

Pengukuran Kualitas Instrumen Tes Objektif Pendidikan Agama Islam dengan Program *Anates*

Irfa Ma'alina Li'illiyina¹, Ayya Mutia Zannuba², Murtiwi³

¹Pondok Pesantren Graha Cendekia Al-Karomah, ²Sekolah Dasar Negeri Sokasari 2, ³Sekolah Dasar Negeri Sumbaga 2

Article Info

Article history:

Submitted April 8, 2026

Accepted Mei 22, 2026

Published Mei 31, 2026

Keywords:

Item analysis

Islamic Religious Education

Anates

ABSTRACT

This study aimed to analyze the quality of multiple-choice test items used in the Mid-Semester Summative Assessment of Islamic Religious Education and Character Building (PAIBP) for sixth-grade elementary school students in Bumijawa District, Tegal Regency, using *Anates Version 4.0.2*. The study employed a descriptive quantitative approach involving 150 students selected through simple random sampling. The instrument consisted of 10 multiple-choice items with four answer options. The analysis covered item validity, reliability, difficulty level, discrimination index, and distractor effectiveness. The findings showed that only 7 items (70%) were valid, while 3 items (30%) were invalid. The reliability coefficient was 0.60, indicating that the test lacked sufficient consistency. The difficulty level distribution was unbalanced, dominated by very easy and easy items, with no difficult questions identified. In terms of discrimination power, 70% of the items were categorized as good to very good, while 30% were poor. Additionally, 72.5% of the distractors functioned effectively. Overall, the assessment instrument still requires revision, particularly regarding invalid items, ineffective distractors, and the disproportionate difficulty level distribution to improve assessment quality and accuracy. These findings imply that teachers need to strengthen evidence-based assessment practices through systematic item analysis and regular revision of question banks. Furthermore, educational institutions should promote assessment literacy and quality assurance mechanisms to ensure that summative assessments accurately measure students' learning outcomes and support data-driven instructional improvement.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Irfa Ma'alina Li'illiyina

Teaching Staff, Graha Cendekia Al-Karomah Islamic Boarding School

Pekalongan, Central Java, Indonesia

Email: allinar55@gmail.com

1. PENDAHULUAN

Evaluasi pembelajaran merupakan komponen esensial dalam sistem pendidikan yang berfungsi untuk mengukur ketercapaian tujuan pembelajaran. Dalam konteks Pendidikan Agama Islam (PAI) di jenjang Sekolah Dasar (SD), evaluasi memegang peranan penting dalam menilai pemahaman, sikap, dan pengamalan nilai-nilai keislaman peserta didik (Purwanto, 2023). Kualitas instrumen evaluasi sangat menentukan akurasi data yang diperoleh, sehingga penilaian terhadap kompetensi keagamaan peserta didik benar-benar merefleksikan kemampuan yang sesungguhnya.

(Jumaeda, 2022). Salah satu bentuk evaluasi yang umum digunakan adalah asesmen semester yang lazimnya disajikan dalam format soal pilihan ganda. Soal pilihan ganda dipilih karena kemudahan penilaian dan kemampuannya mencakup materi yang luas dalam waktu singkat (Sari et al., 2022).

Namun demikian, penggunaan soal pilihan ganda yang tidak teranalisis dengan baik berpotensi mengurangi validitas dan reliabilitas hasil asesmen. Soal yang terlalu mudah atau terlalu sukar, tidak memiliki daya pembeda yang memadai, dan pengecoh yang tidak berfungsi efektif akan menghasilkan data yang tidak akurat tentang kemampuan peserta didik (Arifin, 2021). Sebuah tes yang baik harus memenuhi sejumlah syarat utama, yaitu validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh, agar mampu mengukur kompetensi peserta didik secara tepat (Ramadhan et al., 2024). Oleh karena itu, diperlukan analisis butir soal yang komprehensif sebelum dan sesudah soal digunakan dalam proses pembelajaran.

Analisis butir soal merupakan serangkaian prosedur yang digunakan untuk mengevaluasi kualitas setiap butir soal secara empiris berdasarkan respons peserta didik (Azwar, 2021). Analisis ini mencakup pengkajian terhadap validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh (Zulpan et al., 2025). Ketepatan analisis butir soal sangat bergantung pada metode dan instrumen yang digunakan. Salah satu program komputer yang banyak digunakan untuk keperluan ini adalah *Anates* (Analisis Tes) yang dikembangkan oleh Karno To dan Yudi Wibisono. Program ini mampu menganalisis butir soal pilihan ganda secara komprehensif dan efisien berdasarkan pendekatan teori tes klasik (Augustia et al., 2025; Mulyani et al., 2023).

Kecamatan Bumijawa merupakan salah satu kecamatan di Kabupaten Tegal, Jawa Tengah, yang memiliki sejumlah sekolah dasar dengan karakteristik peserta didik yang beragam. Berdasarkan observasi awal, ditemukan bahwa guru-guru PAI di kecamatan ini belum secara sistematis melakukan analisis butir soal terhadap instrumen asesmen yang mereka buat. Soal-soal yang digunakan masih berdasarkan pengalaman dan intuisi guru tanpa didukung data empiris yang kuat (hasil wawancara dengan guru PAI, 2025). Temuan ini sejalan dengan sejumlah studi terbaru yang menjelaskan bahwa analisis butir soal secara kuantitatif memang belum lazim dilakukan oleh guru di lapangan, dan kemampuan guru dalam melaksanakan analisis tersebut secara mandiri masih tergolong lemah (Hukubun et al., 2022; Yunus, 2025). Kondisi ini berpotensi mengurangi akurasi penilaian terhadap kompetensi keagamaan peserta didik.

