

Analisis Kegagalan Evaluatif Guru Pendidikan Agama Islam Melalui Butir Soal Pendidikan Agama Islam di Sekolah Dasar

Nopita Sari¹ Harriska² Komalasari³ Fitri Muthmainnah⁴ Asmiati⁵

Pendidikan Agama Islam, FTIK, IAIN Pontianak, Pontianak, Kalimantan Barat, Indonesia¹

Pendidikan Islam Anak Usia Dini, FTIK, IAIN Pontianak, Pontianak, Kalimantan Barat, Indonesia²

Tadris Matematika, FTIK, IAIN Pontianak, Pontianak, Kalimantan Barat, Indonesia³

Pendidikan Guru Madrasah Ibtidaiyah, FTIK, IAIN Pontianak, Pontianak, Kalimantan Barat, Indonesia^{4,5}

Email: sarievitta@gmail.com¹ bharriska@gmail.com² rie.komalasari@iainptk.ac.id³
fitriponti12@gmail.com⁴ asmiatiugm@gmail.com⁵

Abstrak

Penilaian bukan sekadar angka. Di dalam soal ulangan, tersembunyi cara berpikir seorang guru, harapan lembaga, dan nilai-nilai yang ingin diwariskan. Penelitian ini menelaah kualitas soal ulangan umum semester ganjil mata pelajaran Pendidikan Agama Islam di SDN 12 Teluk Pakedai, mencakup lima aspek: validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan pengecoh. Pendekatan yang digunakan adalah kuantitatif deskriptif, dengan data berupa naskah soal, kunci jawaban, dan lembar jawaban siswa. Hasilnya menunjukkan bahwa hanya 50% soal yang valid. Meski demikian, reliabilitasnya tinggi (0,996), menandakan konsistensi yang kuat. Tingkat kesukaran didominasi soal kategori sedang (60%) dan mudah (40%). Namun, daya pembeda menunjukkan ironi: tak satu pun soal berkategori sangat baik, hanya 2 soal (10%) yang baik, dan sisanya tergolong cukup hingga sangat jelek. Dari 40 pengecoh yang dianalisis, hanya 60% tergolong baik hingga sangat baik. Temuan ini menunjukkan bahwa evaluasi dalam pendidikan agama tidak bisa dianggap remeh. Soal ujian bukan sekadar alat seleksi, melainkan representasi dari mutu pendidikan itu sendiri. Ketika instrumen pengukur tidak tajam, nilai-nilai keislaman yang ingin diajarkan pun menjadi kabur. Maka, validitas dan reliabilitas bukan sekadar angka statistik, melainkan cermin dari ketulusan dan ketelitian kita dalam membentuk generasi yang beriman dan berpikir jernih.

Kata Kunci: Analisis Butir Soal, Pendidikan Agama Islam, Sekolah Dasar



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

PENDAHULUAN

Dalam dunia yang semakin ditata oleh angka dan indikator, evaluasi pendidikan seharusnya menjadi penjernih: membantu kita melihat dengan terang apakah yang diajarkan memang sampai, dan apakah yang dipelajari benar-benar tinggal. Di ruang kelas, ujian bukan hanya ritual tahunan atau syarat administratif. Ia adalah cermin. Di balik setiap soal yang diajukan, tersimpan pengakuan diam-diam tentang apa yang dianggap penting oleh seorang guru dan apa yang dianggap mampu dikuasai oleh murid. Idealnya, setiap evaluasi adalah jembatan yang menyambungkan niat mendidik dengan hasil belajar yang bisa ditelusuri, diukur, dan ditindaklanjuti. Namun kenyataan sering tak berjalan serapi ideal. Di banyak ruang kelas sekolah dasar, terutama pada mata pelajaran Pendidikan Agama Islam, evaluasi kerap berubah menjadi rutinitas mekanis yang miskin makna. Soal-soal disusun sekadarnya, sering diambil dari lembar kerja siswa, lalu diuji tanpa melalui proses analisis kualitas. Guru, yang seharusnya menjadi arsitek soal, justru tak punya bekal memadai untuk membaca mutu butir yang ia buat. Di SDN 12 Teluk Pakedai misalnya, seorang guru mengaku hanya mengandalkan intuisi: soal dinilai baik bila mudah dipahami, atau cukup sulit untuk “mengaduk otak” siswa. Tak ada validitas. Tak ada reliabilitas. Tak ada daya pembeda atau analisis pengecoh. Yang ada

hanyalah kepercayaan bahwa dengan bertanya, ia telah mengukur. Dalam konteks inilah, penting kiranya kita mengajukan kembali pertanyaan mendasar: apa yang sebenarnya diuji saat kita memberikan soal kepada siswa? Bila soal tak pernah diperiksa kualitasnya apakah ia benar-benar mengukur kompetensi, membedakan penguasaan, atau sekadar menguji ingatan? Maka yang kita hadapi bukan sekadar kelemahan teknis, melainkan kegagalan evaluatif. Sebab setiap soal yang buruk bukan hanya soal yang salah; ia adalah cermin buram dari cara berpikir seorang pengajar. Dan bila dibiarkan, ia menjadi warisan kesalahan yang berulang, dari satu generasi ke generasi berikutnya.

