



Analisis Butir Soal Bahasa Arab Kelas 3 Di Smpit Al-khiraat Condet

¹Ahmad Wildan Sahuri Ramdani ²Muammar Qadavi Umasugi ³Ubaid Ridlo ⁴Raswan

¹⁻⁴Universita Islam Negeri Syarif Hidayatullah Jakarta, Indonesia

Email: asahuriramdani@gmail.com

Abstrak

Evaluasi pembelajaran merupakan komponen penting dalam mengukur ketercapaian tujuan pembelajaran, termasuk dalam pembelajaran Bahasa Arab. Instrumen evaluasi yang berkualitas harus mampu mengukur kemampuan peserta didik secara tepat, objektif, dan konsisten. Namun, dalam praktiknya, tidak semua butir soal yang digunakan telah memenuhi kriteria kualitas yang baik. Oleh karena itu, penelitian ini bertujuan untuk menganalisis kualitas butir soal Bahasa Arab kelas III di SMPIT Al-Khiraat Condet ditinjau dari aspek validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan daya pengecoh. Penelitian ini menggunakan pendekatan kuantitatif dengan jenis penelitian deskriptif evaluatif. Data diperoleh dari 25 butir soal pilihan ganda dan hasil jawaban peserta didik kelas III. Analisis data dilakukan dengan bantuan aplikasi SPSS untuk uji validitas dan reliabilitas menggunakan korelasi Product Moment dan koefisien Alpha Cronbach, sedangkan analisis tingkat kesukaran, daya pembeda, dan daya pengecoh diolah menggunakan Microsoft Excel. Hasil penelitian menunjukkan bahwa dari 25 butir soal, sebanyak 21 butir soal dinyatakan valid dan reliabel, sedangkan 4 butir soal tidak memenuhi kriteria tersebut. Analisis tingkat kesukaran menunjukkan 3 butir soal berkategori sedang dan 22 butir soal berkategori sukar. Selain itu, sebagian besar butir soal memiliki daya pembeda dan pengecoh yang berfungsi dengan baik, meskipun beberapa butir soal masih memerlukan revisi. Dengan demikian, instrumen tes secara umum layak digunakan, namun perlu perbaikan agar kualitas evaluasi pembelajaran Bahasa Arab menjadi lebih optimal.

Kata Kunci : Analisis Butir Soal, Bahasa Arab, Evaluasi Pembelajaran, SMPIT

Abstract

Learning evaluation is an essential component in measuring the achievement of instructional objectives, including in Arabic language learning. A high-quality evaluation instrument must be able to measure students' abilities accurately, objectively, and consistently. However, in practice, not all test items used meet the required quality standards. Therefore, this study aims to analyze the quality of Arabic language test items for Grade III students at SMPIT Al-Khiraat Condet in terms of validity, reliability, difficulty level, discriminating power, and distractor effectiveness. This research employed a quantitative approach with a descriptive-evaluative design. The data were obtained from 25 multiple-choice test items and students' response results. Data analysis was conducted using SPSS to test validity and reliability through Product Moment correlation and Cronbach's Alpha coefficient, while the analysis of difficulty level, discriminating power, and distractor effectiveness was processed using Microsoft Excel. The results of the study indicate that out of 25 test items, 21 items were found to be valid and reliable, whereas 4 items did not meet these criteria. The analysis of difficulty levels shows that 3 items fall into the moderate category and 22 items into the difficult category. Furthermore, most test items demonstrate adequate discriminating power and well-functioning distractors, although several items still require revision. Thus, the test instrument is generally suitable for use; however, improvements are necessary to enhance the quality of Arabic language learning evaluation.

Keywords: Item Analysis, Arabic Language, Learning Evaluation, Islamic Junior High School

PENDAHULUAN

Bahasa merupakan alat komunikasi utama manusia yang berbentuk sistem lambang bunyi dan digunakan oleh suatu masyarakat untuk menyampaikan ide, gagasan, serta nilai-nilai budaya. Bahasa bersifat sistematis, disepakati bersama, dan dihasilkan oleh alat ucapan manusia, sehingga membedakannya dari bentuk lambang lain seperti isyarat maupun tulisan. Tulisan pada hakikatnya bukan bahasa itu sendiri, melainkan representasi visual dari bahasa lisan. Keberagaman bahasa di dunia lahir dari perbedaan masyarakat dan kebudayaan, yang oleh para ahli linguistik diklasifikasikan ke dalam beberapa rumpun bahasa, salah satunya rumpun bahasa Semit, tempat bahasa Arab berasal (Sudijono, 2021).