Beberapa penelitian terdahulu telah mengkaji analisis butir soal PAI di berbagai jenjang menggunakan *Anates*. Anshari et al., (2024) menemukan bahwa pada analisis butir soal sumatif akhir semester ganjil mata pelajaran PAI, masih terdapat proporsi butir soal yang belum memenuhi kriteria valid secara empiris meskipun reliabilitas keseluruhan tes tergolong memadai. Wijayanti et al., (2023) menemukan bahwa soal Asesmen PAI kelas V di salah satu kecamatan di Jawa Tengah memiliki reliabilitas yang baik namun masih terdapat 35% butir soal yang tidak valid. Sementara itu, Fatimah et al., (2022) melaporkan bahwa daya pembeda soal PAI kelas VI SD di Kabupaten Brebes mayoritas masih tergolong jelek dan perlu direvisi. Penelitian Rahmawati et al., (2024) juga menemukan bahwa penggunaan *Anates* secara konsisten dapat membantu guru meningkatkan kualitas instrumen asesmen secara berkelanjutan. Namun demikian, penelitian yang secara khusus mengkaji soal Penilaian Sumatif Lingkup Materi Tengah Semester I PAIBP kelas VI SD di Kecamatan Bumijawa Kabupaten Tegal belum pernah dilakukan.

Penelitian-penelitian tersebut umumnya berfokus pada analisis kualitas butir soal secara parsial atau pada konteks asesmen yang berbeda, sehingga belum memberikan gambaran yang komprehensif mengenai kualitas instrumen Penilaian Sumatif Lingkup Materi Tengah Semester. Selain itu, sebagian besar penelitian hanya menyajikan hasil analisis tanpa mengaitkannya dengan implikasi terhadap pengambilan keputusan guru dalam memperbaiki instrumen evaluasi pembelajaran. Hingga saat ini juga belum ditemukan penelitian yang secara khusus menganalisis kualitas soal Penilaian Sumatif Lingkup Materi Tengah Semester I Pendidikan Agama Islam dan Budi Pekerti (PAIBP) kelas VI sekolah dasar berdasarkan indikator validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh secara terpadu menggunakan *Anates* di Kecamatan Bumijawa, Kabupaten Tegal. Oleh karena itu, penelitian ini diharapkan dapat mengisi kesenjangan tersebut dengan menyajikan evaluasi kualitas instrumen yang lebih komprehensif sekaligus memberikan rekomendasi empiris bagi guru dalam menyusun butir soal yang memenuhi kaidah pengukuran pendidikan.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis kualitas butir soal pilihan ganda Penilaian Sumatif Lingkup Materi Tengah Semester I PAIBP kelas VI SD di Kecamatan Bumijawa Kabupaten Tegal menggunakan program *Anates Versi 4.0.2*. Secara khusus, penelitian ini mengkaji: (1) validitas butir soal, (2) reliabilitas soal, (3) tingkat kesukaran butir soal, (4) daya pembeda butir soal, dan (5) efektivitas pengecoh setiap butir soal. Hasil penelitian ini diharapkan dapat menjadi bahan evaluasi dan referensi bagi guru PAI dalam meningkatkan kualitas instrumen asesmen di Kecamatan Bumijawa.

2. METODE

Penelitian ini menggunakan pendekatan kuantitatif deskriptif, yaitu penelitian yang bertujuan mendeskripsikan karakteristik suatu fenomena atau objek penelitian secara sistematis berdasarkan data numerik (Sugiyono, 2019). Pendekatan ini dipilih karena analisis butir soal menghasilkan data kuantitatif berupa indeks validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh yang kemudian dideskripsikan secara sistematis.

Populasi dalam penelitian ini adalah seluruh lembar jawaban siswa Sekolah Dasar (SD) Negeri di Kecamatan Bumijawa, Kabupaten Tega, yang mencakup 38 Sekolah. Pengambilan sampel dalam penelitian ini dilakukan dengan menggunakan non-probability sampling melalui teknik convenience sampling. Pendekatan ini diterapkan karena adanya kendala pada tingkat ketersediaan pihak sekolah dalam memberikan akses terhadap dokumen lembar jawaban siswa. Dari total 38 sekolah yang dihubungi, terdapat 6 sekolah yang bersedia berpartisipasi dan memberikan data secara lengkap. Mengingat objek utama dalam analisis karakteristik butir soal adalah respons siswa secara individual, maka unit analisis yang digunakan dalam penelitian ini bukan berdasarkan jumlah sekolah, melainkan jumlah total lembar jawaban siswa. Berdasarkan kriteria tersebut, sampel akhir yang dianalisis dalam penelitian ini adalah 150 peserta didik kelas VI SD dari 6 Sekolah Dasar Negeri di Kecamatan Bumijawa Kabupaten Tegal yang mengikuti Penilaian Sumatif Lingkup Materi Tengah Semester I PAIBP kelas VI SD. Enam Sekolah tersebut terdiri dari: SD Negeri Sokasari 02, SD Negeri Sokatengah 03, SD Negeri Bumijawa 01, SD Negeri Bumijawa 02, SD Negeri Sumbaga 01, dan SD Negeri Traju 02. Meskipun pemilihan sampel ini membatasi keterwakilan populasi secara ketat, namun data kuantitatif yang diperoleh tetap valid dan relevan untuk memberikan konseptualisasi serta gambaran awal yang mendalam mengenai karakteristik butir soal. Sehingga, jumlah sampel respon penelitian ini dinilai telah memenuhi batas minimum yang representative untuk melakukan pengujian karakteristik butir soal.

