



Analisis Butir Soal Penilaian Akhir Tahun Pelajaran Bahasa Arab Kelas VII MTs Al-Ma'arif Rakit Banjarnegara Jawa Tengah

Muhammad Ali Hasymi Rafsanjani Tanjung*¹, Alif Anharul Fahmi², Fatia Rahmanita³,
Ihda Filzafatin Habibah⁴, Nur Qomari⁵

¹Madrasah Tsanawiyah Negeri 2 Aceh Timur, Indonesia.

²Madrasah Tsanawiyah Rakit Banjarnegara, Indonesia.

^{3,4,5}Arabic Education Study Universitas Islam Negeri Maulana Malik Ibrahim Malang, Indonesia.

Correspondence Address: Malihasyimirt@gmail.com

Received: 11-11-2023

Revised: 29-01-2024

Accepted: 31-01-2024

Abstract

Analysis of question items is very important in learning evaluation because it determines the quality of the evaluation tool used. The aim of this research is to analyze the results of the final semester Arabic evaluation at the Al-Ma'arif Junior High School. The researchers conducted a series of tests consisting of measurements of validity, reliability, differential strength, difficulty level, and debilitating power. The method used in this research is quantitative descriptive method. The population of the study was 25 students of the 7th grade in the IT High School Building Amal School Year 2022/2023. As for the sample used, the final evaluation of the Arabic semester of the teaching year 2022/2023 was documentation and observation. The result of this study is; in the validity test found 30% of questions can be said to be valid. Then the remaining 70% is declared invalid because it has a significance value $<0,05$ and the value of R count is smaller than R Table. Then in the reliability test found using Cronbach's Alpha method, and the element of this question is worth 0.625 with the number of questions (N of items) amounting to 40. Difference shows 6 questions in the category "Bad once in distinction, 25 questions in category "badly once in differentiation", 6 issues in the categories "Sufficient enough in distinguishment", 3 questions with category "Good once in distinctness" whereas with the question of category "good once in discrimination" was not found. Difficulty level there are 25 questions with very easy questions category, 14 questions with easy categories, 1 questions with difficult categories. The fractional strength of the answer option A indicates that the question is not working well with the student as the data indicates more fractionals of the question are rejected than accepted. The same applies to options B, C, and D. Thus, a question that does not meet the revised criteria can be concluded or a new question created and tested again. This research can be used as evaluation material for teachers to create quality evaluation tools. Further research can focus on discussing the development of superior evaluation instruments.

Keywords: Analysis of Question Items, Arabic Language, Final Semester Assessment

ملخص

تحليل عناصر السؤال يعتبر أمرًا هامًا جدًا في تقييم التعلم لأنه يساعد في تحديد جودة أداة التقييم المستخدمة. يهدف هذا البحث إلى تحليل عناصر اختبار نهاية الفصل لمادة اللغة العربية في مدرسة المتوسطة الإسلامية المعارف في راکیت. قام الباحث بإجراء سلسلة اختبارات تتألف من قياسات الصحة، والموثوقية، وفعالية الاختبار، ومستوى الصعوبة، وفعالية الخداع. الأسلوب المستخدم في هذا البحث هو الأسلوب الوصفي الكمي. يبلغ عدد سكان البحث ٢٥ طالبًا في الصف السابع في المدرسة الإسلامية المعارف في راکیت للعام الدراسي ٢٠٢٢/٢٠٢٣. والعينة المستخدمة هي اختبار نهاية الفصل لمادة اللغة العربية للعام الدراسي ٢٠٢٢/٢٠٢٣. الأسلوب المستخدم في جمع البيانات هو التوثيق والمراقبة. نتائج هذا البحث هي: في اختبار الصحة، تبين أن ٣٠% من الأسئلة يمكن اعتبارها صالحة. ثم يتم إعلان البقية، أي ٧٠%، غير صالحة لأنها تحمل قيمة إحصائية أقل من ٠.٠٥. وقيمة R الحسابية أقل من R

الجدولية. ثم في اختبار الموثوقية، تم العثور على قيمة 0.625. باستخدام طريقة ألفا كرونباخ، وعدد الأسئلة (N of item) هو 40. يظهر الفرق في 6 أسئلة بتصنيف "ضعيف جدًا في التمييز"، و 25 سؤالاً بتصنيف "ضعيف في التمييز"، و 6 أسئلة بتصنيف "كافٍ في التمييز"، و 3 أسئلة بتصنيف "جيد في التمييز"، بينما لم يتم العثور على أسئلة تصنيفها "جيد جدًا في التمييز". في مستوى الصعوبة، يوجد 25 سؤالاً بتصنيف "سهل جدًا"، و 14 سؤالاً بتصنيف "سهل"، وسؤال واحد بتصنيف "صعب". يظهر الخداع في الخيار A أن السؤال لا يعمل بشكل جيد على الطلاب لأن البيانات تظهر رفض الخيارات الضالة أكثر من قبولها. وهكذا هو الحال مع الخيارات B و C و D وبالتالي، يمكن استنتاج أن الأسئلة التي لا تفي بالمعايير يجب مراجعتها أو إعداد أسئلة جديدة وإجراء اختبار جديد. يمكن لهذا البحث أن يكون مادة لتقييم الأداء للمعلمين في إنشاء أدوات تقييم عالية الجودة. يمكن أن يركز البحث المستقبلي على مناقشة تطوير أدوات التقييم الأفضل.

الكلمات المفتاحية: تحليل فقرات الأسئلة، اللغة العربية، التقييم الفصلي النهائي

© 2024 Muhammad Ali Hasymi Rafsanjani Tanjung, Alif Anharul Fahmi, Fatia Rahmanita, Ihda Filzafatin Habibah, Nur Qomari



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Pendahuluan

Tes menjadi sebuah hal yang tak terelakkan dalam dunia pendidikan. Sebagai alat atau prosedur pengukuran dan penilaian, tes memainkan peran krusial dalam mengevaluasi prestasi siswa. Dalam praktiknya, tes mengambil bentuk pemberian tugas atau serangkaian aktivitas yang harus diselesaikan oleh siswa. Dari hasil tes ini, dapatlah diperoleh gambaran yang lebih jelas tentang pencapaian siswa dalam bidang pendidikan. Hal ini mencerminkan kompleksitas interaksi antara kurikulum, pengajaran, dan belajar, serta memengaruhi sebagian besar aspek system pendidikan.¹ Testing mewakili suatu proses penting yang terjadi saat pelaksanaan pengukuran dan penilaian. Proses ini menjadi momen kritis di dalam dunia pendidikan, di mana siswa menjalani serangkaian tugas atau kegiatan untuk mengukur dan menilai prestasi mereka. Proses testing mencakup berbagai aspek, seperti desain tugas, pelaksanaan evaluasi, dan interpretasi hasil. Proses ini menandai titik penentuan dalam perjalanan pembelajaran, di mana hasil dari pengukuran tersebut memberikan wawasan mendalam tentang pencapaian siswa. Sebagai suatu peristiwa berlangsungnya pengukuran, testing menciptakan dinamika yang memengaruhi proses pendidikan secara keseluruhan dan menjadi tonggak evaluasi yang esensial.² Kiranya dapat dipahami dalam dunia pendidikan tes juga bisa dikatakan cara yang dapat dipergunakan dalam rangka pengukuran dan penilaian di bidang pendidikan, yang berbentuk pemberian tugas baik berupa pertanyaan-pertanyaan yang harus dijawab atau perintah-perintah yang harus dikerjakan sehingga dapat dihasilkan nilai yang melambangkan

¹ Rina Febriana, *Evaluasi Pembelajaran*, ed. Bunga Sari Fatmawati, (Jakarta: Bumi Aksara, 2019), hlm. 50-57.

² Ubaid Ridho, "Evaluasi Dalam Pembelajaran Bahasa Arab," *An Nabighoh: Jurnal Pendidikan dan Pembelajaran Bahasa Arab*. 20, no. 01 (2018): 19–26. DOI: <https://doi.org/10.32332/an-nabighoh.v20i01.1124>

tingkah laku atau prestasi tes.³ Tes bahasa, khususnya tes bahasa Arab merupakan alat untuk mengukur kemampuan dan performansi bahasa Arab siswa.⁴ Menurut Sri⁵ pengukuran didefinisikan sebagai penetapan suatu angka terhadap suatu subjek dengan cara sistematis. Hasil pengukuran yang berupa angka/skor diharap mencerminkan kemampuan peserta tes sebenarnya.⁶

Evaluasi mencerminkan suatu proses kunci dalam dunia pendidikan yang menjadi acuan oleh pendidik untuk mengetahui keberhasilan dan tidaknya suatu proses pembelajaran yang telah dilakukan.⁷ Sebagai kegiatan atau proses penilaian, evaluasi berfungsi sebagai instrumen untuk menentukan nilai pendidikan dan mengungkapkan mutu serta hasil-hasil yang dicapai.⁸ Sebagai kegiatan krusial, evaluasi memberikan gambaran mendalam tentang prestasi siswa dan efektivitas pembelajaran.⁹ Aktivitas pendidikan menuntut adanya pengalaman belajar dari peserta didik, yang dimaksudkan untuk mencapai tujuan (menguasai kompetensi tertentu).¹⁰ Di sinilah penilaian dibutuhkan untuk melihat sejauh mana kompetensi yang telah dikuasai oleh peserta dalam bentuk hasil belajar yang diperlihatkan setelah mereka menempuh pengalaman belajar.¹¹ Apabila penilaian merupakan salah satu bagian penting dalam rangkaian proses pembelajaran dalam pendidikan,¹² maka setiap pembelajaran perlu dilakukan penilaian.¹³ Evaluasi diperlukan dalam rangka memperoleh informasi dan data yang diperlukan yang selanjutnya dijadikan dasar dalam pengambilan suatu keputusan, khususnya dari pihak sekolah.¹⁴ Oleh karenanya penilaian hasil belajar dapat menentukan baik tidaknya pendidikan. Sehingga, ketepatan penilaian hasil belajar memberikan dampak yang sangat signifikan terhadap upaya peningkatan mutu pendidikan di sekolah.¹⁵

³ A. Kadir, "Menyusun dan Menganalisis Tes Hasil Belajar," *Al-TA'DIB: Jurnal Kajian Ilmu Kependidikan* 8, no. 2 (2015): 70-81. DOI: <https://dx.doi.org/10.31332/atdb.v8i2.411>

