْرَأُ فِيْ الْبَيْتِ وَ فَاطِمَةُ تَقْرَأُ فِيْ الْمَدْرَسَةِ حُسَيْن تَقْرَأُ فِيْ الْمَدْرَسَةِ وَ فَاطِمَةُ يَقْرَأُ فِيْ الْبَيْتِ حُسَيْن يَقْرَأُ فِيْ الْبَيْتِ وَ فَاطِمَةُ تَقْرَأُ فِيْ الْبَيْتِ فَاطِمَةُ تَقْرَأُ فِيْ الْبَيْتِ وَ حُسَيْن يَقْرَأُ فِيْ الْمَدْرَسَةِ
16.	3.4 Menganalisis gagasan dari teks naratif sederhana yang erkaitan dengan tema: يومياتنا dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal الجملة الاسمية	• Teks naratif tema يومياتنا Bentuk, makna dan fungsi dari susunan gramatikal الجملة الاسمية	Menguraikan kalimat berbentuk الجملة الاسمية dalam teks naratif	يَعْمَلُ أَحْمَدُ الوَجِبَ الْمُنْزِلِيّ Kalimat diatas adalah jumlah fi'liyah, jika dirubah menjadi jumlah ismiyah, menjadi .... يَعْمَلُ أَحْمَدُ الوَجِبَ الْمُنْزِلِيّ أَحْمَدُ يَعْمَلُ الوَجِبَ الْمُنْزِلِيّ الوَجِبَ الْمُنْزِلِيّ يَعْمَلُ أَحْمَدُ يَعْمَلُ الوَجِبَ أَحْمَدُ الْمُنْزِلِيّ
17.	3.6 Menganalisis gagasan dari teks sederhana yang berkaitan dengan tema: الهواية	• Teks tema الهواية Bentuk, makna dan fungsi dari susunan gramatikal	Mengklasifikasi macam-macam hobi dalam teks tema الهواية	إِسْمِيْ أَحْمَدُ، عِنْدِيْ أَصْدِقَائِيْ. هُمْ يُحِبُّوْنَ الرِّيَاضَةَ الْمُخْتَلِفَةَ. إِسْمَاعِيْلُ يُحِبُّ كُرَةَ الرِّيْشَةِ، وَحَمْدَان يُحِبُّ التَّصْوِيْرَ. وَإِبْرَاهِيْمُ يُحِبُّ كُرَةَ القَدَمِ، وَ يَحْيَى يُحِبُّ السِّبَاحَةَ. الهِوَايَةُ الْفَرْدِيَّةُ (Individual) مِنْ هَذِهِ الْفَقْرَةِ هِيَ...

	dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal			كُرَةَ الرِّيْشَةِ وَ التَّصْوِيْرَ كُرَةَ القَدَمِ وَ السِّبَاحَةَ التَّصْوِيْرَ وَ السِّبَاحَةَ كُرَةَ الرِّيْشَةِ وَ كُرَةَ القَدَم
18.	3.6 Menganalisis gagasan dari teks sederhana yang berkaitan dengan tema: الهواية dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal	• Teks tema الهواية Bentuk, makna dan fungsi dari susunan gramatikal	Mengklasifikasi macam-macam hobi dalam teks tema الهواية	إِسْمِيْ أَحْمَدُ، عِنْدِيْ أَصْدِقَائِيْ. هُمْ يُحِبُّوْنَ الرِّيَاضَةَ الْمُخْتَلِفَةَ. إِسْمَاعِيْلُ يُحِبُّ كُرَةَ الرِّيْشَةِ، وَحَمْدَان يُحِبُّ التَّصْوِيْرَ. وَإِبْرَاهِيْمُ يُحِبُّ كُرَةَ القَدَمِ، وَ يَحْيَى يُحِبُّ السِّبَاحَةَ. الهِوَايَةُ الْجَمَاعِيَّةُ (Kelompok) مِنْ هَذِهِ الْفَقْرَةِ هِيَ ... كُرَةَ الرِّيْشَةِ وَ التَّصْوِيْرَ كُرَةَ القَدَمِ وَ السِّبَاحَةَ التَّصْوِيْرَ وَ السِّبَاحَةَ كُرَةَ الرِّيْشَةِ وَ كُرَةَ القَدَم
