



Research Article

Analisis Kualitas Butir Soal Penilaian Akhir Semester (PAS) Al-Qur'an Hadis Kelas VIII MTS

Indah Rahmawati¹, Dina Hermina²

1. Universitas Islam Negeri Antasari Banjarmasin, Indonesia; indahrahmaawati44@gmail.com
2. Universitas Islam Negeri Antasari Banjarmasin, Indonesia; dinahermina@uin-antasari.ac.id

Copyright © 2026 by Authors, Published by **Kasyafa: Jurnal Pendidikan Agama Islam**. This is an open access article under the CC BY License <https://creativecommons.org/licenses/by/4.0/>

Received : March 18, 2026
Accepted : May 17, 2026

Revised : April 24, 2026
Available online : June 12, 2026

How to Cite: Indah Rahmawati, & Dina Hermina. (2026). Analysis of the Quality of Final Semester Assessment (PAS) Questions for Al-Qur'an Hadith Class VIII MTS. *Kasyafa: Jurnal Pendidikan Agama Islam*, 3(2), 232-247. <https://doi.org/10.61166/kasyafa.v3i2.167>

Analysis of the Quality of Final Semester Assessment (PAS) Questions for Al-Qur'an Hadith Class VIII MTS

Abstract. This study aims to analyze the quality of the items in the Final Semester Examination (Penilaian Akhir Semester/PAS) of Qur'an Hadith for Grade VIII MTs in terms of item validity, reliability, difficulty level, discrimination index, and distractor effectiveness. This research employed a quantitative descriptive approach. The instrument analyzed consisted of 40 multiple-choice items with 21 student respondents. Data analysis was conducted using item analysis techniques supported by Microsoft Excel. The results show that 25 items were classified as valid, while 15 items were invalid. The instrument reliability was categorized as very high, with a KR-20 coefficient of 0.932. Most items were in the moderate difficulty category (85%), and the discrimination index was predominantly in the good to very good category. In addition, distractor effectiveness analysis indicated that most distractors functioned properly, and no items were found in which all distractors were non-functional. Overall, the instrument demonstrates fairly good quality, although several items still require revision. Furthermore, 14 items met all quality criteria and can serve as references for future assessments.

Keywords: Item Analysis, Validity, Reliability, Difficulty Level, Discrimination Index, Distractor Effectiveness

Abstrak. Penelitian ini bertujuan untuk menganalisis kualitas butir soal Penilaian Akhir Semester (PAS) mata pelajaran Al-Qur'an Hadis kelas VIII MTs ditinjau dari aspek validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Penelitian ini menggunakan pendekatan kuantitatif dengan jenis deskriptif. Instrumen yang dianalisis terdiri dari 40 butir soal pilihan ganda dengan responden sebanyak 21 siswa. Teknik analisis data dilakukan melalui analisis butir soal menggunakan bantuan Microsoft Excel. Hasil penelitian menunjukkan bahwa 25 butir soal dinyatakan valid dan 15 tidak valid. Reliabilitas instrumen berada pada kategori sangat tinggi dengan nilai KR-20 sebesar 0,932. Sebagian besar butir soal berada pada tingkat kesukaran sedang (85%), dengan daya pembeda didominasi kategori baik hingga sangat baik. Selain itu, efektivitas pengecoh menunjukkan bahwa sebagian besar pengecoh berfungsi dengan baik dan tidak ditemukan butir soal yang seluruh pengecohnya tidak berfungsi. Secara keseluruhan, instrumen memiliki kualitas yang cukup baik, meskipun masih terdapat beberapa butir soal yang perlu direvisi. Sebanyak 14 butir soal memenuhi seluruh indikator kualitas dan dapat menjadi acuan untuk evaluasi berikutnya.

Kata Kunci: Analisis Butir Soal, Validitas, Reliabilitas, Tingkat Kesukaran, Daya Pembeda, Efektivitas Pengecoh

PENDAHULUAN

Evaluasi merupakan bagian integral dalam proses pembelajaran, termasuk dalam Pendidikan Agama Islam (PAI). Melalui evaluasi, pendidik dapat mengetahui sejauh mana tujuan pembelajaran telah tercapai serta mengidentifikasi keberhasilan maupun kelemahan peserta didik dalam memahami materi yang diajarkan (Hazirman dkk., 2026, hlm. 971). Evaluasi yang baik berfungsi sebagai alat ukur dan sebagai dasar dalam pengambilan keputusan untuk perbaikan pembelajaran selanjutnya.

Dalam kegiatan evaluasi diperlukan alat atau teknik penilaian agar pelaksanaannya lebih terarah dan sistematis. Alat evaluasi dalam pendidikan yang digunakan untuk mengumpulkan data dapat berupa tes maupun non tes. Adapun definisi tes adalah suatu cara penilaian yang berbentuk tugas atau serangkaian tugas yang harus dikerjakan oleh siswa, baik secara individu maupun kelompok, sehingga menghasilkan nilai yang menggambarkan perilaku atau prestasi peserta didik (Magdalena dkk., 2021, hlm. 200).

Dalam praktiknya, alat ukur berupa tes tersebut juga perlu dievaluasi kembali untuk mengetahui kualitasnya sebagai instrumen pengukuran. Analisis terhadap kualitas butir soal diperlukan untuk memastikan bahwa setiap item benar-benar berfungsi sebagaimana mestinya. Hasil analisis ini diharapkan menjadi umpan balik bagi pendidik untuk melakukan perbaikan dan penyempurnaan butir-butir soal sehingga pada masa mendatang instrumen yang digunakan memiliki kualitas yang lebih baik. Kegiatan ini dikenal dengan istilah analisis butir soal (*item analysis*) (Sudijono, 2009, hlm. 369–370).