⁴ Ahmad Syamsul Ma'arif, "Sighab Iktibar Al-'Arabiyyah Fi Dui Al-Iktibar Al-Mutqn (TOAFL)," *Jurnal Al Bayan: Jurnal Jurusan Pendidikan Bahasa Arab* 9, no. 2 (2017): 160-173. DOI: <http://dx.doi.org/10.24042/albayan.v9i2.2233>

⁵ Sri Suharti, "Kualitas Tes Bahasa Arab dan Prestasi Peserta Didik Madrasah Tsanawiyah Kabupaten Bantul (Analisis Butir Soal UAMBN Tahun Ajaran 2013/2014)," *Jurnal Pendidikan Madrasah* 2, no. 1 (2017): 185-196. DOI: <https://ejournal.uin-suka.ac.id/tarbiyah/JPM/article/view/1439>

⁶ Muhib Abdul Wahab, "Pembelajaran Bahasa Arab di Era Posmetode," *Arabiyat: Jurnal Pendidikan Bahasa Arab dan Kebahasaan* 2, no. 1 (2015): 59-74. DOI: <https://dx.doi.org/10.15408/a.v2i1.1519>

⁷ Sherly Yustuti, Masrun Masrun, and Hikmah Hikmah, "Development of Listening Skills Evaluation Instruments | Pengembangan Instrumen Evaluasi Keterampilan Menyimak," *Mantiq Tayr: Journal of Arabic Language* 3, no. 1 (2023): 1-10. DOI: <https://doi.org/10.25217/mantiqtayr.v3i1.2614> "Analisis Validitas dan Reliabilitas Tes Mata Pelajaran Pendidikan Agama Islam Tingkat SMA," *Edukasia: Jurnal Penelitian Pendidikan Islam*, 9, no. 2 (2014): 253-266. DOI: <https://dx.doi.org/10.21043/edukasia.v9i2.775>

⁸ Moh. Raswan Matsna, *Evaluasi Pembelajaran Bahasa Arab*, (Ciputat: UIN Jakarta Press, 2015), hlm. 20-25.

⁹ Ubaid Ridho, "Evaluasi Dalam Pembelajaran Bahasa Arab," *An Nabighoh: Jurnal Pendidikan dan Pembelajaran Bahasa Arab* 20, no. 01 (2018): 19-26. DOI: <https://doi.org/10.32332/an-nabighoh.v20i01.1124>

¹⁰ Yayat Hidayat, "Teori Perolehan dan Perkembangan Bahasa untuk Jurusan Pendidikan Bahasa Arab," *Maharaat: Jurnal Pendidikan Bahasa Arab* 1, no. 01 (2018): 24-40. DOI: <https://doi.org/10.18196/mht.113>

¹¹ M. Zulkifli, "Analisis Bentuk Evaluasi Kurikulum 2013 Mata Pelajaran Bahasa Arab di MI," *Al-Madrasah: Jurnal Pendidikan Madrasah Ibtidaiyah* 2, no. 2 (2018): 125-143. DOI: <https://dx.doi.org/10.35931/am.v0i0.29>

¹² Zainal Arifin, *Penilaian Pembelajaran Bahasa Berbasis Kompetensi, Remaja Rosdakarya*, (Bandung: Remaja Rosdakarya Offset, 2014), hlm. 35-39.

¹³ Muhib Abdul Wahab, "Standarisasi Kurikulum Pendidikan Bahasa Arab di Perguruan Tinggi Keagamaan Islam Negeri," *Arabiyat: Jurnal Pendidikan Bahasa Arab dan Kebahasaan* 3, no. 1 (2016): 32-51. DOI: <https://doi.org/10.15408/a.v3i1.3187>

¹⁴ Annisa Elita Rahmawati and Umi Hijriyah, "Development of the Quizizz Application-Based Evaluation Tool for Learning Arabic for MTs | Pengembangan Alat Evaluasi Berbasis Aplikasi Quizizz pada Pembelajaran Bahasa Arab untuk MTs," *Mantiq Tayr: Journal of Arabic Language* 3, no. 2 (2023): 135-150. DOI: <https://doi.org/10.25217/mantiqtayr.v3i2.3484>

¹⁵ Zaenal Khafidin,

Keterkaitan antara tes, pengukuran dan penilaian adalah hasil belajar baru dapat dilakukan dengan baik dan benar jika menggunakan informasi yang diperoleh melalui pengukuran hasil belajar dengan menggunakan tes sebagai alat ukurnya.¹⁶ Tes yang baik harus memenuhi validitas rasional yaitu validitas isi dan validitas konstruk dan validitas empirik.¹⁷ Selain validitas, tes yang baik juga harus memiliki reliabilitas, tingkat kesukaran dan juga daya pembeda.¹⁸ Validitas juga dapat diartikan sebagai ketepatan interpretasi yang dihasilkan dari skor tes atau instrument evaluasi.¹⁹ Validitas isi menuntut adanya kesesuaian isi antara kemampuan yang ingin diukur dengan tes yang digunakan untuk mengukurnya.²⁰ Cara untuk mengkaji validitas isi adalah dengan pendekatan rasional, yaitu membandingkan antara soal dengan kisi-kisi soal atau kurikulum yang telah diajarkan.²¹ Langkah-langkah untuk mengetahui validitas isi adalah dengan membandingkan proporsi kurikulum dengan soal, membandingkan kesesuaian butir soal dengan kurikulum, dan membandingkan cakupan materi tes dengan cakupan kurikulum. Kegunaan tes pengukuran dan penilaian dalam pendidikan antara lain adalah untuk seleksi, penempatan, diagnosa, remedial, umpan balik, motivasi dan membimbing, perbaikan kurikulum, program pendidikan serta pengembangan ilmu.²²

Sebuah instrumen evaluasi dianggap memiliki reliabilitas tinggi ketika tes yang dirancang mampu memberikan hasil yang konsisten dalam mengukur variabel yang diinginkan. Hal ini menciptakan dasar keyakinan dalam interpretasi hasil evaluasi, menegaskan bahwa instrumen tersebut serta dapat diandalkan untuk memberikan gambaran yang akurat tentang prestasi atau kemampuan yang diuji. Reliabilitas menjadi landasan penting dalam pengembangan dan penggunaan tes, memastikan bahwa instrumen tersebut memberikan hasil yang dapat dipercaya dan tidak terpengaruh oleh faktor-faktor yang tidak diinginkan.²³ Tingkat kesukaran adalah peluang untuk menjawab benar suatu soal pada tingkat kemampuan tertentu yang biasanya dinyatakan dalam bentuk indeks. Indeks tingkat kesukaran ini pada umumnya dinyatakan dalam bentuk proporsi yang besarnya berkisar 0,00-1,00. Daya pembeda soal adalah kemampuan suatu butir soal dapat membedakan antara warga belajar/siswa yang telah menguasai materi yang ditanyakan dan warga

¹⁶ Amir Rudin, "Analisis Butir Soal Ujian Kenaikan Kelas Mata Pelajaran Bahasa Arab Kelas VIII di MTs Modern Al-Azhary Ajibarang Tahun Pelajaran 2015/2016," Skripsi thesis, Institut Agama Islam Negeri Purwokerto. <https://repository.uinsaizu.ac.id/2117/>

¹⁷ Obi Saepul Bahri, "Analisis Isi Soal UAMBN Bahasa Arab Madrasah Aliyah Tabun Pelajaran 2016/2017 dan 2017/2018," *Educare: Jurnal Pendidikan dan Pembelajaran* 17, no. 1 (2019): 50–54. <https://jurnal.fkip.unla.ac.id/index.php/educare/article/view/241>

¹⁸ Tanwir Tanwir, "Dasar-Dasar dan Ruang Lingkup Evaluasi Pendidikan," *Al-Ishlah: Jurnal Pendidikan Islam* 13, no. 1 (2015): 47–59. DOI: <https://doi.org/10.35905/alishlah.v13i1.490>

¹⁹ Leni Fitrianti, "Prinsip Kontinuitas dalam Evaluasi Proses Pembelajaran," *Al-Ishlah: Jurnal Pendidikan* 10, no. 1 (2018): 89–102. DOI: <https://doi.org/10.35445/alishlah.v10i1.68>

²⁰ Miftahun Ni'mah Suseno, "Pengembangan Pengujian Validitas Isi dan Validitas Konstruk: Interpretasi Hasil Pengujian Validitas," in *Seminar Nasional Psikometri* (Surakarta: Universitas Muhammadiyah Surakarta, 2014), hlm. 70–83. <https://publikasiilmiah.ums.ac.id/xmliui/handle/11617/6399>

²¹ Budi Utomo, "Analisis Validitas Isi Butir Soal Sebagai Salah Satu Upaya Peningkatan Kualitas Pembelajaran di Madrasah Berbasis Nilai-Nilai Islam," *Jurnal Pendidikan Matematika (Kudus)* 1, no. 2 (2022): 155-175. DOI: <https://dx.doi.org/10.21043/jmtk.v1i2.4868>

²² Purwadhi Purwadhi, *Pengembangan Kurikulum dalam Pembelajaran Bahasa Arab*, MIMBAR PENDIDIKAN: Jurnal Indonesia untuk Kajian Pendidikan, vol. 4 no. 2 (2019): 103-112. DOI: <https://doi.org/10.17509/mimbardik.v4i2.22201>

²³ Adhi Setiawan, "Faktor-Faktor yang Mempengaruhi Reliabilitas Tes," *AN NUR: Jurnal Studi Islam* 6, no. 2 (2014): 341–54. <https://jurnalannur.ac.id/index.php/An-Nur/article/view/53>

belajar/siswayang tidak/kurang/belum menguasai materi yang ditanyakan. Indeks daya pembeda berkisar antara -1,00 sampai +1,00.²⁴

Dalam konteks evaluasi pembelajaran, butir soal menjadi suatu hal yang penting sebagai instrumen penilaian, terutama dalam penilaian akhir semester.²⁵ Analisis butir, sebagai uji lapangan dari suatu tes atau soal, menonjol sebagai hal yang mendasar. Kegiatan ini melibatkan proses pengumpulan, peringkasan, dan pemanfaatan informasi dari jawaban peserta didik untuk mengambil keputusan terkait setiap penilaian.²⁶ Keberadaannya sebagai alat konfirmasi kualitas soal. Setelah diketahui bahwa soal tersebut lemah dalam hal daya beda, terlalu sukar atau terlalu mudah serta pengecohnya buruk maka itu harus jadi umpan balik bagi pembuat tes untuk melihat ulang kesalahan apa yang terjadi pada soal tersebut.²⁷ Analisis butir dapat dilakukan secara kualitatif, dalam kaitannya dengan ciri-ciri statistiknya.²⁸ Butir soal objektif dapat dianalisis secara lebih akurat dan bertanggung jawab sehingga dapat diketahui kelemahannya secara tepat. Butir soal tes objektif dapat digunakan berulang-ulang, asalkan tidak dalam perangkat tes yang sama.²⁹ Oleh karena itu, ada manfaat atau kegunaan analisis butir soal, kemudian direvisi sehingga butir soal yang kurang baik konstruksinya dapat diperbaiki. Akhirnya akan diperoleh butir soal yang telah diuji dan secara akurat mengukur hasil belajar yang ingin diukur. Untuk mempermudah menganalisis suatu tes tersebut maka diperlukan alat bantu bagi para guru dan calon guru untuk mempermudah menganalisis hasil butir soal. Salah satu alat bantu tersebut adalah software ANATES yang bisa digunakan untuk menganalisa kualitas butir soal tersebut.