Kajian tentang mutu butir soal memang bukan hal baru. Namun kebanyakan berpusat pada pelajaran eksakta atau jenjang menengah, seperti matematika di SMP dan SMA. Sedikit yang menyoroti Pendidikan Agama Islam di sekolah dasar padahal di sinilah nilai-nilai dasar ditanamkan. Terlebih lagi, guru-guru di wilayah pinggiran seperti Teluk Pakedai, Kalimantan Barat, berhadapan dengan keterbatasan sumber daya dan pelatihan. Maka meneliti kualitas soal yang mereka susun bukan hanya soal teknis pengukuran, tapi menyentuh dimensi yang lebih dalam: bagaimana pendidikan agama ditanamkan, dan sejauh mana ia betul-betul dimengerti, bukan hanya diajarkan. Penelitian ini mengambil latar SDN 12 Teluk Pakedai, sebuah sekolah dasar negeri di daerah yang secara fasilitas tergolong cukup, namun masih terbatas dari sisi pengembangan kompetensi guru, khususnya dalam hal evaluasi pembelajaran. Di sini, guru PAI menyusun sendiri soal ulangan umum semester, namun belum pernah melakukan analisis butir soal secara sistematis. Penelitian ini hadir untuk mengisi celah tersebut: menganalisis kualitas soal pilihan ganda dari segi validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan pengecoh. Kajian ini penting bukan hanya untuk menilai soal, tetapi untuk melihat bagaimana soal itu mewakili cara guru memahami proses mengajar dan belajar. Sebab pada akhirnya, setiap soal bukan sekadar alat ukur, tetapi adalah pernyataan pedagogis dan mungkin juga, pengakuan diam-diam tentang sejauh mana kita memahami murid kita.

Penelitian Terdahulu yang Relevan

No	Penulis/ Tahun	Judul	Tujuan	Metode	Hasil	Kontribusi dan relevansi
1	Agustin (2019)	<i>Analisis Soal Ujian Akhir Semester Ganjil Mata Pelajaran Pendidikan Agama Islam Ditinjau dari Tingkat Kesukaran dan Daya Pembeda pada Kelas VIII di SMP Islam Terpadu Bina Insani Metro</i>	Mengetahui kualitas butir soal UAS PAI ditinjau dari tingkat kesukaran dan daya pembeda.	Penelitian kuantitatif dengan analisis butir soal.	Analisis butir soal digunakan untuk mengetahui apakah soal berada pada kategori mudah, sedang, atau sukar serta kemampuan soal dalam membedakan peserta didik berdasarkan tingkat kemampuannya. (UIN Jurai Siwo Repository)	Sangat relevan karena sama-sama mengkaji kualitas butir soal PAI. Perbedaan, penelitian Nopita Sari menganalisis lebih komprehensif dengan lima aspek, yaitu validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh.
2	Atika et al. (2023)	<i>Formative Test Analysis of Islamic Religious Education Learning Evaluation Practices (Study at SMP Negeri 1 Muara Kemumu)</i>	Menganalisis pelaksanaan asesmen formatif dalam pembelajaran PAI.	Kualitatif deskriptif, dengan pengumpulan data melalui observasi, wawancara, dan dokumentasi. (CV Media)	Guru PAI di SMP Negeri 1 Muara Kemumu melaksanakan asesmen formatif setelah menyelesaikan materi atau bab. Namun, asesmen yang digunakan masih terbatas pada post-	Relevan karena menunjukkan bahwa instrumen tes merupakan bagian penting dari proses evaluasi PAI. Penelitian Nopita Sari melanjutkan kajian tersebut dengan menguji kualitas teknis butir soal yang digunakan dalam evaluasi PAI.