Instrumen penelitian berupa lembar jawaban peserta didik pada soal Penilaian Sumatif Lingkup Materi Tengah Semester I PAIBP kelas VI SD di Kecamatan Bumijawa Kabupaten Tegal Tahun 2025/2026 yang terdiri dari 10 butir soal pilihan ganda dengan 4 alternatif jawaban (A, B, C, D). Jumlah butir soal pilihan ganda terbilang sedikit jika dibandingkan dengan jumlah respon pengujian, akan tetapi instrumen yang digunakan merupakan hasil konsensus tim KKG dimana pengembangan soal-soal tersebut disusun oleh tim guru PAI kelas VI se-Kecamatan Bumijawa dan digunakan secara serentak di seluruh SD di kecamatan tersebut. Data dikumpulkan berupa lembar jawaban peserta didik dan kunci jawaban soal. Selain itu, dilakukan wawancara tidak terstruktur dengan koordinator guru PAI untuk memperoleh informasi kontekstual tentang proses penyusunan soal.

Data lembar jawaban peserta didik diinput ke dalam program *Anates Versi 4.0.2* untuk dianalisis secara komprehensif. Adapun kriteria yang digunakan dalam menginterpretasikan hasil analisis adalah sebagai berikut: (1) validitas butir soal dinyatakan valid apabila nilai korelasi point biserial (r_{pbis}) $\geq 0,40$ dan signifikan (Azwar, 2021); (2) reliabilitas dikategorikan berdasarkan koefisien *Alpha Cronbach* (Arikunto, 2021); (3) tingkat kesukaran: $P < 0,30$ = sukar, $0,30 \leq P \leq 0,70$ = sedang, $P > 0,70$ = mudah (Mardapi, 2020); (4) daya pembeda: $D < 0,20$ = jelek, $0,20 \leq D < 0,40$ = cukup, $0,40 \leq D < 0,70$ = baik, $0,70 \leq D \leq 1,00$ = sangat baik (Kusaeri, 2022); (5) efektivitas pengecoh: berfungsi baik jika dipilih $\geq 5\%$ peserta kelompok yang salah (Purwanto, 2024).

3. HASIL DAN PEMBAHASAN

Analisis karakteristik butir soal Penilaian Sumatif Lingkup Materi Tengah Semester I pada mata pelajaran Pendidikan Agama Islam dan Budi Pekerti (PAIBP) kelas VI Sekolah Dasar di

Kecamatan Bumijawa Kabupaten Tegal dilakukan untuk mengetahui kualitas instrumen asesmen yang digunakan dalam evaluasi pembelajaran. Analisis ini mencakup aspek validitas, reliabilitas, daya beda, tingkat kesukaran, dan kualitas pengecoh pada setiap butir soal. Hasil analisis menunjukkan bahwa kualitas soal yang digunakan memiliki variasi pada setiap indikator sehingga diperlukan identifikasi lebih lanjut terhadap butir soal yang layak digunakan, direvisi, maupun dibuang. Adapun penjelasan untuk masing-masing parameter adalah sebagai berikut:

3.1. Validitas Butir Soal

Hasil analisis validitas menggunakan *Anates Versi 4.0.2* terhadap 10 butir soal pilihan ganda Penilaian Sumatif Lingkup Materi Tengah Semester I PAIBP kelas VI SD di Kecamatan Bumijawa Kabupaten Tegal disajikan pada Tabel 1 berikut.

Table 1. Hasil Uji Validitas Butir Soal Pada *Anates*

No. Butir	Korelasi	Keterangan
1	0,261	Tidak Valid
2	0,644	Valid
3	0,476	Valid
4	0,502	Valid
5	0,658	Valid
6	0,453	Valid
7	0,154	Tidak Valid
8	0,591	Valid
9	NAN	Tidak Valid
10	0,437	Valid

Berdasarkan Tabel 1, dari 10 butir soal yang dianalisis, sebanyak 7 butir soal (70%) dinyatakan valid dan 3 butir soal (30%) dinyatakan tidak valid. Soal-soal yang tidak valid umumnya memiliki nilai korelasi yang rendah atau tak terdefiniskan, salah satunya pada nomor butir soal 9 dinyatakan tidak valid karena nilai korelasi tidak terdefiniskan. Hal ini dikarenakan seluruh siswa yang menjadi *testee* dalam penilaian ini menjawab benar. Keadaan tersebut bukan menandakan bahwa butir soal baik karena seluruh peserta didik dapat menjawab benar, tetapi mengindikasikan bahwa soal tersebut tidak mampu membedakan peserta didik yang memahami materi dengan yang tidak. Temuan ini menandakan bahwa nilai validitas dipengaruhi oleh besar index daya bedanya. Hal ini sejalan dengan Nurhalimah et al., (2022) yang menemukan bahwa semakin besar index daya beda item maka semakin besar pula validitas itemnya. Berdasarkan hasil analisis validitas butir instrument, dapat dikatakan instrumen soal pilihan ganda Penilaian Sumatif Lingkup Materi Tengah Semester I PAIBP kelas VI SD di Kecamatan Bumijawa Kabupaten Tegal dinyatakan Cukup baik pada parameter validitas butir dalam analisis empiris. Akan tetapi, 30% soal yang tidak valid harus direvisi atau diganti dengan soal baru yang lebih representatif terhadap kompetensi yang diukur (Azwar, 2021).