Analisis butir secara kuantitatif digunakan untuk menentukan karakteristik butir soal meliputi: (1) tingkat kesukaran, (2) daya beda, (3) berfungsi atau tidaknya pilihan jawaban atau pengecoh (distractors), termasuk perhitungan tentang (4) reliabilitas skor tes.³⁰ Tujuan analisis butir soal juga untuk membantu meningkatkan tes melalui revisi atau membuang soal yang tidak efektif, serta untuk mengetahui informasi diagnostik pada peserta didik apakah mereka sudah/belum memahami materi yang telah diajarkan. Pada dasarnya, penelitian tentang analisis butir soal sudah pernah dilakukan oleh peneliti sebelumnya yaitu: analisis butir soal bahasa Arab Ujian Akhir Madrasah Bersama Daerah (UAMBD) Madrasah Ibtidaiyah Tahun 2017-2018.³¹ Analisis validitas dan reliabilitas butir

²⁴ Sri Lestari Handayani and Khairil Iba, "Karakteristik Tes Keterampilan Proses Sains: Validitas, Reliabilitas, Tingkat Kesukaran dan Daya Pembeda Soal," Publikasi Pendidikan : Jurnal Pemikiran, Penelitian dan Pengabdian Masyarakat Bidang Pendidikan 10, no. 2 (2020): 100–106. DOI: <https://doi.org/10.26858/publikan.v10i2.13051>

²⁵ Umi Machmudah, Khuzaimah Khuzaimah, and Mufidatus Sholihah, "The Characteristics of Arabic Language Textbooks of Madrasah Ibtidaiyah," Jurnal Al Bayan: Jurnal Jurusan Pendidikan Bahasa Arab 11, no. 1 (2019):179–93. DOI: <https://dx.doi.org/10.24042/albayan.v11i1.4091>

²⁶ Susan M Brookhart and Anthony J Nitko, *Educational Assessment of Students*, 8th ed., (United States: Pearson Education Inc., 2019), hlm. 411.

²⁷ Zahratul Mahfudhah, *Tablil Iktibarat Al-Lughab Al-Arabiyyah (Dirasah Taqwimiyyah Bil Madrasah Muthawasithah Sarusy Syari'ah Bi Banda Aceh*, Skripsi thesis, UIN Ar-Raniry Banda Aceh, 2017. <https://repository.ar-raniry.ac.id/id/eprint/5658/>

²⁸ Fatimah Depi Susanty, "Analisis Validasi Soal Tes Hasil Belajar pada Pelaksanaan Pembelajaran Bahasa Arab di Pusat Pengembangan Bahasa (P3B) UIN Suska Riau," Kutubkhanah: Jurnal Penelitian Sosial Keagamaan 19, no. 2 (2016): 112-132. DOI: <https://dx.doi.org/10.24014/kutubkhanah.v19i2.2546>

²⁹ Ahmad Bukhari Muslim, "Al-Iktibarat Al-Lughawiyah: Tasmimuba Wa Anwa'uba," Jurnal Al Bayan: Jurnal Jurusan Pendidikan Bahasa Arab 8, no. 1 (2016): 23–39. DOI: <https://dx.doi.org/10.24042/albayan.v8i1.359>

³⁰ Abdul Mutholib, "Analisis Butir Soal Bahasa Arab Ujian Akhir Madrasah Berstandar Nasional (UAMBN) Bagi Siswa Madrasah Ibtidaiyah (MI) Tahun Pelajaran 2012-2013," Arabia: Jurnal Pendidikan Bahasa Arab 5, no. 2 (2016): 141-161. DOI: <https://dx.doi.org/10.21043/arabia.v5i2.1387>

³¹ Deni Maulana and Anwar Sanusi, "Analisis Butir Soal Bahasa Arab Ujian Akhir Madrasah Bersama Daerah (UAMBD) Madrasah Ibtidaiyah Tahun 2017-2018," Ta'lim Al-'Arabiyyah: Jurnal Pendidikan Bahasa Arab & Kebahasaan 4, no. 1 (2020): 12–24. DOI: <https://doi.org/10.15575/jpba.v4i1.8054>

soal ulangan akhir semester bidang studi bahasa Indonesia kelas X.D SMAN 1 terhadap pencapaian kompetensi.³² Analisis validitas, reliabilitas, tingkat kesulitan dan daya beda butir soal ujian akhir semester tema 7 kelas III SDN Karet 1 Sepatan.³³ Analisis kesukaran soal, daya pembeda dan fungsi distraktor³⁴ validitas dan reliabilitas dalam analisis butir soal.³⁵ Dari hasil review literature ditemukan gab penelitian ini dengan penelitian yang sebelumnya bahwa penelitian ini menjadi pelengkap dan menyempurnakan penelitian-penelitian sebelumnya dengan menambahkan daya pengecoh dalam analisis butir soalnya. Penelitian ini bertujuan untuk menganalisis butir soal berupa validitas, reliabilitas, daya beda soal, tingkat kesukaran, dan daya pengecoh pada soal penilaian akhir semester bahasa Arab di MTs Al-Ma'arif Rakit Banjarnegara Jawa Tengah.

Berdasarkan hasil observasi penulis di MTS Al-Ma'arif Rakit sering dilakukan tahap evaluasi pembelajaran melalui tes tulisan, akan tetapi kegiatan menganalisis soal tes ini jarang dilakukan sehingga kualitas dari soal tidak diketahui secara pasti. Hasil wawancara dengan salah satu guru pemangku mata pelajaran bahasa Arab juga menjelaskan, kegiatan menganalisis soal juga jarang dilaksanakan. Kegiatan evaluasi hasil pembelajaran hanya sampai pada pemberian skor pada peserta didik. Faktor kualitas tes yang belum diketahui, akan berpengaruh terhadap kemampuan testee dalam mengerjakan tes. Kelemahan ini akan berdampak pada sulitnya menentukan kemampuan testee yang sebenarnya.³⁶

Berdasarkan berbagai masalah di atas permasalahan utama dalam penelitian ini, peneliti akan meneliti tentang analisis butir soal penilaian akhir semester pelajaran bahasa Arab siswa kelas VII MTs Al-Ma'arif Rakit tahun ajaran 2022/2023 (validitas, reliabilitas, daya beda, tingkat kesukaran, dan daya pengecoh). Fokus Penelitian ini adalah kualitas butir-butir dan soal-soal mata pelajaran bahasa Arab yang digunakan pada penilaian akhir semester tahun ajaran 2022-2023. Ditinjau dari segi validitas, reliabilitas, daya beda, tingkat kesukaran, dan daya pengecoh pada soal.

Metode

Penelitian ini menggunakan pendekatan kuantitatif. Penelitian kuantitatif merupakan penelitian yang dilakukan mulai dari mengumpulkan data lalu menganalisisnya hingga menyampaikan data tersebut dalam bentuk gambar, tabel, grafik, atau yang lain dengan menerapkan prinsip-prinsip objektivitas.³⁷ Fokus penelitian dalam menganalisis butir soal melalui IBM SPSS Statistics 25 dan *Microsoft Excel* adalah untuk menjadi alat pengukur kemampuan guru kelas VII dalam membuat soal untuk peserta didik. Metode yang dipakai ialah metode deskriptif. Pada penelitian kuantitatif deskriptif, peneliti selaku perencana penelitian, pelaksana penelitian, pengumpul data, penganalisis

³² Sri Dwipayani, "Analisis Validitas dan Reliabilitas Butir Soal Ulangan Akhir Semester Bidang Studi Bahasa Indonesia Kelas XD SMAN 1 Terhadap Pencapaian Kompetensi," *Jurnal Pendidikan Bahasa dan Sastra Indonesia Undiksha* 1, no. 5 (2013): 1-18. DOI: <https://doi.org/10.23887/jjpbs.v1i5.578>

³³ Ina Magdalena et al., "Analisis Validitas, Reliabilitas, Tingkat Kesulitan dan Daya Beda Butir Soal Ujian Akhir Semester Tema 7 Kelas III SDN Karet 1 Sepatan," *BINTANG: Jurnal Pendidikan dan Sains* 3, no. 2 (2021): 198-214. <https://ejournal.stitpn.ac.id/index.php/bintang/article/view/1291>

³⁴ Laela Umi Fatimah and Khairuddin Alfath, "Analisis Kesukaran Soal, Daya Pembeda dan Fungsi Distraktor," *AL-MANAR: Jurnal Komunikasi dan Pendidikan Islam* 8, no. 2 (2019): 37-64. DOI: <https://doi.org/10.36668/jal.v8i2.115>

³⁵ Farida Farida and Anna Musyarofah, "Validitas dan Reliabilitas dalam Analisis Butir Soal," *AL-MU'ARRIB: Journal of Arabic Education* 1, no. 1 (2021): 34-44. DOI: <https://doi.org/10.32923/al-muarrib.v1i1.2100>

³⁶ Pardimin Pardimin, Sri Adi Widodo, and Indriyati Eko Purwaningsih, "Analisis Butir Soal Tes Pemecahan Masalah Matematika," *Wacana Akademika: Majalah Ilmiah Kependidikan* 1, no. 1 (2017): 69-76. DOI: <https://doi.org/10.30738/wa.v1i1.1084>

³⁷ Sudaryono, *Metodologi Penelitian Kuantitatif, Kualitatif, dan Mixed Method*, 2nd ed., (Depok: PT Raja Grafindo Persada, 2021), hlm. 177.

data dan pelaporan hasil penelitian.³⁸ Data yang akan dikumpulkan berupa penilaian akhir semester peserta didik kelas VII. Sumber data pada penelitian ini adalah dokumen soal pilihan ganda penilaian akhir semester pada pembelajaran bahasa Arab kelas VII MTs Al-Ma'arif. Penelitian ini dilaksanakan pada semester kedua tahun ajaran 2022/2023. Populasi penelitian berjumlah 25 peserta didik kelas VII di MTs Al-Ma'arif Rakit tahun ajaran 2022/2023. Teknik pengumpulan data yang dipakai yaitu dokumentasi dan observasi. Teknik dokumentasinya berupa perangkat soal penilaian akhir semester kelas VII tahun ajaran 2022/2023 yang berupa file soal penilaian akhir semester, kisi-kisi soal, kunci jawaban. Jawaban seluruh peserta didik kelas VII. Analisis yang dilakukan dalam penelitian ini adalah teknik analisis kuantitatif. Dengan skor benar = 1 skor salah = 0 (dikotomi).³⁹ Peneliti dalam menganalisis data dan untuk mencari tingkat validitas, reabilitas, tingkat kesukaran, daya pembeda, daya pengecoh menggunakan aplikasi *Microsoft Excel* 2018 dan aplikasi software IBM SPSS Statistics 25.