Bahasa Arab memiliki kedudukan yang sangat penting karena merupakan bahasa Al-Qur'an dan sumber ajaran Islam. Bahasa ini digunakan secara luas oleh ratusan juta penutur dan menjadi bahasa resmi di banyak negara. Al-Qur'an secara tegas menyatakan bahwa wahyu diturunkan dalam bahasa Arab agar dapat dipahami oleh manusia. Oleh karena itu, Islam dan bahasa Arab tidak dapat dipisahkan. Para ulama, seperti Ibnu Taimiyah, menegaskan bahwa mempelajari bahasa Arab merupakan kewajiban bagi umat Islam, karena pemahaman terhadap Al-Qur'an dan Sunnah tidak dapat dicapai tanpa penguasaan bahasa Arab (Arikunto, 2020).

Di Indonesia, bahasa Arab hadir bersamaan dengan masuknya Islam dan menjadi bagian penting dalam kehidupan keagamaan umat Muslim. Pada awalnya, pembelajaran bahasa Arab berorientasi pada kebutuhan ibadah, seperti membaca Al-Qur'an dan dzikir. Namun, seiring perkembangan zaman, pembelajaran bahasa Arab meluas tidak hanya untuk memahami ajaran Islam, tetapi juga sebagai alat komunikasi aktif. Pengajaran bahasa Arab kini berkembang di berbagai lembaga pendidikan formal, nonformal, dan pesantren, sehingga diperlukan pengelolaan kurikulum yang sistematis serta

evaluasi pembelajaran yang terstruktur untuk menjamin mutu dan pencapaian tujuan pendidikan (Djiwandono, 2021).

Evaluasi pembelajaran merupakan bagian yang tidak terpisahkan dari proses pendidikan, karena melalui evaluasi dapat diketahui tingkat ketercapaian tujuan pembelajaran yang telah ditetapkan. Salah satu bentuk evaluasi yang paling sering digunakan dalam pembelajaran adalah tes, khususnya tes tertulis dalam bentuk pilihan ganda. Dalam pembelajaran Bahasa Arab, tes berfungsi tidak hanya untuk mengukur penguasaan peserta didik terhadap aspek kebahasaan seperti kosakata, struktur, dan pemahaman makna, tetapi juga untuk menilai efektivitas proses pembelajaran yang telah berlangsung. Oleh karena itu, kualitas butir soal menjadi faktor penentu dalam menghasilkan data evaluasi yang akurat dan dapat dipercaya (Widayanti, 2022).

Butir soal yang baik harus memenuhi kriteria tertentu, antara lain validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Soal yang tidak memenuhi kriteria tersebut berpotensi menghasilkan penilaian yang bias dan tidak mencerminkan kemampuan peserta didik yang sesungguhnya. Dalam konteks pembelajaran Bahasa Arab di tingkat Sekolah Menengah Pertama Islam Terpadu (SMPIT), penyusunan soal yang berkualitas menjadi tantangan tersendiri, mengingat Bahasa Arab bagi sebagian besar peserta didik merupakan bahasa asing yang memiliki karakteristik linguistik yang berbeda dengan bahasa ibu mereka (Hasanah, 2020).

SMPIT Al-Khiraat Condet sebagai lembaga pendidikan Islam memiliki peran strategis dalam membekali peserta didik dengan kemampuan Bahasa Arab sebagai sarana memahami ajaran Islam dan mengembangkan kompetensi akademik. Namun demikian, efektivitas pembelajaran Bahasa Arab perlu didukung oleh sistem evaluasi yang tepat dan terstandar. Oleh

