

تحليل بنود أسئلة متعددة إختيارات اللغة العربية

## TAHLILU BUNUD ASILAH MUTA'DDIDAH IKHTIYARAT AL-LUGHAH AL-ARABIYAH

Ali Fuddin Nasution<sup>1</sup>, Baiq Laili Rahmawati<sup>2\*</sup>, Madian Fitri Salma<sup>3</sup>, Nur Qomari<sup>4</sup>, M. Syarifuddin<sup>5</sup>  
<sup>1,2,3,4</sup>Universitas Maulana Malik Ibrahim Malang

<sup>5</sup>Universitas Islam Negeri Mataram

<sup>1</sup>[alifuddinnasution@gmail.com](mailto:alifuddinnasution@gmail.com), <sup>2\*</sup>[bqlailirahmawati@gmail.com](mailto:bqlailirahmawati@gmail.com), <sup>3</sup>[madian.fitri.salma@gmail.com](mailto:madian.fitri.salma@gmail.com),  
<sup>4</sup>[qomari@uin-malang.ac.id](mailto:qomari@uin-malang.ac.id), <sup>5</sup>[muhammadsyarifuddin@uinmataram.ac.id](mailto:muhammadsyarifuddin@uinmataram.ac.id)

### Abstract:

This study discussed about the importance of vocabulary analysis in multiple choice, because look at the facts occur, the teacher was asked questions only based on the book without considering whether a different strength, level of difficulty or deceptive force. The purpose of this study was to analyzed in the form of the differential power, difficulty and the deceptive power found among the students of the second grade-A at Khairuddin High School Gondanglegi Malang. This research uses a mixed research method. Data collection methods using documentation and data analysis using content analysis techniques. Data sources in this study are primary data sources in the form of documents on the results of student assessments and secondary data sources in the form of supporting references from research articles and related books. The results of this study are: 1) The discriminatory power of 60% is very low and 40% is low, medium and difficult category must be reviewed but the difficulty without be reviewed. 2) The difficulty level of the questions is 46.66% easy, 26.66% medium, 26.66% difficult. 3) The deceptive power, there was 16 options that were accepted so that they could be used immediately without review.

Keywords: *deceptive power, discriminatory power, Item analysis, test difficulty level*

Received: January 6, 2023

Revised: February 4, 2023

Accepted: February 20, 2023

Published: February 23, 2023

### المقدمة

وكة نحفي عالم التعليم، لا غنى عن التقييم والتقييم ولا يمكن فصلهما، لأنهما مرتبطان ارتباطاً وثيقاً ببعضهما البعض في عملية التعلم. التقييمات والاختبارات التي يتم إجراؤها بشكل صحيح وصحيح ستشجع المعلمين على التدريس بشكل أفضل وتحفيز الطلاب ليكونوا أكثر نشاطاً في التعلم. يبدأ التقييم عادة بقياس لتحديد مستوى فهم الطلاب وقدرتهم أثناء عملية التعلم. القياس هو فرع من الإحصائيات التطبيقية يهدف إلى بناء الأسس لتطوير اختبار جيد، بحيث يكون الاختبار الناتج موضوعياً ويعمل بشكل مثالي وصحيح وموثوق.

هناك اعتباران رئيسان في اختيار نوع اختبار جيد للاستخدام. أولاً، طبيعة مخرجات التعلم. يجب أن تقيس عناصر الاختبار المعطاة للطلاب نتائج التعلم مباشرة. ثانياً، نوعية المواد التي تم<sup>1</sup>. لذلك، يعد هذا الاختبار الموضوعي مناسباً جداً لتقييم المدى الذي تتطلب فيه قدرات الطلاب عمليات عقلية ليست عالية جداً، مثل تذكر مبادئ الاختبار الموضوعي ككل والتعرف عليها وفهمها وتطبيقها. ويتم إجراء هذا الاختبار الموضوعي بهذه الطريقة، بحيث يمكن تقييم نتائج الاختبار بموضوعية من قبل أي شخص وستنتج نفس النتيجة. في هذه الحالة، هناك عدة أشكال من الاختبارات الموضوعية، مثل اختبارات الإكمال (الاختبارات الكاملة)، والملاءم (ملء الفراغات في الجمل)، والاختلاف في هذا الاختبار في اختبار الملاءم، والمادة التي يتم اختبارها هي قصة واحدة، بينما عند الانتهاء من الأسئلة يمكن أن تكون مختلفة عن بعضها البعض. ثم صح خطأ (صح خطأ)، واختيار من متعدد (اختيار من متعدد)، ومطابقة (مطابقة). في كل شكل من أشكال هذا الاختبار، لكل منها مزاياها وعيوبها<sup>2</sup>.