Hasil dan Pembahasan

A. Validitas

Validitas butir soal mengacu pada sejauh mana setiap item atau pertanyaan yang disajikan dalam tes benar-benar mengukur apa yang dimaksudkan untuk diukur.⁴⁰ Itu dapat melibatkan analisis apakah butir soal tersebut relevan, apakah itu mencakup aspek yang dimaksud, dan apakah jawaban yang diberikan mencerminkan kondisi yang diukur. Instrumen dinyatakan valid manakala nilai R Hitung $>$ R Tabel dan Nilai Sig $<$ 0,05.

Pada penelitian ini berdasarkan analisis data menggunakan aplikasi SPSS 25, dengan menggunakan R Tabel sebagai alat uji validitas dan berfungsi untuk menguji validitas suatu instrumen yang digunakan, pada penelitian ini lembar jawaban dan butir soal ujian akhir semester bahasa Arab siswa kelas 7 MTs Al-Ma'arif Rakit menjadi instrumen untuk pengumpulan data. Tinggi rendahnya suatu instrumen menunjukkan sejauh mana data yang didapat sesuai dengan variabel yang dimaksud.⁴¹ Validitas butir soal pada analisis ini menggunakan teknik Korelasi Biserial. Interpretasi indeks korelasi Point Biserial yang menunjukkan nilai validitas pada kriteria tabel 1.

Tabel 1. Tabel Validitas

Kategori Validitas Tes	Nilai Koefisiensi Korelasi
Sangat Tinggi	0,800 – 1,000
Tinggi	0,600 – 0,799
Sedang	0,400 – 0,599
Rendah	0,200 – 0,399
Sangat Rendah	0.000 – 0,199

Berdasarkan tabel 1 tentang table validitas dijelaskan bahwasanya untuk mencapai kategori validitas tes yang sangat tinggi, nilai koefisiensi korelasi berkisar antara 0,800 sampai 1,000. Sedangkan valiitas tes yang tinggi berkisar di nilai antara 0,600 – 0,799. Validitas yang sedang, nilai

³⁸ Sugiyono, *Metode Penelitian Kuantitatif*, (Bandung: Alfabeta, 2022), hlm. 69.

³⁹ Fuad Hasyim Purwono et al., *Metodologi Penelitian (Kuantitatif, Kualitatif dan Mixed Method)*, (Jakarta: Guepedia Group, 2019), hlm. 33-40.

⁴⁰ Kadek Ayu Astiti, *Evaluasi Pembelajaran*, (Yogyakarta: Penerbit Andi, 2017), hlm. 98.

⁴¹ Sri Nurhalimah et al., "Hubungan Antara Validitas Item dengan Daya Pembeda dan Tingkat Kesukaran Soal Pilihan Ganda Pas," *Natural Science Education Research* 4, no. 3 (2022): 249-257. DOI: <https://doi.org/10.21107/nser.v4i3.8682>

koefisiensi korelasi antara 0,400 – 0,599. Adapun validitas tes yang berkategori rendah dan sangat rendah manakala nilai koefisiensi korelasi berkisar antara 0,200 – 0,399 dan 0,000 – 0,199.⁴²

Berikut adalah hasil penghitungan validitas dengan menggunakan aplikasi software IBM SPSS Statistics 25.

Tabel 2. Hasil Penghitungan IBM SPSS Statistics 25

No. Soal	<i>Pearson Correlation</i>	<i>R</i> Tabel	<i>SIG (2-Tailed)</i>	Keterangan	Interpretasi
1	0.101	0.396	0.629	Tidak Valid	-
2	0.101	0.396	0.629	Tidak Valid	-
3	0.335	0.396	0.102	Tidak Valid	-
4	0.38	0.396	0.061	Tidak Valid	-
5	0.036	0.396	0.863	Tidak Valid	-
6	0.497	0.396	0.011	Valid	Cukup
7	0.232	0.396	0.265	Tidak Valid	-
8	0.101	0.396	0.629	Tidak Valid	-
9	-0.114	0.396	0.586	Tidak Valid	-
10	0.492	0.396	0.013	Valid	Cukup
11	-0.094	0.396	0.656	Tidak Valid	-
12	0.101	0.396	0.629	Tidak Valid	-
13	0.101	0.396	0.629	Tidak Valid	-
14	0.232	0.396	0.265	Tidak Valid	-
15	-0.132	0.396	0.530	Tidak Valid	-
16	0.089	0.396	0.671	Tidak Valid	-
17	0.492	0.396	0.013	Valid	Cukup
18	0.492	0.396	0.013	Valid	Cukup
19	0.241	0.396	0.247	Tidak Valid	-
20	0.346	0.396	0.090	Tidak Valid	-
21	-0.159	0.396	0.449	Tidak Valid	-
22	-0.193	0.396	0.356	Tidak Valid	-
23	0.167	0.396	0.426	Tidak Valid	-
24	0.492	0.396	0.013	Valid	Cukup
25	0.622	0.396	0.001	Valid	Tinggi
26	0.297	0.396	0.150	Tidak Valid	-
27	0.167	0.396	0.426	Tidak Valid	-
28	0.301	0.396	0.143	Tidak Valid	-
29	0.241	0.396	0.247	Tidak Valid	-
30	0.396	0.396	0.050	Tidak Valid	-
31	0.576	0.396	0.003	Valid	Cukup
32	0.232	0.396	0.265	Tidak Valid	-
33	0.622	0.396	0.001	Valid	Tinggi
34	0.151	0.396	0.472	Tidak Valid	-

⁴² Suharsimi Arikunto, *Prosedur Penelitian Suatu Pendekatan Praktik*, Ed Cet 14, (Jakarta: Rineka Cipta, 2016). hlm. 214.

No. Soal	<i>Pearson Correlation</i>	<i>R Tabel</i>	<i>SIG (2-Tailed)</i>	Keterangan	Interpretasi
35	0.614	0.396	0.001	Valid	Tinggi
36	0.565	0.396	0.003	Valid	Cukup
37	0.147	0.396	0.484	Tidak Valid	-
38	0.297	0.396	0.150	Tidak Valid	-
39	0.569	0.396	0.003	Valid	Cukup
40	0.492	0.396	0.013	Valid	Cukup

Berdasarkan tabel 2 tentang hasil penghitungan validitas tes mengguakan aplikasi IBM SPSS Statistics 25 ditemukan bahwa terdapat 12 Soal (30%) dapat dikatakan valid yakni ada pada nomor 6, 10, 17, 18, 24, 25, 31, 33, 35, 36, 39, dan 40. Kemudian sisanya yakni 28 Soal (70%) dinyatakan tidak valid yakni pada nomor 1, 2, 3, 4, 5, 7, 8, 9, 11, 12, 13, 14, 15, 16, 19, 20, 21, 22, 23, 26, 27, 28, 29, 30, 32, 34, 37, dan 38. Ini berarti mayoritas soal yang diujikan di MTs Al-Ma'arif Rakit memiliki kategori soal tes yang tidak valid.

Penentuan status kevalidan suatu butir tes didasarkan pada nilai signifikasi yang harus lebih kecil dari 0,05 atau nilai R hitung lebih besar dari R Tabel.⁴³ Berdasarkan hal ini, data butir soal yang tidak valid dalam penelitian ini disebabkan karena nilai signifikasi yang lebih besar dari 0,05 seperti pada nomor 3 dan 4 yang memiliki nilai signifikasi 0,102 dan 0,061. Selain itu nilai r hitung yang dihasilkan lebih kecil dari r tabel sebesar 0,396, seperti ada dinomor 6 dan 7 yang memiliki nilai r hitung sebesar 0,497 dan 0,232. Adapun butir soal yang berkategori valid dalam penelitian ini merupakan soal yang memiliki nilai signifikasi lebih kecil dari 0,05 seperti yang terdapat pada nomor 6 dan 10 yang memiliki nilai sigifikasi 0,011 dan 0,013. Nilai r hitung juga lebih besar dari r tabel (0,396) seperti yang terdapat dalam nomor 17 dan 25 yang bernilai r hitung sebesar 0,492 dan 0,622.

B. Reliabilitas

Reliabilitas butir soal menunjukkan seberapa konsisten dan andal butir soal dalam mengukur suatu konsep atau variabel. Ini mengacu pada sejauh mana butir soal tersebut dapat diandalkan dan memberikan hasil yang konsisten ketika diuji berkali-kali.⁴⁴ Tujuan dari reliabilitas adalah untuk memahami bahwa alat tersebut cukup dapat diandalkan untuk digunakan sebagai alat pengumpul data karena alat tersebut baik.⁴⁵

Koefesiensi reliabilitas secara wajar bergerak pada interval 0,00 – 1,00 dengan kriteria-kriteria. Adapun kriteria-kriteria tersebut dijadikan sebagai acuan untuk menginterpretasikan hasil penghitungan reliabilitas tes dengan aplikasi IBM SPSS Statistics 25.⁴⁶

⁴³ Sugiono, *Metode Penelitian Pendidikan*, (Bandung: Alfabeta, 2012), hlm. 261.

⁴⁴ Imam Ghozali, *Aplikasi Analisis Multivariate Dengan IBM SPSS 25*, (Semarang: Badan Penerbit Universitas Diponegoro, 2018), hlm. 45.