				Inti Teknologi	test berupa soal esai dan pilihan ganda. (CV Media Inti Teknologi)	
3	Marjuah et al. (2023)	<i>The Role of the Islamic Religious Education Teacher Working Group in Increasing Islamic Religious Education Teachers' Competence</i>	Mengetahui kompetensi profesional guru PAI serta peran Kelompok Kerja Guru (KKG) dalam meningkatkan kompetensi guru PAI di sekolah dasar.	Penelitian lapangan mengenai kompetensi profesional guru dan peran KKG.	KKG memiliki peran dalam meningkatkan kompetensi profesional guru PAI melalui kegiatan pengembangan dan kerja sama profesional. (IJCIED)	Relevan karena kualitas penyusunan dan analisis instrumen evaluasi sangat berkaitan dengan kompetensi guru. Hasil penelitian Nopita Sari yang menemukan masih lemahnya validitas dan daya pembeda soal dapat menjadi dasar perlunya peningkatan kompetensi evaluatif guru melalui KKG.
4	Naofa et al. (2023)	<i>Learning Objectives of Islamic Religious Education in Schools: The Role of the Teacher and Its Implications Based on Relevant Study</i>	Mengkaji tujuan pembelajaran PAI dan peran guru dalam mencapai tujuan pembelajaran tersebut.	Kajian terhadap penelitian yang relevan mengenai tujuan pembelajaran PAI.	Tujuan PAI tidak hanya berorientasi pada penguasaan pengetahuan, tetapi juga pada pembentukan sikap dan perilaku Islami peserta didik.	Relevan karena memberikan landasan bahwa evaluasi PAI harus sesuai dengan kompetensi yang hendak dicapai. Dengan demikian, kualitas butir soal perlu diperhatikan agar soal benar-benar mengukur kompetensi peserta didik, bukan sekadar hafalan. Penelitian ini memperkuat urgensi analisis validitas dan karakteristik butir soal dalam penelitian Nopita Sari.
5	Huda et al. (2024)	<i>Islamic Religious Education Learning Media in the Technology Era: A Systematic Literature Review</i>	Mengkaji perkembangan media pembelajaran PAI pada era teknologi serta implikasinya terhadap pembelajaran PAI.	Systematic Literature Review (SLR).	Perkembangan teknologi memberikan peluang bagi pembelajaran PAI, tetapi juga menuntut adanya penyesuaian terhadap media dan proses pembelajaran.	Relevan untuk menunjukkan bahwa evaluasi PAI perlu berkembang mengikuti perubahan pembelajaran. Penelitian Nopita Sari memberikan kontribusi berbeda karena berfokus pada kualitas instrumen tes PAI secara empiris, bukan pada media pembelajaran.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif, yakni pendekatan yang bertumpu pada angka dan analisis statistik untuk memahami sebuah fenomena secara objektif. Creswell menyebut bahwa penelitian kuantitatif bertujuan menjelaskan realitas melalui data numerik yang dianalisis secara matematis. (Creswell, 2014) Sejalan dengan itu, Sugiyono menegaskan bahwa pendekatan ini berakar pada paradigma positivisme, yang mengedepankan objektivitas dan generalisasi berdasarkan data yang terkumpul dari populasi dan sampel. (Sugiyono, 2013) Angka-angka, dalam penelitian kuantitatif, bukan sekadar simbol matematis. Ia adalah cara untuk menangkap dunia secara terukur membekukan kenyataan dalam satuan yang bisa dicermati. Di dalam penelitian ini, pendekatan yang digunakan adalah deskriptif kuantitatif. Bukan untuk menguji sebab-akibat, bukan pula untuk menelusuri kait-mengait antarvariabel, melainkan untuk melihat, secara jernih dan sistematis, bagaimana sebuah gejala hadir dalam realitas. Di situlah penelitian deskriptif menjadi alat: ia membantu peneliti membaca keadaan sebagaimana adanya apa yang berlangsung, dalam konteks yang khas. Seperti yang ditegaskan

Sugiyono (2013), pendekatan ini bertujuan memahami nilai sebuah variabel secara mandiri, tanpa harus menautkannya pada variabel lain. Sebuah cara untuk mendengar fakta berbicara tanpa diganggu oleh prasangka hubungan.

Penelitian ini dilaksanakan di SDN 12 Teluk Pakedai, sebuah sekolah dasar yang terletak di Kuala Karang, Kecamatan Teluk Pakedai, Kabupaten Kubu Raya, Kalimantan Barat. Teknik pengumpulan data dilakukan melalui dokumentasi, yaitu dengan menghimpun naskah soal, kunci jawaban, dan lembar jawaban siswa. Sugiyono (2013) mendefinisikan dokumentasi sebagai catatan peristiwa yang telah terjadi, baik dalam bentuk tulisan, gambar, maupun karya lainnya. Data yang terkumpul kemudian dianalisis menggunakan lima indikator utama: validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Kelima aspek ini menjadi tolok ukur untuk menilai kualitas instrumen evaluasi, sekaligus menyingkap sejauh mana kemampuan guru dalam merancang alat ukur yang tepat bagi siswa.

