

EVALUASI KUALITAS BUTIR SOAL ASESMEN PENDIDIKAN PANCASILA MELALUI APLIKASI ANATES DI SDN TRITIH WETAN 04

Adrianus Anjani

SDN Tritih Wetan 04 Jeruklegi Cilacap

anjaniadrianus@gmail.com

Istikomah

SDN Cinangsi 03 Gandrungmangu Cilacap

istikomah.28.iis.kk@gmail.com

Abstrak

Penelitian ini bertujuan untuk mengevaluasi kualitas soal pada asesmen Pendidikan Pancasila melalui analisis butir soal dengan bantuan aplikasi ANATES. Subjek penelitian adalah hasil Asesmen Sumatif Lingkup Materi II dari 18 siswa kelas V SDN Tritih Wetan 04 tahun pelajaran 2024/2025. Instrumen berupa 25 butir soal pilihan ganda dianalisis menggunakan pendekatan deskriptif kuantitatif dengan fokus pada reliabilitas, daya pembeda, tingkat kesukaran, validitas, dan kualitas pengecoh. Hasil penelitian menunjukkan bahwa instrumen asesmen memiliki koefisien reliabilitas sangat tinggi sebesar 0,92. Sebanyak 72% butir soal memiliki daya pembeda baik dan sangat baik, sedangkan 64% butir soal berada pada kategori tingkat kesukaran sedang. Analisis validitas menunjukkan 92% butir soal valid, sementara 8% tidak valid. Dari segi pengecoh, mayoritas soal berada pada kategori baik dan sangat baik, meskipun masih terdapat beberapa pengecoh yang kurang berfungsi optimal. Secara keseluruhan, instrumen asesmen Pendidikan Pancasila ini tergolong baik, tetapi perlu revisi pada butir soal yang tidak valid atau memiliki daya pembeda rendah. Hasil penelitian ini menegaskan pentingnya analisis butir soal secara berkala untuk meningkatkan kualitas asesmen pembelajaran di sekolah dasar.

Kata Kunci : Analisis Butir Soal, ANATES, Asesmen, Pendidikan Pancasila.

Abstract

This study aims to evaluate the quality of a Civic Education assessment instrument through item analysis using the ANATES application. The research subjects were the results of the summative assessment of Learning Scope II from 18 fifth-grade students at SDN Tritih Wetan 04 in the 2024/2025 academic year. The instrument consisted of 25 multiple-choice items, analyzed using a descriptive quantitative approach focusing on reliability, discrimination index, difficulty level, validity, and distractor effectiveness. The findings show that the instrument achieved a very high reliability coefficient of 0.92. A total of 72% of the items had good to very good discrimination power, while 64% of the items were categorized as moderate in difficulty. The validity analysis revealed that 92% of the items were valid, whereas 8% were invalid. In terms of distractors, most items were classified as good or very good, although some distractors did not function optimally. Overall, the Civic Education assessment instrument is considered to be of good quality but requires revisions for items that are invalid or have

low discrimination indices. These results highlight the importance of conducting regular item analysis to improve the quality of assessment in elementary school learning.

Keywords: ANATES, assessment, Civic Education, item analysis.

PENDAHULUAN

Pendidikan memiliki peran penting dalam membantu seseorang mengembangkan kemampuan dan potensi yang dimilikinya. Melalui pendidikan, peserta didik tidak hanya memperoleh pengetahuan, tetapi juga belajar membangun karakter, berpikir secara kritis, dan berinteraksi dengan lingkungan. Untuk mengetahui keberhasilan proses tersebut, diperlukan evaluasi yang dilakukan secara tepat. Hasil evaluasi dapat membantu guru mengetahui perkembangan belajar siswa serta menentukan bagian pembelajaran yang masih perlu diperbaiki. Dengan demikian, asesmen tidak hanya digunakan untuk mengetahui hasil belajar, tetapi juga menjadi dasar bagi guru dalam melihat ketercapaian tujuan pembelajaran.

Jaminan atas hak pendidikan bagi seluruh warga negara telah tercantum secara eksplisit dalam Pasal 31 ayat (1) Undang-Undang Dasar 1945. Penegasan lebih lanjut mengenai hal tersebut termuat dalam Undang-Undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional, yang menempatkan pengembangan potensi peserta didik sebagai tujuan utama pendidikan — mencakup dimensi keimanan, ketakwaan, akhlak mulia, kesehatan, keilmuan, kecakapan, kreativitas, kemandirian, serta kesiapan menjadi warga negara yang demokratis dan bertanggung jawab. Konsekuensinya, seluruh rangkaian proses pembelajaran — termasuk kegiatan evaluasi di dalamnya — dituntut untuk dirancang secara terukur dan dapat dipertanggungjawabkan.

Fungsi evaluasi pembelajaran tidak terbatas pada penilaian hasil belajar siswa semata, melainkan turut berperan sebagai landasan dalam pengambilan keputusan di bidang pendidikan. Evaluasi yang baik membantu guru mengetahui ketercapaian pembelajaran, mengidentifikasi kesulitan belajar, serta menyusun strategi perbaikan (Fuadiy, 2021). Agar fungsi tersebut berjalan optimal, instrumen asesmen yang digunakan guru perlu memenuhi kriteria validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan efektivitas pengecoh (Permatasari, 2025). Apabila instrumen yang digunakan tidak memenuhi kriteria yang ditetapkan, informasi yang dihasilkan berisiko menyesatkan, sehingga dapat menurunkan kualitas pembelajaran sekaligus merugikan peserta didik (Permatasari, 2025).