⁴⁵ Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan*, (Jakarta: Bumi Aksara, 2018), hlm. 203.

⁴⁶ Mutholib, "Analisis Butir Soal Bahasa Arab Ujian Akhir Madrasah Berstandar Nasional (UAMBN) Bagi Siswa Madrasah Ibtidaiyah (MI) Tahun Pelajaran 2012-2013," *Arabia: Jurnal Pendidikan Bahasa Arab*, 5(2): 141-161. DOI: <http://dx.doi.org/10.21043/arabia.v5i2.1387>

Tabel 3. Interpretasi Kategori Tes

Kategori Reliabilitas Tes	Nilai Koefisiensi Korelasi
Sangat Rendah	0,00 – 0,20
Rendah	0,20 – 0,40
Sedang	0,40 – 0,60
Tinggi	0,60 – 0,80
Sangat Tinggi	0,80 – 1,00

Berdasarkan tabel 3 tentang interpretasi kategori tes dapat dijelaskan bahwasanya untuk mencapai kategori reliabilitas tes yang sangat tinggi, nilai koefisiensi korelasi berkisar antara 0,80 sampai 1,00. Sedangkan reliabilitas tes yang tinggi berkisar dinilai antara 0,60 – 0,80. Reliabilitas yang sedang, nilai koefisiensi korelasi antara 0,40 – 0,60. Adapun reliabilitas tes yang berkategori rendah dan sangat rendah manakala nilai koefisiensi korelasi berkisar antara 0,20 – 0,40 dan 0,00 – 0,20.⁴⁷ Uji reliabilitas pada penelitian ini menggunakan aplikasi IBM SPSS Statistics 25. Berikut hasil reliabilitas dengan menggunakan penghitungan melalui aplikasi IBM SPSS Statistics 25.

Tabel 4. Case Processing Summary

		N	%
Cases	Valid	25	100.0
	Excluded ^a	0	.0
	Total	25	100.0

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

Reliability Statistics

Cronbach's	
Alpha	N of Items
.625	40

Berdasarkan tabel 4 di atas, ditemukan bahwasanya nilai Cronbach's Alpha sebesar 0,625 dengan jumlah N of Item sebesar 40 yang menunjukkan jumlah soal yang diujikan dan dijawab oleh peserta didik.

Tabel 5. Hasil Penghitungan Reliabilitas Soal

R_{xy}	R Tabel	Keterangan
0.625	0,396	Reliabel

Berdasarkan tabel 5 tentang hasil penghitungan reliabilitas soal ditemukan bahwasanya dalam mencari realibilitas soal Penilaian Akhir Semester (PAS) bahasa Arab kelas 7 MTs Al-Ma'arif Rakit peneliti menggunakan metode Cronbach's Alpha dalam melakukan analisis data yang ada, kemudian didapatkan hasil bahwa nilai Cronbach's Alpha bernilai 0,625 dengan jumlah butir soal (N of item) berjumlah 40.

⁴⁷ Suharsimi Arikunto, *Prosedur Penelitian Suatu Pendekatan Praktik, Ed Cet 14*, (Jakarta: Rineka Cipta, 2016), hlm. 224.

Pengambilan keputusan uji reliabilitas didasarkan nilai Cronbach's Alpha. Jika nilai Cronbach's Alpha $>$ r tabel maka soal dinyatakan reliabel, jika nilai Cronbach's Alpha $<$ r tabel maka tidak reliabel.⁴⁸ Berdasarkan output "reliability statistic" di atas di ketahui nilai Cronbach's Alpha sebesar 0,625. Nilai tersebut lebih besar di dibandingkan dengan nilai r tabel yakni sebesar 0,396 dengan nilai N= 40 dan tingkat signifikasi 0,05. Ini bisa diinterpretasikan bahwasanya soal Penilaian Akhir Semester (PAS) bahasa Arab kelas 7 di MTs Al-Ma'arif Rakit bisa dikatakan reliabel dengan kategori tinggi. Hasil ini menunjukkan bahwa butir soal penilaian Akhir Semester (PAS) bahasa Arab tersebut dapat dipercaya karena cenderung memberikan hasil yang tetap.

C. Daya Beda Soal

Daya beda soal adalah kemampuan suatu soal untuk membedakan antara siswa yang pandai (berkemampuan tinggi) dengan siswa yang kurang pintar (berkemampuan rendah). Angka yang menunjukkan besarnya daya beda disebut indeks diskriminasi, yang mana berkisar antara 0,00 sampai 1,00. Dengan klasifikasi sebagai berikut:

D= 0,00 – 0,20 dikategorikan Jelek dalam membedakan

D= 0,20 – 0,40 dikategorikan Cukup dalam membedakan

D= 0,40 – 0,70 dikategorikan Baik dalam membedakan

D= 0,70 – 1,00 dikategorikan Baik sekali dalam membedakan

Apabila nilai D lebih rendah dari 0 atau negative maka dikategorikan sangat jelek. Untuk mendapatkan nilai daya beda menggunakan rumus $D = (BA/JA) - (BB/JB)$.

Keterangan

BA = Jumlah siswa kelompok atas yang menjawab benar

BB = Jumlah siswa kelompok bawah yang menjawab benar

JA = Jumlah siswa kelompok atas

JB = Jumlah siswa kelompok bawah

Setelah dilakukan penghitungan dan analisis 40 soal ujian PAS di MTs Al-Ma'arif Rakit didapatkan hasil sebagai berikut:

Tabel 6. Nilai Indeks Daya Beda Soal

No Soal	Indeks	Kriteria	No. Soal	Indeks	Kriteria
1	0,08	Jelek	21	-0,08	Jelek Sekali
2	0,08	Jelek	22	-0,19	Jelek Sekali
3	0,15	Jelek	23	0,08	Jelek
4	0,23	Cukup	24	0,08	Jelek
5	0,08	Jelek	25	0,54	Baik
6	0,23	Cukup	26	0,08	Jelek
7	0,08	Jelek	27	0,08	Jelek
8	0,08	Jelek	28	0,23	Cukup
9	-0,03	Jelek Sekali	29	0,15	Jelek
10	0,08	Jelek	30	0,13	Jelek
11	-0,08	Jelek Sekali	31	0,23	Cukup
12	0,08	Jelek	32	0,08	Jelek

⁴⁸ Joko Widiyanto, *SPSS for Windows untuk Analisis Data Statistik dan Penelitian*, (Surakarta: BP-FKIP UMS, 2018), hlm. 108.

No Soal	Indeks	Kriteria	No. Soal	Indeks	Kriteria
13	0,08	Jelek	33	0,54	Baik
14	0,08	Jelek	34	0,03	Jelek
15	-0,11	Jelek Sekali	35	0,62	Baik
16	-0,10	Jelek Sekali	36	0,31	Cukup
17	0,08	Jelek	37	0,15	Jelek
18	0,08	Jelek	38	0,08	Jelek
19	0,15	Jelek	39	0,15	Jelek
20	0,21	Cukup	40	0,08	Jelek

Tabel 6 merupakan hasil dari analisis daya pembeda soal ujian penilaian akhir semester yang selanjutnya dibuatkan tabel 7 yakni status soal dalam membedakan.

Tabel 7. Status Soal dalam Membedakan

Kategori Soal dalam Membedakan	Nomor Butir Soal
Sangat Jelek	9, 11, 15, 16, 21, 22
Jelek	1, 2, 3, 5, 7, 8, 10, 12, 13, 14, 17, 18, 19, 20, 23, 24, 26, 27, 29, 30, 32, 34, 37, 38, 39, 40
Cukup	4, 6, 20, 28, 31, 36
Baik	25, 33, 35
Baik Sekali	0

Hasil temuan pada tabel 7 menjelaskan bahwa terdapat 6 Soal dengan Kategori “Jelek Sekali dalam membedakan, 25 Soal dengan kategori “Jelek dalam membedakan”, 6 Soal dengan kategori “Cukup dalam Membedakan”, 3 Soal dengan kategori “Baik dalam membedakan”, sedangkan dengan soal kategori Baik Sekali dalam membedakan tidak ditemukan.

Sebagai tindak lanjut atas hasil analisis daya pembeda adalah:

1. Butir-butir item yang sudah memiliki daya pembeda baik (*satis factory, good, dan excellent*) hendaknya dimasukkan dalam bank soal dan dapat dikeluarkan lagi pada tes berikutnya karena kualitasnya sudah cukup memadai.
2. Butir-butir item yang daya pembedanya masih rendah (*poor*), ada 2 kemungkinan:
 - a. Ditelusuri untuk kemudian diperbaiki, dan setelah diperbaiki dapat diajukan lagi dalam tes hasil belajar, yang kemudian dianalisis lagi apakah meningkat atau tidak.
 - b. Dibuang atau didrop dan tidak dikeluarkan lagi untuk tes hasil belajar.
3. Khusus butir-butir item yang angka indeks diskriminasi itemnya bertanda negatif, sebaiknya pada tes hasil belajar tidak usah dikeluarkan lagi, sebab butir yang demikian kualitasnya sangat jelek (*testee* yang pandai lebih banyak menjawab salah ketimbang *testee* yang tidak pandai, justru hanya sedikit yang menjawab salah.⁴⁹

D. Tingkat kesukaran soal

Pengukuran tingkat kesulitan suatu soal adalah penilaian sejauh mana tingkat kesulitan soal tersebut. Jika suatu soal memiliki tingkat kesulitan yang seimbang secara proporsional, maka dapat

⁴⁹ Anas Sudijono, *Pengantar Evaluasi Pendidikan*, (Jakarta: Raja Grafindo Persada, 2011), hlm. 385-386.

disimpulkan bahwa soal tersebut dinilai baik.⁵⁰ Guna mendapatkan nilai indeks kesukaran dapat digunakan rumus:

$$P = B/JS$$

Keterangan:

P = Nilai indeks kesukaran

B = Jumlah siswa yang menjawab soal dengan benar

JS = Jumlah seluruh peserta tes

Guna menentukan kategori tingkat Kesukaran butir soal dapat dilihat dari prosentase (indeks kesukaran butir/IKB) dapat bernilai 0,00-1,00; dengan Kriteria IKB sebagai berikut: 0,00-0,20 adalah sangat sukar 0,20-0,40 sukar 0,40-0,60 sedang 0,60-0,90 mudah 0,90-1,00 sangat mudah.