HASIL PENELITIAN DAN PEMBAHASAN

Hasil penelitian

Validitas

Analisis validitas pada soal pilihan ganda ulangan umum semester ganjil tahun ajaran 2024/2025 pada mata pelajaran Pendidikan Agama Islam kelas II dengan jumlah sebanyak 26 peserta didik di SDN 12 Teluk Pakedai dengan menggunakan rumus Point Biserial didapatkan hasil sebagai berikut:

Tabel 1. Analisis Validitas

No.	R Hitung	R Tabel	Keterangan
1	-0.315	0.388	Tidak Valid
2	0.464	0.388	Valid
3	0.078	0.388	Tidak Valid
4	0.479	0.388	Valid
5	0.401	0.388	Valid
6	-0.242	0.388	Tidak Valid
7	0.504	0.388	Valid
8	0.450	0.388	Valid
9	-0.208	0.388	Tidak Valid
10	0.497	0.388	Valid
11	-0.073	0.388	Tidak Valid
12	0.185	0.388	Tidak Valid
13	0.445	0.388	Valid
14	-0.265	0.388	Tidak Valid
15	0.404	0.388	Valid
16	0.049	0.388	Tidak Valid
17	0.425	0.388	Valid
18	-0.037	0.388	Tidak Valid
19	0.394	0.388	Valid
20	-0.015	0.388	Tidak Valid

Sumber: Hasil Analisis Validitas

Hasil analisis soal ulangan umum semester pada mata pelajaran Pendidikan Agama Islam kelas II SDN 12 Teluk Pakedai tahun ajaran 2024/2025 dari 20 soal pilihan ganda terdapat 10 soal (50%) dinyatakan valid dan 10 soal (50%) dinyatakan tidak valid. 10 soal (50%) yang dinyatakan valid tersebut berada pada soal nomor: 2, 4, 5, 7, 8, 10, 13, 15, 17 dan 19. Sedangkan 10 soal (50%) yang tidak valid berada pada nomor: 1, 3, 6, 9, 11, 12, 14, 16, 18 dan 20.

Reliabilitas

Hasil pengujian reliabilitas pada soal pilihan ganda ulangan umum semester ganjil tahun ajaran 2024/2025 pada mata pelajaran Pendidikan Agama Islam kelas II SDN 12 Teluk Pakedai dengan menggunakan rumus Alpha Cronbach didapatkan hasil sebagai berikut:

Tabel 2. Analisis Reliabilitas

Reliability Statistics	
Cronbach's Alpha	N of Items
0.996	20

Sumber: Hasil SPSS

Berdasarkan hasil analisis di atas didapat hasil sebesar 0,996 dengan jumlah item sebanyak 20 soal, berdasarkan koefisien reliabilitas yang menunjukkan bahwa hasil ini memiliki tingkat reliabilitas sangat tinggi.

Tingkat Kesukaran

Berdasarkan soal pilihan ganda ulangan umum semester ganjil tahun ajaran 2024/2025 pada mata pelajaran Pendidikan Agama Islam kelas II SDN 12 Teluk Pakedai sebagai berikut:

Tabel 3. Analisis Tingkat Kesukaran

No Butir	Jumlah Betul	Tingkat Kesukaran	Kategori
1	24	0.92	Mudah
2	13	0.50	Sedang
3	23	0.88	Mudah
4	14	0.54	Sedang
5	17	0.65	Sedang
6	16	0.62	Sedang
7	20	0.77	Mudah
8	17	0.65	Sedang
9	21	0.81	Mudah
10	12	0.46	Sedang
11	18	0.69	Sedang
12	13	0.50	Sedang
13	25	0.96	Mudah
14	14	0.54	Sedang
15	12	0.46	Sedang
16	22	0.85	Mudah
17	16	0.62	Sedang
18	17	0.65	Sedang
19	20	0.77	Mudah
20	22	0.85	Sedang

Sumber: Hasil Anates V4

Dari tabel tersebut diketahui bahwa tidak terdapat soal dengan kategori sukar (0%), untuk soal dengan kategori sedang terdapat 12 soal (60%) dan untuk soal dengan kategori mudah terdapat 8 soal (40%).

Tabel 4. Kategori Indeks Kesukaran

No.	Indeks Kesukaran	Butir Soal	Persen
1	0,00 – 0,30 (Sukar)	0	0%
2	0,31 – 0,70 (Sedang)	2,4,5,6,8,10,11,12,14,15,17,18	60%
3	0,71- 1,00 (Mudah)	1,3,7,9,13,16,19,20	40%

Sumber: Hasil Tingkat Kesukaran

Daya Pembeda

Berdasarkan hasil analisis soal pilihan ganda ulangan umum semester ganjil tahun ajaran 2024/2025 pada mata pelajaran Pendidikan Agama Islam kelas II dapat diketahui bahwa untuk soal dengan kategori sangat baik tidak ada atau 0 (0%), untuk soal baik terdapat 2 soal (10%), untuk soal cukup terdapat 6 soal (30%), untuk soal jelek terdapat 6 soal (30%) dan untuk soal sangat jelek terdapat 6 soal (30%).