Analisis butir soal umumnya mencakup lima aspek utama, yaitu validitas item, reliabilitas tes, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Validitas item menunjukkan ketepatan suatu butir soal dalam mengukur kompetensi yang

seharusnya diukur. Reliabilitas menunjukkan tingkat konsistensi hasil pengukuran. Tingkat kesukaran menggambarkan proporsi peserta didik yang dapat menjawab benar suatu butir soal. Daya pembeda menunjukkan kemampuan soal dalam membedakan peserta didik yang berkemampuan tinggi dan rendah. Sementara itu, efektivitas pengecoh menunjukkan sejauh mana opsi jawaban yang salah mampu berfungsi sebagai distraktor yang efektif (Rahayu & Djazari, 2016, hlm. 86).

Dalam perspektif Islam, proses analisis kembali terhadap instrumen evaluasi dapat dikaitkan dengan konsep *tabayyun*. *Tabayyun* merupakan prinsip verifikasi dan kehati-hatian dalam menerima serta menetapkan suatu informasi agar tidak menimbulkan kesalahan dalam pengambilan keputusan. Allah Swt. berfirman dalam QS. Al-Hujurât ayat 6:

يَا أَيُّهَا الَّذِينَ آمَنُوا إِن جَاءَكُمْ فَاسِقٌ بِنَبَأٍ فَتَبَيَّنُوا أَن تُصِيبُوا قَوْمًا بِجَهَالَةٍ فَتُصْبِحُوا عَلَىٰ مَا فَعَلْتُمْ
نَدِيمِينَ ۖ

Artinya: “Wahai orang-orang yang beriman, jika seorang fasik datang kepadamu membawa berita penting, maka telitilah kebenarannya agar kamu tidak mencelakakan suatu kaum karena ketidaktahuan(-mu) yang berakibat kamu menyesali perbuatanmu itu.” (Q.S Al-Hujurât/49: 6)

Dalam Tafsir Imam al-Alusi, Kata *tabayyun* pada ayat ini secara bahasa bermakna mencari kejelasan dan memastikan kebenaran suatu informasi. Dalam beberapa penafsiran, *tabayyun* memiliki kedekatan makna dengan *tatsabbut* (ketelitian), sebagaimana dalam qira'ah lain yaitu “*fatasabbatû*” yang bermakna sikap berhati-hati, tidak tergesa-gesa, dan menunggu hingga keadaan menjadi jelas (Al-Alusi, 1994, hlm. 297). Makna ini menekankan pentingnya proses verifikasi sebelum mengambil keputusan berdasarkan suatu informasi. Prinsip *tabayyun* ini selaras dengan konsep verifikasi dalam evaluasi pembelajaran, yaitu memastikan bahwa instrumen tes benar-benar sahih dan layak digunakan sebelum dijadikan dasar dalam penilaian peserta didik. Dengan demikian, analisis kualitas butir soal merupakan bentuk implementasi nilai kehati-hatian dan verifikasi dalam praktik evaluasi Pendidikan Agama Islam.

Berdasarkan hasil wawancara awal, ditemukan bahwa di salah satu MTs swasta di Kabupaten Barito Kuala yang menjadi lokasi penelitian ini, guru PAI belum pernah melakukan analisis butir soal terhadap instrumen evaluasi yang digunakan, khususnya pada Penilaian Akhir Semester (PAS).

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk menganalisis kualitas butir soal Penilaian Akhir Semester (PAS) mata pelajaran Al-Qur'an Hadis kelas VIII MTs ditinjau dari aspek validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan jenis penelitian deskriptif. Penelitian ini bertujuan untuk menganalisis kualitas butir soal Penilaian Akhir Semester (PAS) mata pelajaran Al-Qur'an Hadis.

Penelitian ini dilaksanakan di salah satu Madrasah Tsanawiyah (MTs) swasta di Kabupaten Barito Kuala. Subjek penelitian adalah peserta didik kelas VIII yang berjumlah 21 siswa.

Instrumen penelitian yang dianalisis adalah soal Penilaian Akhir Semester (PAS) mata pelajaran Al-Qur'an Hadis semester genap tahun ajaran 2025/2026. Soal berbentuk pilihan ganda sebanyak 40 butir. Selain itu, terdapat pula soal esai berbasis ayat Al-Qur'an yang meminta peserta didik untuk mengidentifikasi atau menandai hukum tajwid pada potongan ayat. Namun, analisis dalam penelitian ini difokuskan pada soal pilihan ganda.

Data diperoleh melalui dokumentasi hasil jawaban peserta didik terhadap soal PAS. Teknik analisis data menggunakan analisis butir soal yang meliputi lima aspek, yaitu validitas butir soal, reliabilitas tes, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Perhitungan setiap indikator dilakukan dengan bantuan Microsoft Excel untuk mempermudah proses pengolahan data dan perhitungan statistik.

HASIL DAN PEMBAHASAN

Validitas Item

Validitas item adalah tingkat ketepatan suatu butir soal dalam mengukur apa yang seharusnya diukur sebagai bagian dari suatu tes secara keseluruhan. Setiap tes hasil belajar pada hakikatnya merupakan kumpulan dari sejumlah butir soal yang disusun untuk mengungkap capaian belajar peserta didik setelah mengikuti proses pembelajaran. Oleh karena itu, setiap butir soal merupakan bagian yang tidak terpisahkan dari tes sebagai suatu totalitas (Sudijono, 2009, hlm. 182–183).

Menurut Creswell dalam Maryana, Validitas bertujuan untuk meningkatkan keakuratan dan kredibilitas hasil penelitian. Dengan memastikan bahwa alat ukur memiliki validitas yang tinggi, peneliti dapat yakin bahwa hasil yang diperoleh benar-benar mencerminkan variabel yang diteliti. Adapun peran validitas adalah meningkatkan integritas hasil, memastikan interpretasi data yang tepat, serta berperan penting untuk pengambilan Keputusan (Maryana & Alawiyah, 2025, hlm. 875).

Secara statistik, suatu butir item dinyatakan memiliki validitas yang tinggi apabila skor pada butir tersebut berkorelasi positif dan signifikan dengan skor total tes. Dalam hal ini, skor item berperan sebagai variabel bebas (independen), sedangkan skor total sebagai variabel terikat (dependen). Oleh karena itu, pengujian validitas item dilakukan menggunakan teknik korelasi (Sudijono, 2009, hlm. 184).