sebab itu, analisis butir soal Bahasa Arab kelas 3 di SMPIT Al-Khairaat Condet menjadi penting untuk dilakukan, guna mengetahui kualitas soal yang digunakan dalam evaluasi pembelajaran. Hasil analisis ini diharapkan dapat menjadi dasar perbaikan dan pengembangan instrumen evaluasi yang lebih valid, reliabel, dan sesuai dengan kemampuan peserta didik, sehingga tujuan pembelajaran Bahasa Arab dapat tercapai secara optimal.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan jenis penelitian deskriptif evaluatif, yang bertujuan untuk menganalisis kualitas butir soal Bahasa Arab kelas 3 di SMPIT Al-Khairaat Condet. Subjek penelitian adalah seluruh butir soal Bahasa Arab berbentuk pilihan ganda yang digunakan dalam evaluasi pembelajaran, dengan jumlah 25 butir soal, sedangkan objek penelitian difokuskan pada karakteristik kualitas soal. Data penelitian diperoleh dari hasil jawaban peserta didik kelas 3 serta dokumen soal ujian yang digunakan oleh guru. Teknik pengumpulan data dilakukan melalui studi dokumentasi terhadap perangkat evaluasi dan hasil tes peserta didik (Fadli, 2021).

Analisis data dilakukan dengan menggunakan teknik analisis butir soal yang meliputi uji validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh (Hilalludin et al., 2025). Uji validitas dan reliabilitas dianalisis dengan bantuan aplikasi SPSS menggunakan korelasi Product Moment dan koefisien Alpha Cronbach, sedangkan tingkat kesukaran, daya pembeda, dan analisis pengecoh diolah menggunakan aplikasi Microsoft Excel. Hasil analisis digunakan untuk mengklasifikasikan butir soal ke dalam kategori layak digunakan, perlu direvisi, atau tidak layak digunakan, sehingga dapat menjadi dasar perbaikan instrumen evaluasi pembelajaran Bahasa Arab di SMPIT Al-Khairaat Condet (Abdurrozaq Abdurrozaq, 2025).

HASIL DAN PEMBAHASAN

Validitas Butir Soal

Suatu instrumen tes dapat dikatakan valid apabila mampu mengukur secara tepat dan konsisten apa yang seharusnya diukur. Validitas instrumen menunjukkan sejauh mana butir soal benar-benar merepresentasikan kompetensi atau kemampuan yang ingin dinilai. Dalam penelitian ini, penentuan validitas butir soal dilakukan untuk memastikan bahwa setiap item mampu membedakan peserta didik yang memiliki tingkat penguasaan materi yang berbeda (Farida & Setiawan, 2022).

$$r_{xy} = \frac{N \sum XY (\sum X)(\sum Y)}{\sqrt{\{N \sum X^2 - (\sum X)^2\} \{N \sum Y^2 - (\sum Y)^2\}}}$$

Keterangan

r_{xy} : Koefisien validitas

N : Banyaknya subjek

X : Nilai pembanding

Y : Nilai dari instrument yang akan dicari validitasnya

Pengujian validitas butir soal dilakukan dengan menggunakan uji Product Moment (Pearson Correlation) yang dianalisis melalui aplikasi SPSS. Uji validitas ini dilakukan dengan cara mengorelasikan skor setiap butir soal dengan skor total yang diperoleh responden. Dasar pengambilan keputusan dalam uji validitas Product Moment adalah sebagai berikut: (a) membandingkan nilai r hitung dengan nilai r tabel, di mana item dinyatakan valid apabila r hitung $>$ r tabel dan dinyatakan tidak valid apabila r hitung $<$ r tabel; (b) membandingkan nilai signifikansi (*Sig. 2-tailed*) dengan taraf signifikansi 0,05. Item soal dinyatakan valid apabila nilai *Sig. (2-tailed)* $<$ 0,05 dan koefisien korelasi bernilai positif. Adapun secara manual, uji validitas dapat dihitung menggunakan rumus Product Moment Pearson. Dalam penelitian ini, peneliti menggunakan SPSS untuk menguji validitas butir soal yang ada di SDIT Al Islam Kudus Kelas 4. Adapun langkah-langkah pengujian yang dilakukan meliputi:(Slamet & Sri, 2022).