Tabel 5. Kategori Indeks Daya Pembeda

No.	Indeks Daya Pembeda	Butir Soal	Persen
1	0,71 – 1,00 (Sangat Baik)	0	0%
2	0,41 – 0,70 (Baik)	4 dan 10	10%
3	0,21 – 0,40 (Cukup)	2,5,7,12,15,17	30%
4	0,00 – 0,20 (Jelek)	8,11,13,16,19,20	30%
5	(-) Negatif (Sangat Jelek)	1,3,6,9,14,18	30%

Sumber: Hasil Daya Pembeda

Efektifitas Pengecoh

Pengecoh yang baik apabila dipilih oleh peserta didik minimal 5% dari jumlah seluruh peserta didik. Berdasarkan hasil analisis butir soal pilihan ganda mata pelajaran Pendidikan Agama Islam dari segi pengecoh dengan total 40 opsi jawaban dari 20 soal pilihan ganda diperoleh 12 opsi (30%) dengan pengecoh dalam kategori sangat baik, 12 opsi (30%) dalam kategori baik, 7 opsi (18%) dalam kategori kurang baik, 9 opsi (22%) dalam kategori jelek, dan 0 opsi (0%) kategori sangat jelek.

Pembahasan

Validitas

Validitas merupakan salah satu fondasi utama dalam pengembangan dan evaluasi instrumen tes. Djemari Mardapi menyatakan bahwa *validitas adalah dukungan bukti dan teori terhadap penafsiran skor tes yang sesuai dengan tujuan penggunaan tes.* (Mardapi, 2008) Dengan kata lain, validitas menunjukkan sejauh mana sebuah butir soal benar-benar mengukur apa yang seharusnya diukur. Dalam penelitian ini, analisis validitas dilakukan terhadap butir soal pilihan ganda pada ulangan umum semester ganjil tahun ajaran 2024/2025 untuk mata pelajaran Pendidikan Agama Islam kelas II di SDN 12 Teluk Pakedai. Proses analisis dilakukan menggunakan Microsoft Excel, dengan mengacu pada nilai *r tabel* pada taraf signifikansi 5%. Dengan jumlah responden sebanyak 26 siswa, nilai *r tabel* yang digunakan adalah 0,388. Kriteria penilaian yang digunakan cukup sederhana: jika *r hitung* lebih besar dari *r tabel*, maka butir soal dinyatakan valid; sebaliknya, jika *r hitung* lebih kecil, maka soal dinyatakan tidak valid. Hasil analisis menunjukkan bahwa dari total 20 butir soal, hanya 10 soal (50%) yang memenuhi kriteria validitas, sementara 10 soal lainnya (50%) tidak valid. Temuan ini mengindikasikan bahwa tingkat validitas soal berada pada kategori sedang. Karena hanya separuh dari total soal yang layak secara validitas, maka perlu dilakukan revisi atau pengembangan lebih lanjut terhadap butir-butir soal yang tidak memenuhi syarat. Validitas yang lemah menjadi sinyal bahwa beberapa soal mungkin tidak merepresentasikan kompetensi yang hendak diukur, dan hal ini dapat berdampak pada kualitas evaluasi secara keseluruhan.

Reliabilitas

Reliabilitas dalam suatu tes berkaitan dengan tingkat ketetapan atau konsistensi hasil yang diperoleh dari pengukuran. Sebuah instrumen dinyatakan reliabel apabila hasil pengukurannya stabil, meskipun digunakan berulang kali untuk mengukur objek yang sama,

selama tidak terjadi perubahan signifikan. Suharsini Arikunto menjelaskan bahwa reliabilitas menunjukkan sejauh mana instrumen menghasilkan data yang tetap ketika digunakan dalam kondisi yang serupa. (Arikunto, 2018) Berdasarkan hasil analisis terhadap 20 butir soal pilihan ganda pada ulangan umum semester ganjil mata pelajaran Pendidikan Agama Islam kelas II di SDN 12 Teluk Pakedai, diperoleh koefisien reliabilitas sebesar 0,996. Angka ini termasuk dalam kategori reliabilitas sangat tinggi, menunjukkan bahwa instrumen memiliki konsistensi yang kuat dalam mengukur kemampuan peserta didik.

Reliabilitas ini penting sebagai dasar kepercayaan terhadap kualitas instrumen evaluasi. Uji reliabilitas dilakukan untuk memastikan bahwa instrumen soal dapat memberikan hasil yang dapat diandalkan. Zainal Arifin (hlm. 258) menyebutkan bahwa terdapat empat faktor utama yang memengaruhi reliabilitas suatu instrumen, yaitu panjang tes, sebaran skor, tingkat kesukaran, dan tingkat objektivitas dalam penyusunan maupun penilaiannya. Dengan demikian, dapat disimpulkan bahwa soal-soal ulangan umum semester ganjil tahun ajaran 2024/2025 pada mata pelajaran Pendidikan Agama Islam di SDN 12 Teluk Pakedai memiliki tingkat reliabilitas yang sangat tinggi. Temuan ini sejalan dengan pendapat Anas Sudijono, yang menyatakan bahwa suatu instrumen dapat dianggap reliabel apabila memiliki koefisien reliabilitas sebesar 0,70 atau lebih.