Adapun rumus yang bisa digunakan adalah dengan korelasi produk moment dengan angka kasar: (Riinawati, 2021, hlm. 128)

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

Keterangan:

r_{xy} = Koefisien korelasi

X = skor masing-masing item pertanyaan

Y = skor total

Nilai korelasi untuk masing-masing item pertanyaan tersebut selanjutnya dibandingkan dengan nilai kritis yang diperoleh dari r_{tabel} yaitu dengan menentukan nilai $r_{\alpha, n-2}$ yang nilainya dapat dilihat dari r_{tabel} . Kriteria pengujian validitas yaitu instrumen dikatakan valid jika nilai korelasi $r_{hitung} > r_{tabel}$ dengan tingkat signifikansi (α) 5% atau 0.05 (Azizah & Chalimatusadiah, 2025, hlm. 6639).

Tabel 1. Hasil tes

No.	NAMA SISWA	SOAL NO-																																								TOTAL		
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40			
1	AM	1	1	0	0	1	0	0	0	0	0	0	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	14	
2	AL	1	1	0	1	0	0	1	1	1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	31		
3	AZ	1	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	1	0	1	0	1	1	1	1	1	1	0	0	1	0	0	1	0	1	1	0	1	1	0	1	1	18	
4	AR	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0	0	8		
5	G	1	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	9		
6	H	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	1	0	0	0	0	1	0	0	0	0	0	7		
7	MAL	1	0	0	1	1	1	1	0	0	0	1	0	0	1	1	0	1	0	1	0	0	1	0	0	0	1	1	0	1	1	0	1	0	0	1	0	0	1	1	0	0	18	
8	MAZ	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	1	0	1	0	1	0	0	0	1	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	9		
9	MAN	1	1	0	0	0	0	0	1	1	0	0	1	1	1	1	0	1	0	1	0	1	1	0	1	1	0	0	0	0	0	1	0	1	0	0	0	1	0	0	1	1	1	35
10	MAH	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18	
11	MH	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	1	1	0	0	1	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	9	
12	MI	1	0	1	0	1	0	0	0	1	0	0	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	1	0	1	0	1	0	13		
13	MIH	1	1	1	0	1	1	1	0	1	0	1	0	0	1	1	0	0	0	1	0	0	1	0	1	1	1	1	0	0	1	0	1	1	0	1	0	1	0	1	1	22		
14	MIR	0	1	0	0	0	0	1	0	0	1	0	0	1	1	1	1	0	0	0	1	0	0	0	1	0	1	0	1	1	0	0	0	0	1	0	0	0	1	0	0	14		
15	MSR	1	1	0	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	1	1	0	0	0	0	0	0	0	1	0	0	1	0	0	12		
16	N	1	1	1	1	1	0	1	1	1	1	0	1	0	1	1	0	1	0	0	0	1	0	1	0	1	0	0	0	1	1	1	1	1	0	0	1	0	1	1	25			
17	SI	1	0	1	0	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	0	1	1	1	1	1	1	0	1	32		
18	SP	1	1	1	1	0	1	0	1	1	1	1	1	1	1	0	0	1	1	0	0	0	0	1	0	1	0	1	1	0	0	1	1	1	1	0	0	1	1	1	1	27		
19	SA	1	1	1	0	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	35		
20	ZR	1	1	1	0	1	0	1	1	0	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	32		
21	ZA	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	35		

Tabel 2. Hasil Validitas Item

[illegible]

Sementara itu, terdapat 15 butir soal yang dinyatakan tidak valid karena memiliki nilai r_{hitung} lebih kecil daripada r_{tabel} , yaitu butir soal nomor 2, 4, 5, 6, 13, 14, 16, 17, 19, 20, 22, 24, 26, 31, dan 38.

Kasyafa: Jurnal Pendidikan Agama Islam
<https://kasyafa.hellowpustaka.id>

Reliabilitas

Reliabilitas merupakan tingkat konsistensi suatu instrumen dalam memberikan hasil pengukuran yang dapat dipercaya, relatif stabil, dan konsisten. Lebih lanjut, menurut Allen dan Yen dalam Khumaedi, reliabilitas dapat diukur melalui beberapa metode, yaitu test-retest, bentuk paralel (*parallel forms*), dan konsistensi internal (*internal consistency*). Dari ketiga metode tersebut, konsistensi internal merupakan pendekatan yang paling banyak digunakan dalam analisis butir soal (Khumaedi, 2012, hlm. 26).

Dalam penelitian ini, reliabilitas dihitung menggunakan rumus Kuder-Richardson 20 (KR-20) yang termasuk dalam teknik konsistensi internal, karena digunakan untuk instrumen berbentuk pilihan ganda dengan skor dikotomi (0-1).

Adapun rumus KR-20 (Kuder Richardson 20) adalah sebagai berikut:

$$r_{11} = \frac{n}{n-1} \left(1 - \frac{\sum pq}{S^2} \right)$$

Keterangan:

r_{11} = reliabilitas tes

n = jumlah soal 40

p = proporsi siswa yang menjawab benar = $\frac{\text{jumlah benar}}{\text{jumlah siswa (21 Siswa)}}$

q = proporsi siswa yang menjawab salah = $1-p$

$\sum pq$ = jumlah hasil perkalian = $p \times q$

S^2 = varians total skor

$$S^2 = \frac{\sum Y^2 - \frac{(\sum Y)^2}{N}}{N - 1}$$

Tabel 3. Perhitungan Reliabilitas

No Soal	Jumlah Benar	$p = \frac{\text{Jumlah benar}}{21}$	$q = 1-p$	pq	$\sum pq$	S^2
1	16	0,762	0,238	0,181	9,315	$S^2 = \frac{10555 - \frac{(423)^2}{21}}{21 - 1} = 101,729$
2	14	0,667	0,333	0,222		
3	9	0,429	0,571	0,245		
4	8	0,381	0,619	0,236		
5	12	0,571	0,429	0,245		
6	8	0,381	0,619	0,236		
7	13	0,619	0,381	0,236		
8	12	0,571	0,429	0,245		
9	10	0,476	0,524	0,249		
10	9	0,429	0,571	0,245		
11	10	0,476	0,524	0,249		
12	8	0,381	0,619	0,236		
13	6	0,286	0,714	0,204		
14	17	0,810	0,190	0,154		
15	11	0,524	0,476	0,249		
16	4	0,190	0,810	0,154		