Setelah dilakukan penghitungan dan analisis soal ujian PAS MTs Al-Ma'arif Rakit didapatkan hasil sebagai berikut:

Tabel 8. Indeks Kesukaran Soal

No Soal	Jawaban Benar	Tingkat Kesukaran	Kategori Soal
1	24	0,96	Sangat Mudah
2	24	0,96	Sangat Mudah
3	23	0,92	Sangat Mudah
4	22	0,88	Mudah
5	24	0,96	Sangat Mudah
6	22	0,88	Mudah
7	24	0,96	Sangat Mudah
8	24	0,96	Sangat Mudah
9	15	0,6	Mudah
10	24	0,96	Sangat Mudah
11	24	0,96	Sangat Mudah
12	24	0,96	Sangat Mudah
13	24	0,96	Sangat Mudah
14	24	0,96	Sangat Mudah
15	16	0,64	Mudah
16	20	0,8	Mudah
17	24	0,96	Sangat Mudah
18	24	0,96	Sangat Mudah
19	23	0,92	Sangat Mudah
20	16	0,64	Mudah
21	24	0,96	Sangat Mudah
22	15	0,6	Mudah
23	24	0,96	Sangat Mudah
24	24	0,96	Sangat Mudah
25	18	0,72	Mudah
26	24	0,96	Sangat Mudah
27	24	0,96	Sangat Mudah

⁵⁰ Arikunto Suharsimi, *Dasar-Dasar Evaluasi Pendidikan*, (Jakarta: Bumi Aksara, 2018), hlm. 223.

No Soal	Jawaban Benar	Tingkat Kesukaran	Kategori Soal
28	22	0,88	Mudah
29	23	0,92	Sangat Mudah
30	17	0,68	Mudah
31	22	0,88	Mudah
32	24	0,96	Sangat Mudah
33	18	0,72	Mudah
34	8	0,32	Sukar
35	17	0,68	Mudah
36	21	0,84	Mudah
37	23	0,92	Sangat Mudah
38	24	0,96	Sangat Mudah
39	23	0,92	Sangat Mudah
40	24	0,96	Sangat Mudah

Tabel 8 merupakan nilai indeks kesukaran yang didapat setelah melakukan analisis pada soal ujian penilaian akhir semester yang selanjutnya dibuatkan tabel 9 sebagai keterangan status kesukaraan soal.

Tabel 9. Status Kesukaran Soal

Kategori Kesukaran Soal	No Butir Soal
Sangat Mudah	1, 2, 3, 5, 7, 8, 10, 11, 12, 13, 14, 17, 18, 19, 21, 23, 24, 26, 27, 29, 37, 38, 39, 40
Mudah	6, 9, 15, 16, 20, 25, 28, 30, 31, 32, 33, 35, 36
Sedang	-
Sukar	34
Sangat Sukar	-

Berdasarkan data tabel 9 ditemukan bahwasannya soal ujian PAS MTs Al-Ma'arif Rakit terdapat 25 soal dengan kategori soal sangat mudah, 14 Soal dengan kategori mudah, 1 soal kategori sukar serta tidak ditemukan soal dengan kategori sedang dan sangat sukar. Sedangkan soal dengan tingkat kesukaran yang sangat mudah disarankan untuk diganti dengan soal yang lain karena tingkat kesukaran soal menggunakan untuk membedakan siswa yang memiliki kemampuan tinggi, sedang dan rendah.⁵¹

E. Daya Pengecoh

Daya pengecoh, dalam konteks soal atau ujian, merujuk pada opsi jawaban alternatif yang sengaja disusun untuk mengecoh atau membingungkan peserta ujian. Tujuan dari daya pengecoh ini adalah untuk menguji pemahaman dan pengetahuan peserta terhadap materi yang diuji. Opsi jawaban yang menjadi daya pengecoh ini dirancang sedemikian rupa agar terlihat menarik atau masuk akal, namun sebenarnya tidak benar atau tidak relevan dengan pertanyaan yang diajukan.

Dalam penyusunan soal, daya pengecoh ini penting karena mendorong peserta ujian untuk benar-benar memahami materi yang diuji. Daya pengecoh yang efektif akan memaksa peserta untuk menganalisis dengan cermat dan mempertimbangkan pilihan jawaban yang tepat. Secara umum, daya

⁵¹ Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan*, (Jakarta: Bumi Aksara, 2018), hlm. 223.

pengecoh dirancang untuk menguji kemampuan kognitif peserta, mendorong pemikiran kritis, serta mengukur tingkat pemahaman sebenarnya terhadap materi yang diuji.

Rumus yang digunakan dalam pengecoh adalah sebagai berikut:

$$P = \frac{X_i}{X_n} \times 100\%$$

Keterangan:

P = Pengecoh

X_i = Jumlah peserta yang memilih jawaban

X_n = Jumlah seluruh peserta didik yang mengikuti tes

Berikut ini disajikan tabel klasifikasi pengecoh soal menjadi 3 kategori.

Tabel 10. Klasifikasi Efektivitas Pengecoh

Kategori Pengecoh Tes	Nilai Koefisiensi Korelasi
Diterima, Karena Sudah Baik	$\geq 5\%$
Diulang, karena kurang baik	$< 5\%$
Ditolak, karena tidak baik	$= 0,5\%$

Tabel 10 diatas merupakan standard nilai prosentase untuk mengklasifikasikan efektivitas pengecoh dari soal ujian yang telah dilaksanakan. Setelah dilakukan analisis pengecoh pada soal pilihan ganda ujian akhir semester genap kelas 7 pelajaran bahasa Arab di MT's Al Ma'arif Rakit maka didapatkan hasil sebagai berikut :

Tabel 11. Hasil Analisis Pengecoh Soal

No	Jumlah Jawaban				Prosentase				Status Jawaban			
	A	B	c	d	a	b	C	D	A	B	c	d
1	0	1	0	24	0%	4%	0%	96%	ditolak	ditolak	ditolak	diterima
2	1	24	0	0	4%	96%	0%	0%	ditolak	diterima	ditolak	ditolak
3	1	0	23	1	4%	0%	92%	4%	ditolak	ditolak	diterima	ditolak
4	22	1	1	1	88%	4%	4%	4%	diterima	ditolak	ditolak	ditolak
5	24	0	1	0	96%	0%	4%	0%	diterima	ditolak	ditolak	ditolak
6	22	1	1	1	88%	4%	4%	4%	diterima	ditolak	ditolak	ditolak
7	1	24	0	0	4%	96%	0%	0%	ditolak	diterima	ditolak	ditolak
8	0	1	24	0	0%	4%	96%	0%	ditolak	ditolak	diterima	ditolak
9	4	17	1	3	16%	68%	4%	12%	diterima	diterima	ditolak	diterima
10	0	0	1	24	0%	0%	4%	96%	ditolak	ditolak	ditolak	diterima
11	24	0	1	0	96%	0%	4%	0%	diterima	ditolak	ditolak	ditolak
12	24	0	0	1	96%	0%	0%	4%	diterima	ditolak	ditolak	ditolak
13	0	24	1	0	0%	96%	4%	0%	ditolak	diterima	ditolak	ditolak
14	1	0	0	24	4%	0%	0%	96%	ditolak	ditolak	ditolak	diterima
15	16	5	1	3	64%	20%	4%	12%	diterima	diterima	ditolak	diterima
16	4	2	0	19	16%	8%	0%	76%	diterima	diterima	ditolak	diterima
17	1	24	0	0	4%	96%	0%	0%	ditolak	diterima	ditolak	ditolak
18	0	0	24	1	0%	0%	96%	4%	ditolak	ditolak	diterima	ditolak

No	Jumlah Jawaban				Prosentase				Status Jawaban			
	A	B	c	d	a	b	C	D	A	B	c	d
19	1	1	0	23	4%	4%	0%	92%	ditolak	ditolak	ditolak	diterima
20	4	18	0	3	16%	72%	0%	12%	diterima	diterima	ditolak	diterima
21	24	1	0	0	96%	4%	0%	0%	diterima	ditolak	ditolak	ditolak
22	4	3	15	3	16%	12%	60%	12%	diterima	diterima	diterima	diterima
23	1	24	0	0	4%	96%	0%	0%	ditolak	diterima	ditolak	ditolak
24	0	0	24	0	0%	0%	96%	0%	ditolak	ditolak	diterima	ditolak
25	2	3	18	2	8%	12%	72%	8%	diterima	diterima	diterima	diterima
26	1	24	0	0	4%	96%	0%	0%	ditolak	diterima	ditolak	ditolak
27	24	1	0	0	96%	4%	0%	0%	diterima	ditolak	ditolak	ditolak
28	1	1	22	1	4%	4%	88%	4%	ditolak	ditolak	diterima	ditolak
29	0	1	23	1	0%	4%	92%	4%	ditolak	ditolak	diterima	ditolak
30	2	3	3	17	8%	12%	12%	68%	diterima	diterima	diterima	diterima
31	22	2	0	1	88%	8%	0%	4%	diterima	diterima	ditolak	ditolak
32	0	1	24	0	0%	4%	96%	0%	ditolak	ditolak	diterima	ditolak
33	1	2	4	18	4%	8%	16%	72%	ditolak	diterima	diterima	diterima
34	6	7	4	8	24%	28%	16%	32%	diterima	diterima	diterima	diterima
35	4	17	2	2	16%	68%	8%	8%	diterima	diterima	diterima	diterima
36	2	21	1	1	8%	84%	4%	4%	diterima	diterima	ditolak	ditolak
37	1	23	1	0	4%	92%	4%	0%	ditolak	diterima	ditolak	ditolak
38	0	0	1	24	0%	0%	4%	96%	ditolak	ditolak	ditolak	diterima
39	23	1	0	1	92%	4%	0%	4%	diterima	ditolak	ditolak	ditolak
40	0	0	1	24	0%	0%	4%	96%	ditolak	ditolak	ditolak	diterima

Dari temuan tabel 11 di atas maka dapat dijelaskan bahwa pilihan jawaban A dengan kategori diterima berjumlah 18 butir soal atau dengan prosentase 45%. soal yakni nomor 4, 5, 6, 9, 11, 12, 15, 16, 20, 21, 22, 25, 30, 31, 34, 35, 36, dan 39. Pilihan jawaban B dengan kategori diterima berjumlah 19 butir soal atau 47,5% yakni pada soal nomor 2, 7, 9, 13, 15, 16, 17, 20, 22, 23, 25, 26, 30, 31, 33, 34, 35, 36, dan 37. Pilihan jawaban C dengan kategori diterima hanya berjumlah 13 butir soal atau 32,5% dari keseluruhan yakni nomor 3, 8, 18, 22, 24, 25, 28, 29, 30, 32, 33, 34, dan 35. Pilihan jawaban D dengan kategori diterima berjumlah 16 butir soal atau 40% yakni nomor 1, 9, 10, 14, 15, 16, 19, 20, 25, 30, 33, 34, 35, 38, dan 40.