No Butir Soal	Hasil r Hitung	Hasil r Tabel (0,05)	Kriteria Validasi Butir Soal
1	1.75	0.41	Valid
2	4.08	0.41	Valid
3	0.79	0.41	Valid
4	0.91	0.41	Valid
5	2.83	0.41	Valid
6	3.61	0.41	Valid
7	2.04	0.41	Valid
8	3.04	0.41	Valid
9	0.28	0.41	Tidak Valid
10	0.28	0.41	Tidak Valid
11	0.91	0.41	Valid
12	0.00	0.41	Tidak Valid
13	0.55	0.41	Valid
14	2.04	0.41	Valid
15	1.13	0.41	Valid
16	2.04	0.41	Valid
17	4.17	0.41	Valid
18	3.61	0.41	Valid
19	3.35	0.41	Valid
20	1.67	0.41	Valid
21	4.40	0.41	Valid
22	1.38	0.41	Valid

Tabel. 1 Analisis Validitas Butir Soal

Berdasarkan hasil pengujian validitas butir soal Bahasa Arab menggunakan uji Product Moment, diperoleh bahwa dari total 21 butir soal yang diuji, terdapat 4 butir soal yang tidak memenuhi kriteria validitas karena nilai r hitung lebih kecil dari r tabel (0,41) pada taraf signifikansi 0,05. Dengan demikian, 17 butir soal dinyatakan valid dan layak digunakan dalam proses pembelajaran dan evaluasi, sedangkan 4 butir soal yang tidak valid perlu direvisi atau disusun ulang. Hal ini sejalan dengan pendapat Sugiyono yang menyatakan bahwa validasi data merupakan proses sistematis untuk menilai ketepatan instrumen dalam mengukur variabel penelitian. Senada dengan itu, Sutama menegaskan bahwa validitas instrumen mencerminkan kualitas data yang dihasilkan dalam menjelaskan fenomena yang diteliti. Oleh karena itu pengujian validitas butir soal menjadi langkah penting untuk menjamin keabsahan dan kualitas instrumen evaluasi pembelajaran (Fitria, 2022).

Reliabilitas Butir Soal

$$r_{11} = \left[\frac{k}{(k-1)} \right] \left[1 - \frac{\Sigma \sigma_b^2}{\sigma_t^2} \right]$$

Keterangan :

- r_{11} = koefisien reliabilitas instrument (total tes)
- k = jumlah butir pertanyaan yang sah
- $\Sigma \sigma_b^2$ = jumlah varian butir
- σ_t^2 = varian skor total

Uji reliabilitas digunakan untuk mengetahui sejauh mana alat Pengumpul data memiliki tingkat ketelitian, kestabilan, dan konsistensi dalam mengukur suatu gejala pada kelompok individu, meskipun digunakan pada waktu yang berbeda. Menurut Suharsimi Arikunto, reliabilitas menunjukkan bahwa suatu instrumen cukup andal untuk digunakan sebagai alat pengumpul data karena mampu menghasilkan data yang dapat dipercaya. Instrumen yang reliabel akan memberikan hasil pengukuran yang relatif sama apabila digunakan dalam kondisi yang serupa. Oleh karena itu, uji reliabilitas bertujuan untuk memastikan bahwa instrumen penelitian layak dan konsisten digunakan dalam proses pengumpulan data.(Sugiyono, 2018)

No.	No. Item	Statistics Item		
		Prop. Correct	Biser	Point Biser
1	1	0,810	0,793	0,551
2	2	0,667	0,787	0,489
3	3	0,857	0,681	0,518
4	4	0,714	0,842	0,533
5	5	0,619	1,291	0,797
6	6	0,571	1,374	0,853
7	7	0,667	1,202	0,747
8	8	0,286	0,564	0,442
9	9	0,571	1,003	0,623
10	10	0,571	0,731	0,454
11	11	0,286	0,778	0,610
12	12	0,333	0,849	0,623
13	13	0,476	0,538	0,347
14	14	0,667	0,709	0,441
15	15	0,381	0,990	0,689
16	16	0,333	0,585	0,430
17	17	0,429	0,747	0,498
18	18	0,571	1,350	0,838
19	19	0,714	0,763	0,483
20	20	0,571	1,053	0,654
21	21	0,524	1,115	0,703
22	22	0,524	0,560	0,353
23	23	0,429	0,332	0,222
24	24	0,619	-0,001	-0,001
25	25	0,429	0,539	0,360

