



Meningkatkan Mutu Evaluasi Pembelajaran: Analisis Soal Pilihan Ganda Bahasa Arab

Andini Nuraisyah^{1*}

Email: andininuraisyah529@gmail.com¹

¹UIN Syarif Hidayatullah, Jakarta, Indonesia

DOI: <http://dx.doi.org/10.35931/am.v7i2.3353>

Article Info

Received: 20th June 2024

Revised: th July 2024

Accepted: 29th July 2024

Correspondence:

Phone: +6285695884836

Abstract: Learning evaluation is one part of the learning process. Through evaluation, a teacher can find out students' abilities and understanding of the material they have studied. Evaluation instruments require quality tests or questions in terms of validity, reliability, level of difficulty, distinguishing power and effectiveness of distractors. Question items prepared by a teacher to test students' abilities must be recommended to be tested for suitability based on these aspects so that the instruments that have been prepared can be used according to their function. This research was conducted using a descriptive methodology and quantitative research form. The research instrument used was 15 grade 5 PAS Arabic multiple choice questions and the research subjects were 5th grade students at SDIT Al Inayah for the 2023-2024 academic year who were tested by 37 participants. The results of the analysis of the questions in this study in the validity aspect were still not good because only 3 questions (20%) were valid. Then, in terms of the level of difficulty, it is sufficient but not yet balanced because no difficult questions were found. From the aspect of discriminating power, the questions studied were quite good and produced a discriminating power of each question which was 40% and good, reaching 33%. The effectiveness of the distractors in this study was also sufficient because 25% of the distractors were found to be functional, and 25% were rejected. Thus, from the results of this analysis, a teacher can improve and develop question items that are still not good in some of these aspects.

Keywords: *Question analysis, evaluation instrument, multiple choice*

PENDAHULUAN

Dalam suatu pembelajaran bahasa Arab, evaluasi merupakan salah satu alat untuk mengukur suatu pembelajaran dan kemampuan siswa dalam mencapai suatu tujuan (Wisman, 2021). Hasil dari evaluasi dapat dijadikan sebuah acuan yang digunakan dalam proses pembelajaran peserta didik (Sari, 2018). Dengan evaluasi proses pembelajaran dapat diperbaiki dan ditingkatkan kembali dari segi kekurangannya. Karena dalam proses pembelajaran bertujuan yang sama yaitu mencapai suatu tujuan atas keberhasilan dalam pembelajaran. Oleh karena itu, evaluasi menjadi barometer untuk mengukur keadaan atau situasi pendidikan yang ada di lapangan agar jika ditemukan kekurangan dapat segera diperbaiki dan ditingkatkan sehingga mencapai suatu tujuan (Matsna & Mahyudin, 2014). Dengan evaluasi juga dapat memberikan

masukan atau arahan dari kualitasnya suatu pembelajaran baik dari segi metodologi, pendekatan, dan strategi yang dipakai maupun dari segi kemampuan guru tersebut.

Salah satu cara untuk menguji kualitas pembelajaran adalah dengan menganalisis soal yang dijadikan sebagai instrumen pembelajaran pada aspek evaluasi. Analisis tes adalah salah satu penelitian yang berfokus pada aspek penelitian instrument evaluasi berupa tes atau soal yang dijadikan sebagai instrument evaluasi. Tujuan adanya analisis tes adalah agar tes yang dijadikan instrument evaluasi merupakan soal yang termasuk kriteria yang layak atau baik agar dapat memperoleh hasil belajar yang efektif dan efisien (Saputra et al., 2022). Hasil belajar siswa diagi menjadi 3 kategori meliputi hasil belajar kategori kognitif, afektif, dan psikomotorik. Dimana ranah tersebut saling berhubungan

dengan keterampilan (Sayuti, 2023). Adapun definisi dari instrumen evaluasi itu sendiri merupakan metode yang digunakan untuk mengukur, menilai, atau mengevaluasi pencapaian suatu tujuan pembelajaran, keterampilan, atau kemampuan seseorang dalam konteks Pendidikan atau pelatihan (Sugiyono, 2017). Kemudian instrumen dalam evaluasi pembelajaran meliputi evaluasi yang berbentuk tes dan non tes. Tes adalah instrumen yang terkumpul dari beberapa informasi berupa tugas atau soal-soal yang dapat dikerjakan peserta didik (Susilawati, 2018). Dalam evaluasi, instrumen tes yang disusun harus menyesuaikan tujuan atau indikator yang ingin dicapai. Oleh karena itu, terdapat beberapa kriteria dalam menyusun tes bahasa Arab yang berkualitas yaitu pada aspek validitas, reliabilitas, objektivitas, praktibilitas, dan diskriminatif (Matsna & Mahyudin, 2014). Adapun dalam sebuah tes yang berbentuk pilihan ganda, butir soalnya dapat dianalisis dan ditelaah dari aspek tingkat kesulitan, daya pembeda, keberfungsian pengecoh. Hal ini bertujuan agar soal yang disusun dan digunakan dalam mengevaluasi peserta didik dapat diujikan kualitasnya dan dapat diketahui keberfungsian soal tersebut dalam menguji pemahaman peserta didik terhadap materi yang telah dipelajari. Sistem penilaian dan instrument evaluasi juga harus saling melengkapi dengan memngkombinasikan secara sistematis agar dapat mencapai pemahaman total dan komprehensif tentang pemahaman siswa (Roman-Gonzalez et al., 2019).