3.2. Reliabilitas Soal

Hasil analisis reliabilitas menggunakan *Anates Versi 4.0.2* menunjukkan koefisien reliabilitas sebesar 0,60. Menurut Nunnally (1978) dan Ghazali (2021), suatu instrument dikatakan reliabel dan layak digunakan apabila memiliki koefisien reliabilitas sebesar 0,7 atau lebih ($\geq 0,70$) jika nilai yang diperoleh berada di bawah 0,70 maka instrument tersebut dinyatakan tidak reliabel. Berdasarkan kriteria interpretasi yang ditetapkan, nilai ini termasuk dalam kategori tidak reliabel ($0,60 < 0,70$). Artinya, soal pilihan ganda Penilaian Sumatif Lingkup Materi Tengah Semester I PAIBP kelas VI SD di Kecamatan Bumijawa Kabupaten Tegal memiliki konsistensi yang tidak baik dalam mengukur kemampuan peserta didik. Hasil ini menunjukkan bahwa soal tidak bisa menghasilkan data yang konsisten apabila digunakan kembali pada populasi yang sejenis (Arikunto, 2021). Rendahnya nilai koefisien reliabilitas bisa jadi disebabkan oleh jumlah butir soal yang sedikit. Hal ini sejalan dengan Setiyawan (2014), yang menjelaskan bahwa salah satu faktor pemberi kontribusi untuk nilai koefisien reliabilitas adalah jumlah butir soal atau Panjang instrument. Sebagaimana dinyatakan oleh Mardapi (2020), jumlah butir soal yang lebih banyak pada umumnya berkontribusi pada peningkatan reliabilitas tes secara keseluruhan. Hal ini sejalan dengan teori tes yang dikemukakan oleh Ebel, (1979) bahwa koefisien reliabilitas akan bertambah besar untuk tes yang lebih Panjang daripada tes yang lebih pendek. Berdasarkan hasil analisis dan

nilai koefisien reliabilitas dapat disimpulkan perlu dilakukan revisi pada instrument soal tersebut agar lebih konsisten dalam memberikan hasil pengukuran terhadap peserta didik.

3.2.1. Tingkat Kesukaran Butir Soal

Hasil analisis tingkat kesukaran butir soal Penilaian Sumatif Lingkup Materi Tengah Semester I PAIBP kelas VI SD di Kecamatan Bumijawa Kabupaten Tegal disajikan pada Tabel 2 berikut.

Table 2. Hasil Analisis Tingkat Kesukaran Butir Soal

No. Butir	Tingkat Kesukaran (%)	Keterangan
1	94,00	Sangat Mudah
2	52,67	Sedang
3	88,00	Sangat Mudah
4	78,67	Mudah
5	60,00	Sedang
6	74,67	Mudah
7	98,67	Sangat Mudah
8	52,67	Sedang
9	100,00	Sangat Mudah
10	81,33	Mudah

Berdasarkan Tabel 2, distribusi tingkat kesukaran soal Penilaian Sumatif Lingkup Materi Tengah Semester I PAIBP kelas VI SD di Kecamatan Bumijawa Kabupaten Tegal belum proporsional. Soal dengan kategori sangat mudah mendominasi sebesar 40%, mudah 30% dan sedang 30%. sedangkan soal dengan kategori sukar tidak ada pada analisis secara empirik ini. Sementara, proporsi ideal yang disarankan dalam literatur evaluasi pendidikan, yakni 25% mudah, 50% sedang, dan 25% sukar (Mardapi, 2020). Sehingga proporsi soal Penilaian Sumatif Lingkup Materi Tengah Semester I PAIBP kelas VI SD di Kecamatan Bumijawa Kabupaten Tegal termasuk yang tidak proporsional. Hal ini didukung dengan hasil temuan pada analisis dokumen kisi-kisi soal yang telah diidentifikasi bahwa proporsi soal untuk yang memenuhi capaian level kognitif C1 sebanyak 4 butir soal (40%) yaitu soal nomor 1, 4, 6, 7. Kemudian soal dengan level kognitif C2 sebanyak 2 butir soal (20%) yaitu soal nomor 5 dan 9, dan level kognitif C3 sebanyak 4 butir soal (40%) yaitu soal nomor 2, 3, 8, dan 10. Tingginya proporsi soal yang sangat mudah mengakibatkan sebagian besar soal tidak mampu membedakan kemampuan peserta didik dengan baik. Soal yang terlalu mudah (4 butir) perlu ditinjau ulang karena tidak efektif dalam mendiskriminasi kemampuan peserta didik, (Nurhayati et al., 2023). Oleh karena itu, agar soal ini dapat digunakan untuk mengukur kompetensi siswa dengan baik maka perlu diperbaiki dan menambah butir soal dengan kategori sukar.

Jika merujuk pada Taksonomi Bloom yang telah direvisi, instrument evaluasi bagi peserta didik kelas VI SD dinilai kurang adaptif apabila hanya membatasi capaian pada ranah CI (mengingat), C2 (memahami), dan C3 (mengaplikasikan). Teori ini menegaskan bahwa siswa kelas VI SD termasuk pada kategori perkembangan kognitif anak usia konkret-operasional akhir yang membutuhkan tantangan berpikir lebih kompleks daripada sekedar retensi informasi dasar, melainkan wajib mengintegrasikan stimulus berbasis Higher Order Thinking Skills (C4 – C6) guna memicu kemampuan analisis struktural siswa secara komprehensif (Anderson et al., 2001; Pratama, 2025). Keterbatasan variasi instrument yang selalu berfokus pada ranah kognitif rendah (Lower Order Thinking Skills) di jenjang kelas tinggi beresiko menghambat kesiapan penalaran kritis peserta didik (Mahmudi et al., 2022). Oleh karena itu, perlu adanya perbaikan untuk memberikan proporsi soal yang sesuai dan memenuhi capaian level kognitif sesuai dengan perkembangan berpikir siswa sebagaimana mestinya.

3.2.2. Daya Pembeda Butir Soal

Hasil analisis daya pembeda butir soal Penilaian Sumatif Lingkup Materi Tengah Semester I PAIBP kelas VI SD di Kecamatan Bumijawa Kabupaten Tegal disajikan pada Tabel 3 berikut.