Koefisien Alpha Cronbach merupakan salah satu teknik statistik yang paling umum digunakan untuk menguji reliabilitas instrumen penelitian, khususnya instrumen berbentuk tes atau angket. Suatu instrumen dinyatakan memiliki tingkat reliabilitas yang memadai apabila nilai koefisien Alpha Cronbach $\geq 0,60$. Metode ini digunakan untuk mengukur konsistensi internal antarbutir soal dalam satu instrumen. Secara manual, perhitungan reliabilitas dapat dilakukan menggunakan rumus Alpha Cronbach, namun dalam penelitian ini analisis dilakukan dengan bantuan aplikasi **SPSS** untuk memperoleh hasil yang lebih akurat dan efisien. Berdasarkan hasil analisis reliabilitas menggunakan SPSS dengan metode Alpha Cronbach, diperoleh bahwa dari total 21 butir soal yang diuji, terdapat 4 butir soal yang tidak

memenuhi kriteria reliabilitas. Dengan demikian, sebagian besar butir soal telah menunjukkan tingkat konsistensi yang baik dan dapat digunakan sebagai alat evaluasi pembelajaran. Temuan ini sejalan dengan pendapat Suharsimi Arikunto yang menyatakan bahwa reliabilitas berkaitan dengan tingkat kepercayaan terhadap instrumen dalam menghasilkan data yang akurat dan konsisten. Instrumen yang reliabel menjadi fondasi penting dalam penelitian karena menjamin keabsahan data yang dikumpulkan dan mendukung validitas hasil penelitian secara keseluruhan.(Nurbayan et al., 2024)

Analisis Tingkat Kesukaran Butir Soal

Tingkat kesukaran butir soal merupakan salah satu indikator penting dalam evaluasi pembelajaran yang berfungsi untuk mengetahui derajat kesulitan suatu soal dalam mengukur kemampuan peserta didik. Soal yang berkualitas bukanlah soal yang terlalu mudah atau terlalu sulit, melainkan soal yang memiliki tingkat kesukaran proporsional sehingga mampu membedakan peserta didik berdasarkan tingkat penguasaan materi secara tepat. Oleh karena itu, analisis tingkat kesukaran menjadi tahap penting dalam pengembangan dan evaluasi instrumen tes (Nugroho, 2023).

Dalam praktik evaluasi, terdapat beberapa metode yang dapat digunakan untuk menganalisis tingkat kesukaran soal, di antaranya metode Tingkat Kesukaran (TK), tabel batas tingkat kesukaran, dan metode proporsi jawaban benar. Pada penelitian ini, analisis tingkat kesukaran dilakukan menggunakan metode proporsi jawaban benar karena metode ini bersifat sederhana, mudah diterapkan, dan umum digunakan dalam penelitian kuantitatif. Tingkat kesukaran dihitung dengan membandingkan jumlah peserta didik yang menjawab benar terhadap jumlah seluruh peserta didik yang mengikuti tes (Fitriani, 2021).

Rumus yang digunakan untuk menentukan tingkat kesukaran soal adalah sebagai berikut:

$$P = \sum B / N$$

Keterangan:

P = tingkat kesukaran soal

$\sum B$ = jumlah peserta didik yang menjawab benar

N = jumlah seluruh peserta didik

Hasil perhitungan tingkat kesukaran kemudian diinterpretasikan berdasarkan kriteria berikut: nilai P lebih dari 0,70 menunjukkan soal berkategori mudah, nilai P antara 0,30 sampai 0,70 termasuk kategori sedang, sedangkan nilai P kurang dari 0,30 menunjukkan soal berkategori sukar. Berdasarkan hasil analisis terhadap 25 butir soal yang diuji, diperoleh bahwa sebanyak 22 butir soal termasuk dalam kategori sukar, 3 butir soal berada pada kategori sedang, dan tidak terdapat butir soal yang tergolong mudah. Dominasi soal dengan kategori sukar menunjukkan bahwa sebagian besar soal memiliki tingkat kesulitan yang relatif tinggi dibandingkan kemampuan rata-rata peserta didik (Rahman, 2022).

Temuan ini mengindikasikan perlunya perhatian khusus dalam penyusunan instrumen evaluasi. Menurut Anas Sudijono, analisis tingkat kesukaran merupakan tahap penting yang harus dilakukan setelah uji validitas dan reliabilitas, karena tingkat kesukaran berpengaruh langsung terhadap kualitas hasil tes. Butir soal dengan tingkat kesukaran sedang dinilai paling ideal dan layak disimpan dalam bank soal, karena mampu mengukur kemampuan peserta didik secara optimal. Oleh sebab itu, dominasi soal sukar dalam penelitian ini menunjukkan perlunya revisi dan perbaikan butir soal agar distribusi tingkat kesukaran lebih seimbang dan sesuai dengan karakteristik peserta didik (Hamid, 2021).