17	15	0,714	0,286	0,204
18	9	0,429	0,571	0,245
19	12	0,571	0,429	0,245
20	5	0,238	0,762	0,181
21	9	0,429	0,571	0,245
22	14	0,667	0,333	0,222
23	11	0,524	0,476	0,249
24	12	0,571	0,429	0,245
25	12	0,571	0,429	0,245
26	7	0,333	0,667	0,222
27	9	0,429	0,571	0,245
28	10	0,476	0,524	0,249
29	12	0,571	0,429	0,245
30	9	0,429	0,571	0,245
31	10	0,476	0,524	0,249
32	12	0,571	0,429	0,245
33	12	0,571	0,429	0,245
34	11	0,524	0,476	0,249
35	9	0,429	0,571	0,245
36	9	0,429	0,571	0,245
37	12	0,571	0,429	0,245
38	9	0,429	0,571	0,245
39	13	0,619	0,381	0,236
40	13	0,619	0,381	0,236

Berdasarkan hasil perhitungan pada tabel di atas, nilai-nilai tersebut dimasukkan ke dalam rumus Kuder-Richardson 20 (KR-20) sebagai berikut:

$$r_{11} = \frac{40}{40 - 1} \left(1 - \frac{9,315}{101,729} \right) = \frac{40}{39} (1 - 0,092) = \frac{40}{39} (0,908) = 0,932$$

Dari hasil perhitungan diperoleh nilai reliabilitas sebesar 0,932. Selanjutnya, nilai tersebut dikonsultasikan dengan kriteria interpretasi reliabilitas sebagai berikut: (Loka dkk., 2024, hlm. 2049)

Tabel 4. Kriteria Interpretasi Reliabilitas

Interval Koefisien Reliabilitas	Kriteria
0,80 – 1,00	Sangat tinggi
0,60 – 0,79	Tinggi
0,40 – 0,59	Cukup
0,20 – 0,39	Rendah
0,00 – 0,19	Sangat rendah

Berdasarkan kriteria tersebut, nilai reliabilitas sebesar 0,932 berada pada rentang 0,80–1,00 sehingga termasuk dalam kategori sangat tinggi. Maka $r_{11} = 0,932$ termasuk dalam kategori sangat tinggi reliabel.

Jika kita melihat pada analisis validitas item sebelumnya, ditemukan 25 butir soal yang valid dan 15 butir soal yang tidak valid. Sementara itu, hasil uji reliabilitas menunjukkan bahwa instrumen memiliki nilai KR-20 sebesar 0,932 yang termasuk dalam kategori sangat tinggi.

Menurut Scarvia B. Anderson dalam Suharsimi Arikunto, sebuah tes dapat bersifat reliabel tetapi belum tentu valid, namun tes yang valid pada umumnya cenderung reliabel. Selain itu, salah satu faktor yang memengaruhi reliabilitas tes adalah panjang tes, di mana tes yang terdiri dari banyak butir soal cenderung memiliki reliabilitas yang lebih tinggi dibandingkan dengan tes yang hanya terdiri dari sedikit butir soal (Arikunto, 2009, hlm. 87). Dengan demikian, meskipun terdapat beberapa butir soal yang tidak valid, tingginya nilai reliabilitas pada tes ini kemungkinan dapat dipengaruhi oleh jumlah butir soal yang cukup banyak, sehingga secara keseluruhan reliabilitas tes tetap tergolong sangat tinggi.

Tingkat kesukaran

Tingkat kesukaran merupakan ukuran yang menunjukkan derajat mudah atau sulitnya suatu butir soal bagi peserta didik. Soal yang dapat dijawab benar oleh sebagian besar peserta didik termasuk kategori mudah, sedangkan soal yang hanya dapat dijawab oleh sedikit atau tidak ada peserta didik termasuk kategori sulit (Sabilla dkk., 2026, hlm. 858). Adapun rumus Tingkat kesukaran adalah sebagai berikut:

$$P = \frac{B}{N}$$

Keterangan:

P = indeks kesukaran

B = jumlah siswa yang menjawab benar

N = jumlah seluruh siswa

Kemudian kriteria interpretasi tingkat kesukaran menurut Arikunto adalah sebagai berikut: (Arikunto, 2009, hlm. 210)

Tabel 5. Kriteria Tingkat Kesukaran

Nilai P	Kriteria Kesukaran
0,00 – 0,30	Sukar
0,31 – 0,70	Sedang
0,71 – 1,00	Mudah

Maka hasil penghitungan Tingkat kesukaran untuk soal PAS dalam penelitian ini adalah sebagaimana tabel di bawah ini:

Tabel 6. Hasil Perhitungan Tingkat Kesukaran

No Soal	Jumlah Benar	P = Jumlah benar/21	Kriteria
1	16	0,76	Mudah
2	14	0,67	Sedang
3	9	0,43	Sedang
4	8	0,38	Sedang
5	12	0,57	Sedang
6	8	0,38	Sedang
7	13	0,62	Sedang
8	12	0,57	Sedang