باستخدام أداة في شكل اختبار، يمكن للمدرس معرفة نتائج التعلم للطلاب بسهولة. في عالم التعليم، تتكون الاختبارات عادةً من نوعين، هما الاختبارات القياسية والاختبارات غير القياسية<sup>3</sup>. يتم إجراء اختبارات موحدة لتحسين ما لدى هؤلاء الطلاب. في هذا الاختبار القياسي، يجب مراعاة جوانب مختلفة، مثل الصلاحية والموثوقية وجدوى العناصر. بينما في الاختبارات غير القياسية، أي الاختبارات التي يتم إجراؤها وترتيبها بواسطة معلمي المادة، مثل الاختبارات التي يتم ترتيبها في اختبارات منتصف الفصل الدراسي. يمكن أن تكون هذه الاختبارات غير المعيارية في شكل خيارات أو أوصاف متعددة<sup>4</sup>.

تحليل العنصر هو اختبار ميداني لاختبار أو سؤال يتم إجراؤه في مؤسسة تعليمية. وجودها كوسيلة لتأكيد جودة الأسئلة، بعد أن يعرف أن الأسئلة ضعيفة من حيث القوة التمييزية، أو صعوبة للغاية أو سهلة للغاية، وأن المشتتات غير مناسبة، فيجب أن تكون هذه التغذية الراجعة للاختبار أو صانع الأسئلة لمراجعة الأخطاء التي حدثت في الأسئلة التي تم إعدادها<sup>5</sup>. يمكن استخدام العناصر الموضوعية بشكل متكرر، طالما أن الأسئلة ليست في نفس مجموعة الاختبار. لذلك، فإن الفائدة أو

<sup>1</sup> SaifuddinAzwar, *Reliabilitas Dan Validitas* (Yogyakarta: Pustaka Pelajar, 2010), 34.

<sup>2</sup> Muhammad ilyas Ismail, *Evaluasi Pembelajaran: Konsep Dasar, Prinsip, Teknik, Dan Prosedur* (Depok: Rajawali pers, 2020), 37–39.

<sup>3</sup> Suharsimi Arikunto, *Prosedur Penelitian: Suatu Pendekatan Praktik* (jakarta: Rineka cipta, 2013), 159.

<sup>4</sup> Nurul Fikriyah, 'Analisis Butir Soal Ulangan Tengah Semester Mata Pelajaran Bahasa Arab Kelas VII Semester Ganjil SMP Muhammadiyah 1 Yogyakarta Tahun Ajaran 2019/2020', *Maharaat: Jurnal Pendidikan Bahasa Arab* 3, no. 2 (2021): 129, <https://doi.org/10.18196/mht.v3i2.10501>.

<sup>5</sup> dan Mad'ali Muhammad Rizal, Syihabuddin, 'Evaluasi Pembelajaran Bahasa Arab Dengan Memanfaatkan Aplikasi Misk', *Tsaqofiya: Jurnal Pendidikan Bahasa Dan Sastra Arab* Vo.3, no. No.2 (2021).

الفائدة في تحليل هذا العنصر هي مراجعته بحيث يمكن تصحيح العناصر التي لم يتم إنشاؤها بشكل صحيح.