Daya Pembeda Soal

Daya pembeda soal merupakan indikator penting yang menunjukkan kemampuan suatu butir soal dalam membedakan peserta didik yang berkemampuan tinggi dan rendah. Semakin tinggi daya pembeda suatu soal, semakin baik kualitas soal tersebut dalam mengukur penguasaan materi peserta didik secara akurat. Oleh karena itu, analisis daya pembeda menjadi bagian esensial dalam evaluasi kualitas instrumen tes. Dalam penelitian ini, indeks daya pembeda (D) dianalisis menggunakan bantuan aplikasi Microsoft Excel dengan mengacu pada kriteria Djiwandono, yang mengklasifikasikan daya pembeda ke dalam kategori baik sekali, baik, cukup, rendah, dan tidak memiliki daya pembeda (Putra, 2023).

Hasil analisis menunjukkan bahwa dari keseluruhan butir soal yang diuji, terdapat 4 butir soal yang memiliki daya pembeda kategori baik sekali, 10 butir soal termasuk kategori baik, 8 butir soal berada pada kategori cukup, 2 butir soal tergolong jelek, dan 1 butir soal termasuk jelek sekali. Temuan ini mengindikasikan bahwa sebagian besar butir soal telah memiliki kemampuan yang cukup baik dalam membedakan kemampuan peserta didik. Namun demikian, butir soal yang berada pada kategori cukup, jelek, dan jelek sekali perlu dilakukan evaluasi dan revisi agar kualitas instrumen evaluasi pembelajaran menjadi lebih optimal dan mampu memberikan hasil pengukuran yang lebih akurat (Suryani, 2021).

Analisis Pengecoh Soal

Pengecoh merupakan alternatif jawaban pada soal berbentuk pilihan ganda yang berfungsi untuk mengalihkan peserta didik yang belum menguasai materi secara optimal. Butir soal yang baik memiliki pengecoh yang efektif, yaitu dipilih secara relatif merata oleh peserta didik yang menjawab salah. Sebaliknya, pengecoh yang tidak berfungsi dengan baik akan dipilih secara tidak merata atau bahkan tidak dipilih sama sekali. Oleh karena itu, analisis

pengecoh penting dilakukan untuk mengetahui kualitas setiap alternatif jawaban dalam suatu butir soal. Secara teoritis, analisis pengecoh dapat dilakukan dengan menggunakan rumus indeks pengecoh yang memperhatikan jumlah peserta didik yang memilih setiap alternatif jawaban. Namun, dalam penelitian ini peneliti menggunakan **Microsoft Excel** untuk mempermudah pengolahan dan analisis data pengecoh butir soal.(Warju et al., 2020)

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

Keterangan:

IP : Banyaknya peserta tes yang memilih pengecoh tertentu
N : Banyaknya seluruh peserta tes

B : Banyaknya peserta tes yang menjawab benar butir soal yang bersangkutan

N : Banyaknya alternatif jawaban Akan tetapi peneliti memilih menggunakan Excel dalam pengolahan data uji data butir soal pengecoh. Berikut hasil data butir soal pengecoh:

Analisis ini bertujuan untuk melihat apakah setiap opsi jawaban berfungsi dengan baik dalam mengecoh peserta didik yang tidak menjawab benar.(Kafi & Ilma, 2022)

Analisis daya pengecoh dilakukan untuk mengetahui apakah setiap opsi jawaban dalam butir soal mampu berfungsi secara efektif sebagai pengecoh, khususnya bagi peserta didik yang belum menguasai materi. Pengecoh yang baik ditandai dengan adanya pemilihan oleh peserta didik, sehingga dapat menunjukkan bahwa opsi tersebut cukup logis dan mampu membedakan antara peserta didik yang memahami materi dan yang belum. Dengan demikian, analisis daya pengecoh menjadi bagian penting dalam menilai kualitas instrumen tes secara keseluruhan (Mardapi, 2021).