19.	3.8 Menganalisis gagasan dari teks naratif sederhana yang berkaitan dengan tema: الرياضة dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal أن – لن – لـ(+الفعل المضارع)	• Teks naratif tema الرياضة	• Menemukan makna tersirat dalam teks naratif tema الرياضة	عُمَرُ يُحِبُّ كَرَةَ الْقَدَمِ ، يَلْعَبُ كُرَةَ الْقَدَمِ كُلَّ الْمَسَاءِ فِيْ يَوْمِ الْجُمْعَةِ وَ يَومِ السَّبْتِ وَ يَوْمِ الأَحَدِ. كَمْ يَوْمًا عُمَرُ يَلْعَبُ كُرَةَ الْقَدَمِ فِيْ الأُسْبُوْعِ ؟ ثَلاَثُ أَيَّامٍ يَوْمًا وَاحِدًا يَوْمَيْنِ خَمْسَةُ أَيَّامٍ
20.	3.8 Menganalisis gagasan dari teks naratif sederhana yang berkaitan dengan tema: الرياضة dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal أن – لن – لـ(+الفعل المضارع)	• Bentuk, makna dan fungsi dari susunan gramatikal (الفعل + أن – لن – لـ المضارع)	Membedakan penggunaan لن , أن dan لـ dalam kalimat	أَنَا أُرِيْدُ ... أَذْهَبَ إِلَى الْمَكْتَبَةِ لِأَنَّ أُحِبُّ الْقِرَاءَةَ. عَامِلُ النَّصْبِ الْمُنَاسِبِ لِتِلْكَ الْجُمْلَةِ هِيَ؟ لَمْ لَنْ أَنْ لِـ
21.	3.10 Menganalisis gagasan dari teks naratif sederhana yang berkaitan dengan tema: المهنة dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal المصدر الصريح	• Teks naratif tema المهنة	• Menganalisis macam-macam profesi dalam teks tema المهنة	... يَذْهَبُ إِلَى مَوْ قِفِ السَّيَّارَةِ السَّائِقُ الطَّالِبَةُ الْفَلَّاحُ الْمُهَنْدِسَةُ
22.	3.10 Menganalisis gagasan dari teks naratif sederhana yang berkaitan dengan tema: المهنة dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal المصدر الصريح	Bentuk, makna dan fungsi dari susunan gramatikal المصدر الصريح	• Membedakan konsep المصدر الصريحdan المصدر المؤول pada teks naratif	Dibawah ini yang merupakan jenis kalimat dari MASHDAR MUAWWAL adalah ... أُرِيْدُ أَنْ اَقْرَأَ الْقُرْآنَ أُرِيدُ قِرَاءَةَ الْكتابَ أُحِبُّ الْكِتَابَةَ الْكِتَابَ سَرَرْنِيْ نَجَاحُكَ
23.	3.10 Menganalisis gagasan dari teks naratif sederhana yang berkaitan dengan tema: المهنة	Bentuk, makna dan fungsi dari susunan gramatikal المصدر الصريح	• Membedakan konsep المصدر الصريحdan المصدر المؤول pada teks naratif	أُحِبُّ رُكُوبَ السَّيَّارَةَ Kalimat diatas merupakan jenis kalimat MASHDAR SHARIH. BENAR SALAH

	dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal المصدر الصريح			
24.	3.12 Menganalisis gagasan dari teks sederhana yang berkaitan dengan tema: عيادة المرضى dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal الفعل الماضي والجملة الفعلية	Teks tema عيادة المرضى	• Menyimpulkan teks tentang عيادة المرضى	طَلَبَ الطَّبِيْبُ مِنَ الْمَرِيْضِ أَنْ يَبْقِى فِيْ ... عَدَدِ أَيَامِ الْمُسْتَشْفَى الْمَعْمَلِ الْمَدْرَسَةِ الْمَصْنَعِ
25.	3.12 Menganalisis gagasan dari teks sederhana yang berkaitan dengan tema: عيادة المرضى dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal الفعل الماضي والجملة الفعلية	Teks tema عيادة المرضى	• Menyimpulkan teks tentang عيادة المرضى	"شَفَاكَ الله شِفَاءً عَاجِلًا" هِيَ دُعَاءُ .... السَّفَرُ عِيَادَةُ الْمَرِيْضِ قَبْلَ الْمُذَاكَرَةِ الْوُضُوْءِ