9	10	0,48	Sedang
10	9	0,43	Sedang
11	10	0,48	Sedang
12	8	0,38	Sedang
13	6	0,29	Sukar
14	17	0,81	Mudah
15	11	0,52	Sedang
16	4	0,19	Sukar
17	15	0,71	Mudah
18	9	0,43	Sedang
19	12	0,57	Sedang
20	5	0,24	Sukar
21	9	0,43	Sedang
22	14	0,67	Sedang
23	11	0,52	Sedang
24	12	0,57	Sedang
25	12	0,57	Sedang
26	7	0,33	Sedang
27	9	0,43	Sedang
28	10	0,48	Sedang
29	12	0,57	Sedang
30	9	0,43	Sedang
31	10	0,48	Sedang
32	12	0,57	Sedang
33	12	0,57	Sedang
34	11	0,52	Sedang
35	9	0,43	Sedang
36	9	0,43	Sedang
37	12	0,57	Sedang
38	9	0,43	Sedang
39	13	0,62	Sedang
40	13	0,62	Sedang

Berdasarkan hasil analisis tingkat kesukaran, diperoleh 3 butir soal berkategori mudah, yaitu nomor 1, 14, dan 17. Selanjutnya, 34 butir soal berkategori sedang, yaitu nomor 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 15, 18, 19, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, dan 40. Adapun 3 butir soal berkategori sukar, yaitu nomor 13, 16, dan 20.

Dominasi soal kategori sedang menunjukkan bahwa sebagian besar butir soal telah berada pada tingkat kesukaran yang ideal. Menurut Arikunto, soal yang baik adalah soal yang tidak terlalu mudah dan tidak terlalu sukar, yaitu soal dengan indeks kesukaran sedang. Ia juga menegaskan bahwa bukan berarti soal yang terlalu mudah maupun terlalu sukar tidak boleh digunakan. Keberadaan soal mudah dan sukar tetap

diperlukan untuk mengukur kemampuan peserta didik secara lebih beragam (Arikunto, 2009, hlm. 210).

Daya Pembeda

Daya pembeda merupakan kemampuan suatu butir soal untuk membedakan peserta didik yang memiliki kemampuan tinggi dengan peserta didik yang memiliki kemampuan rendah. Suatu butir soal dikatakan memiliki daya pembeda yang baik apabila lebih banyak dijawab dengan benar oleh kelompok berkemampuan tinggi dibandingkan kelompok berkemampuan rendah. Sebaliknya, apabila suatu soal dapat dijawab dengan benar oleh kedua kelompok dalam proporsi yang hampir sama, maka soal tersebut kurang mampu membedakan kemampuan peserta didik sehingga memiliki daya pembeda yang rendah (Arikunto, 2009, hlm. 211).

Menurut Arikunto, untuk jumlah peserta yang relatif kecil, pengelompokan peserta tes dapat dilakukan dengan membagi peserta menjadi dua kelompok yang sama besar, yaitu 50% kelompok atas dan 50% kelompok bawah (Arikunto, 2009, hlm. 212).

Berdasarkan skor total peserta didik yang berjumlah 21 orang, diperoleh 10 siswa sebagai kelompok atas dan 10 siswa sebagai kelompok bawah, sedangkan 1 siswa yang berada di tengah tidak diikutsertakan dalam perhitungan daya pembeda.

Tabel 7. Pembagian Kelompok Atas dan Bawah

Kelompok Atas	
Skor	Nama
35	MAH
35	SA
35	ZA
32	SI
32	ZR
31	AL
27	SP
25	N
22	MIH
18	AZ
Kelompok Bawah	
Skor	Nama
18	MAN
14	AM
14	MIR
13	MI
12	MSR
9	G
9	MAZ
9	MH
8	AR
7	H

Adapun rumus daya pembeda dan kriteria indeks daya pembeda adalah sebagai berikut: (Arikunto, 2009, hlm. 218)

$$D = \frac{B_A}{J_A} - \frac{B_B}{J_B}$$

Keterangan:

D = daya pembeda

B_A = jumlah siswa kelompok atas yang benar

J_A = jumlah siswa kelompok atas

B_B = jumlah siswa kelompok bawah yang benar

J_B = jumlah siswa kelompok bawah

Tabel 8. Kriteria Daya Pembeda

Indeks Daya Pembeda (DB)	Kriteria
0,00 - 0,20	Jelek
0,21 - 0,40	Cukup
0,41 - 0,70	Baik
0,71 - 1,00	Baik Sekali

Maka hasil perhitungan indeks daya pembeda pada soal PAS dalam penelitian ini dijabarkan dalam tabel berikut:

Tabel 9. Hasil Perhitungan Daya Pembeda

No Soal	BA	BA/JA	BB	BB/JB	D	Kriteria
1	10	1	5	0,5	0,50	Baik
2	8	0,8	6	0,6	0,20	Jelek
3	7	0,7	2	0,2	0,50	Baik
4	5	0,5	2	0,2	0,30	Cukup
5	7	0,7	4	0,4	0,30	Cukup
6	5	0,5	2	0,2	0,30	Cukup
7	9	0,9	3	0,3	0,60	Baik
8	9	0,9	3	0,3	0,60	Baik
9	8	0,8	2	0,2	0,60	Baik
10	8	0,8	1	0,1	0,70	Baik
11	8	0,8	1	0,1	0,70	Baik
12	7	0,7	1	0,1	0,60	Baik
13	4	0,4	2	0,2	0,20	Jelek
14	9	0,9	7	0,7	0,20	Jelek
15	7	0,7	3	0,3	0,40	Cukup
16	2	0,2	2	0,2	0,00	Jelek
17	8	0,8	6	0,6	0,20	Jelek
18	7	0,7	2	0,2	0,50	Baik
19	7	0,7	4	0,4	0,30	Cukup
20	3	0,3	2	0,2	0,10	Jelek
21	7	0,7	2	0,2	0,50	Baik
22	8	0,8	5	0,5	0,30	Cukup
23	9	0,9	2	0,2	0,70	Baik