Tingkat Kesukaran

Berdasarkan hasil analisis terhadap tingkat kesukaran butir soal ulangan umum semester ganjil tahun ajaran 2024/2025 untuk mata pelajaran Pendidikan Agama Islam, diketahui bahwa dari 20 butir soal pilihan ganda yang dianalisis, tidak ada satu pun soal yang termasuk kategori sukar (0%), sebanyak 12 soal (60%) tergolong dalam kategori sedang, dan 8 soal (40%) masuk dalam kategori mudah. Menurut pandangan para ahli seperti Arikunto dan Sudjana, soal yang baik umumnya berada pada kategori sedang. Soal dengan tingkat kesukaran sedang memiliki kemampuan untuk membedakan peserta didik yang benar-benar memahami materi dengan yang belum menguasainya secara utuh. Sementara itu, soal kategori mudah berguna untuk membangun rasa percaya diri peserta didik, dan soal yang tergolong sukar ditujukan untuk mengukur kemampuan pada level yang lebih tinggi. Penelitian Nelly Agustin juga menegaskan bahwa soal yang ideal adalah soal dengan tingkat kesukaran sedang—tidak terlalu mudah dan tidak terlalu sukar. Soal yang terlalu mudah cenderung tidak menantang peserta didik dalam berpikir kritis, sedangkan soal yang terlalu sukar justru dapat mematikan semangat belajar karena berada di luar jangkauan kemampuan mereka. (Agustin, 2019)

Berdasarkan pertimbangan tersebut, dapat disimpulkan bahwa kualitas butir soal tergolong cukup baik, mengingat mayoritas soal (60%) berada pada kategori sedang dan sisanya (40%) berada pada kategori mudah. Tindak lanjut yang dapat dilakukan setelah analisis tingkat kesukaran ini adalah sebagai berikut: butir soal yang tergolong sedang, karena memenuhi kriteria kualitas, dapat langsung dimasukkan ke dalam bank soal. Sementara itu, untuk butir soal yang masuk kategori terlalu mudah atau terlalu sukar, terdapat dua opsi tindak lanjut. Pertama, butir soal tersebut dapat dihapus dan tidak digunakan lagi dalam ulangan umum semester berikutnya. Kedua, soal-soal tersebut dapat ditelaah kembali untuk mengidentifikasi penyebab ketidaksesuaian tingkat kesukarannya, lalu dilakukan perbaikan agar dapat digunakan kembali dalam penyusunan soal ulangan umum yang akan datang.

Daya Pembeda

Analisis daya pembeda dilakukan untuk mengetahui sejauh mana sebuah butir soal mampu membedakan antara peserta didik yang telah menguasai materi dengan yang belum. Pengujian ini penting untuk menilai efektivitas soal dalam mengklasifikasikan kemampuan siswa secara objektif dan adil. Hasil analisis terhadap soal pilihan ganda pada ulangan umum

semester ganjil tahun ajaran 2024/2025 mata pelajaran Pendidikan Agama Islam di kelas II SDN 12 Teluk Pakedai menunjukkan bahwa tidak ada satupun soal yang termasuk dalam kategori daya pembeda sangat baik (0%). Soal dengan kategori baik hanya berjumlah 2 soal (10%), sedangkan 6 soal (30%) termasuk kategori cukup, 6 soal (30%) berkategori jelek, dan 6 soal lainnya (30%) tergolong sangat jelek. Mengacu pada pendapat Mardiah Astuti, suatu butir soal dikatakan memiliki daya pembeda yang cukup baik apabila nilai indeksinya berada pada angka $\geq 0,21$.