26.	3.12 Menganalisis gagasan dari teks sederhana yang berkaitan dengan tema: عيادة المرضى dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal الفعل الماضي والجملة الفعلية	Bentuk, makna dan fungsi dari susunan gramatikal	Mengaitkan konsep الفعل الماضي dalam الجملةالفعلية dalam kalimat	... الطَّبِيْبُ الْمَرْضَيْ فِيْ الْمُسْتَشْفَيْ. يُعَالِجُ تُعَالِجُ يُعَالِجُوْنَ تُعَالِجِيْنَ
27.	3.12 Menganalisis gagasan dari teks sederhana yang berkaitan dengan tema: عيادة المرضى dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal الفعل الماضي والجملة الفعلية	Bentuk, makna dan fungsi dari susunan gramatikal	Mengaitkan konsep الفعل الماضي dalam الجملةالفعلية dalam kalimat	تَسْعُرُ ... بِعَلَمِ سَدِيْدٍ فِيْ الرَّأْسِهَا أَحْمَدُ فَاطِمَةُ حُسَيْنُ فَرْحَانُ
28.	3.2 Menganalisis gagasan dari teks naratif sederhana yang berkaitan dengan tema: رأس السنة الهجرية dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal تصريف الفعل الماضي المجرد	• Teks naratif tema رأس السنة الهجرية	• Menyimpulkan isi teks naratif tema رأس السنة الهجرية	أَمَرَ اللهُ النَّبِيَّ صَلَّى اللهُ عَلَيْهِ وَسَلَّمَ وَ أَصْحَابَهُ بِالْهِجْرَةِ إِلَى الْمَدِيْنَةِ. ذٰلِكَ لِأَنَّ كُفَّارَ مَكَّةَ يُؤْذُوْنَهُمْ. فَهَاجَرُوْا اِلَى الْمَدِيْنَةِ بِالْخَوْفِ، خَوْفًا مِنْ أَنْ يَمْنَعَهُمْ كُفَّارُ مَكَّةَ. لَكِنْ قَدِمَ النَّبِيُّ صَلَّى اللهُ عَلَيْهِ وَسَلَّمَ و أَصْحَابُهُ فِي الْمَدِيْنَةِ سَالِمِيْنَ وَ اِسْتَقْبَلَهُمْ الْأَنْصَارَ اِسْتِقْبَالًا حَارًّا. مَنْ أَمَرَ النَّبِيَّ صَلَّى اللهُ عَلَيْهِ وَسَلَّمَ وَ أَصْحَابَهُ بِالْهِجْرَةِ ؟ اللهُ أَصْحَابُ رَسُوْلِ الله رَسُوْلُ الله اللهُ وَرَسُوْلُهُ

29.	3.2 Menganalisis gagasan dari teks naratif sederhana yang berkaitan dengan tema: رأس السنة الهجرية dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal تصريف الفعل الماضي المجرد	• Bentuk, makna dan fungsi dari susunan gramatikal تصريف الفعل الماضي المجرد	• Mendeteksi berbagai bentuk dhomir dalam تصريف الفعل الماضي المجرد	الْمُدَرِّسُوْنَ ... التِّلاَمِيْذَ فِى الْمَدْرَسَةِ. عَلَّمْنَ عَلَّمُوْا عَلَّمْتُمْ عَلَّمْنَا
30.	3.2 Menganalisis gagasan dari teks naratif sederhana yang berkaitan dengan tema: رأس السنة الهجرية dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal تصريف الفعل الماضي المجرد	• Bentuk, makna dan fungsi dari susunan gramatikal تصريف الفعل الماضي المجرد	• Mendeteksi berbagai bentuk dhomir dalam تصريف الفعل الماضي المجرد	+ = هَلْ ... الْوَاجِبَ الْمَنْزِلِيّ ؟ – = نَعَمْ، عَمِلْتُ. عَمِلْنَا عَمِلْتُ عَمِلْتَ عَمِلْتُمْ