24	7	0,7	5	0,5	0,20	Jelek
25	9	0,9	3	0,3	0,60	Baik
26	4	0,4	2	0,2	0,20	Jelek
27	7	0,7	1	0,1	0,60	Baik
28	7	0,7	3	0,3	0,40	Cukup
29	8	0,8	3	0,3	0,50	Baik
30	9	0,9	0	0	0,90	Sangat Baik
31	5	0,5	4	0,4	0,10	Jelek
32	10	1	2	0,2	0,80	Sangat Baik
33	9	0,9	3	0,3	0,60	Baik
34	8	0,8	2	0,2	0,60	Baik
35	8	0,8	1	0,1	0,70	Baik
36	7	0,7	2	0,2	0,50	Baik
37	8	0,8	3	0,3	0,50	Baik
38	4	0,4	4	0,4	0,00	Jelek
39	10	1	3	0,3	0,70	Baik
40	10	1	3	0,3	0,70	Baik

Berdasarkan hasil analisis daya pembeda, diperoleh 10 butir soal berkategori jelek, yaitu nomor 2, 13, 14, 16, 17, 20, 24, 26, 31, dan 38. Selanjutnya, terdapat 7 butir soal berkategori cukup, yaitu nomor 4, 5, 6, 15, 19, 22, dan 28. Adapun 21 butir soal berkategori baik, yaitu nomor 1, 3, 7, 8, 9, 10, 11, 12, 18, 21, 23, 25, 27, 29, 33, 34, 35, 36, 37, 39, dan 40, serta 2 butir soal berkategori sangat baik, yaitu nomor 30 dan 32.

Dengan demikian, sebagian besar butir soal memiliki daya pembeda yang baik hingga sangat baik, yaitu sebanyak 23 butir soal (57,5%). Hasil ini menunjukkan bahwa mayoritas soal telah mampu membedakan peserta didik yang berkemampuan tinggi dan peserta didik yang berkemampuan rendah. Meskipun demikian, masih terdapat 10 butir soal yang memiliki daya pembeda jelek dan 7 butir soal yang memiliki daya pembeda cukup, sehingga butir-butir soal tersebut perlu mendapat perhatian untuk diperbaiki pada penggunaan berikutnya.

Efektivitas Pengecoh

Efektivitas pengecoh merupakan ukuran yang menunjukkan sejauh mana alternatif jawaban salah (*distractor*) pada suatu butir soal mampu berfungsi dengan baik dalam menarik peserta didik yang belum menguasai materi untuk memilihnya. Pengecoh yang efektif adalah pilihan jawaban yang tampak masuk akal dan memiliki kemiripan dengan jawaban benar sehingga dapat mengecoh peserta didik yang kurang memahami materi, namun tidak mengecoh peserta didik yang benar-benar menguasai materi (Zulpan dkk., 2025, hlm. 97).

Adapun rumus efektivitas pengecoh adalah sebagai berikut:

$$IP = \frac{P}{N} \times 100\%$$

Keterangan:

IP= indeks pengecoh

P= jumlah siswa yang memilih opsi

N= jumlah seluruh siswa (21 orang)

Tabel 10. Kriteria Efektivitas Pengecoh

Presentase Pemilih	Kriteria
$\geq 5\%$	Pengecoh berfungsi
$< 5\%$	Pengecoh tidak berfungsi

Maka hasil perhitungan efektivitas pengecoh pada soal PAS dalam penelitian ini adalah sebagaimana disajikan dalam tabel berikut:

Tabel 11. Hasil Perhitungan Efektivitas Pengecoh

Soal	Opsi	Jumlah Pemilih	Presentase	Keterangan	Soal	Opsi	Jumlah Pemilih	Presentase	Keterangan
1	a	3	14,29%	Berfungsi	21	b	7	33,33%	Berfungsi
	b	0	0,00%	Tidak Berfungsi		c	4	19,05%	Berfungsi
	d	2	9,52%	Berfungsi		d	1	4,76%	Tidak Berfungsi
2	a	4	19,05%	Berfungsi	22	b	1	4,76%	Tidak Berfungsi
	b	3	14,29%	Berfungsi		c	3	14,29%	Berfungsi
	d	0	0,00%	Tidak Berfungsi		d	3	14,29%	Berfungsi
3	a	6	28,57%	Berfungsi	23	b	5	23,81%	Berfungsi
	b	3	14,29%	Berfungsi		c	3	14,29%	Berfungsi
	c	3	14,29%	Berfungsi		d	2	9,52%	Berfungsi
4	a	6	28,57%	Berfungsi	24	a	3	14,29%	Berfungsi
	c	3	14,29%	Berfungsi		b	5	23,81%	Berfungsi
	d	4	19,05%	Berfungsi		d	1	4,76%	Tidak Berfungsi
5	b	2	9,52%	Berfungsi	25	a	3	14,29%	Berfungsi
	c	3	14,29%	Berfungsi		c	2	9,52%	Berfungsi
	d	4	19,05%	Berfungsi		d	4	19,05%	Berfungsi
6	a	7	33,33%	Berfungsi	26	a	13	61,90%	Berfungsi
	c	2	9,52%	Berfungsi		c	1	4,76%	Tidak Berfungsi
	d	4	19,05%	Berfungsi		d	0	0,00%	Tidak Berfungsi
7	b	5	23,81%	Berfungsi	27	b	6	28,57%	Berfungsi
	c	3	14,29%	Berfungsi		c	3	14,29%	Berfungsi
	d	0	0,00%	Tidak Berfungsi		d	3	14,29%	Berfungsi
8	b	5	23,81%	Berfungsi	28	a	5	23,81%	Berfungsi
	c	3	14,29%	Berfungsi		b	3	14,29%	Berfungsi
	d	1	4,76%	Tidak Berfungsi		c	3	14,29%	Berfungsi
9	b	8	38,10%	Berfungsi	29	a	5	23,81%	Berfungsi
	c	0	0,00%	Tidak Berfungsi		c	1	4,76%	Tidak Berfungsi
	d	3	14,29%	Berfungsi		d	3	14,29%	Berfungsi
10	a	7	33,33%	Berfungsi	30	a	7	33,33%	Berfungsi
	c	4	19,05%	Berfungsi		c	3	14,29%	Berfungsi
	d	1	4,76%	Tidak Berfungsi		d	2	9,52%	Berfungsi
11	a	3	14,29%	Berfungsi	31	b	6	28,57%	Berfungsi
	b	4	19,05%	Berfungsi		c	3	14,29%	Berfungsi
	d	4	19,05%	Berfungsi		d	2	9,52%	Berfungsi
12	a	3	14,29%	Berfungsi	32	a	2	9,52%	Berfungsi
	b	6	28,57%	Berfungsi		b	4	19,05%	Berfungsi
	c	4	19,05%	Berfungsi		c	3	14,29%	Berfungsi
13	b	9	42,86%	Berfungsi	33	a	1	4,76%	Tidak Berfungsi
	c	3	14,29%	Berfungsi		b	2	9,52%	Berfungsi
	d	9	42,86%	Berfungsi		c	6	28,57%	Berfungsi
14	b	0	0,00%	Tidak Berfungsi	34	b	4	19,05%	Berfungsi
	c	1	4,76%	Tidak Berfungsi		c	2	9,52%	Berfungsi
	d	3	14,29%	Berfungsi		d	4	19,05%	Berfungsi
15	a	4	19,05%	Berfungsi	35	a	1	4,76%	Tidak Berfungsi
	b	6	28,57%	Berfungsi		b	6	28,57%	Berfungsi
16	a	6	28,57%	Berfungsi		c	5	23,81%	Berfungsi
	b	8	38,10%	Berfungsi	36	a	4	19,05%	Berfungsi
	c	3	14,29%	Berfungsi		c	3	14,29%	Berfungsi
17	a	1	4,76%	Tidak Berfungsi		d	5	23,81%	Berfungsi
	b	4	19,05%	Berfungsi	37	b	5	23,81%	Berfungsi
	d	1	4,76%	Tidak Berfungsi		c	2	9,52%	Berfungsi
18	a	7	33,33%	Berfungsi		d	2	9,52%	Berfungsi
	c	2	9,52%	Berfungsi	38	a	4	19,05%	Berfungsi
	d	3	14,29%	Berfungsi		c	8	38,10%	Berfungsi
19	a	5	23,81%	Berfungsi		d	0	0,00%	Tidak Berfungsi
	b	3	14,29%	Berfungsi	39	a	4	19,05%	Berfungsi
	d	1	4,76%	Tidak Berfungsi		b	2	9,52%	Berfungsi
20	a	7	33,33%	Berfungsi		c	2	9,52%	Berfungsi
	c	6	28,57%	Berfungsi	40	a	2	9,52%	Berfungsi
	d	3	14,29%	Berfungsi		c	3	14,29%	Berfungsi
						d	3	14,29%	Berfungsi

Berdasarkan hasil analisis efektivitas pengecoh, diperoleh 23 butir soal yang seluruh pengecohnya berfungsi dengan baik, yaitu nomor 3, 4, 5, 6, 11, 12, 13, 15, 16, 18, 20, 23, 25, 27, 28, 30, 31, 32, 34, 36, 37, 39, dan 40. Selanjutnya, terdapat 15 butir soal yang memiliki dua pengecoh berfungsi dan satu pengecoh tidak berfungsi, yaitu nomor 1, 2, 7, 8, 9, 10, 17, 19, 21, 22, 24, 29, 33, 35, dan 38. Adapun 2 butir soal, yaitu nomor 14 dan 26, hanya memiliki satu pengecoh yang berfungsi, sedangkan dua pengecoh lainnya tidak berfungsi.

Namun demikian, apabila ditinjau dari segi konstruksi soal, ditemukan bahwa pada butir soal nomor 15 terdapat dua kunci jawaban. Kondisi ini tidak sesuai dengan kaidah penyusunan soal pilihan ganda. Menurut Arikunto, soal pilihan ganda seharusnya hanya memiliki satu jawaban yang benar atau paling tepat sebagai kunci jawaban (Arikunto, 2009, hlm. 170). Oleh karena itu, meskipun secara statistik pengecoh pada soal nomor 15 berfungsi dengan baik, butir soal tersebut tetap perlu direvisi dari segi konstruksinya.

Dengan demikian, secara umum efektivitas pengecoh pada instrumen PAS sudah tergolong baik. Hal ini ditunjukkan oleh sebagian besar butir soal yang memiliki seluruh pengecoh berfungsi atau setidaknya sebagian besar pengecohnya masih mampu dipilih oleh peserta didik dan tidak ada soal yang seluruh pengecohnya tidak berfungsi.

KESIMPULAN

Berdasarkan hasil analisis kualitas butir soal Penilaian Akhir Semester (PAS) Al-Qur'an Hadis kelas VIII MTs, diperoleh bahwa dari 40 butir soal yang dianalisis, sebanyak 25 butir soal (62,5%) memenuhi kriteria valid, sedangkan 15 butir soal (37,5%) tidak valid. Hasil uji reliabilitas menggunakan rumus KR-20 memperoleh koefisien sebesar 0,932 yang menunjukkan bahwa instrumen memiliki reliabilitas sangat tinggi.

Dari aspek tingkat kesukaran, sebagian besar butir soal berada pada kategori sedang, yaitu sebanyak 34 butir soal (85%), sedangkan 3 butir soal (7,5%) berkategori mudah dan 3 butir soal (7,5%) berkategori sukar. Pada aspek daya pembeda, sebanyak 23 butir soal (57,5%) memiliki daya pembeda baik hingga sangat baik. Adapun pada aspek efektivitas pengecoh, terdapat 23 butir soal yang seluruh pengecohnya berfungsi dengan baik, 15 butir soal memiliki dua pengecoh berfungsi dan satu pengecoh tidak berfungsi, serta 2 butir soal hanya memiliki satu pengecoh yang berfungsi. Selain itu, tidak ditemukan butir soal yang seluruh pengecohnya tidak berfungsi.