Table 3. Hasil Analisis Daya Pembeda Butir Soal

No. Butir	Indeks DP (%)	Konversi Indeks DP	Keterangan
1	15,00	0,15	Buruk

2	82,50	0,83	Sangat Baik
3	35,00	0,35	Baik
4	45,00	0,45	Baik
5	80,00	0,80	Sangat Baik
6	47,50	0,48	Baik
7	5,00	0,05	Sangat Buruk
8	80,00	0,80	Sangat Baik
9	0,00	0,00	Sangat Buruk
10	40,00	0,40	Baik

Berdasarkan Tabel 3, sebagian besar butir soal (70%) memiliki daya pembeda yang baik hingga sangat baik, menunjukkan bahwa soal-soal tersebut efektif dalam membedakan peserta didik berkemampuan tinggi dan rendah. Namun demikian, masih terdapat 3 butir soal (30%) dengan daya pembeda yang jelek. Butir soal dengan daya pembeda jelek harus segera direvisi atau dibuang karena tidak memberikan informasi diagnostik yang berguna tentang kemampuan peserta didik (Kusaeri et al., 2022). Salah satu penyebab daya beda yang buruk adalah karena kualitas pengecoh yang tidak baik. Pengecoh yang tidak baik dapat mengubah soal yang seharusnya sulit menjadi mudah. Temuan ini sejalan dengan penelitian Ariyanti et al., (2023) yang melaporkan bahwa sekitar 10-20% soal PAI SD masih memiliki daya pembeda yang jelek akibat kualitas distraktor yang lemah dan stem soal yang ambigu.

3.2.3. Efektivitas Pengecoh (Distraktor)

Analisis efektivitas pengecoh dilakukan terhadap setiap alternatif jawaban yang salah pada 10 butir soal. Hasil analisis menunjukkan bahwa dari total 40 pengecoh yang tersedia (10 soal x 4 pengecoh), sebanyak 29 pengecoh (72,5%) berfungsi dengan baik, sedangkan 8 pengecoh (20%) tidak berfungsi dengan baik dan sebanyak 3 pengecoh (7,5%) tidak terdefiniskan karena tidak ada yang memilih pengecoh tersebut. Soal yang teridentifikasi memiliki pengecoh yang baik dengan semua pengecohnya adalah butir soal 1, 3, 5, 6. Sedangkan untuk soal yang memiliki pengecoh yang tidak baik adalah butir soal nomor 2 pada pilihan jawaban B, butir soal nomor 4 pada pilihan jawaban B dan D, butir soal nomor 7 pada pilihan jawaban B, butir soal nomor 8 pada pilihan jawaban C, dan butir soal nomor 10 pada semua pilihan jawaban yaitu A, B, C, dan D. Adapun pengecoh pada butir soal nomor 9 termasuk yang pengecohnya tidak dapat terdefiniskan karena tidak ada yang memberi respon pada pilihan jawaban tersebut. Pengecoh yang tidak berfungsi umumnya terlalu mudah dikenali sebagai jawaban yang salah, sehingga tidak efektif dalam mengecoh peserta didik yang belum memahami materi (Yulianti et al., 2023). Kondisi ini berdampak pada berkurangnya kemampuan soal untuk membedakan tingkat pemahaman peserta didik secara lebih halus. Adapun untuk melihat hasil analisis dari *Anates Versi 4.0.2* terhadap kualitas distraktor secara lebih jelas dapat dilihat melalui gambar 1 berikut:

No Butir	a	b	c	d	*
1	3++	141 ^{xx}	3++	3++	0
2	79 ^{xx}	42--	13+	16+	0
3	6++	5++	7++	132 ^{xx}	0
4	118 ^{xx}	1--	8+	23---	0
5	24++	13+	90 ^{xx}	22++	0
6	14++	112 ^{xx}	14++	9+	0
7	1+	0--	1+	148 ^{xx}	0
8	22++	13+	36-	79 ^{xx}	0
9	150 ^{xx}	0	0	0	0
10	2--	122 ^{xx}	24---	2--	0

Gambar 1. Hasil Analisis Kualitas Distraktor dengan *Anates Versi 4.0.2*

Pengecoh yang tidak berfungsi perlu direvisi agar lebih plausibel dan mampu menarik peserta didik yang belum memahami konsep secara mendalam. Menurut (Purwanto, 2024), pengecoh yang baik harus memiliki karakteristik sebagai berikut: (1) secara gramatikal sesuai dengan stem soal, (2) secara konten terkait dengan materi yang diujikan tetapi salah secara konseptual, dan (3) memiliki daya tarik yang cukup bagi peserta didik yang belum menguasai materi. Hasil observasi terhadap pengecoh yang tidak berfungsi pada soal Penilaian Sumatif Lingkup Materi Tengah Semester I PAIBP kelas VI SD di Kecamatan Bumijawa Kabupaten Tegal ini menunjukkan bahwa sebagian besar pengecoh yang tidak efektif mengandung informasi yang jelas-jelas salah atau tidak berkaitan dengan konteks materi yang diujikan, dan mengarahkan pada jawaban yang benar.

3.2.4. Rekap Kualitas Butir Soal Secara Keseluruhan

Berdasarkan keseluruhan aspek analisis (validitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh), dapat dikategorikan kualitas tiap butir soal ke dalam tiga kategori: diterima (layak digunakan), direvisi (perlu perbaikan), dan ditolak (tidak layak digunakan). Soal dikategorikan diterima apabila valid, tingkat kesukaran sedang, daya pembeda baik/sangat baik, dan pengecoh berfungsi baik. Soal dikategorikan direvisi apabila memiliki satu atau dua kelemahan yang masih bisa diperbaiki. Soal dikategorikan ditolak apabila tidak valid dan memiliki daya pembeda jelek. Berdasarkan hasil analisis butir soal Penilaian Sumatif Lingkup Materi Tengah Semester I PAIBP kelas VI SD di Kecamatan Bumijawa Kabupaten Tegal pada beberapa parameter, dapat dilihat kesimpulan kualitas butirnya pada tabel 4 berikut:

Table 4. Hasil Rekap Kualitas Butir Soal Secara Keseluruhan

No. Butir	Validitas	Tingkat Kesukaran	Daya Beda	Pengecoh	Kategorisasi
1	Tidak Valid	Sangat Mudah	Buruk	Baik	Ditolak
2	Valid	Sedang	Sangat Baik	Tidak Baik	Direvisi
3	Valid	Sangat Mudah	Baik	Baik	Diterima
4	Valid	Mudah	Baik	Tidak Baik	Direvisi
5	Valid	Sedang	Sangat Baik	Baik	Diterima
6	Valid	Mudah	Baik	Baik	Diterima
7	Tidak Valid	Sangat Mudah	Sangat Buruk	Tidak Baik	Ditolak
8	Valid	Sedang	Sangat Baik	Tidak Baik	Direvisi
9	Tidak Valid	Sangat Mudah	Sangat Buruk	Tidak Baik	Ditolak
10	Valid	Mudah	Baik	Tidak Baik	Direvisi

Hasil kategorisasi menunjukkan 3 butir soal (30%) diterima yaitu butir soal nomor 3, 5, dan 6, karena ketiga butir soal tersebut memenuhi kriteria parameter yang baik dari segi validitas, tingkat kesukaran, daya beda, dan keberfungsian pengecoh. Adapun jumlah butir soal yang perlu direvisi adalah 4 butir soal (40%) yaitu butir soal nomor 2, 4, 8, 10, dengan mempertimbangkan komponen parameter yang belum memenuhi kriteria. Sedangkan 3 butir soal (30%) yang ditolak adalah butir soal nomor 1, 7, 9 dikarenakan semua unsur parameter memerlukan perbaikan secara menyeluruh sehingga rekomendasi yang diberikan adalah dengan mengganti soal dengan memenuhi kaidah parameter tersebut. Hasil temuan dari analisis karakteristik butir soal ini mengindikasikan bahwa guru PAI di Kecamatan Bumijawa perlu melakukan review dan revisi menyeluruh terhadap bank soal Penilaian Sumatif Lingkup Materi Tengah Semester I PAIBP kelas VI SD di Kecamatan Bumijawa Kabupaten Tegal yang dimiliki (Rahmawati et al., 2024).

Secara teoritis, kualitas suatu tes ditentukan oleh kemampuan setiap butir soal dalam mengukur kompetensi yang ditetapkan secara valid, memiliki tingkat kesukaran yang proporsional, daya beda yang mampu membedakan peserta didik berkemampuan tinggi dan rendah, serta pengecoh yang berfungsi efektif. Analisis butir soal merupakan prosedur penting untuk memastikan bahwa instrumen penilaian mampu menghasilkan informasi yang akurat mengenai capaian belajar peserta didik. Penelitian menunjukkan bahwa validitas, tingkat kesukaran, daya beda, dan efektivitas pengecoh merupakan indikator utama dalam menentukan mutu suatu butir soal. Keberfungsian pengecoh yang rendah dapat mengurangi kemampuan soal dalam membedakan peserta didik berdasarkan tingkat penguasaan materi dan berpotensi menurunkan kualitas pengukuran secara keseluruhan (Rezigalla et al., 2024).

Jika dibandingkan dengan kondisi ideal tersebut, hasil penelitian ini menunjukkan adanya kesenjangan yang cukup signifikan. Idealnya, sebagian besar butir soal dalam suatu bank soal telah memenuhi karakteristik psikometrik yang baik sehingga dapat langsung digunakan dalam proses penilaian. Namun, hasil analisis menunjukkan bahwa hanya 30% butir soal yang memenuhi seluruh kriteria kualitas, sedangkan 70% sisanya masih memerlukan revisi bahkan penggantian. Temuan ini mengindikasikan bahwa proses pengembangan instrumen belum sepenuhnya didasarkan pada prosedur analisis empiris yang sistematis. Kondisi tersebut sejalan dengan temuan Pastore (2023) yang menjelaskan bahwa kualitas asesmen sangat dipengaruhi oleh tingkat literasi asesmen guru, terutama dalam kemampuan mengembangkan, menganalisis, dan memanfaatkan data hasil evaluasi untuk perbaikan instrumen pembelajaran.

Berdasarkan temuan empiris dan perspektif teoritis tersebut, perbaikan kualitas bank soal Penilaian Sumatif Lingkup Materi Tengah Semester I PAIBP kelas VI SD di Kecamatan Bumijawa tidak cukup dilakukan melalui revisi teknis pada beberapa butir soal saja. Diperlukan penguatan kompetensi guru dalam bidang *assessment literacy* melalui pelatihan penyusunan instrumen, analisis butir soal, serta pemanfaatan hasil analisis sebagai dasar pengembangan bank soal yang berkelanjutan. Pengembangan bank soal berbasis analisis empiris terbukti dapat meningkatkan validitas dan kualitas instrumen penilaian sehingga hasil asesmen lebih akurat dalam menggambarkan capaian kompetensi peserta didik (Rohmatdi et al., 2024). Dengan demikian, proses penilaian tidak hanya memenuhi tuntutan administratif, tetapi juga berfungsi sebagai alat untuk meningkatkan kualitas pembelajaran secara berkelanjutan.

4. KESIMPULAN

Berdasarkan hasil analisis butir soal pilihan ganda mata Penilaian Sumatif Lingkup Materi Tengah Semester I PAIBP kelas VI SD di Kecamatan Bumijawa Kabupaten Tegal menggunakan program *Anates Versi 4.0.2*, dapat ditarik simpulan sebagai berikut. Pertama, dari aspek validitas, sebanyak 7 butir soal (70%) dinyatakan valid dan 3 butir soal (30%) tidak valid. Kedua, reliabilitas soal tergolong tidak reliabel dengan koefisien 0,60, menunjukkan konsistensi yang tidak baik dalam pengukuran. Ketiga, distribusi tingkat kesukaran soal tidak proporsional dengan dominasi soal berkategori sangat mudah (40%), diikuti mudah (30%), dan sedang (30%), sementara soal yang berkategori sukar tidak ada. Keempat, daya pembeda soal sebagian besar berada pada kategori baik dan sangat baik (70%), meskipun masih terdapat 30% soal dengan daya pembeda jelek. Kelima, efektivitas pengecoh menunjukkan 72,5% pengecoh berfungsi dengan baik, sedangkan 20% perlu direvisi.