31.	3.4 Menerapkan kaidah tentang bentuk, makna dan fungsi dari gramatikal كان واسمها وخبرها المفرد dalam menyusun teks naratif sederhana yang berkaitan dengan tema: الحفل بمولد الرسول	• Teks naratif tema الحفل بمولد الرسول	• Menyimpulkan rangkaian acara peringatan maulid Nabi dalam teks naratif tema الحفل بمولد الرسول	اِحْتَفَلَ الْمُسْلِمُوْنَ بِذِكْرَى مَوْلِدِ النَّبِيِّ صَلَّى اللهُ عَلَيْهِ وَسَلَّمَ فِي السَّاعَةِ السَّابِعَة وَالنِّصْف لَيْلًا بَعْدَ صَلاَةِ الْعِشَاء. قَامَ أَحْمَد كَرَئِيْسِ الْجَلْسَةِ وَقَرَأَ بَرَامِجَ الْحَفْلِ تَرْتِيْبًا، وَهِيَ الْإِفْتِتَاح بِقِرَاءَةِ سُوْرَةِ الْفَاتِحَةِ، و قِرَاءَةُ الْقُرْآنِ الْكَرِيْمِ التى قَرَأَهُ يُوسُفْ، وَكَلِمَاتٌ تَرْحِيْبِيَّة قَدَّمَهَا مَرْوَان كَرَئِيْس اللَجْنَةِ، و الْمُحَاضَرَةُ الدِّيْنِيَّة أَلْقَاهَا اْلأُسْتَاذ عَبْدُ الرَّحِيْم، وَ الْإِخْتِتَام بِالدُّعَاء. فِي أَيِّ السَّاعَة اِحْتَفَلَ الْمُسْلِمُوْنَ بِذِكْرَى مَوْلِدِ النَّبِيِّ صَلَّى اللهُ عَلَيْهِ وَسَلَّمَ؟ فِي السَّاعَة.... ٦:٣٠ ٣:٣٠ ٧:٣٠ ٤:٣٠
32.	1.5. Menerapkan kaidah tentang bentuk, makna dan fungsi dari gramatikal كان واسمها وخبرها المفرد dalam menyusun teks naratif sederhana yang berkaitan dengan tema: الحفل بمولد الرسول	• Bentuk, makna dan fungsi dari susunan gramatikal كان واسمها وخبرها المفرد	Mendeteksi fungsi كان وخبرها المفرد واسمها dalam kalimat	١. كَانَتْ خَدِيْجَةُ تِلْمِيْذَةً مَاهِرَةً ٢. الْمُسْلِمُوْنَ كُنْتُمْ صَالِحِيْنَ ٣. كَانَتْ الْمَدْرَسَةُ جَمِيْلٌ ٤. كَانَ الْفَصْلُ وَاسِعًا الْجُمْلَة الصَّحِيْحَة نَمْرَة .... ١، ٣ ١، ٤ ٢، ٤ ٢، ٣
33.	berkaitan dengan tema: نزول القرآن والعيدان yang melibatkan tindak tutur melarang melakukan suatu pekerjaan dengan memperhatikan susunan gramatikal لا الناهية/لم + الفعل (الفعل المضارع - المزيد من الثلاثي)ماض ومضارع ومصدر	(bunyi, kata, dan makna) dari teks sederhana yang berkaitan dengan tema: نزول القرآن والعيدان	Mengidentifikasi unsur kebahasaan dalam teks dengan tema نزول القرآن والعيدان	... الْطَّالِبَاتُ ... فِي دُرُوْسِهِنَّ. كُنَّ – مُجْتَهِدَاتٌ كَانَتْ - مُجْتَهِدَاتٌ كَانَتْ – مُجْتَهِدَاتٍ كَانَ – مُجْتَهِدِيْنَ

34.	3.6 Menganalisis gagasan dari teks naratif sederhana yang berkaitan dengan tema: نزول القرآن والعيدان dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal - لا الناهية/لم + الفعل المضارع المزيد من الفعل الثلاثي)ماض ومضارع ومصدر(	Teks naratif tema نزول القرآن والعيدان Bentuk, makna dan fungsi dari struktur kalimat + لا الناهية/لم - الفعل المضارع المزيد من الفعل الثلاثي)ماض ومضارع ومصدر(	• Menentukan pokok pikiran