في كل سؤال، في هذه الحالة، يجب أن يكون للاختيار من متعدد قوة مختلفة، ومستوى صعوبة، وقوة خادعة. القوة المميزة هي قياس مدى قدرة العنصر على التمييز بين الطلاب الذين أتقنوا كفاءة السؤال عن الطلاب الذين لم يعتمدوا على معايير معينة. مستوى صعوبة السؤال هو قياس مدى صعوبة درجة صعوبة السؤال. وقوة التشتيت هي إجابة بديلة وهي مشتت للانتباه في كل عنصر من عناصر السؤال. أسئلة جيدة، سيتم اختيار الإجابات بالتساوي من قبل الطلاب الذين يجيبون بشكل غير صحيح. من ناحية أخرى، العناصر غير الجيدة، سيتم اختيار المشتت بشكل غير متساو.<sup>6</sup>

تركز المشكلة في هذه الدراسة على هذه الأشياء الثلاثة، وهي تحليل أسئلة اللغة العربية في شكل قوة التفريق، ومستوى الصعوبة، وقوة التشتيت الموجودة في طلاب الصف الثاني (أ) بمدرسة خير الدين الثانوية جوندنليجي مالانج. بناءً على المعلم. أن المعلومات في المدرسة، قاموا بطرح أسئلة تستند فقط إلى مادة الكتاب المستخدمة في تعلم اللغة العربية دون النظر فيما إذا كان لكل عنصر بالفعل قوة مميزة أو مستوى صعوبة أو قوة خادعة. وهذا بالطبع مخالف لمبدأ طرح الأسئلة، أي أنه يجب طرح الأسئلة من خلال الاهتمام بنوعية الأسئلة، وليس فقط الاهتمام بكمية الأسئلة واكتمال طرحها. لذلك، يهتم الباحثون بإجراء بحث بعنوان تحليل عناصر أسئلة الاختيار من متعدد باللغة العربية لدى طلاب الصف الثاني (أ) بمدرسة خير الدين الثانوية جوندنليجي مالانج.

هناك العديد من الدراسات التي غطت هذا الموضوع سابقًا بما في ذلك البحث الذي أجراه (رشماد رمضان وفاسيش نور فردوس) "تحليل بنود إختبار الفصلي الدراسي الأوسط للغة العربية الصف الثاني عشر في مدرسة العالية العزة IBS مالانج". في هذا الاختبار، يوجد 25 سؤالاً يحتوي كل سؤال فيها على 5 اختيارات للإجابة. في السؤال الأول يطرح المعلم أسئلة حول الفهم. حتى لا يعرف الطلاب قواعد اللغة العربية فحسب، بل يحتاجون أيضًا إلى فهم معنى عناصر اللغة العربية التي تتم دراستها. في السؤال الثاني، يعطي المعلم أسئلة تتطلب من الطلاب تحليل الأسئلة للعثور على إجابات. في السؤال، يعطي المعلم آيات من القرآن جملة لتحليلها وتحديدتها من خلال الخيارات المتعددة المتاحة. وفي السؤال الثالث يطرح المعلم أسئلة تتطلب من الطلاب فهم قواعد اللغة العربية. في هذا

<sup>6</sup> Arifin, *Evaluasi Pembelajaran*.

السؤال، يطرح المعلم أسئلة لتكوين جمل عربية، ولكن لأن هذا السؤال في شكل اختيار من متعدد، يحتاج الطلاب فقط إلى اختيار الإجابة التي طرحها المعلم<sup>7</sup>.

أجرى البحث التالي أكبر إسكندر ومحمد ريزال بعنوان "تحليل جودة الأسئلة في التعليم العالي بناءً على تطبيق TAP". أظهرت النتائج بالنسبة لصحة المحتوى أن قيمة صدق المحتوى (6) كانت 0.42 متضمنة الفئة المتوسطة. علاوة على ذلك، يبدو أن قيمة معامل موثوقية الأداة هي 0.514. عدد العناصر الصعبة 57.5٪، الفئة المتوسطة 42.5٪ ولا توجد أسئلة فئة سهلة. بالإضافة إلى ذلك، فإن العناصر التي تتمتع بقوة تمييز ممتازة هي 5٪، جيدة بقدر 20٪، وتحتاج إلى مراجعة تصل إلى 13.75٪، وليست جيدة بقدر 61.25٪. في حين أن الخيار الذي لا يعمل بشكل صحيح عند استخدامه كمشتت هو 5 أسئلة، ولكن هناك 40 سؤالاً يجب مراجعتها لأن خيار التشتيت يعتبر مفتاح الإجابة من قبل المشاركين الأذكاء<sup>8</sup>.