Jenis tes yang seringkali dijadikan sebagai instrument evaluasi adalah jenis tes pilihan ganda. Dari jenis soal ulangan harian sampai soal seleksi masuk perguruan tinggi atau disebut SNBT juga menggunakan tes soal pilihan ganda. Soal pilihan ganda adalah bentuk pertanyaan dalam evaluasi atau tes yang mengajukan pertanyaan atau pernyataan dan memberikan dari 3 hingga 5 opsi jawaban dan diantaranya merupakan jawaban yang paling tepat, sedangkan opsi yang lainnya merupakan pengecoh yang dirancang untuk menilai pemahaman, kemampuan, atau pengetahuan peserta tes tentang materi yang diujikan (N, 2019). Selain itu, melalui analisis tes juga, seorang guru dapat mengetahui cara bagaimana merancang pertanyaan yang memungkinkan agar ketika mengukur tingkat pengetahuan siswa dapat diuji secara akurat (Cantabrana et al., 2019). Karena tidak sedikit di zaman teknologi dan AI saat ini seorang guru memanfaatkannya dalam ranah pendidikan baik yang bersifat positif maupun negative salah satunya adalah menyusun soal dengan bantuan AI (Dimitriadou & Lanitis, 2023). Dimana jika soal disusun oleh AI sepenuhnya, soal tidak akan mengukur kemampuan siswa dengan tepat. Hal itu disebabkan AI tidak mnegetahui perkembangan kemampuan dan kognitif siswa terkait materi yang sudah dipelajari sebelumnya. Oleh karena itu, dalam rangka meningkatkan mutu pembelajaran penulis akan melakukan peningkatan pembelajaran da Pendidikan

melalui pengujian instrument evaluasi berupa soal pilihan ganda. Karena dari hasil evaluasi menunjukkan hasil belajar siswa, dan sangat dibutuhkan suatu kualitas dalam sebuah instrument evaluasi (Zain et al., 2024).

Dari uraian di atas, bahwa perlunya dilakukan penelitian dalam menelaah butir soal agar mengetahui kualitas sebuah soal. Dalam penelitian ini, tes atau soal yang akan dianalisis adalah soal Penilaian Akhir Semester (Selanjutnya disebut PAS) bahasa Arab kelas 5. Adapun soal yang akan diteliti sebelumnya belum dilakukan dan diketahui kualitasnya baik dari aspek validitas, reliabilitas, daya pembeda, tingkat kesukaran, serta efektivitas pengecoh. Oleh karena itu, penelitian ini berfokus pada analisis kualitas butir soal berdasarkan dari segi aspek hasil uji validitas, realibilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Penelitian ini sebelumnya sudah dikaji oleh beberapa peneliti. *Pertama*, oleh Iswatul Khasanah dkk. yang berjudul "Analisis Soal Ulangan Harian Matematika Bentuk Pilihan Ganda" pada tahun 2023 (Khasanah et al., 2023). Penelitian tersebut meneliti soal ulangan harian mata Pelajaran Matematika yang berbentuk pilihan ganda. *Kedua*, penelitian Habibah Ahmad dan rekannnya yang berjudul "Analisis Soal Tes Kompetensi Kebahasaan pada Penilaian Tengah Semester Bahasa Arab Sekolah Menengah Pertama" (Ahmad & Auliya, 2024). Penelitian ini menganalisis tes kompetensi kebahasaaraban meliputi tes struktur kalimat, tes struktur morfologi tanpa teks, tes pemahaman mufrodzat dalam teks dan tanpa teks serta tes penyimpanan mufrodzat dalam teks yang disajikan dalam soal pilihan ganda dan esai. *Ketiga*, penelitian yang dilakukan oleh Nusrotus Sa'idah dkk. yang berjudul "Analisis Instrume Tes *Higher Order Thinking* Matematika SMP". Penelitian ini juga sama-sama bertujuan untuk mengetahui karakteristik dari instrument tes matematika berbenuk pilihan ganda melalui analisis aspek kesukaran dan daya pembeda soal. Adapun jenis soal yang diteliti adalah soal yang berkriteria HOTS (Sa'idah et al., 2019). Dari ketiga penelitian di atas, kebaruan diantara ketiganya adalah dari objek penelitiannya yaitu soal PAS Ganjil bentuk pilihan ganda Bahasa Arab pada tingkat lebih rendah yaitu tingkat sekolah dasar sebanyak 15 butir soal.

Penelitian ini dilakukan bertujuan untuk memberikan informasi tentang kualitas butir soal Penilaian Akhir Semester (PAS) Ganjil bahasa Arab dari beberapa aspek yaitu aspek validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Hal ini bertujuan agar guru peduli dalam menyusun soal yang berkualitas, dan agar dapat mengukur pemahaman peserta didik pada materi yang diujikan, sehingga guru dapat meningkatkan dan mengevaluasi kualitas soal sesuai dengan masalah yang dihadapi peserta didik.

METODE PENELITIAN

Penelitian ini dilakukan menggunakan pendekatan kuantitatif dengan jenis deskriptif. Hasil data yang diperoleh berupa angka-angka, kemudian dianalisis dan disesuaikan berdasarkan kaidah ilmiah yang objektif, konkrit, terukur, sistematis, dan rasional. Adapun pendekatan deskriptif berfungsi sebagai penjelasan terkait data hasil penelitian (Sugiyono, 2017). Subjek yang digunakan dalam penelitian ini adalah peserta didik kelas 5 SD SDIT Al Inayah Pamulang pada tahun pelajaran 2023-2024 sebanyak 37 siswa. Adapun instrumen penelitiannya yaitu tes objektif atau tes berbentuk pilihan ganda berjumlah 15 butir soal dari 20 soal meliputi 15 soal pilihan ganda dan 5 soal esai sesuai dengan ketentuan yang berlaku di sekolah tersebut. Tes objektif yang digunakan adalah Soal PAS Ganjil bahasa Arab kelas 5 SD yang membahas tentang kosa kata dan ungkapan-ungkapan mengenai anggota tubuh, kebun binatang, dan profesi. Selain itu, di dalam tes tersebut diujikan terkait kaidah nahwu tentang "an-Na'û" dan "mubtada khabar". Adapun tes objektif sendiri adalah tes yang disusun sekaligus alternatif jawabannya telah disediakan (Lumbantobing, n.d.). Tes objektif memiliki beberapa karakteristik antara lain disediakan alternatif atau pilihan jawaban, jawaban yang bukan jawabannya disebut dan harus berfungsi sebagai pengecoh, tipe soalnya adalah pertanyaan secara langsung atau kalimat tidak lengkap. Kemudian dalam tes pilihan ganda terdapat dua jenis yaitu berupa jawaban yang benar dan jawaban terbaik (Wartoni & Benyamin, 2020).