pada teks tema نزول القرآن والعيدان Mendeteksi penggunaan لم النفي dan لا الناهية dalam kalimat	إِخْتَفَلَ الطُّلاَّبُ بِذِكْرَى نُزُوْلِ الْقُرْآنِ فِى الْأُسْبُوْعِ الْمَاضِى. إِخْتَفَلُوْا بِهَا فِى السَّابِعِ عَشْرَةَ مِنْ شَهْرِ رَمَضَان. الْفَقْرَةُ الْأُوْلَى الْإِفْتِتَاح, وَالْفَقْرَة الثَّانِيَة كَلِمَة تَرْجِيْبِيَّة أَلْقَاهَا رَئِيْسُ اللَّجْنَة. وَالْفَقْرَة الثَّالِثَة الْمُحَاضَرَة الدِّيْنِيَّة الْفَقْرَة الرَّابِعَةِ الْإِخْتِتَام بِالدُّعَاء. مَتَى إِخْتَفَلَ الطُّلاَّبُ بِذِكْرَى نُزُوْلِ الْقُرْآن؟ فِي الْأُسْبُوْعِ الْقَادِم فِي السَّابِعِ عَشْرَة مِنْ شَهْرِ رَمَضَان فِي السَّابِعِ وَعِشْرِيْنَ مِنْ شَهْرِ رَمَضَان فِي الْيَوْمِ الْأَوَّلِ مِنْ شَهْرِ شَوَّال
35.	3.7 Memahami fungsi sosial, struktur teks dan unsur kebahasaan (bunyi, kata, dan makna) dari teks sederhana yang	Fungsi sosial, struktur teks dan unsur kebahasaan (bunyi, kata, dan	Menganalisis teks dengan tema جمال الطبيعة	اللهُ الَّذِي أَنْزَلَ ... مِنَ السَّمَآءِ. الْمَطَر الأَشْجَار أَزْهَار النُّجُوْم
36.	berkaitan dengan tema: جمال الطبيعة yang melibatkan tindak tutur memberi instruksi dengan memperhatikan susunan gramatikal فعل الأمر وتصريفه	makna) dari teks sederhana yang berkaitan dengan tema: جمال الطبيعة	Mengidentifikasi unsur kebahasaan dalam teks dengan tema جمال الطبيعة	جَبَلُ Semeru مُرْتَفِعٌ. بَلْ ... الجِبَالِ فِي جَاوَا أَوْسَعُ أَكْبَرُ أَطْوَلُ أَرْفَعُ
37.	3.8 Menganalisis gagasan dari teks sederhana yang berkaitan dengan tema: جمال الطبيعة dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal فعل الأمر وتصريفه	Teks tema جمال الطبيعة Bentuk, makna dan fungsi dari susunan gramatikal فعل الأمر وتصريفه	• Menyimpulkan isi teks tema جمال الطبيعة Merinci perubahan dalam تصريف فعل الأمر	أَكْرَمْتُ التَلَامِيْذُ الدِّيْنَ يَتَخَلَّقُوْنَ بِأَخْلَاقِ كَرِيْمَةِ. الكَلِمَة التي تَحْتَهَا خَطٌّ ... إِسْم الْمَوْصُوْل مَوْصُوْلَة صِلَة الْمَوْصُوْل عَائِد
38.	3.10 Menganalisis gagasan dari teks naratif sederhana yang berkaitan dengan tema: خالق العالم dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal الموصول)الذي، التي، الذين، اللاتي(	• Teks naratif tema خالق العالم • Bentuk, makna dan fungsi dari susunan gramatikal الموصول)الذي، التي، الذين، اللاتي(	Mengumpulkan bukti ciptaan Allah dalam teks naratif tentang tema Menelaah perubahan bentuk pada اسم الموصول	الْمُؤْمِنُونَ ... يُصَلُّوْنَ فِي الْمَسْجِدِ خَاشِعُوْنَ فِي صَلَاتِهِم. الَّذِيْ الَّتِي الَّذِيْنَ اللَّاتي