تم إجراء دراسة أخرى بواسطة بيبريان ترميزي، وبانوت سيتيونو، ويولي أمالية، وعريف أجريان بعنوان "تحليل أسئلة الاختيار من متعدد حول الموضوعات الصحية مهم لمدرسة الصف الخامس الابتدائية 04 مدينة بنجكولو". أظهرت نتائج الدراسة أن 90٪ من الأسئلة تم تصنيفها على أنها صحيحة و 10٪ من الأسئلة مصنفة على أنها غير صالحة، ومستوى الموثوقية مرتفع مع R1 من 0.83، ومستوى صعوبة الأسئلة بنسبة 15٪. من الأسئلة بما في ذلك متوسط و 85٪ من الأسئلة بما في ذلك فئة عالية جداً، وعلى مستوى قوة التفاضل، هناك 35٪ أو ما يصل إلى 7 أسئلة مصنفة على أنها قوة تمييز جيدة، و 55٪ أو ما يصل إلى 11 سؤالاً في الفئة المتوسطة و 10٪ أو ما يصل إلى سؤالين في فئة "سيئة". في غضون ذلك، يُظهر تحليل وظيفة المشتتات أن 10٪ من المشتتات تعمل بشكل جيد، والتي توجد في 18 عنصراً و 90٪ من المشتتات لا تعمل، والتي تنتشر على 20 عنصراً<sup>9</sup>.

تم إجراء بحث لاحق بواسطة سيتي فاطمة الفتحة بعنوان "تحليل الأسئلة في دروس اللغة العربية في مدرسة العالية روضة العلوم بيغك مالاج". 3 بنود فقط (2.85٪) وتم تضمينها في فئة الأسئلة غير الصالحة 12 بنداً (97.15٪). (2) أظهرت نتائج اختبار الموثوقية أن مستوى موثوقية اختبار الاختيار من متعدد كان 0.54، مما يشير إلى الفئة المتوسطة. (3) تُعرف نتائج تحليل مستوى

<sup>7</sup> Rachmad Ramadhan and Fasich Nur Firdaus, 'Analisis Butir Soal Ujian Tengah Semester Bahasa Arab Kelas XII Di SMA Al-Izzah IIBS Malang', *Tsaqofiya : Jurnal Pendidikan Bahasa Dan Sastra Arab* 4, no. 1 (2022): 126–35, <https://doi.org/10.21154/tsaqofiya.v4i1.49>.

<sup>8</sup> Akbar Iskandar and Muhammad Rizal, 'Analisis Kualitas Soal Di Perguruan Tinggi Berbasis Aplikasi TAP', *Jurnal Penelitian Dan Evaluasi Pendidikan* 22, no. 1 (2018): 12–23, <https://doi.org/10.21831/pep.v22i1.15609>.

<sup>9</sup> Pebrian Tarmizi et al., 'Analisis Butir Soal Pilihan Ganda Tema Sehat Itu Penting Kelas V SD Negeri 04 Kota Bengkulu', *ELSE (Elementary School Education Journal) : Jurnal Pendidikan Dan Pembelajaran Sekolah Dasar* 4, no. 2 (2021): 124, <https://doi.org/10.30651/else.v4i2.7090>.

الصعوبة أنه بين الأسئلة رقم من 1 إلى 15 يوجد 8 عناصر ضمن فئة السؤال "السهل"، و 3 عناصر ضمن فئة السؤال "المتوسط"، و 4 عناصر مدرجة في فئة الأسئلة "الصعبة". (4) أظهرت نتائج الاختبار التمييزي أن هناك 3 عناصر تم تضمينها في فئة جيدة، و 3 عناصر تم تضمينها في الفئة المتوسطة، والباقي تنتمي إلى فئة الأسئلة التي لا يمكنها التمييز بين الطلاب من الفئة العليا. المجموعة والمجموعة الدنيا. حتى فرق مستوى الطاقة هو 0 (صفر) لذلك لا يوجد تمييز. المعنى الضمني هو أن العنصر تمت مراجعته أو عدم استخدامه. (5) تم العثور على جودة الأسئلة المشتتة للانتباه، ولا يزال هناك العديد من الأسئلة المشتتة التي تحتاج إلى المراجعة<sup>10</sup>.