Pada penelitian ini akan dianalisis dan dicari hasil data terkait kualitas dan keberfungsian pada butir soal PAS Ganjil mata pelajaran bahasa Arab kelas 5 SD yaitu dengan cara menganalisis butir-butir soal dari aspek validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Adapun data didapatkan dan dikumpulkan dari lembar soal dan jawaban peserta didik dan kunci jawaban. Kemudian menganalisis data yang terkumpul menggunakan Microsoft excel dengan ketentuan formula khusus sehingga dapat menghasilkan validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan kualitas pengecoh butir soal. Teknik analisis data yang dilakukan adalah sebagai berikut:

1) Analisis Validitas dan reabilitas tes

Validitas tes yang akan diterapkan dalam penelitian ini adalah dari tes internal, yang didasarkan pada instrumen tes itu sendiri, atau dari validitas eksternal tes yang didasarkan pada hasil tes tes atau hasil tes lain di luar instrumen. Validitas diartikan sebagai validitas yang dinilai berdasarkan koefisien korelasi antara skor total instrumen dengan skor butirnya. Semakin tinggi koefisien korelasinya, maka semakin tinggi validitasnya. Dan hipotesis yang dihasilkan dapat digunakan untuk memperkirakan konsistensi dan hasil pengujian masing-masing instrumen dengan hasil pengujian instrumen itu

sendiri. Adapun teknik pengumpulan data untuk menganalisis validitas tes menggunakan rumus titik biner yang digunakan dalam analisis ini adalah sebagai berikut:

$$r_{pbi} = \frac{M_p - M_t}{SD_t} \sqrt{\frac{p}{q}}$$

Keterangan :

- r : koefisien korelasi point bisareal
- M_p : jumlah responden yang menjawab benar
- M_t : mean skor total yang berhasil dicapai oleh seluruh peserta
- SD_t : standar deviasi untuk semua item
- P : proporsi responden yang menjawab benar
- Q : proporsi responden yang menjawab salah

Adapun berdasarkan kategori reliabilitas uji analisisnya menggunakan rumus KR 20 yaitu salah satu rumus yang mirip dengan *Cronbach alpha*. Formula KR-20 adalah formula yang dapat digunakan untuk menguji butir dikotomi dengan memberikan skor yang benar 1 dan skor 0 jika salah (Retnawati, n.d.). Adapun rumus tersebut adalah:

$$r_{11} = \left(\frac{k}{k-1} \right) \left(\frac{S_{t^2} - \sum pq}{S_{t^2}} \right)$$

Keterangan :

- r₁₁ : reliabilitas instrumen
- k : jumlah item
- k-1 : jumlah item dikurangi 1
- S_t² : varians total
- P : proporsi siswa yang menjawab benar
- Q : proporsi siswa yang menjawab salah (1-p)

2) Analisis Tingkat Kesukaran

Dari hasil yang telah dihitung, tingkat kesukaran memiliki 3 kategori yaitu kategori mudah, sedang, dan sulit. Hasil penelitian dan penghitungan dari soal PAS bahasa Arab menggunakan rumus pada Microsoft excel. Adapun rumus atau formula yang digunakan adalah sebagai berikut (Matsna & Mahyudin, 2014):

$$P = \frac{R}{T}$$

Keterangan :

- P : Indeks kesukaran
- R : Banyaknya siswa yang menjawab soal itu dengan benar
- T : Jumlah seluruh peserta tes

3) Analisis Daya Pembeda

Daya pembeda soal adalah salah satu faktor yang bertujuan untuk membedakan antara siswa yang memiliki pemahaman tinggi (pandai) dengan siswa yang memiliki pemahaman rendah pada setiap butir soal (Solichin, 2017). Adapun simbol dari besaran daya pembeda atau disebut dengan indeks diskriminasi adalah *D*. Kemudian dalam menganalisis butir soal ini dapat menggunakan

rumus (Matsna & Mahyudin, 2014) atau formula pada Microsoft excel yaitu:

$$D = \frac{RH}{TH} - \frac{RL}{TL} = PH - PL$$

Keterangan :

D : Daya pembeda

RH : Banyaknya peserta yang menjawab benar dari kelompok atas

RL : Banyaknya peserta yang menjawab benar dari kelompok bawah

TH : Jumlah peserta tes kelompok atas

TL : Jumlah peserta tes kelompok bawah

PH : Proporsi peserta yang menjawab benar dari kelompok atas

PL : Proporsi peserta yang menjawab benar dari kelompok bawah

Sebelum menganalisis dengan rumus di atas, peneliti mengelompokkan peserta berdasarkan dengan skor yang didapat. Adapun kelompok atas adalah peserta tes yang skornya mencapai 61-100, dan kelompok bawah adalah peserta tes yang skornya hanya mencapai 0-60.

4) Analisis Efektivitas Pengecoh

Salah satu aspek yang menjadi dasar dalam menelaah soal adalah pengecoh pada butir soal yang memiliki peran penting dalam pilihan jawaban. Dengan analisis efektivitas pengecoh, dapat diketahui pilihan jawaban selain kunci jawaban itu yang berfungsi dan tidak. Adapun pengecoh yang berfungsi adalah pengecoh yang dipilih oleh 5% peserta tes (Alwi, 2012). Rumus digunakan untuk menghitung presentasi dari setiap pilihan jawaban pada setiap butir soal. Adapun rumus yang digunakan dalam menguji efektivitas pengecoh pada butir soal ini adalah:

$$\text{Efektivitas Pengecoh} = \frac{\text{Jumlah pemilih jawaban}}{\text{Jumlah semua peserta}} \times 100$$