Adapun pilihan jawab A yang memiliki kategori ditolak sebanyak 22 butir soal atau sejumlah 55% yakni yang terdapat pada soal nomor 1, 2, 3, 7, 8, 10, 13, 14, 17, 18, 19, 23, 24, 26, 27, 28, 29, 32, 33, 37, 38, dan 40. Pilihan jawab B yang memiliki kategori ditolak sebanyak 21 butir soal atau sejumlah 52,5% yakni yang terdapat pada soal nomor 1, 3, 4, 5, 6, 8, 10, 11, 12, 14, 18, 19, 21, 24, 27, 28, 29, 32, 38, 39, dan 40. Pilihan jawab C yang memiliki kategori ditolak sebanyak 27 butir soal atau sejumlah 68,5% yakni yang terdapat pada soal nomor 1, 2, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15, 16, 17, 19, 20, 21, 23, 26, 27, 31, 36, 37, 38, 39, dan 40. Pilihan jawab D yang memiliki kategori ditolak sebanyak 24 butir soal atau sejumlah 60% yakni yang terdapat pada soal nomor 2, 3, 4, 5, 6, 7, 8, 11, 12, 13, 17, 18, 21, 22, 23, 24, 26, 27, 28, 29, 31, 32, 36, 37, dan 39.

Sedangkan untuk soal yang ditolak karena tidak memenuhi standard minimal pemilih opsi tersebut yaitu sebesar persentase 5% dari jumlah peserta yang mengikuti tes, maka pilihan ganda

tersebut disarankan untuk direvisi karena soal yang baik adalah yang memiliki pilihan ganda yang mampu mengecoh.⁵²

Kesimpulan

Kesimpulan yang didapatkan dari hasil analisis butir soal ini tentunya dari beragam data yang telah diperoleh. Tentunya juga dari beragam cara dari hasil uji data dari ranah validitas, reliabilitas, daya beda, tingkat kesukaran, hingga, daya pengecoh. Didalam Validitas butir soal mengacu pada sejauh mana setiap item atau pertanyaan yang disajikan dalam tes benar-benar mengukur apa yang dimaksudkan untuk diukur dengan hasil yang telah dijabarkan diatas dengan hasil 30% soal valid dan 70% soal tidak valid karena memiliki nilai signifikansi $<0,05$ dan nilai R hitung lebih kecil dari R Tabel. Selanjutnya pada uji reliabilitas butir soal menunjukkan seberapa konsisten dan andal butir soal dalam mengukur suatu konsep atau variabel. Ini mengacu pada sejauh mana butir soal tersebut dapat diandalkan dan memberikan hasil yang konsisten ketika diuji berkali-kali, dan hasil yang ditemukan bahwa nilai Cronbach's Alpha sebesar 0,868 dan $> 0,60$ maka butir soal bahasa Arab siswa kelas 7 MTs Al-Ma'arif Rakit reliabel dan nilai item pada butir soal ini bernilai 0,625 dengan jumlah butir soal (N of item) berjumlah 40. Daya beda menunjukan 6 Soal dengan Kategori "Jelek Sekali dalam membedakan, 25 Soal dengan kategori "jelek dalam membedakan", 6 Soal dengan kategori "Cukup dalam membedakan", 3 Soal dengan kategori "Baik dalam membedakan". Sedangkan dengan soal kategori baik sekali dalam membedakan tidak ditemukan. Tingkat kesukaran terdapat 25 soal dengan kategori soal sangat mudah, 14 Soal dengan kategori mudah, 1 soal kategori sukar. Daya pengecoh, dalam konteks soal atau ujian, merujuk pada opsi jawaban alternatif yang sengaja disusun untuk mengecoh atau membingungkan peserta ujian. Tujuan dari daya pengecoh ini adalah untuk menguji pemahaman dan pengetahuan peserta terhadap materi yang diuji. Pada akhirnya ditemukan hasil bahwa pilihan jawaban A menunjukkan pengecoh soal kurang berfungsi dengan baik terhadap peserta didik karena lebih banyak pengecoh soal yang ditolak daripada pengecoh soal yang diterima. Begitu pula dengan pilihan jawaban B, C, dan D. Dengan demikian, dapat disimpulkan soal yang tidak memenuhi kriteria direvisi atau membuat soal baru dan dilakukan uji coba lagi.

Dari penjabaran diatas kami selaku peneliti berharap semoga dengan hadirnya penelitian ini bisa berguna bagi para pendidik sebagai referensi dalam menganalisis butir soal agar lebih efektif dan efisien kedepannya, terutama pada pembelajaran bahasa Arab di era saat ini dan di era mendatang.

Ucapan Terima Kasih

Penyelesaian artikel penelitian ini tidak mungkin tercapai tanpa dukungan, bimbingan, dan dorongan dari berbagai individu dan lembaga. Kami mengungkapkan rasa terima kasih banyak kepada mereka yang telah berkontribusi pada keberhasilan proyek penelitian ini. Pertama-tama, kami ingin menyampaikan berterima kasih dan penghargaan setinggi-tingginya kepada pembimbing kami, yakni Dr. Nur Qomari, M.Pd., yang keahliannya, bimbingan, dan dukungan tak tergantikan sepanjang proses penelitian. Masukan dan kritik yang mendalam dari beliau secara signifikan meningkatkan kualitas penelitian ini. Kami juga berterima kasih kepada Madrasah Tsanawiyah Al

⁵² Meita Fitriawanati, "Peran Analisis Butir Soal Guna Meningkatkan Kualitas Butir Soal, Kompetensi Guru dan Hasil Belajar Peserta Didik," *Prosiding Seminar Nasional Pendidikan PGSD UMS & HDPGSDI Wilayah Jawa* 5, no. 3 (2015): 282–95. <https://publikasiilmiah.ums.ac.id/xmlui/handle/11617/9117>

Ma'arif Rakit dan juga Ibu Tria Khoerunisa, S.Pd., yang telah memberikan izin kepada peneliti untuk melakukan penelitian dilembaga ini serta menyediakan lingkungan penelitian yang kondusif dan akses ke fasilitas yang kami diperlukan.

Daftar Pustaka

- Arifin, Zainal. *Penilaian Pembelajaran Bahasa Berbasis Kompetensi. Remaja Rosdakarya*. Bandung: Remaja Rosdakarya Offset, 2014.
- Arikunto, Suharsimi. *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara, 2018.
- Arikunto, Suharsimi. *Prosedur Penelitian Suatu Pendekatan Praktik, Ed Cet 14*. Jakarta: Rineka Cipta, 2016.
- Astiti, Kadek Ayu. *Evaluasi Pembelajaran*. Yogyakarta: Penerbit Andi, 2017.
- Bahri, Obi Saepul. "Analisis Isi Soal UAMBN Bahasa Arab Madrasah Aliyah Tabun Pelajaran 2016/2017 dan 2017/2018." *Educare: Jurnal Pendidikan dan Pembelajaran* 17, no. 1 (2019): 50–54. <https://jurnal.fkip.unla.ac.id/index.php/educare/article/view/241>
- Brookhart, Susan M, and Anthony J Nitko. *Educational Assessment of Students*. 8th ed. United States: Pearson Education Inc, 2019.
- Dwipayani, Sri. "Analisis Validitas dan Reliabilitas Butir Soal Ulangan Akhir Semester Bidang Studi Bahasa Indonesia Kelas XD SMAN 1 Terhadap Pencapaian Kompetensi." *Jurnal Pendidikan Bahasa dan Sastra Indonesia Undiksha* 1, no. 5 (2013): 1-18. DOI: <https://doi.org/10.23887/jjpbs.v1i5.578>
- Fatimah, Laela Umi, and Khairuddin Alfath. "Analisis Kesukaran Soal, Daya Pembeda dan Fungsi Distraktor." *AL-MANAR: Jurnal Komunikasi dan Pendidikan Islam* 8, no. 2 (2019): 37-64. DOI: <https://doi.org/10.36668/jal.v8i2.115>
- Febriana, Rina. *Evaluasi Pembelajaran*. Edited by Bunga Sari Fatmawati. Jakarta: Bumi Aksara, 2019.
- Fitrianawati, Meita. "Peran Analisis Butir Soal Guna Meningkatkan Kualitas Butir Soal, Kompetensi Guru dan Hasil Belajar Peserta Didik." *Prosiding Seminar Nasional Pendidikan PGSD UMS & HDPGSDI Wilayah Jawa* 5, no. 3 (2015): 282–95. <https://publikasiilmiah.ums.ac.id/xmlui/handle/11617/9117>
- Fitrianti, Leni. "Prinsip Kontinuitas dalam Evaluasi Proses Pembelajaran." *Al-Ishlah: Jurnal Pendidikan* 10, no. 1 (2018): 89–102. DOI: <https://doi.org/10.35445/alishlah.v10i1.68>
- Handayani, Sri Lestari, and Khairil Iba. "Karakteristik Tes Keterampilan Proses Sains: Validitas, Reliabilitas, Tingkat Kesukaran dan Daya Pembeda Soal." *Publikasi Pendidikan : Jurnal Pemikiran, Penelitian dan Pengabdian Masyarakat Bidang Pendidikan* 10, no. 2 (2020): 100–106. DOI: <https://doi.org/10.26858/publikan.v10i2.13051>
- Hidayat, Yayat. "Teori Perolehan dan Perkembangan Bahasa untuk Jurusan Pendidikan Bahasa Arab." *Maharaat: Jurnal Pendidikan Bahasa Arab* 1, no. 01 (2018): 24-40. DOI: <https://doi.org/10.18196/mht.113>