تم إجراء دراسة أخرى من قبل نور فكرية "تحليل أسئلة اختبار الفصل الدراسي الأوسط لموضوعات اللغة العربية الفصل السابع الفردي مدرسة المتوسطة المحمدية 1 يوجياكارتا العام الأكاديمي 2020/2019 وكانت نتائج الدراسة: (1) صحة الأسئلة تصل إلى 31 فقرة (89٪) وليس هناك 4 عناصر صالحة (11٪)، (2) مستوى صعوبة الأسئلة السهلة 3 بنود (10٪) بينما المتوسط 25 فقرة (80٪) والصعوبة هي 3 بنود (10٪)، (3) القوة التمييزية للأسئلة اللائقة هي 25 فقرة (81٪) ولا تساوي الكمية (19٪). لذلك، يمكن الاستنتاج أن جدوى وجودة عناصر اختبار منتصف الفصل الدراسي للصف السابع مدرسة المتوسطة المحمدية 1 يوجياكارتا للعام الدراسي 2020/2019 مدرجة في الفئة الكافية.<sup>11</sup>

المساواة بين الدراسات السابقة وهذه الدراسة يعني كلهم يبحثون في تحليل بنود الأسئلة في المدرسة الرسمية. وأما الإختلاف بينهم يمكن النظر من ميدان البحث وأهدافها. يجري هذا البحث بحث المدرسة المتوسطة خير الدين جوندنجليجي مالانج وأهدافه فقط لتحليل الأسئلة الموجودة في تلك المدرسة من جوانب الصعوبة والتمييز والتشتت. يرجى هذا البحث أن يكون مرجعا عند استعداد الأسئلة لكي يكون الأسئلة أسئلة جيدة لتحسين جودة التعلم والأسئلة في المدرسة.

### منهجية البحث

تستخدم هذه الدراسة أسلوب بحث مختلط، أو طريقة مختلطة بين النوعية والكمية للحصول على صورة أكثر اكتمالا لتحليل عنصر اللغة العربية، حيث تكون نتائج البحث في شكل بيانات كمية

<sup>10</sup> Siti Fathimah Al Fathiyah, 'Analisis Butir Soal Pelajaran Bahasa Arab Di MA Roudlotul Ulum Pagak Malang', *Tarbiyatuna* 4, no. 1 (2019): 78.

<sup>11</sup> Fikriyah, 'Analisis Butir Soal Ulangan Tengah Semester Mata Pelajaran Bahasa Arab Kelas VII Semester Ganjil SMP Muhammadiyah 1 Yogyakarta Tahun Ajaran 2019/2020'.

وبيانات نوعية<sup>12</sup>. تأتي البيانات الكمية من نتائج حساب اختبار التمييز، الصعوبة، والتشتت، وبعد ذلك يصف الباحثون البيانات نوعياً. لذلك نوع البيانات في هذا البحث هو الوصف عن التمييز والصعوبة والتشتت لأسئلة اللغة العربية عن الأسئلة الإختبار لدى الطلاب. مصدر بيانات البحث هنا نتائج الإختبار من طلاب الصف الثاني (أ) بمدرسة خير الدين الثانوية جوندنليجي مالانج. لجمع البيانات يستخدم الباحثون الوثائق فقط ولتحليل البيانات يستخدم الباحثون تحليل المحتوى<sup>13</sup>. لتحليل البيانات يستخدم الباحثون الرمز لدى عارفين في كتابه "تقويم التعليم"<sup>14</sup>. وأما تفصيله كما يلي:

الرمز من الصعوبة:<sup>15</sup>

$$TK = \frac{(WL + WH)}{(nL + nH)} \times 100\%$$

الرمز من التمييز:<sup>16</sup>

$$DP = \frac{(WL - WH)}{n}$$

الرمز من التشتت:<sup>17</sup>

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

### تحليلها ومناقشتها

بالتأكيد أن التقويم لا يتخلع عن التقييم والقياس. يستخدم كلاهما لمعرفة مدى الاختصاص على كفاءة الطلاب في التعلم<sup>18</sup>. عند إجراء التقويم، هناك حاجة إلى أدوات البحث التي لا تحتوي على الكمية فحسب، بل تتميز أيضاً بالجودة التي يمكن إثباتها من حيث الصلاحية والموثوقية تمتلك فيها التمييز، الصعوبة والتشتت. كما أوضحنا سابقاً، يركز هذا البحث فقط على تحليل التمييز،

<sup>12</sup> John W. Creswell, *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research* (boston: MA: Pearson, 2012).

<sup>13</sup> sandu siyoto dan M. Ali Sodik, *Dasar Metodologi Penelitian* (Yogyakarta: Literasi Media, 2015).

<sup>14</sup> Zainal Arifin, *Evaluasi Pembelajaran*, ed. Pipih Latifah (Bandung: PT. Remaja Rosdakarya, 2012).

<sup>15</sup> WL= عدد فرقة السفلى، nL= عدد الطلاب الذي يجيبون الأسئلة خاطئاً من فرقة السفلى، WH= عدد الطلاب الذي يجيبون الأسئلة خاطئاً من فرقة الأعلى، nH= عدد فرقة الأعلى

<sup>16</sup> DP= قوة التمييز، WL= عدد فرقة السفلى، WH= عدد الطلاب خائب من فرقة الأعلى، n= 27% x N

<sup>17</sup> IP= مؤشر التشتت، P= عدد الطلاب الذين يشاركون التشتت، N= عدد الطلاب الذين يشتركون في الإختبار، B= عدد الطلاب الصحيحة في إجابة الأسئلة، I= عدد التثبيت، n= لكل رقم

<sup>18</sup> Ibadullah Malawai dan Endang Sari Maruti, *Evaluasi Pendidikan* (Jawa timur: Cv. Ae Media Grafika, n.d.).

الصعوبة والتشتت في أسئلة الاختيار من متعدد باللغة العربية في مدرسة جوندنليجي، جاوة الشرقية. وبناءً على نتائج تحليل الأسئلة يكسب الباحثون إلى البيانات التالية:

## 1. التمييز

القوة المميزة بقدرة السؤال على التمييز بين الكفاءات التي يمتلكها الطلاب، من الأقل ذكاءً إلى الأكثر ذكاءً. بناءً على نتائج تحليل العناصر باستخدام Microsoft Excel، وجد الباحثون أن عنصرًا واحدًا كان في فئة جيدة، 3 أسئلة في فئة متوسطة، 8 أسئلة في فئة ضعيفة، 18 سؤالًا في فئة ضعيفة جدًا. لذلك، هناك 3 متابعات محتملة وهي: (1) يمكن قبول الأسئلة في فئة جيدة أو جيدة جدًا ويمكن تخزينها في بنك الأسئلة كأداة تقييم أخرى لأنها تتمتع بجودة مناسبة، (2) أسئلة في فئة سيئة يمكن رفض الفئة أو تجاهلها لأنه لا يمكن استخدامها لمزيد من التقييم، ولكن بصرف النظر عن تجاهلها، يمكن أيضًا مراجعة هذه الأسئلة بحيث يمكن إعادة استخدامها، (3) قد لا يتم استخدام الأسئلة ذات الفئات غير المرضية في التقييمات المستقبلية، مما يعني أن يجب استبدال الأسئلة<sup>19</sup>

بناءً على هذا البيان، في الأسئلة التي حلها الباحث، كان هناك 12 سؤالًا يمكن استخدامها مرة أخرى كأداة تقييم، على الرغم من أنه تم مراجعة 8 منهم مرة أخرى لإنتاج أسئلة الجودة. بينما الأسئلة الـ 18 المتبقية عبارة عن فئة ضعيفة جدًا بحيث لا يمكن استخدامها كأداة لإعادة التقييم أو يجب استبدالها. يشير هذا إلى أن 60٪ من الأسئلة ذات جودة رديئة وأن الـ 40٪ الأخرى ذات جودة جيدة ومتوسطة وضعيفة ولكن لا يزال من الممكن مراجعتها لإعادة استخدامها.