HASIL DAN PEMBAHASAN

Permendikbud Nomor 23 Tahun 2016 menjelaskan persyaratan untukevaluasi pada Pasal 13 ayat 2 angka 4 yang menyatakan bahwa salah satu evaluasi utama pendidikan adalah mengkoordinasikan kegiatan peserta didik dengan menganalisis kualitas instrumen sebagai sarana penilaian hasil pembelajaran. Pasal 13 ayat 2 angka 4 yang menyatakan bahwa salah satu tujuan pokok pendidikan adalah mengkoordinasikan kegiatan peserta didik dengan menganalisis kualitas instrumen sebagai sarana penilaian hasil belajar. Tes yang baik adalah tes yang dapat memberikan perbedaan hasil antara peserta tes di kelas, di sekolah, bahkan secara nasional. Ada beberapa aspek yang harus diperhatikan agar hasil tes dapat diandalkan dan digunakan sebagai dasar penilaian

dan pembelajaran. Aspek tersebut adalah reliabilitas dan validitas yang berupa analisis teoritis atau disebut juga sebagai analisis kualitatif dan kuantitatif. Analisis kualitatif biasanya menganalisis sebelum soal diujicobakan (Suseno & Susongko, 2021). Soal yang digunakan adalah soal PAS Ganjil Bahasa Arab dengan rincian melalui gambar berikut:

Gambar 1.
Soal PAS Ganjil Bahasa Arab

I. Berilah tanda silang (X) pada salah satu huruf ا, ب, ج, د, sebagai jawaban yang kamu anggap benar!

١. عُنْدِي عَيْنَانِ وَأُذُنٌ وَفَمٌ

Arti dari kata yang digarisbawahi adalah

ا - mata
ب - hidung
ج - telinga
د - mulut

٢. هَذِهِ ...



ا - شَعْرٌ
ب - فَمٌ
ج - أُذُنٌ
د - يَدٌ

٣. هَذِهِ مَدْرَسَةٌ إِتْمَمَهَا حَسَنَةٌ

Gambar yang sesuai dengan arti dari kata yang digarisbawahi adalah






٤. + : مَنْ هَذِهِ؟
- : هَذِهِ ...

Kata yang sesuai dengan gambar untuk mengisi kalimat rumpang di atas adalah

٥. هَذَا فَيْسَاحٌ

ا - مَا بَيْنَهُ
ب - مَكَاتِبٌ
ج - مَسْرَعٌ
د - فَيْسَاحَةٌ

Arti dari kata yang digarisbawahi adalah

ا - buaya
ب - gajah
ج - monyet
د - singa

٦. + : مَاذَا فِي الْحَدِيثَةِ الْحَيَوَانِيَّةِ؟
- : فِي الْحَدِيثَةِ الْحَيَوَانِيَّةِ ...



ا - اَزْبَتٌ
ب - فَيْسَاحٌ
ج - اَسْتَدٌ
د - فَيْزٌ

٧. اذْهَبْ إِلَى حَدِيثَةِ الْحَيَوَانِيَّةِ

Arti dari kata yang digarisbawahi adalah

ا - museum
ب - taman bermain
ج - kebun binatang
د - margasatwa

٨. قَاطِنَةٌ : مَا وَطِنُهُ الْأُدُنُّ؟
أَحْمَدُ : وَطِنُهُ الْأُدُنُّ ...

Kalimat yang tepat untuk melengkapi percakapan di atas adalah

ا - لِلشَّعْرِ
ب - لِلرَّيْزِ
ج - لِلأَخْلِ
د - لِلنَّعَاسِ

٩. فِي اللِّغَةِ لِإِنْسَانٍ طَوِيلٌ

Arti dari kali kalimat di atas adalah

ا - di dalam mulut terdapat lidah yang pendek
ب - di dalam mulut terdapat lidah yang panjang
ج - di dalam mulut terdapat lidah yang bagus
د - di dalam mulut terdapat lidah yang tajam

١٠. هَذَا فَلَاحٌ إِتْمَمَ أَحْمَدُ

Arti dari kalimat yang di atas adalah

ا - ini pegawai namanya ahmad
ب - ini guru namanya ahmad
ج - ini petani namanya ahmad
د - ini polisi namanya ahmad

١١. هَذَا طَيِّبٌ ...
 أ- جَيِّلَةٌ ب- مَاهِرَةٌ ج- نَيْبَةٌ د- نَيْبٌ

١٢. فَاطِمَةُ : مَنْ هَذِهِ؟
 أَحْمَدُ : هَذِهِ أُمِّي إِسْمَاءُ خَدِيجَةٌ، هِيَ ...

Kata yang tepat untuk melengkapi kalimat di atas adalah
 أ- طَيِّبٌ ب- تَاجِرٌ ج- تَاجِرَةٌ د- طَيِّبَاتٌ

١٣. تِلْكَ الزَّوْجَةُ جَيِّلَةٌ
 Dari kalimat di atas yang termasuk na'at adalah ...
 أ- تِلْكَ ب- الزَّوْجَةُ ج- جَيِّلَةٌ د- جَيِّلَاتٌ

Kata yang tepat untuk melengkapi kalimat di atas adalah
 أ- إِسْمَةٌ ب- إِسْمَاءُ ج- إِسْمَاتٌ د- إِسْمٌ

١٤. هَذَا طَيِّبٌ ... أَحْمَدُ
 في خَدِيقَةِ الْحَيَوَانَاتِ حَيَوَانَاتٌ مُتَنَوِّعَةٌ

Dari kalimat di atas yang termasuk man'ut adalah
 أ- خَدِيقَةٌ ب- الْحَيَوَانَاتِ ج- حَيَوَانَاتٌ د- مُتَنَوِّعَةٌ

Adapun pada penelitian ini, akan dijabarkan hasil analisis tes dari analisa kuantitatif.

Analisis Validitas Tes

Adapun ringkasan hasil dari rumus korelasi tersebut untuk menguji validitas soal PAS bahasa Arab kelas 5 yang dipakai pada Microsoft excel sebagai berikut:

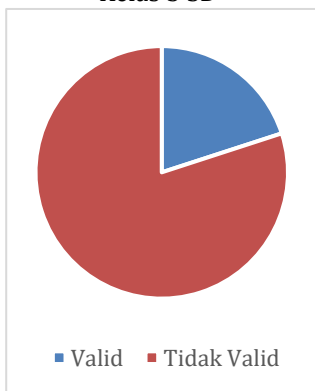
Tabel 1.

Hasil Analisa Validitas Soal PAS Bahasa Arab kelas 5

Jenis Soal	Jumlah	Presentase	Nomor Soal
Valid	3	20.0%	5, 6, 14
Tidak Valid	12	80.0%	1, 2, 3, 4, 7, 8, 9, 10, 11, 12, 13, 15

Gambar 2.