Hasil analisis menunjukkan adanya keterkaitan antar-aspek kualitas butir soal. Dari 25 butir soal yang valid, sebanyak 23 butir soal juga memiliki daya pembeda baik hingga sangat baik, yaitu nomor 1, 3, 7, 8, 9, 10, 11, 12, 18, 21, 23, 25, 27, 29, 30, 32, 33, 34, 35, 36, 37, 39, dan 40. Menariknya, 22 dari 23 butir soal tersebut berada pada kategori tingkat kesukaran sedang, sedangkan hanya 1 butir soal yang berkategori mudah. Temuan ini menunjukkan bahwa butir soal dengan tingkat kesukaran sedang cenderung memiliki kualitas yang lebih baik karena mampu mengukur kemampuan

peserta didik secara lebih tepat sekaligus membedakan peserta didik berkemampuan tinggi dan rendah.

Selain itu, terdapat 14 butir soal, yaitu nomor 3, 11, 12, 18, 23, 25, 27, 30, 32, 34, 36, 37, 39, dan 40, yang memenuhi seluruh indikator kualitas butir soal, yakni valid, memiliki daya pembeda baik hingga sangat baik, tingkat kesukaran sedang, serta seluruh pengecohnya berfungsi dengan baik. Oleh karena itu, butir-butir soal tersebut dapat direkomendasikan untuk dimasukkan ke dalam bank soal dan dijadikan acuan dalam penyusunan instrumen evaluasi pada periode berikutnya. Adapun butir soal lainnya yang belum memenuhi seluruh kriteria dapat direvisi atau diperbaiki sesuai dengan aspek kelemahan yang ditemukan.

Terakhir, pada kesempatan ini, penulis mengucapkan terima kasih kepada Ibu Dr. Dina Hermina, M.Pd. selaku dosen pengampu mata kuliah Pengembangan Evaluasi PAI, Bapak Muhammad Fadhilah selaku guru Al-Qur'an Hadis, serta pihak MTs Swasta di Kabupaten Barito Kuala yang telah memberikan dukungan dalam pelaksanaan penelitian ini.

DAFTAR PUSTAKA

- Al-Alusi, S. M. (1994). *Rûh al-Ma'ânî* (Vol. 13). Dâr al-Kutub al-Ilmiyyah. Maktabah al-Syamilah
- Arikunto, S. (2009). *Dasar-Dasar Evaluasi Pendidikan*. Bumi Aksara.
- Azizah, N., & Chalimatusadiah, C. (2025). Uji Validitas dan Uji Reliabilitas Instrumen Penelitian Pemahaman Konsep Dasar Aljabar. *Jurnal Pendidikan Tambusai*, 9(1), 6637–6643. <https://doi.org/10.31004/jptam.v9i1.25533>
- Hazirman, H., Nilawati, N., Lona, D., Putri, M., Susmita, D., & Wahyuni, S. (2026). Evaluasi Pembelajaran Pendidikan Agama Islam: Konsep, Prinsip, dan Landasan Evaluasi dalam Perspektif Islam. *Jurnal Pendidikan Tambusai*, 10(1), 970–979. <https://doi.org/10.31004/jptam.v10i1.36044>
- Khumaedi, M. (2012). Reliabilitas Instrumen Penelitian Pendidikan. *Jurnal Pendidikan Teknik Mesin*, 12(1). <https://doi.org/10.15294/jptm.v12i1.5273>
- Loka, I. N., Sofia, B. F. D., Husnawati, H., & Anwar, Y. A. S. (2024). Identifikasi Tingkat Kemandirian Belajar Siswa dan Korelasinya dengan Hasil Belajar di Masa Covid 19. *Jurnal Ilmiah Profesi Pendidikan*, 9(3), 2047–2055. <https://doi.org/10.29303/jipp.v9i3.2455>
- Magdalena, I., Fauziah, S. N., Faziah, S. N., & Nupus, F. S. (2021). Analisis Validitas, Reliabilitas, Tingkat Kesulitan dan Daya Beda Butir Soal Ujian Akhir Semester Tema 7 Kelas III SDN Karet 1 Sepatan. *BINTANG*, 3(2), 198–214. <https://doi.org/10.36088/bintang.v3i2.1291>
- Maryana, I., & Alawiyah, G. T. (2025). The Importance of Validity and Reliability in Research. *Al-Afkar, Journal For Islamic Studies*, 8(4), 872–879. <https://doi.org/10.31943/afkarjournal.v8i4.1699>
- Rahayu, R., & Djazari, M. (2016). Analisis Kualitas Soal Pra Ujian Nasional Mata Pelajaran Ekonomi Akuntansi. *Jurnal Pendidikan Akuntansi Indonesia*, 14(1). <https://doi.org/10.21831/jpai.v14i1.11370>

- Riinawati. (2021). *Pengantar Evaluasi Pendidikan*. Thema Publishing. <https://idr.uin-antasari.ac.id/16681/1/Pengantar%20Evaluasi%20Pendidikan%20ISI.pdf>
- Sabilla, W., Hanif, M., Febriana, I. A., Gunawan, S., & Lana, K. B. (2026). Mengenal Tiga Pilar Kualitas Soal: Daya Beda, Tingkat Kesukaran, Dan Pengecoh. *Pendas : Jurnal Ilmiah Pendidikan Dasar*, 11(02), 850–865. <https://doi.org/10.23969/jp.viii02.47029>
- Sudijono, A. (2009). *Pengantar Evaluasi Pendidikan*. Rajawali Press. <https://www.scribd.com/document/614546986/Pengantar-Evaluasi-Pendidikan-Oleh-Anas-Sudijono>
- Zulpan, Z., Syafi', A., Alfansuri, D. U., Salsabila, S., Hasibuan, S. W., Azizah, N., & Mustikawati, D. L. (2025). Kualitas Soal: Tingkat Kesukaran Daya Pembeda Pengecoh Soal. *Teaching and Learning Research Journal*, 1(4), 93–101. <https://doi.org/10.66770/tlrj.vii4.81>