39.	3.11 Memahami fungsi sosial, struktur teks dan unsur kebahasaan (bunyi, kata, dan makna) dari teks sederhana yang berkaitan dengan tema: الحفاظ على البيئة yang melibatkan tindak tutur pernyataan perbandingan dengan memperhatikan susunan gramatikal	Fungsi sosial, struktur teks dan unsur kebahasaan (bunyi, kata, dan makna) dari teks sederhana yang berkaitan dengan tema: الحفاظ على البيئة	Mengidentifikasi unsur kebahasaan dalam teks dengan tema جمال الطبيعة	جَعَلَ اللهُ القَمَرَ وَ النُجُوْمَ لِتُزَيِّنَ ... فِي اللَّيْلِ، فَكَانَ اللَّيْلُ نُوْرًا وَ جَمِيْلًا الأَرْضُ السَّمَاءُ البِحَارُ الْمَدِيْنَةُ

	اسم التفضيل)أفعل + من/ جمع(			
40.	3.12 Menganalisis gagasan dari teks sederhana yang berkaitan dengan tema: الحفاظ على البيئة dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal اسم التفضيل)أفعل + من/ جمع(	Teks tema الحفاظ على البيئة Bentuk, makna dan fungsi dari susunan gramatikal / اسم التفضيل)أفعل + من جمع(	• Menganalisis fakta dalam teks tema الحفاظ على البيئة Menelaah perubahan bentuk اسم التفضيل dalam kalimat	نَهْرُ Kapuas .... الأَنْهَار فِي إِنْدُونِيْسِيَّا أَكْبَرُ أَطْوَلُ أَوْسَعُ أَقَلُّ
41.	3.2 Menganalisis gagasan dari teks naratif sederhana yang berkaitan dengan tema: رأس السنة الهجرية dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal تصريف الفعل الماضي المجرد	• Bentuk, makna dan fungsi dari susunan gramatikal تصريف الفعل الماضي المجرد	• Mendeteksi berbagai bentuk dhomir dalam تصريف الفعل الماضي المجرد	هَذَا أَخِي اِسْمُهُ تَوْفِيق. هُوَ لَاعِبُ كُرَةِ السَّلَّة . يَلْعَبُ كُرَةَ السَّلَّة كُلَّ يَوْم فِي القَاعَةِ الرِّيَاضِيَّة. مَا مَعْنَى كُرَّة السَّلَّة؟ Sepak bola Bola voli Bola basket Bulu tangkis
42.	3.4 Menerapkan kaidah tentang bentuk, makna dan fungsi dari gramatikal كان واسمها وخبرها المفرد dalam menyusun teks naratif sederhana yang berkaitan dengan tema: الحفل بمولد الرسول	• Teks naratif tema الحفل بمولد الرسول	• Menyimpulkan rangkaian acara peringatan maulid Nabi dalam teks naratif tema الحفل بمولد الرسول	الجُمْلَةُ الَّتِي فِيْهَا مَصْدَر هِيَ . . . نَأْكُلُ الرُّزَّ مَعًا فِي غُرْفَةِ الْأَكْلِ يَتَعَلَّمُ الطلاَّبُ ثَمَانِيَ حِصَصٍ فِي الْمَدْرَسَة نُرِيْدُ مُشَاهَدَةَ الْمُبَارَة فِي التِّلْفَازِ يَسْتَرِيْحُ الطُّلَّابُ فِي السَّاعَةِ الْعَاشِرَةِ
43.	1.6. Menerapkan kaidah tentang bentuk, makna dan fungsi dari gramatikal كان واسمها وخبرها المفرد dalam menyusun teks naratif sederhana yang berkaitan dengan tema: الحفل بمولد الرسول	• Bentuk, makna dan fungsi dari susunan gramatikal كان واسمها وخبرها المفرد	Mendeteksi fungsi كان وخبرها المفرد واسمها dalam kalimat	فَرِيدُ لَمْ يَدْخُلْ إِلَى الْمَدْرَسَةِ لِأَنَّهُ مَرِيْضٌ. فَيَذْهَبُ إِلَى الْمُسْتَشْفَى لِيَفْحَصَهُ الطَّبِيْبُ وَ يَقُوْلُ لَهُ " قَدْ أَصَابَكَ الزُّكَامُ وَ سَيَزُوْلُ آلَمُكَ بَعْدَ تَنَاوُلِ هَذِهِ الْأَدْوِيَّةِ." مَا أصَابه فَرِيْدُ ؟ الْمَرِيْضُ الصُّدَاعُ الزُّكَامُ آلَمٌ