- Ida, Farida Far, and Anna Musyarofah. “Validitas dan Reliabilitas dalam Analisis Butir Soal.” AL-MU’ARRIB: Journal of Arabic Education 1, no. 1 (2021): 34-44. DOI: <https://doi.org/10.32923/al-muarrib.v1i1.2100>
- Imam Ghozali. *Aplikasi Analisis Multivariate Dengan IBM SPSS 25*. Semarang: Badan Penerbit Universitas Diponegoro, 2018.
- Kadir, A. “Menyusun dan Menganalisis Tes Hasil Belajar.” Al-TA’DIB: Jurnal Kajian Ilmu Kependidikan 8, no. 2 (2015): 70-81. DOI: <https://dx.doi.org/10.31332/atdb.v8i2.411>
- Khafidin, Zaenal. “Analisis Validitas dan Reliabilitas Tes Mata Pelajaran Pendidikan Agama Islam Tingkat SMA.” Edukasia: Jurnal Penelitian Pendidikan Islam, 9, no. 2 (2014): 253-266. DOI: <https://dx.doi.org/10.21043/edukasia.v9i2.775>
- Ma’arif, Ahmad Syamsul. “Sighab Iktibarat Al-’Arabiyah Fi Dui Al-Iktibar Al-Mutqn (TOAFL).” Jurnal Al Bayan: Jurnal Jurusan Pendidikan Bahasa Arab 9, no. 2 (2017): 160-173. DOI: <http://dx.doi.org/10.24042/albayan.v9i2.2233>
- Machmudah, Umi, Khuzaimah Khuzaimah, and Mufidatus Sholihah. “The Characteristics of Arabic Language Textbooks of Madrasah Ibtidaiyah.” Jurnal Al Bayan: Jurnal Jurusan Pendidikan Bahasa Arab 11, no. 1 (2019):179–93. DOI: <https://dx.doi.org/10.24042/albayan.v11i1.4091>
- Magdalena, Ina, Septy Nurul Fauziah, Siti Nur Faziah, and Fika Sulachatur Nupus. “Analisis Validitas, Reliabilitas, Tingkat Kesulitan dan Daya Beda Butir Soal Ujian Akhir Semester Tema 7 Kelas III SDN Karet 1 Sepatan.” BINTANG: Jurnal Pendidikan dan Sains 3, no. 2 (2021): 198-214. <https://ejournal.stitpn.ac.id/index.php/bintang/article/view/1291>
- Mahfudhah, Z. *Tablil Iktibarat Al-Lughab Al-Arabiyah (Dirasab Taqwimiyah Bil Madrasah Muthawasithab Sarusy Syari’ah Bi Banda Aceh*. Skripsi thesis. UIN Ar-Raniry Banda Aceh, 2017. <https://repository.ar-raniry.ac.id/id/eprint/5658/>
- Matsna, Moh. Raswan. *Evaluasi Pembelajaran Bahasa Arab*. Ciputat: UIN Jakarta Press, 2015.
- Maulana, Deni, and Anwar Sanusi. “Analisis Butir Soal Bahasa Arab Ujian Akhir Madrasah Bersama Daerah (UAMBD) Madrasah Ibtidaiyah Tahun 2017-2018.” Ta’lim Al-’Arabiyah: Jurnal Pendidikan Bahasa Arab & Kebahasaaraban 4, no. 1 (2020): 12–24. DOI: <https://doi.org/10.15575/jpba.v4i1.8054>
- Muslim, Ahmad Bukhari. “Al-Iktibarat Al-Lughawiyah: Tasmimuba Wa Anwa’uba.” Jurnal Al Bayan: Jurnal Jurusan Pendidikan Bahasa Arab 8, no. 1 (2016): 23–39. DOI: <https://dx.doi.org/10.24042/albayan.v8i1.359>
- Mutholib, Abdul. “Analisis Butir Soal Bahasa Arab Ujian Akhir Madrasah Berstandar Nasional (UAMBN) Bagi Siswa Madrasah Ibtidaiyah (MI) Tahun Pelajaran 2012-2013.” Arabia: Jurnal Pendidikan Bahasa Arab 5, no. 2 (2016): 141-161. DOI: <https://dx.doi.org/10.21043/arabia.v5i2.1387>
- Nurhalimah, Sri, Yunin Hidayati, Irsad Rosidi, and Wiwin Puspita Hadi. “Hubungan Antara Validitas Item dengan Daya Pembeda dan Tingkat Kesukaran Soal Pilihan Ganda Pas.” Natural Science Education Research 4, no. 3 (2022): 249-257. DOI: <https://doi.org/10.21107/nser.v4i3.8682>

- Pardimin, Pardimin, Sri Adi Widodo, and Indriyati Eko Purwaningsih. "Analisis Butir Soal Tes Pemecahan Masalah Matematika." *Wacana Akademika: Majalah Ilmiah Kependidikan* 1, no. 1 (2017): 69–76. DOI: <https://doi.org/10.30738/wa.v1i1.1084>
- Purwadhi, Purwadhi. *Pengembangan Kurikulum dalam Pembelajaran Bahasa Arab*. MIMBAR PENDIDIKAN: Jurnal Indonesia untuk Kajian Pendidikan, vol. 4 no. 2 (2019): 103-112. DOI: <https://doi.org/10.17509/mimbardik.v4i2.22201>
- Purwono, Fuad Hasyim, Annida Unatiq Ulya, Nurwulan Purnasari, and Ronnawan Juniatmoko. *Metodologi Penelitian (Kuantitatif, Kualitatif dan Mixed Method)*. Jakarta: Guepedia Group, 2019.
- Rahmawati, Annisa Elita, and Umi Hijriyyah. "Development of the Quizizz Application-Based Evaluation Tool for Learning Arabic for MTs | Pengembangan Alat Evaluasi Berbasis Aplikasi Quizizz pada Pembelajaran Bahasa Arab untuk MTs." *Mantiq Tayr: Journal of Arabic Language* 3, no. 2 (2023): 135-150. DOI: <https://doi.org/10.25217/mantiqtayr.v3i2.3484>
- Ridho, Ubaid. *Evaluasi Dalam Pembelajaran Bahasa Arab*. An Nabighoh: Jurnal Pendidikan dan Pembelajaran Bahasa Arab. 20, no. 01 (2018): 19–26. DOI: <https://doi.org/10.32332/an-nabighoh.v20i01.1124>
- Rudin, Amir. "Analisis Butir Soal Ujian Kenaikan Kelas Mata Pelajaran Bahasa Arab Kelas VIII di MTs Modern Al-Azhary Ajibarang Tahun Pelajaran 2015/2016." Skripsi thesis. Institut Agama Islam Negeri Purwokerto. <https://repository.uinsaizu.ac.id/2117/>
- Setiyawan, Adhi. "Faktor-Faktor yang Mempengaruhi Reliabilitas Tes." *AN NUR: Jurnal Studi Islam* 6, no. 2 (2014): 341–54. <https://jurnalannur.ac.id/index.php/An-Nur/article/view/53>
- Sudaryono. *Metodologi Penelitian Kuantitatif, Kualitatif, dan Mixed Method*. 2nd ed. Depok: PT Raja Grafindo Persada, 2021.
- Sudijono, Anas. *Pengantar Evaluasi Pendidikan*. Jakarta: Raja Grafindo Persada, 2011.
- Sugiono. *Metode Penelitian Pendidikan*. Bandung: Alfabeta, 2012.
- Sugiyono. *Metode Penelitian Kuantitatif*. Bandung: Alfabeta, 2022.
- Suharti, Sri. "Kualitas Tes Bahasa Arab dan Prestasi Peserta Didik Madrasah Tsanawiyah Kabupaten Bantul (Analisis Butir Soal UAMBN Tahun Ajaran 2013/2014)." *Jurnal Pendidikan Madrasah* 2, no. 1 (2017): 185-196. DOI: <https://ejournal.uin-suka.ac.id/tarbiyah/JPM/article/view/1439>
- Susanty, Fatimah Depi. "Analisis Validasi Soal Tes Hasil Belajar pada Pelaksanaan Pembelajaran Bahasa Arab di Pusat Pengembangan Bahasa (P3B) UIN Suska Riau." *Kutubkhanah: Jurnal Penelitian Sosial Keagamaan* 19, no. 2 (2016): 112-132. DOI: <https://dx.doi.org/10.24014/kutubkhanah.v19i2.2546>
- Suseno, Miftahun Ni'mah. "Pengembangan Pengujian Validitas Isi dan Validitas Konstruk: Interpretasi Hasil Pengujian Validitas." in *Seminar Nasional Psikometri* (Surakarta: Universitas Muhammadiyah Surakarta, 2014), hlm. 70–83. <https://publikasiilmiah.ums.ac.id/xmlui/handle/11617/6399>
- Tanwir, Tanwir. "Dasar-Dasar dan Ruang Lingkup Evaluasi Pendidikan." *Al-Ishlah: Jurnal Pendidikan Islam* 13, no. 1 (2015): 47–59. DOI: <https://doi.org/10.35905/alishlah.v13i1.490>

- Utomo, Budi. “*Analisis Validitas Isi Butir Soal Sebagai Salah Satu Upaya Peningkatan Kualitas Pembelajaran di Madrasah Berbasis Nilai-Nilai Islam.*” *Jurnal Pendidikan Matematika (Kudus)* 1, no. 2 (2022): 155-175. DOI: <https://dx.doi.org/10.21043/jmtk.v1i2.4868>
- Wahab, Muhib Abdul. “*Pembelajaran Bahasa Arab di Era Posmetode.*” *Arabiyat: Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban* 2, no. 1 (2015): 59-74. DOI: <https://dx.doi.org/10.15408/a.v2i1.1519>
- Wahab, Muhib Abdul. “*Standarisasi Kurikulum Pendidikan Bahasa Arab di Perguruan Tinggi Keagamaan Islam Negeri.*” *Arabiyat: Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban* 3, no. 1 (2016): 32–51. DOI: <https://doi.org/10.15408/a.v3i1.3187>
- Widiyanto, Joko. *SPSS for Windows untuk Analisis Data Statistik dan Penelitian*. Surakarta: BP-FKIP UMS, 2018.
- Yustuti, Sherly, Masrun Masrun, and Hikmah Hikmah. “*Development of Listening Skills Evaluation Instruments | Pengembangan Instrumen Evaluasi Keterampilan Menyimak.*” *Mantiqutayr: Journal of Arabic Language* 3, no. 1 (2023): 1-10. DOI: <https://doi.org/10.25217/mantiqutayr.v3i1.2614>
- Zulkifli, M. “*Analisis Bentuk Evaluasi Kurikulum 2013 Mata Pelajaran Bahasa Arab di MI.*” *Al-Madrasah: Jurnal Pendidikan Madrasah Ibtidaiyah* 2, no. 2 (2018): 125-143. DOI: <https://dx.doi.org/10.35931/am.v0i0.29>