Berdasarkan hasil analisis daya pengecoh terhadap 25 butir soal Bahasa Arab yang dikerjakan oleh 21 peserta didik, diketahui bahwa sebagian

besar butir soal telah memiliki pengecoh yang berfungsi dengan baik. Butir soal nomor 1, 2, 3, 4, 5, 6, 7, 9, 10, 13, 14, 15, 17, 18, 19, 20, 21, 22, 23, dan 25 dinyatakan layak digunakan karena opsi jawabannya mampu mengecoh peserta didik secara proporsional dan mengukur kemampuan secara tepat. Namun demikian, beberapa butir soal, yaitu nomor 8, 11, 12, 16, dan 24, perlu direvisi karena berdasarkan hasil uji validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan daya pengecoh, butir-butir tersebut belum memenuhi kriteria soal yang baik. Revisi diperlukan agar kualitas instrumen evaluasi pembelajaran Bahasa Arab menjadi lebih optimal dan akurat dalam mengukur hasil belajar peserta didik (Azizah, 2022).

KESIMPULAN

Berdasarkan hasil pengujian validitas menggunakan uji Product Moment terhadap butir soal Bahasa Arab kelas III SMPIT Al Khairaat, diperoleh bahwa dari total 25 butir soal, sebanyak 21 butir soal dinyatakan valid dan 4 butir soal dinyatakan tidak valid. Suatu butir soal dinyatakan valid apabila nilai r hitung lebih besar dari 0,41 pada taraf signifikansi 0,05, sedangkan butir soal dinyatakan tidak valid apabila nilai r hitung lebih kecil dari 0,41. Dengan demikian, sebagian besar butir soal layak digunakan dalam proses evaluasi pembelajaran, sementara butir soal yang tidak valid perlu dilakukan evaluasi dan perbaikan agar memenuhi kriteria validitas. Hasil uji reliabilitas menggunakan aplikasi SPSS dengan koefisien Alpha Cronbach menunjukkan bahwa dari 25 butir soal yang dianalisis, sebanyak 21 butir soal dinyatakan reliabel dan 4 butir soal dinyatakan tidak reliabel. Hal ini menunjukkan bahwa instrumen tes secara umum memiliki tingkat konsistensi yang baik dan dapat dipercaya sebagai alat pengumpul data. Selanjutnya, hasil analisis tingkat kesukaran menunjukkan bahwa 3 butir soal berada pada kategori sedang, 22 butir soal berada pada kategori sukar, dan tidak terdapat butir soal yang berada pada kategori mudah, sehingga distribusi tingkat kesukaran soal masih perlu disesuaikan agar lebih proporsional.

Berdasarkan hasil analisis daya pembeda menggunakan Microsoft Excel, butir soal diklasifikasikan ke dalam lima kategori, yaitu baik sekali, baik, cukup, jelek, dan jelek sekali. Selain itu, hasil analisis daya pengecoh menunjukkan bahwa butir soal nomor 1, 2, 3, 4, 5, 6, 7, 9, 10, 13, 14, 15, 17, 18, 19, 20, 21, 22, 23, dan 25 dinyatakan dapat diterima karena pengecohnya berfungsi dengan baik dan mampu mengukur kemampuan peserta didik secara efektif. Sementara itu, butir soal selain nomor tersebut dinyatakan belum memenuhi kriteria yang baik dan disarankan untuk direvisi sebelum digunakan kembali.

DAFTAR PUSTAKA

- Abdurrozak Abdurrozak, H. H. (2025). *Jurnal Penelitian Multidisiplin Indonesia*. 1(3), 171–206.
- Arikunto, S. (2020). Analisis Kualitas Instrumen Tes dalam Evaluasi Pembelajaran. *Jurnal Penelitian Pendidikan*, 21(1). <https://doi.org/10.21831/jpp.v21i1.31245>
- Azizah, N. (2022). Evaluasi Pembelajaran Bahasa Arab Berbasis HOTS. *Alsinatuna*, 8(2). <https://doi.org/10.28918/alsinatuna.v8i2.5619>
- Djiwandono, M. S. (2021). Daya Pembeda dan Tingkat Kesukaran dalam Tes Bahasa. *Indonesian Journal of Applied Linguistics*, 10(3). <https://doi.org/10.17509/ijal.v10i3.31702>
- Fadli, M. R. (2021). Memahami desain metode penelitian kualitatif. *Humanika, Kajian Ilmiah Mata Kuliah Umum*, 21(1), 44. <https://doi.org/10.21831/hum.v21i1>.
- Farida, I., & Setiawan, D. (2022). Business strategies and competitive advantage: the role of performance and innovation. *Journal of Open Innovation: Technology, Market, and Complexity*, 8(3), 163.
- Fitria, dkk. (2022). Pembelajaran Statistika Menggunakan Software SPSS untuk Uji Validitas dan Reliabilitas'. *Jurnal Basicedu*, 6(4), 6491–6504.
- Fitriani, S. (2021). Analisis Daya Pengecoh dalam Tes Pilihan Ganda. *Jurnal Pendidikan Dan Evaluasi*, 6(2). <https://doi.org/10.26740/jpe.v6n2.p120->