44.	berkaitan dengan tema: نزول القرآن والعيدان yang melibatkan tindak tutur melarang melakukan suatu pekerjaan dengan memperhatikan susunan gramatikal لا الناهية/لم + الفعل (الفعل المضارع - المزيد من الثلاثي)ماض ومضارع ومصدر	(bunyi, kata, dan makna) dari teks sederhana yang berkaitan dengan tema: نزول القرآن والعيدان	Mengidentifikasi unsur kebahasaan dalam teks dengan tema نزول القرآن والعيدان	عُمَرُ بْن الْخَطَّاب هُوَ أَوَّلُ مَنْ وَضَعَ بِالسَّنَة الْجَدِيْدَة فِي الْإِسْلَام وَ يُسَمَّى بِالسَّنَةِ الْهِجْرِيَّة. وَ احْتَفَلَ الْمُسْلِمُوْنَ فِي الْعَالَم بِالسَّنَةِ الْجَدِيْدَة الْهِجْرِيَّة بِالسُّرُوْرِ وَ الْفَرَح، هُمْ يُهَنِّئُوْنَ بَعْضًا كُلَّ عَامٍ وَ أَنْتُمْ بِخَيْر. الْمَوْضُوْعُ الْمُنَاسِبَ بِالْفَقْرَة السَّابِقَةِ هي؟ الأَعْيَادُ فِي الْإِسْلَامِ السَّنَةُ الْهِجْرِيَةِ نُزُوْلُ الْقُرْآنُ السَّنَةُ الْمِيْلَادِيَةُ
45.	3.6 Menganalisis gagasan dari teks naratif sederhana yang berkaitan dengan tema: نزول القرآن والعيدان dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal - لا	Teks naratif tema نزول القرآن والعيدان Bentuk, makna dan fungsi dari struktur kalimat لا الناهية/لم + الفعل المضارع - المزيد من الفعل الثلاثي	• Menentukan pokok pikiran pada teks tema نزول القرآن والعيدان Mendeteksi penggunaan لم النفي dan لا الناهية dalam kalimat	كَانَ الْمُسْلِمانَ ... مُخْلِصَيْنِ مُخْلِصٌ مُخْلِصِيْنَ مُخَلِصَاتِ

	الناهية/لم + الفعل المضارع المزيد من الفعل الثلاثي)ماض ومضارع ومصدر(	)ماض ومضارع ومصدر(		
46.	3.7 Memahami fungsi sosial, struktur teks dan unsur kebahasaan (bunyi, kata, dan makna) dari teks sederhana yang	Fungsi sosial, struktur teks dan unsur kebahasaan (bunyi, kata, dan	Menganalisis teks dengan tema جمال الطبيعة	زَيْدٌ : كَمَا عَرَفْنَا عِنْدَنَا مُنَاسِبَةٌ دِيْنِيَّةٌ عَظِيْمَةٌ فِى هَذَا الشَّهْرِ وَهِيَ نُزُوْل ألْقُرْآنِ خَلِيْلٌ: نَعَمْ صَحِيْح يَاأَخِى، إِذَنْ مَتَى نَحْتَفِلُ بِذِكْرَى نُزُوْل ألْقُرْآنِ؟ زَيْدٌ : نَحْتَفِلُ بِذِكْرَى نُزُوْل ألْقُرْآنِ فِى ... الْأَوَّلِ مِنْ شَوَّال الثَّانِي عَشَرَ مِنْ رَبِيْعِ الْأول السَّابِعِ وَعِشْرِيْنَ مِنْ رَجَب السَّابِــعِ عَشَرَ مِنْ رَمَضَـان
47.	berkaitan dengan tema: جمال الطبيعة yang melibatkan tindak tutur memberi instruksi dengan memperhatikan susunan gramatikal فعل الأمر وتصريفه	makna) dari teks sederhana yang berkaitan dengan tema: جمال الطبيعة	Mengidentifikasi unsur kebahasaan dalam teks dengan tema جمال الطبيعة	حميد : يا فَائِز أُنْظُرْ هَذِهِ الصُّوْرَة ! مَاذَا يعْمَل الْمُسْلِمُون؟ فائز : هُمْ ... يُصَلُّوْنَ الْعِيْد يُقَدِّمُوْنَ الزَّكَاة يُكَبِّرُوْنَ اللّٰهَ وَ يَحْمَدُوْنَهُ يَتَصَافَحُوْنَ بَعْضَهُمْ بَعْضًا