Diagram Hasil Analisis Validitas Soal PAS bahasa Arab Kelas 5 SD



Berdasarkan data yang dihasilkan dengan rumus *point biserial* pada tabel di atas ditemukan 3 butir soal (20%) yang berstatus valid korelasinya, kemudian 12 butir soal (80%) yang berstatus tidak valid. Semakin tinggi angka

yang dihasilkan dalam analisis validitas maka semakin tepat dan baik sebuah tes dalam mengukur kemampuan siswa (Antari et al., n.d.). Dengan hasil tersebut, menunjukkan bahwa angka tersebut merupakan hasil tingkat korelasi antara kedua deretan skor yang dikorelasikan. Adapun kriteria di atas berdasarkan dari urutan tabel *r product moment* yang indeks kevalidannya adalah 0,514 dari 15 soal tersebut. Adapun angka yang dibawah indeks tersebut maka status soalnya bernilai tidak valid. Dengan adanya hasil validitas dalam aspek kriteria ini, butir soal yang masih rendah dapat dievaluasi dan ditingkatkan kembali agar dapat menjadi butir soal dan tes yang lebih baik dan lebih berhubungan dengan kemampuan siswa.

Kemudian data yang berhasilkkan dikumpulkan pada aspek reliabilitas dihasilkan dengan angka 0,73. Dan apabila nilai yang dihasilkan lebih besar dari 0,60 maka instrumen dikatakan reliabel (Gunawan & Sunardi, 2016). Dengan demikian, soal PAS bahasa arab yang diuji sudah reliabel karena nilai yang dihasilkan lebih besar dari indeksnya yaitu $0,73 > 0,60$.

Analisis Tingkat Kesukaran

Hasil dari penghitungan atau telaah butir soal PAS bahasa Arab adalah sebagai berikut:

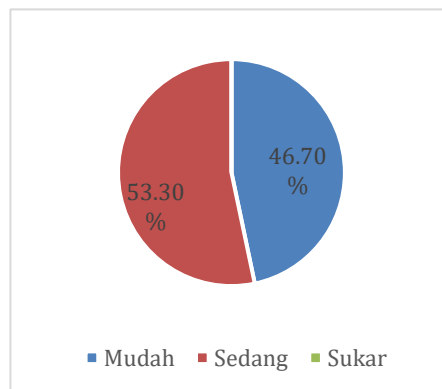
Tabel 2.

Hasil Analisis Tingkat Kesukaran Soal PAS Bahasa Arab Kelas 5 SD

Jenis Soal	Indeks Kesukaran	Jumlah	Presentase	Nomor Soal
Soal Mudah	0,71- 1,00	7	46.7%	1, 2, 3, 5, 6, 7, 9
Soal Sedang	0,31 – 0,70	8	53.3%	4, 8, 10, 11, 12, 13, 14, 15
Soal Sukar	0,00 – 0,30	0	0.0%	0

Gambar 3.

Diagram Hasil Analisis Tingkat Kesukaran Soal PAS



bahasa Arab Kelas 5 SD

Berdasarkan tabel 4 aspek tingkat kesukaran hanya ditemukan 2 kategori, yaitu kategori mudah dan sedang.

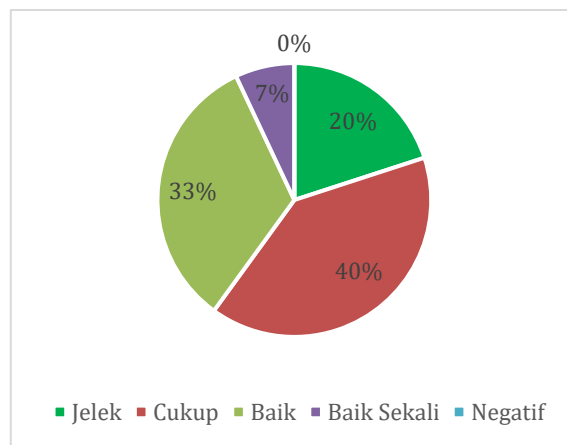
Kategori mudah dengan indeks kesukarannya dari 0,71-1,00 terdapat 7 soal (46,7 %) yaitu pada soal nomor 1, 2, 3, 5, 6, 7, 9. Dan kategori sedang dengan indeks kesukarannya dari 0,31-0,70 terdapat pada 8 butir soal (53,3%) yaitu pada nomor 4, 8, 10, 11, 12, 13, 14, 15. Kemudian, analisis pada aspek tingkat kesukaran tidak ditemukan butir soal yang dikategorikan sulit, karena hasil indeks kesukarannya belum memenuhi dari 0,00-0,30. Oleh karena itu, tingkat kesukaran soal tersebut belum seimbang, karena perbandingan soal yang baik diantara 3 kategori tersebut adalah 30% soal mudah, 40% soal sedang, dan 30% soal sukar (N, 2019). Karena soal adalah alat untuk menjadi acuan dari hasil belajar siswa, dan hal itulah yang menyebabkan soal yang disusun harus berkualitas yaitu dapat memberikan hasil pengukuran yang dijadikan acuan secara tepat dan cermat (Nugroho, 2016). Berikut contoh butir soal sesuai tingkat kesukarannya:

Tabel 3.
Contoh Butir Soal Sesuai Tingkat Kesukaran

Mudah
Sedang

Analisis Daya Pembeda

Berikut hasil yang diperoleh dari analisis daya pembeda pada soal PAS bahasa Arab tersebut:



Gambar 4.
Diagram Hasil Analisis Daya Pembeda Soal PAS Bahasa Arab Kelas 5 SD

Berdasarkan data di atas yaitu pada tabel 6, hasil dari telaah butir soal PAS dalam aspek daya pembeda didapatkan beberapa soal yang berkategori sesuai dengan tabel di atas dan tidak ada soal yang berkategori negatif atau soal yang dianggap tidak baik. Hasil data tersebut didapatkan setelah perhitungan dengan rumus di atas dan menyesuaikan dengan indeks diskriminasi yang telah ditentukan (Matsna & Mahyudin, 2014). Soal yang dikategorikan tidak baik terdapat 3 butir soal (20%) dengan indeks diskriminasi dari nilai 0,00-0,20. Kemudian, butir soal yang dikategorikan cukup terdapat 6 butir (40%) dengan indeks deskriminasi dari 0,21-0,40. Adapun soal yang dikategorikan baik dengan indeks diskriminasi dari 0,41-0,70 ditemukan pada 5 butir soal (33%). Dan hanya ditemukan 1 butir soal yang dikategorikan baik sekali (7%) dengan indeks diskriminasi 0,71-1,00. Dari hasil yang didapatkan, butir soal semakin baik jika semakin tinggi indeks yang dimiliki oleh butir soal tersebut. Hal ini mengakibatkan butir soal tersebut berfungsi dan dapat membedakan kemampuan peserta tes yang pandai dan kurang pandai. Sebaliknya, jika semakin rendah indeks yang didapatkan, maka soal tersebut akan sulit mendapatkan pembeda dari kemampuan peserta didik karena semakin rendahnya indeks pada butir soal tersebut (Nurhalimah et al., 2022). Pada penelitian ini, tidak ditemukan butir soal yang berkategori negatif. Dengan demikian, butir soal PAS bahasa Arab kelas 5 ini tidak ditemukan soal yang tidak sesuai yaitu ketika peserta didik dari kelompok pandai menjawab salah dan kelompok kurang pandai menjawab benar. Dan soal yang terindeks negatif dianjurkan untuk diubah atau dihilangkan. Adapun soal yang terindeks tidak

baik yaitu terdapat pada nomor soal 3, 7, dan 9 dianjurkan untuk dikembangkan jenis butir soalnya agar soal tersebut dapat menjadi butir soal yang lebih berkualitas. Adapun contoh butir soal sesuai dengan kategorinya sebagai berikut:

Daya Pembeda	Indeks Deskriminasi	Jumlah	Presentase	Nomor Soal
Tidak baik	0,00 – 0,20	3	20%	3, 7, 9
Cukup	0,21 – 0,40	6	40%	1, 2, 4, 11, 13, 15
Baik	0,41 – 0,70	5	33%	5, 6, 8, 10, 12
Baik Sekali	0,71 – 1,00	1	7%	14
Negatif	-	0	0%	0

Tabel 5.

Contoh Butir Soal Sesuai Kategori Daya Pembeda

Tidak baik
Cukup
Baik
Baik Sekali

Dari soal di atas yang mendapatkan kategori soal yang tidak baik ada pada soal nomor 3. Dimana pada soal tersebut, ditanyakan gambar yang sesuai dengan kata yang digarisbawahi. Soal tersebut diperlukan adanya perbaikan. Peneliti memberikan opsi perbaikan untuk soal tersebut dengan menambahkan kalimat penjelas pada soal

dan mengganti pengecoh dengan karakter Perempuan semua. Contoh:

هذه مُدْرِسَةٌ، اسْمُهَا حَسَنَةٌ. هِيَ تَعْمَلُ فِي الْمَدْرَسَةِ.

Gambar yang sesuai dengan arti kata yang digarisbawahi adalah...

- A. Gambar dokter wanita
- B. Gambar guru wanita
- C. Gambar petani wanita
- D. Gambar pegawai wanita

Adapun untuk soal yang berkategori cukup terdapat pada butir soal nomor 11. Dimana soal tersebut menanyakan kata yang cocok untuk melanjutkan kalimat tersebut. Menurut peneliti, soal tersebut tidak seimbang khususnya dalam segi pengecohnya. Peneliti memberikan opsi perbaikan dengan menambahkan keterangan penjelas pada kalimat tersebut juga menyajikan gambar yang sesuai. Contoh:

Disajikan gambar dokter (laki-laki)

اسمي مَحْمُودٌ، أَنَا ...

ب- طَبِيبٌ نَشِيطٌ

أ- طَبِيبَةٌ

نَشِيطَةٌ

ج- مُوَضَّفٌ نَشِيطٌ

د- مُوَضَّفَةٌ نَشِيطَةٌ

Dengan diberikannya opsi perbaikan di atas, peneliti berharap soal yang akan dijadikan instrument evaluasi dapat dikembangkan dan diperbaiki terus menerus sehingga dapat menghasilkan hasil evaluasi yang efisien dan relevan dengan kemampuan yang ada pada siswa.

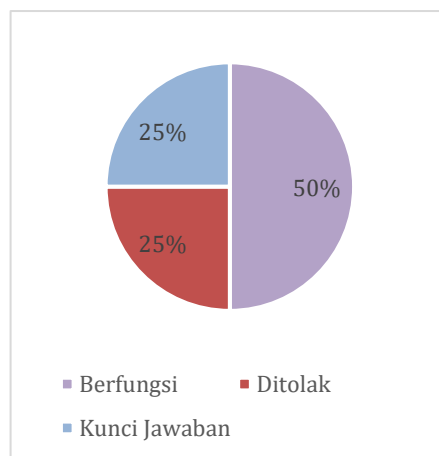
Analisis Efektivitas Pengecoh

Setelah dilakukan uji data, didapatkan hasil dari analisis butir soal PAS bahasa Arab ini dalam aspek efektivitas pengecoh adalah sebagai berikut:

Tabel 6.
Hasil Analisis Efektivitas Pengecoh Soal PAS Bahasa Arab Kelas 5

Soal	1	2	3
A	Ditolak (2,7%)	Ditolak (2,7%)	Ditolak (2,7%)
B	Berfungsi (16,2%)	Ditolak (2,7%)	Kunci Jawaban (91,9%)
C	Ditolak (2,7%)	Berfungsi (13,5%)	Ditolak (2,7%)
D	Kunci Jawaban (78,4%)	Kunci Jawaban (81,1%)	Ditolak (2,7%)
Soal	4	5	6
A	Berfungsi (18,9%)	Kunci Jawaban (75,7%)	Ditolak (0,0%)
B	Berfungsi (16,2%)	Berfungsi (10,8%)	Berfungsi (13,5%)
C	Kunci Jawaban (56,8%)	Ditolak (0,0%)	Kunci Jawaban (73%)
D	Ditolak (2,7%)	Berfungsi (13,5%)	Berfungsi (10,8%)
Soal	7	8	9
A	Ditolak (0,0%)	Kunci Jawaban (45,9%)	Berfungsi (24,3%)
B	Berfungsi (5,4%)	Berfungsi (18,9%)	Kunci Jawaban (70,3%)
C	Kunci Jawaban (91,9%)	Berfungsi (21,6%)	Ditolak (2,7%)
D	Ditolak (2,7%)	Berfungsi (13,5%)	Ditolak (2,7%)
Soal	10	11	12
A	Berfungsi (29,7%)	Berfungsi (27%)	Berfungsi (18,9%)
B	Berfungsi (10,8%)	Berfungsi (18,9%)	Berfungsi (18,9%)
C	Kunci Jawaban (56,8%)	Berfungsi (16,2%)	Kunci Jawaban (54,1%)
D	Ditolak (2,7%)	Kunci Jawaban (35,1%)	Berfungsi (5,4%)
Soal	13	14	15
A	Berfungsi (24,3%)	Kunci Jawaban (43,2%)	Berfungsi (21,6%)
B	Berfungsi (24,3%)	Berfungsi (32,4%)	Berfungsi (24,3%)
C	Kunci Jawaban (45,9%)	Berfungsi (16,2%)	Kunci Jawaban (37,8%)
D	Berfungsi (5,4%)	Berfungsi (8,1%)	Berfungsi (16,2%)

Gambar 5.
Diagram Hasil Analisis Efektivitas Pengecoh Soal PAS Bahasa Arab Kelas 5



Berdasarkan data pada tabel 8, hasil analisis menunjukkan efektivitas pengecoh pada butir soal PAS bahasa Arab kelas 5 SD dan ditemukan pilihan jawaban yang berfungsi dan tidak berfungsi atau ditolak. Pilihan jawaban yang berfungsi berjumlah 30 pilihan jawaban (50%) dengan indeks mencapai lebih dari 5% yaitu pada nomor 1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15. Sedangkan pilihan jawaban yang ditolak atau tidak berfungsi yang indeksnya masih kurang dari 5% berjumlah 15 pilihan jawaban (25%) yaitu pada nomor 1, 2, 4, 5, 6, 7, 9, 10. Berikut contoh butir soal sesuai kategori efektivitas pengecoh:

Tabel 7.
Contoh Butir Soal Sesuai Kategori Efektivitas Pengecoh

Contoh Soal semua pilihan berfungsi
<p>١٥. في حديقة الحيوانات حيوانات متنوعة</p> <p>Dari kalimat di atas yang termasuk man'ut adalah ...</p> <p>أ- حديقة ب- الحيوانات ج- حيوانات د- متنوعة</p>
Contoh Soal semua pilihan ditolak kecuali kunci jawaban
<p>٣. هذه مدرستة إسمها حسنة</p> <p>Gambar yang sesuai dengan arti dari kata yang digarisbawahi adalah</p> <p>أ- ب- ج- د- </p>

Oleh karena itu, dalam menyusun sebuah tes, keberfungsian pengecoh sangat berperan penting, karena hal ini dapat berhubungan dengan tingkat kesukaran dan

daya pembeda soal yang berarti jika tingkat kesukaran dan daya pembeda pada setiap butir soal rendah, hal itu dipicu karena efektivitas pengecohnya rendah. Sebaliknya, jika efektivitas pengecoh sudah baik, maka tingkat kesukaran dan daya pembeda menjadi baik (Akhmadi, 2021).

SIMPULAN

Dari hasil yang telah diuji dan dianalisis, soal PAS bahasa Arab kelas 5 SD dapat disimpulkan dari beberapa aspek, yaitu aspek validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan keberfungsian pengecoh. Adapun ditinjau berdasarkan hasil analisis validitas menggunakan rumus *point bisareal* masih sangat kurang, karena dari 15 butir soal yang masuk kriteria valid hanya 3 butir soal saja atau hanya 20%. Sedangkan ditemukan 12 butir soal yang tidak valid dan persentasenya mencapai 80%. Dengan demikian, butir soal yang tidak valid agar diperbaiki dan dikembangkan agar instrumen tes dapat lebih berkualitas. Kemudian berdasarkan hasil analisis dari aspek reliabilitas menggunakan rumus KR 20, soal PAS bahasa Arab kelas 5 telah reliabel karena nilai yang dihasilkan lebih besar daripada indeks reliabilitas tersebut yaitu $0,73 > 0,60$. Ditinjau dari segi tingkat kesukaran, soal tersebut tidak seimbang dan belum dikatakan soal yang baik karena butir soal hanya masuk 2 tingkat kriteria yaitu kriteria mudah dan sedang. Kriteria mudah sebanyak 7 soal dan persentase mencapai 46,7%. Sedangkan kriteria sedang sebanyak 8 soal dan mencapai 53,3%. Adapun pada aspek ini tidak ditemukan butir soal yang masuk kriteria sukar. Jika dilihat dari aspek daya pembeda dari setiap soal, dihasilkan butir-butir soal yang berstatus tidak baik sebanyak 3 soal (20%), cukup sebanyak 6 soal (40%), baik sebanyak 5 soal (33%), dan baik sekali hanya 1 soal (7%). Dengan demikian, soal PAS bahasa Arab kelas 5 dalam aspek daya pembeda sudah cukup baik, tetapi dapat diperbaiki dan dikembangkan kembali butir soal yang masih berstatus tidak baik agar dapat menjadi butir soal yang jauh lebih baik. Dari hasil analisis pada aspek efektivitas atau keberfungsian pengecoh sudah cukup baik. Karena opsi pilihan atau pengecoh yang berfungsi mencapai 25%, ditolak mencapai 25%, dan 50% adalah kunci jawaban setiap soal. Kendati demikian, ditemukan butir soal yang semua pengecohnya ditolak yaitu pada nomor 3. Oleh karena itu, butir soal yang pengecohnya belum berfungsi agar diperbaiki dan dikembangkan kembali.