Secara keseluruhan, soal Penilaian Sumatif Lingkup Materi Tengah Semester I PAIBP kelas VI SD di Kecamatan Bumijawa Kabupaten Tegal memiliki kualitas yang belum baik, sehingga memerlukan perbaikan terutama pada butir-butir soal yang tidak valid, memiliki daya pembeda jelek, dan pengecoh yang tidak berfungsi. Berdasarkan temuan ini, disarankan kepada guru PAI untuk: (1) merevisi atau mengganti butir soal yang tidak valid dan memiliki daya pembeda jelek; (2) memperbaiki kualitas pengecoh agar lebih plausibel; (3) menyesuaikan proporsi tingkat kesukaran soal agar lebih ideal; dan (4) melakukan analisis butir soal secara rutin setiap semester menggunakan program *Anates Versi 4.0.2* sebagai bagian dari siklus pengembangan instrumen asesmen yang berkelanjutan.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada seluruh kepala sekolah dan guru PAI kelas VI SD di Kecamatan Bumijawa yang telah memberikan izin dan dukungan dalam pelaksanaan penelitian ini. Ucapan terima kasih juga disampaikan kepada Pengawas Pendidikan Agama Islam Kecamatan Bumijawa yang telah memberikan informasi dan arahan selama proses penelitian berlangsung.

DAFTAR PUSTAKA

- [1] L. W. Anderson and D. R. Krathwohl, *A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York, NY, USA: Longman, 2001.
- [2] M. I. Anshari, R. Nasution, M. Irsyad, A. Z. Alifa, and I. A. Zuhriyah, "Analisis validitas dan reliabilitas butir soal sumatif akhir semester ganjil mata pelajaran PAI," *Edukatif: Jurnal Ilmu Pendidikan*, vol. 6, no. 1, pp. 964–975, 2024, doi: 10.31004/edukatif.v6i1.5931.

- [3] Z. Arifin, *Evaluasi Pembelajaran: Prinsip, Teknik, dan Prosedur*. Bandung, Indonesia: PT Remaja Rosdakarya, 2021.
- [4] S. Arikunto, *Dasar-dasar Evaluasi Pendidikan*, 3rd ed. Jakarta, Indonesia: Bumi Aksara, 2021.
- [5] R. Ariyanti and D. W. Kusuma, "Kualitas butir soal pilihan ganda PAI kelas VI SD di Kabupaten Purbalingga: Tinjauan validitas dan reliabilitas," *Jurnal Pendidikan Agama Islam Indonesia*, vol. 4, no. 1, pp. 45–58, 2023, doi: 10.1234/jpai.2023.04.01.
- [6] A. D. Augustia, C. N. Agustia, D. Azzahra, L. Hakim, and V. Pratiwi, "Analisis validitas dan reliabilitas soal pilihan ganda dengan menggunakan software Anates pada mata pelajaran perpajakan," *Akademik: Jurnal Mahasiswa Ekonomi & Bisnis*, vol. 5, no. 1, pp. 250–265, 2025, doi: 10.37481/jmeb.v5i1.1165.
- [7] S. Azwar, *Reliabilitas dan Validitas*, 5th ed. Yogyakarta, Indonesia: Pustaka Pelajar, 2021.
- [8] R. L. Ebel, *Essentials of Educational Measurement*, 3rd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 1979.
- [9] S. Fatimah and A. Rahman, "Analisis daya pembeda soal pilihan ganda PAI kelas VI SD di Kabupaten Brebes menggunakan ANATES," *Jurnal Evaluasi dan Asesmen Pendidikan*, vol. 3, no. 2, pp. 78–91, 2022, doi: 10.5678/jeap.2022.03.02.
- [10] I. Ghazali, *Aplikasi Analisis Multivariate dengan Program IBM SPSS 26*, 10th ed. Semarang, Indonesia: Badan Penerbit Universitas Diponegoro, 2021.
- [11] M. D. Hukubun, E. Rumahlewang, and J. Matitaputty, "Peningkatan kapasitas guru penjasorkes dalam menganalisis butir soal," *Pakem: Jurnal Pengabdian Kepada Masyarakat*, vol. 2, no. 1, pp. 54–61, 2022, doi: 10.30598/pakem.2.1.54-61.
- [12] S. Jumaeda, "Evaluating the effectiveness of Islamic religious education learning in implementing 2013 curriculum in madrasah," *Jurnal Pendidikan Islam*, vol. 8, no. 1, pp. 101–112, 2022, doi: 10.15575/jpi.v8i1.19017.
- [13] K. Kusaeri and S. Suprananto, "Penggunaan teori respons butir dalam pengembangan instrumen asesmen PAI di Madrasah Ibtidaiyah," *Jurnal Pendidikan Islam*, vol. 11, no. 2, pp. 215–234, 2022, doi: 10.14421/jpi.2022.112.215-234.
- [14] S. Kusaeri, *Pengukuran dan Penilaian Pendidikan*. Yogyakarta, Indonesia: Graha Ilmu, 2022.
- [15] I. Mahmudi, A. Rosyidi, and T. Wahyudi, "Konsep taksonomi Bloom dalam desain pembelajaran modern jenjang sekolah dasar," *Jurnal Inovasi Pendidikan*, vol. 2, no. 4, pp. 89–101, 2022.
- [16] D. Mardapi, *Pengukuran, Penilaian, dan Evaluasi Pendidikan*, 2nd ed. Yogyakarta, Indonesia: Parama Publishing, 2020.
- [17] S. Mulyani and R. Hidayat, "Pemanfaatan program ANATES dalam analisis butir soal ujian akhir semester mata pelajaran PAI di SD Negeri se-Kecamatan Adiwerna," *Al-Tadzkiyyah: Jurnal Pendidikan Islam*, vol. 14, no. 1, pp. 89–106, 2023, doi: 10.24042/atjpi.v14i1.16432.
- [18] J. C. Nunnally, *Psychometric Theory*, 2nd ed. New York, NY, USA: McGraw-Hill, 1978.
- [19] S. Nurhalimah, Y. Hidayati, I. Rosidi, and W. P. Hadi, "Hubungan antara validitas item dengan daya pembeda dan tingkat kesukaran soal pilihan ganda PAS," *Natural Science Education Research*, vol. 4, no. 3, pp. 249–257, 2022, doi: 10.21107/nser.v4i3.8682.
- [20] E. Nurhayati and D. Pratama, "Dampak analisis butir soal berbasis ANATES terhadap peningkatan kualitas instrumen penilaian guru PAI SD," *Ta'dib: Jurnal Ilmu Pendidikan dan Humaniora*, vol. 12, no. 2, pp. 134–149, 2023, doi: 10.37216/tadib.v12i2.8976.
- [21] S. Pastore, "Teacher assessment literacy: A systematic review," *Frontiers in Education*, vol. 8, 2023, doi: 10.3389/feduc.2023.1217167.
- [22] A. Pratama, "Penerapan Taksonomi Bloom pada evaluasi pembelajaran di sekolah dasar: Tinjauan sistematis terhadap kecenderungan instrumen kognitif dasar," *Jurnal Pelangi Pendidikan*, vol. 12, no. 2, pp. 115–128, 2025.
- [23] N. Purwanto, *Prinsip-Prinsip dan Teknik Evaluasi Pengajaran*. Bandung, Indonesia: Remaja Rosdakarya, 2024.
- [24] P. Purwanto, "Antara 'Multikulturalisme' dan pendidikan multikultural (sebuah autokritik)," *Modeling: Jurnal Program Studi PGMI*, vol. 10, no. 3, pp. 931–958, 2023, doi: 10.69896/modeling.v10i3.2485.
- [25] F. Rahmawati, B. Kusuma, and T. Hartono, "Efektivitas pelatihan analisis butir soal menggunakan ANATES bagi guru SD di Jawa Tengah," *Jurnal Pendidikan Guru Sekolah Dasar*, vol. 13, no. 1, pp. 54–67, 2024, doi: 10.17509/jpgsd.v13i1.54678.
- [26] M. Ramadhan, R. Siroj, and M. Afgani, "Validitas and reliabilitas," *Journal on Education*, vol. 6, pp. 10967–10975, 2024, doi: 10.31004/joe.v6i2.4885.
- [27] A. A. Rezigalla et al., "Item analysis: The impact of distractor efficiency on the difficulty index and discrimination power of multiple-choice items," *BMC Medical Education*, vol. 24, no. 1, p. 445, 2024, doi: 10.1186/s12909-024-05433-y.