## 2. الصعوبة

يتعلق مستوى صعوبة الأسئلة بجودة العناصر من حيث مستوى صعوبة كل عنصر. بناءً على نتائج تحليل الأسئلة باستخدام Microsoft Excel، وجد الباحثون أن عنصرًا واحدًا كان في الفئة السهلة جدًا ( 3.33٪ )، و 13 سؤالًا في الفئة السهلة ( 43.33٪ )، و 8 أسئلة في الفئة المتوسطة. ( 26.66٪ ) و 8 أسئلة في الفئة الصعبة ( 26.66٪ ). للحصول على إنجاز تعليمي جيد، من الأفضل أن تكون النسبة بين مستويات صعوبة الأسئلة موزعة بشكل طبيعي. يمكن ترتيب حساب هذه النسب على النحو التالي: (1) 25٪ أسئلة صعبة و 50٪ متوسطة و 25٪ سهلة أو (2)

<sup>19</sup> Muhammad Rizal, Syihabuddin, 'Evaluasi Pembelajaran Bahasa Arab Dengan Memanfaatkan Aplikasi Misk'.

20٪ أسئلة صعبة و 60٪ متوسطة و 20٪ سهلة أو 3) 15٪ صعبة سؤال متوسط 70٪ وسهل  
15٪<sup>20</sup>.

بناءً على معايير نسبة الأسئلة، يمكن الاستنتاج أن الأسئلة لم يكن بها نسبة جيدة من الأسئلة لأن نسبة الأسئلة التي تحتوي على أسئلة كانت 26.66٪ في الفئة الصعبة، و 26.66٪ في الفئة المتوسطة، و 46.66٪ من الأسئلة في الفئة السهل. وتشير هذه النتائج إلى أن الأسئلة في الفئة السهلة تتجاوز النسبة التي ينبغي أن تكون 25٪ وأسئلة الفئة المتوسطة وهي أقل من النسبة التي ينبغي أن تكون وهي 26.66٪. ستجعل الأسئلة السهلة للغاية الطلاب لا يأخذونها على محمل الجد أو لا تحفز جهود الطلاب لحل المشكلة المطروحة، كما أن الأسئلة الصعبة للغاية ستجعل الطلاب يائسين في حل المشكلة.

### 3. التثنت

مثل القوة التمييزية، يجب أن تستحق أسئلة متعددة الخيارات أيضًا قدرة التثنت في كل خيار من الخيارات المتاحة في كل عناصر. وبناءً على نتائج تحليل البيانات باستخدام Microsoft Excel، وجد الباحثون أن هناك 16 خيارًا تم رفضها في أسئلة البند، ولتسهيل قراءتها صنفها الباحثون في شكل جدول على النحو التالي:

رقم السؤال	خيار السؤال المرفوض	رقم السؤال	خيار السؤال المرفوض
2	1	16	1
4	1	17	1
6	2	18	1
7	1	19	1
12	1	20	1
13	1	21	1

<sup>20</sup> Arifin, *Evaluasi Pembelajaran*.



15	1	28	1
		29	1

استنادًا إلى الجدول، يستنتج أن هناك 16 خيارًا (10.66٪) سؤالًا يجب استبدالها لأنها مدرجة في الفئة المرفوضة وتشير إلى أن الخيار به قدرة تشتيت ضعيفة وأن 134 (89.33٪) خيارًا مقبولاً.

### الخاتمة

بناءً على نتائج البحث أعلاه، وجد الباحثون أن الأسئلة يحتاج إلى الإصلاح حتى يكون الأسئلة أداة جيدة لقياس كفاءة الطلاب، يمكن النظر هذا من خلال نتائج احتساب جودة الأسئلة من جوانب الصعوبة والتمييز والتشتت. هناك الأسئلة التي ليس لها قوة التمييز جيداً، ومستوى الصعوبة غير مناسبة بمقياس الموجودة، وكذلك مؤشر التشتت هناك بعض الخيارات التي يحتاج إلى الإصلاح قبل استعمالها.