DAFTAR PUSTAKA

- Ahmad, H., & Auliya, A. N. (2024). Analisis Soal Tes Kompetensi Kebahasaan Pada Penilaian Tengah Semester Bahasa Arab Sekolah Menengah Pertama. *Al Mi'yar*, 7(1), 157–176. <https://doi.org/10.35931/am.v7i1.3157>
- Akhmadi, M. N. (2021). Analisis Butir Soal Evaluasi Tema 1 Kelas 4 SDN Plumbungan. *Ed-Humanistics : Jurnal Ilmu Pendidikan*, 6(1), 799–806. <https://doi.org/10.33752/ed-humanistics.v6i1.1464>
- Alwi, I. (2012). Kriteria Empirik Dalam Menentukan Ukuran Sampel Pada Pengujian Hipotesis Statistika Dan Analisis Butir. *Jurnal Formatif*, 2(2), 140–148. <http://dx.doi.org/10.30998/formatif.v2i2.95>
- Antari, B., Susanta, A., & Siagian T.A. (n.d.). Analisis Soal Penilaian Akhir Semester Matematika SMA Negeri 8 Kota Bengkulu Tahun Ajaran 2020/2021. *Jurnal Edukasi Matematika Dan Sains*, 9(2), 299–312.
- Cantabrana, J. L. L., Rodriguez, M. U., & Cervera, M. G. (2019). Assessing Teacher Digital Competence: The Construction of an Instrument for Measuring the Knowledge of Pre-Service Teachers. *Journal of New Approaches in Educational Research*, 8(1), 73–78. <https://doi.org/10.7821/naer.2019.1.370>
- Dimitriadou, E., & Lanitis, A. (2023). A critical evaluation, challenges, and future perspectives of using artificial intelligence and emerging technologies in smart classrooms. *Smart Learning Environments*, 10(12), 1–26. <https://doi.org/10.1186/s40561-023-00231-3>
- Gunawan, A. A., & Sunardi, HP. (2016). Pengaruh Kompensasi Dan Disiplin Kerja Terhadap Kinerja Karyawan Pada PT Gesit Nusa Tangguh. *Jurnal Ilmiah Manajemen Bisnis*, 16(1), 1–12.
- Khasanah, I., Fuady, A., & Sunismi. (2023). Analisis Soal Ulangan Harian Matematika Bentuk Pilihan Ganda. *Mathema Journal*, 5(2), 110–125.
- Lumbantobing, L. R. (n.d.). *Kelebihan dan Keterbatasan Dari Bentuk-Bentuk Penilaian Tes Tertulis*. Pusdiklat Keuangan Umum.
- Matsna, Moh., & Mahyudin, E. (2014). *Pengembangan Evaluasi dan Tes Bahasa Arab*. Alkitabah.
- N, S. (2019). *Penilaian Hasil Proses Belajar Mengajar*. PT Remaja Rosdakarya.
- Nugroho, M. A. (2016). Studi Analisis Butir Soal Latihan Buku Ajar Bahasa Arab Madrasah Tsanawiyah. *Arabia*, 8(2), 215–238.
- Nurhalimah, S., Hidayati, Y., Rosidi, I., & Hadi, W. P. (2022). Hubungan antara Validitas Item Dengan Daya Pembeda Dan Tingkat Kesukaran Soal Pilihan Ganda PAS. *Jurnal Natural Science Education Research*, 4(3), 249–257.

- Retnawati, H. (n.d.). *Reliabilitas Instrumen Penelitian*. Retrieved January 30, 2023, from <https://staffnew.uny.ac.id/upload/132255129/pengabdian/8%20Reliabilitas3%20alhamdulillah.pdf>
- Roman-Gonzalez, M., Moreno-Leon, J., & Robles, G. (2019). *Combining Assessment Tools for a Comprehensive Evaluation of Computational Thinking Interventions*. Springer Open.
- Sa'idah, N., Yulistiana, H. D., & Megawati, E. (2019). Analisis Instrumen Tes Higher Order Thinking Matematika SMP. *Jurnal Pendidikan Matematika*, 13(1), 41–54.
- Saputra, H. D., Purwanto, W., Setiawan, D., Fernandez, D., & Putra, R. (2022). Hasil Belajar Mahasiswa: Analisis Butir Soal Tes. *Edukasi: Jurnal Pendidikan*, 20(1), 15–27.
- Sari, L. M. (2018). Evaluasi Dalam Pendidikan Islam. *Al-Tadzkiyyah : Jurnal Pendidikan Islam*, 9(2), 211–231.
- Sayuti, M. (2023). Pengembangan Penilaian HOTS dalam Pembelajaran Maharah Istima' Pada Buku Bahasa Arab MI Kelas VI Kementerian Agama RI. *Al Mi'yar*, 6(2), 663–680. <https://doi.org/10.35931/am.v6i2.2646>
- Solichin, M. (2017). Analisis Daya Beda Soal, Taraf Kesukaran, Validitas Butir Tes, Interpretasi Hasil Tes dan Validitas Ramalan Dalam Evaluasi Pendidikan. *Dirasat : Jurnal Manajemen Dan Pendidikan Islam*, 2(2), 192–213. <https://doi.org/10.26594/dirasat.v2i2.879>
- Sugiyono. (2017). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Alfabeta.
- Suseno, E., & Susongko, P. (2021). *Mengukur Validitas Tes* (1st ed.). Pameran Edukreatif.
- Susilawati, D. (2018). *Tes dan Pengukuran*. UPI Sumedang Press.
- Wartoni, & Benyamin, P. I. (2020). Strategi Pengembangan Tes Objektif (Pilihan Ganda). *Diegesis: Jurnal Teologi*, 5(1), 1–8.
- Wisman, Y. (2021). Penerapan Konsep Instrumen Evaluasi Hasil Belajar. *Jurnal Ilmiah Kanderang Tingang*, 12(1), 1–9.
- Zain, A. H. N., Fauzy, A. N., & Hermawan, A. (2024). Analisis Keterampilan Bersastra dalam Pembelajaran Bahasa Arab di Madrasah Tsanawiyah. *Al Mi'yar*, 7(1), 177–198. <https://doi.org/10.35931/am.v7i1.3195>