Berdasarkan hasil tersebut, dapat disimpulkan bahwa kualitas butir soal pada ulangan umum semester ini masih perlu banyak perbaikan. Rendahnya jumlah soal dengan daya pembeda baik—dan bahkan tidak adanya soal dengan kategori sangat baik—menunjukkan bahwa mayoritas soal belum efektif dalam membedakan peserta didik dengan kemampuan tinggi dan rendah. Dengan kata lain, fungsi evaluatif dari soal-soal tersebut belum berjalan optimal. Sebagai tindak lanjut, perlu dilakukan langkah strategis terhadap hasil analisis ini. Pertama, butir soal yang memiliki daya pembeda baik sebaiknya disimpan dalam bank soal dan dapat digunakan kembali pada ulangan semester mendatang. Kedua, untuk soal-soal yang masuk dalam kategori cukup, jelek, dan sangat jelek, terdapat dua kemungkinan penanganan: (1) soal-soal tersebut ditelaah kembali, dianalisis penyebab lemahnya daya pembeda, lalu diperbaiki dan diuji ulang pada evaluasi berikutnya; atau (2) jika tidak memungkinkan diperbaiki, soal tersebut dapat dibuang dan tidak digunakan lagi dalam penyusunan soal ke depan. Pendekatan ini penting untuk memastikan bahwa instrumen evaluasi yang digunakan benar-benar mampu mencerminkan kemampuan riil peserta didik.

Efektifitas Pengecoh

Dalam soal pilihan ganda, opsi atau alternatif jawaban umumnya terdiri dari tiga hingga lima pilihan, dengan satu jawaban benar dan sisanya merupakan pengecoh atau *distraktor*. Kehadiran pengecoh bertujuan untuk menguji pemahaman peserta didik secara lebih mendalam—dengan harapan bahwa peserta yang belum benar-benar menguasai materi akan terdorong memilih opsi pengecoh. Analisis pengecoh hanya relevan untuk soal berbentuk pilihan ganda, karena hanya jenis soal ini yang memiliki beberapa alternatif jawaban. Pada ulangan umum semester ganjil tahun ajaran 2024/2025 untuk mata pelajaran Pendidikan Agama Islam kelas II SDN 12 Teluk Pakedai, setiap butir soal memiliki tiga pilihan jawaban: A, B, dan C, dengan satu sebagai kunci jawaban dan dua lainnya sebagai pengecoh. Sebuah pengecoh dikatakan berfungsi dengan baik apabila dipilih oleh minimal 5% dari total peserta.

Berdasarkan hasil analisis terhadap 20 soal (dengan total 40 pengecoh), diperoleh distribusi sebagai berikut: 12 opsi (30%) tergolong pengecoh sangat baik, 12 opsi (30%) tergolong baik, 7 opsi (18%) tergolong kurang baik, dan 9 opsi (22%) tergolong jelek. Tidak ditemukan pengecoh dalam kategori sangat jelek. Dengan demikian, dapat disimpulkan bahwa 60% pengecoh berfungsi efektif dalam menarik perhatian peserta didik yang belum memahami materi secara utuh. Namun, masih terdapat 40% pengecoh yang tidak optimal, yang berpotensi menurunkan kualitas soal dalam mengukur kemampuan secara valid. Oleh karena itu, diperlukan revisi dan penyempurnaan terhadap pengecoh yang tidak efektif agar fungsi evaluatif soal menjadi lebih tajam dan adil.

KESIMPULAN

Berdasarkan hasil analisis validitas, dapat disimpulkan bahwa kualitas instrumen soal pilihan ganda dalam ulangan umum semester mata pelajaran Pendidikan Agama Islam di SDN 12 Teluk Pakedai masih berada dalam kategori sedang. Dari 20 butir soal yang dianalisis, hanya 50% dinyatakan valid, sementara sisanya tidak memenuhi syarat validitas. Ini berarti bahwa hanya setengah dari soal yang benar-benar mengukur kompetensi yang seharusnya diukur,

sedangkan sisanya berpotensi menimbulkan distorsi dalam interpretasi hasil belajar siswa. Secara reliabilitas, butir soal menunjukkan performa yang sangat tinggi. Nilai koefisien reliabilitas sebesar 0,996 menjadi indikator bahwa instrumen soal tersebut secara statistik menghasilkan hasil yang konsisten. Namun reliabilitas yang tinggi ini menjadi kurang berarti apabila tidak dibarengi dengan validitas yang kuat—karena reliabilitas tanpa validitas hanya akan memperkuat kekeliruan dalam pengukuran.