48.	3.8 Menganalisis gagasan dari teks sederhana yang berkaitan dengan tema: جمال الطبيعة dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal فعل الأمر وتصريفه	Teks tema جمال الطبيعة Bentuk, makna dan fungsi dari susunan gramatikal فعل الأمر وتصريفه	• Menyimpulkan isi teks tema جمال الطبيعة Merinci perubahan dalam تصريف فعل الأمر	Tentukan Fiil Amar dari Fiil mudhori' berikut dengan benar! مضارع | يَسْجُدُ | يُسَاعِدُ | يَسْتَمِعُ أمر | ... | ... | ... اِسْجُدْ – اِسَاعِدْ – اِسْتَمِعْ أُسْجُدْ – سَاعِدْ – اَسْتَمِعْ اَسْجُدْ – أُسَاعِدْ – اِسْتَمِعْ أُسْجُدْ – سَاعِدْ – اِسْتَمِعْ
49.	3.10 Menganalisis gagasan dari teks naratif sederhana yang berkaitan dengan tema: خالق العالم dengan memperhatikan bentuk, makna dan fungsi dari susunan gramatikal الموصول)الذي، التي، الذين، اللاتي(	• Teks naratif tema خالق العالم • Bentuk, makna dan fungsi dari susunan gramatikal الموصول)الذي، التي، الذين، اللاتي(	Mengumpulkan bukti ciptaan Allah dalam teks naratif tentang tema Menelaah perubahan bentuk pada اسم الموصول	يَا فَاطِمَةُ، ... عَلَى الْكُرْسِي مِنْ فَضْلِكِ! اِجْلِسِيْ اِجْلِسْنَ اِجْلِسُوْا اِجْلِسْ
50.	3.11 Memahami fungsi sosial, struktur teks dan unsur kebahasaan (bunyi, kata, dan makna) dari teks sederhana yang berkaitan dengan tema: الحفاظ على البيئة	Fungsi sosial, struktur teks dan unsur kebahasaan dari teks sederhana yang berkaitan dengan tema: الحفاظ على البيئة	Mengidentifikasi unsur kebahasaan dalam teks dengan tema جمال الطبيعة	Tentukan Isim Tafdhil dari Isim Sifat berikut dengan benar! صَغِيْرٌ | نَشِيْطٌ | ... | ... | مَاهِرَةٌ – نَشِيْطَةٌ – صَغِيْرَةٌ مَهَرُ –نَشَطُ – صَغَرُ أَمَاهِرُ – أَنَاشِطُ – أَصَاغِرُ أَمْهَرُ – أَنْشَطُ – أَصْغَرُ

LAMPIRAN 2:

DAFTAR NILAI ASESMEN
MADRASAH BAHASA ARAB
TA.2022/2023 MTs. N. Kota Batu

Jumlah Sampel : 313 Siswa
Jumlah Soal : 50 Soal

السيرة الذاتية

أ. المعلومات الشخصية

اسم	:	محمد جندي
تاريخ الميلاد	:	نامبو بادانج، ٧ نوفمبر ١٩٩٩
كلية	:	كلية الدراسات العليا
قسم	:	تعليم اللغة العربية
عنوان	:	نامبو-بانجاي-سولاويسي الوسطى
البريد الإلكتروني	:	jundijundi10@gmail.com

ب. السيرة التعليمية

المستوى الدراسي	
المدرسة الإبتدائية الحكومية ٣ نامبو (SDN 3 Nambo)	٢٠٠٥-٢٠١١م
المدرسة المتوسطة الإسلامية الحكومية لووّك (MTs. N. Luwuk)	٢٠١١-٢٠١٤م
مدرسة إنسان جيندیكيا الثانوية الإسلامية الحكومية غورونتالو (MAN Insan Cendekia Gorontalo)	٢٠١٤-٢٠١٧م
جامعة سلطان أماي الإسلامية الحكومية غورونتالو (IAIN Sultan Amai Gorontalo)	٢٠١٧-٢٠٢١م