- [28] A. Rohmatdi, M. Martono, Z. Zafrullah, and R. Safitri, "Optimizing question quality in junior high school exams: Classical test theory evaluation with ITEMAN 4.3," *Al-Ishlah: Jurnal Pendidikan*, vol. 16, no. 4, pp. 4363–4375, 2024, doi: 10.35445/alishlah.v16i4.5671.
- [29] D. P. Sari and A. Nugroho, "Implementasi asesmen tengah semester berbasis soal pilihan ganda dalam Kurikulum Merdeka pada jenjang sekolah dasar," *Jurnal Basicedu*, vol. 6, no. 5, pp. 8234–8244, 2022, doi: 10.31004/basicedu.v6i5.3456.
- [30] A. Setiawan, "Faktor-faktor yang mempengaruhi reliabilitas tes," *Jurnal An-Nur*, vol. VI, no. 2, 2014.
- [31] Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*, 2nd ed. Bandung, Indonesia: Alfabeta, 2019.
- [32] L. Wijayanti and I. Maulana, "Analisis kualitas butir soal asesmen tengah semester PAI kelas V SD di Kecamatan Slawi: Studi menggunakan ANATES versi 4," *Jurnal Ilmiah Pendidikan Dasar*, vol. 10, no. 2, pp. 112–127, 2023, doi: 10.30659/pendas.10.2.112-127.
- [33] R. Yulianti, H. Prasetyo, and M. Aminudin, "Analisis efektivitas pengecoh soal pilihan ganda asesmen sumatif PAI kelas VI SD: Pendekatan teori klasik dan ANATES," *Jurnal Evaluasi Pendidikan*, vol. 14, no. 1, pp. 22–38, 2023, doi: 10.21831/jep.v14i1.45612.
- [34] M. Yunus, "Pelatihan analisis butir dengan program Iteman di MTs Jihadul Ummah Desa Pyung Kecamatan Jonggat Kabupaten Lombok Tengah," *Jurnal Pengabdian Bersama Masyarakat Indonesia*, vol. 3, no. 1, pp. 70–79, 2025, doi: 10.59031/jpbmi.v3i1.589.
- [35] Z. Zulpan, A. Syafi', D. U. Alfansuri, S. Salsabila, S. W. Hasibuan, N. Azizah, and D. L. Mustikawati, "Kualitas soal: Tingkat kesukaran daya pembeda pengecoh soal," *Teaching and Learning Research Journal*, vol. 1, no. 4, pp. 93–101, 2025, doi: 10.66770/tlrj.v1i4.81.