Tingkat kesukaran soal menunjukkan bahwa mayoritas berada dalam kategori sedang (60%), sementara sisanya tergolong mudah (40%) dan tidak ada yang tergolong sukar. Ini memberi gambaran bahwa soal-soal tersebut cukup terjangkau oleh sebagian besar siswa, tetapi mungkin belum cukup menantang untuk mengukur kapasitas nalar tingkat tinggi. Dari aspek daya pembeda, hanya 10% soal yang tergolong baik, sedangkan 60% soal berada dalam kategori cukup hingga sangat jelek. Ini menandakan lemahnya kemampuan soal untuk membedakan antara siswa berkemampuan tinggi dan rendah. Sementara itu, pengecoh menunjukkan performa campuran: 60% berfungsi baik, namun 40% sisanya masih belum optimal dalam menyesatkan siswa yang belum memahami materi dengan baik. Dari semua data tersebut, terlihat bahwa kelemahan paling mencolok terletak pada aspek validitas dan daya pembeda. Ini memperkuat argumen bahwa banyak guru belum memiliki kapasitas evaluatif yang memadai dalam menyusun instrumen penilaian yang tajam dan akurat. Ketika guru tidak memahami prinsip-prinsip dasar evaluasi—seperti bagaimana merancang soal dengan daya pembeda tinggi atau validitas yang sah—maka proses penilaian hanya menjadi ritual administratif yang kehilangan makna pedagogisnya. Evaluasi berubah dari alat diagnosis menjadi sekadar formalitas. Dalam diskursus Pendidikan Agama Islam, hasil penelitian ini menunjukkan signifikansi yang mendalam. Evaluasi dalam pendidikan agama bukan sekadar menguji hafalan dalil atau istilah, melainkan menilai pemahaman, penghayatan, dan penerapan nilai-nilai. Jika instrumen penilaiannya tidak mampu memotret kompetensi secara valid dan adil, maka pendidikan agama akan gagal dalam fungsi transformasinya. Maka, memperkuat literasi evaluatif guru PAI menjadi agenda strategis untuk memastikan bahwa nilai-nilai keislaman yang diajarkan benar-benar mengakar dalam pemahaman dan perilaku peserta didik.

DAFTAR PUSTAKA

- Agustin, N. (2019). Analisis Soal Ujian Akhir Semester Ganjil Mata Pelajaran Pendidikan Agama Islam Ditinjau dari Tingkat Kesukaran dan Daya Pembeda Pada Kelas VIII di SMP Islam Terpadu Bina Insani Metro [Skripsi]. Institut Agama Islam Negeri Metro.
- Arikunto, S. (2018). Dasar-Dasar Evaluasi Pendidikan. Sinar Grafika Offset.
- Atika, Y., Rahmaniar, Y., Muslim, M., Wanto, D., & Daheri, M. (2023). Formative Test Analysis of Islamic Religious Education Learning Evaluation Practices (Study at SMP Negeri 1 Muara Kemumu). *Ijopate*. <https://doi.org/10.58723/ijopate.v1i2.113>
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed). SAGE Publications.
- Evendi, E. (2022). Teacher Professional Education Program in Islamic Religious Education: Bibliometric Analysis and Review. *Prisma Sains Jurnal Pengkajian Ilmu Dan Pembelajaran Matematika Dan Ipa Ikip Mataram*. <https://doi.org/10.33394/j-ps.v10i3.5322>
- Huda, M., Arif, M., Abdul Rahim, M. M., & Anshari, M. (2024). Islamic Religious Education Learning Media in the Technology Era: A Systematic Literature Review. *At-Tadzkir*. <https://doi.org/10.59373/attadzkir.v3i2.62>
- Mansir, F. (2022). Problems of Islamic Religious Education in the Digital Era. *At Ta Dib*. <https://doi.org/10.21111/at-tadib.v17i2.8405>

- Mardapi, D. (2008). Teknik Penyusunan Instrumen Tes dan Nontes. Mitra Cendikia.
- Marjuah, H., Kamaruddin, K., Sidik, S., & Yusra, Y. (2023). The Role of the Islamic Religious Education Teacher Working Group in Increasing Islamic Religious Education Teachers' Competence. *International Journal of Contemporary Islamic Education*. <https://doi.org/10.24239/ijcied.vol5.iss1.68>
- Naofa, M. I., Zahari, I., Maq, M. M., Sidik, J., Kholik, N., & Muhammad, H. Z. (2023). Learning Objectives of Islamic Religious Education in Schools: The Role of the Teacher and Its Implications Based on Relevant Study. *Dharmas Education Journal (De_journal)*. <https://doi.org/10.56667/dejournal.v4i2.1170>
- Sahri, S. (2023). Transformation of the Independent Curriculum: Strengthening the Islamic Dimension in Islamic Education. *Attanwir Jurnal Keislaman Dan Pendidikan*. <https://doi.org/10.53915/jurnalkeislamandanpendidikan.v14i2.419>
- Sugiyono. (2013). Metode penelitian kuantitatif, kualitatif dan R&D. Alfabeta.
- Yahyani, W. A., Kurnianto, R., & Ariyanto, A. (2020). The Role of Integrated Schools in Improving Islamic Education in Muslim Minority Areas of Cambodia. *Al-Hayat Journal of Islamic Education*. <https://doi.org/10.35723/ajie.v4i2.123>
- Zumzianah, Usman, U., & Wahab, W. (2024). Islamic Religious Education Based on Local Culture: Strategy for Developing Islamic Religious Education Materials in Madrasah Ibtidaiyah. *Ijgie (International Journal of Graduate of Islamic Education)*. <https://doi.org/10.37567/ijgie.v5i1.2727>