- Hamid, A. (2021). Evaluasi Pembelajaran Bahasa Arab di Sekolah Menengah. *Al-Ta'rib*, 10(1). <https://doi.org/10.23971/altarib.v10i1.3120>
- Hasanah, U. (2020). Evaluasi Pembelajaran Bahasa Arab Berbasis Tes Objektif. *Lisanul Arab*, 11(2). <https://doi.org/10.15294/la.v11i2.38562>
- Hilalludin, S. H., Tinggi, S., Tarbiyah, I., Yogyakarta, M., Alma, U., & Yogyakarta, A. (2025). *Kesetaraan Akses Pendidikan Teknologi Tantangan dan Peluang di Indonesia dan Dunia*. 1(1), 44–56.
- Kafi, F. A., & Ilma, Z. (2022). Analisis Soal Tematik Bahasa Arab Berbasis HOTS Pada Kelas XII MAN 3 Jember. *Al-Fusha: Arabic Language Education Journal*, 4(1), 1–9.
- Mardapi, D. (2021). Pengukuran dan Penilaian Pendidikan Abad 21. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 25(1). <https://doi.org/10.21831/pep.v25i1.38647>
- Nugroho, A. (2023). Item Analysis in Language Testing: Difficulty and Discrimination. *Journal of Language Assessment*, 4(1). <https://doi.org/10.1080/09588221.2023.2174501>
- Nurbayan, Y., Sanusi, A., Wulan, N. S., & bin Ismail, Z. (2024). Pedagogical Competence Design: Arabic Teaching as a Foreign Language based on a Praxeological Approach. *ALSUNIYAT: Jurnal Penelitian Bahasa, Sastra, Dan Budaya Arab*, 7(2), 266–290.
- Putra, D. (2023). Analisis Butir Soal Pilihan Ganda Menggunakan Excel. *Jurnal Teknologi Pendidikan*, 23(2). <https://doi.org/10.21009/jtp.v23i2.28617>
- Rahman, A. (2022). Measurement Quality in Islamic School Assessment. *Journal of Islamic Education Studies*, 5(1). <https://doi.org/10.14421/jies.2022.05103>
- Slamet, R., & Sri, W. (2022). Validitas dan Reliabilitas Terhadap Instrumen 'Kepuasan Kerja'. *Jurnal Manajemen & Bisnis*, 17(2), 51–58.
- Sudijono, A. (2021). Evaluasi Pembelajaran dalam Perspektif Pendidikan Modern. *Jurnal Pendidikan Dan Evaluasi*.

<https://doi.org/10.21009/jpe.2021.014>

Sugiyono, P. (2018). *Metode penelitian kuantitatif kualitatif dan R&D*. ALFABETA.

Suryani, L. (2021). Reliability Testing in Educational Measurement. *International Journal of Instruction*, 14(3). <https://doi.org/10.29333/iji.2021.14321a>

Warju, W., Ariyanto, S. R., Soeryanto, S., & Trisna, R. A. (2020). Analisis kualitas butir soal tipe HOTS pada kompetensi sistem REM di sekolah menengah kejuruan. *Jurnal Pendidikan Teknologi Dan Kejuruan*, 17(1), 95–104.

Widayanti, R. (2022). Analisis Validitas dan Reliabilitas Instrumen Evaluasi Bahasa Arab. *Arabiyat: Jurnal Pendidikan Bahasa Arab*, 9(1). <https://doi.org/10.15408/a.v9i1.22471>