بناءً على نتائج هذا البحث، يأمل الباحثون في إلهام المعلمين في طرح المزيد من الأسئلة لتحسين جودة التعليم في المستقبل. بالطبع، لكل بحث حدوده الخاصة التي تصبح عقبات في البحث عن معلومات أكثر تفصيلاً تتعلق بتحليل العناصر. من بين قيود الدراسة أن الباحثين لم يتمكنوا من استكشاف عوامل أوضح تتعلق بالصعوبات التي يواجهها اختصاصيو التوعية في طرح أسئلة التقييم لأن المسافة بين مواقع البحث لم تكن ممكنة. لذلك تأمل الباحثة أن يكون هناك بحث حول الصعوبات التي يواجهها المعلمون في طرح أسئلة التقييم التربوي في المستقبل من أجل إنتاج أدوات التقييم الكمي والكمي.

### قائمة المراجع والمصادر

- Arifin, Zainal, *Evaluasi Pembelajaran*, ed. by Pipih Latifah (Bandung: PT. Remaja Rosdakarya, 2012)
- Arikunto, Suharsimi, *Prosedur Penelitian: Suatu Pendekatan Praktik* (Jakarta: Rineka cipta, 2013)
- Creswell, John W., *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research* (Boston: MA: Pearson, 2012)

- Al Fathiyah, Siti Fathimah, 'Analisis Butir Soal Pelajaran Bahasa Arab Di MA Roudlotul Ulum Pagak Malang', *Tarbiyatuna*, 4.1 (2019), 78
- Fikriyah, Nurul, 'Analisis Butir Soal Ulangan Tengah Semester Mata Pelajaran Bahasa Arab Kelas VII Semester Ganjil SMP Muhammadiyah 1 Yogyakarta Tahun Ajaran 2019/2020', *Maharaat: Jurnal Pendidikan Bahasa Arab*, 3.2 (2021), 128–40 <<https://doi.org/10.18196/mht.v3i2.10501>>
- Iskandar, Akbar, and Muhammad Rizal, 'Analisis Kualitas Soal Di Perguruan Tinggi Berbasis Aplikasi TAP', *Jurnal Penelitian Dan Evaluasi Pendidikan*, 22.1 (2018), 12–23 <<https://doi.org/10.21831/pep.v22i1.15609>>
- Ismail, Muhammad Ilyas, *Evaluasi Pembelajaran: Konsep Dasar, Prinsip, Teknik, Dan Prosedur* (Depok: Rajawali pers, 2020)
- Maruti, Ibadullah Malawai dan Endang Sari, *Evaluasi Pendidikan* (Jawa timur: Cv. Ae Media Grafika)
- Muhammad Rizal, Syihabuddin, dan Mad'ali, 'Evaluasi Pembelajaran Bahasa Arab Dengan Memanfaatkan Aplikasi Misk', *Tsaqofiya: Jurnal Pendidikan Bahasa Dan Sastra Arab*, Vo.3.No.2 (2021)
- Ramadhan, Rachmad, and Fasich Nur Firdaus, 'Analisis Butir Soal Ujian Tengah Semester Bahasa Arab Kelas XII Di SMA Al-Izzah IIBS Malang', *Tsaqofiya : Jurnal Pendidikan Bahasa Dan Sastra Arab*, 4.1 (2022), 126–35 <<https://doi.org/10.21154/tsaqofiya.v4i1.49>>
- SaifuddinAzwar, *Reliabilitas Dan Validitas* (Yogyakarta: Pustaka Pelajar, 2010)
- Sodik, sandu siyoto dan M. Ali, *Dasar Metodologi Penelitian* (Yogyakarta: Literasi Media, 2015)
- Tarmizi, Pebrian, Panut Setiono, Yuli Amaliyah, and Arief Agrian, 'Analisis Butir Soal Pilihan Ganda Tema Sehat Itu Penting Kelas V SD Negeri 04 Kota Bengkulu', *ELSE (Elementary School Education Journal) : Jurnal Pendidikan Dan Pembelajaran Sekolah Dasar*, 4.2 (2021), 124 <<https://doi.org/10.30651/else.v4i2.7090>>