Pada kenyataannya, penyusunan soal oleh guru di lapangan masih sering dilakukan tanpa melalui proses analisis yang mendalam. Hal ini menyebabkan kualitas asesmen tidak selalu dapat dipastikan. Perkembangan teknologi sebenarnya memberikan peluang bagi guru untuk meningkatkan kualitas asesmen melalui aplikasi analisis butir soal. Salah satu aplikasi yang banyak digunakan adalah ANATES (Analisis Tes). Melalui aplikasi ini, proses analisis terhadap reliabilitas, validitas, daya pembeda, tingkat kesukaran, dan efektivitas pengecoh dapat berlangsung secara cepat sekaligus akurat (Ulfahyana, 2022). Pemanfaatan aplikasi berbasis komputer juga terbukti meningkatkan kecepatan dan ketepatan guru dalam merevisi instrumen asesmen (Hasyim et al., 2025).

Sejumlah penelitian menegaskan pentingnya analisis butir soal dalam meningkatkan kualitas instrumen asesmen. Sebagian besar soal UAS di sekolah dasar belum memenuhi kriteria validitas dan reliabilitas, sehingga hasil evaluasi menjadi kurang *representative* (Paskalin & Susanti, 2020). Pemanfaatan perangkat lunak analisis tes mampu meningkatkan kualitas instrumen asesmen yang digunakan guru (Luciano, 2025). Soal yang melalui proses analisis kuantitatif terbukti menghasilkan daya pembeda yang lebih unggul dibandingkan soal yang disusun semata-mata berdasarkan intuisi guru (Hermaya et al., 2024).

Namun demikian, fokus kajian-kajian terdahulu cenderung terpusat pada mata pelajaran matematika, IPA, atau bahasa, sementara penelitian mengenai mutu instrumen asesmen untuk Pendidikan Pancasila di jenjang sekolah dasar tergolong masih jarang dilakukan, masih terbatas. Padahal, Pendidikan Pancasila memiliki peran penting dalam membentuk karakter bangsa, menanamkan nilai kebangsaan, dan menyiapkan peserta didik menjadi warga negara yang demokratis dan bertanggung jawab (Reinita & Syahroh, 2024).

Belum banyaknya studi yang secara khusus menyoroti kualitas instrumen asesmen pada mata pelajaran Pendidikan Pancasila. Padahal, instrumen asesmen dalam mata pelajaran ini sering digunakan sebagai dasar penilaian ranah kognitif yang berhubungan dengan pemahaman nilai-nilai Pancasila. Kondisi ini juga ditemukan di SDN Tritih Wetan 04, di mana asesmen sumatif kelas V masih menggunakan soal yang belum teruji kualitasnya. Akibatnya, muncul kekhawatiran bahwa asesmen tersebut tidak sepenuhnya mencerminkan kemampuan siswa dan berpotensi

menghambat pencapaian tujuan pembelajaran (Fajaruddin et al., 2021).

Kajian ini diarahkan untuk mengkaji mutu instrumen asesmen mata pelajaran Pendidikan Pancasila, dengan menerapkan analisis butir soal berbantuan aplikasi ANATES. Fokus analisis mencakup reliabilitas, validitas, daya pembeda, tingkat kesukaran, dan efektivitas pengecoh. Melalui hasil kajian ini, diharapkan tersedia sumbangan teoretis bagi perkembangan evaluasi pendidikan, sekaligus manfaat praktis yang dapat dimanfaatkan guru dan pihak sekolah dasar guna meningkatkan mutu asesmen dalam proses pembelajaran.

METODE

Pendekatan deskriptif kuantitatif dengan jenis penelitian evaluatif digunakan dalam kajian ini, dengan tujuan memaparkan fenomena secara sistematis dan akurat melalui data numerik, sekaligus menakar mutu instrumen asesmen yang diteliti. Metode ini sesuai untuk penelitian yang menitikberatkan pada analisis kualitas butir soal asesmen Pendidikan Pancasila (Fajaruddin et al., 2021).

Lokasi penelitian bertempat di SDN Tritih Wetan 04, yang berada di Kecamatan Jeruklegi, Kabupaten Cilacap pada Maret-April 2025. Waktu penelitian bertepatan dengan pelaksanaan asesmen sumatif Lingkup Materi II tahun ajaran 2024/2025. Subjek penelitian adalah lembar jawaban asesmen Pendidikan Pancasila dari 18 siswa kelas V. Penentuan sampel dalam penelitian ini menerapkan teknik sampel jenuh, di mana keseluruhan anggota populasi diikutsertakan sebagai sampel mengingat jumlahnya yang tergolong kecil (Hidayati & Itidiyati, 2025).

Data dikumpulkan menggunakan teknik dokumentasi, mencakup lembar jawaban siswa, naskah soal asesmen, kunci jawaban, dan dokumen

pendukung lainnya. Instrumen yang digunakan berupa 25 butir soal pilihan ganda dengan empat pilihan jawaban, yang disusun oleh guru mata pelajaran Pendidikan Pancasila mengacu pada kompetensi dalam Kurikulum Merdeka. Selain berfungsi mengukur capaian belajar siswa, instrumen tersebut juga dianalisis mutunya dengan bantuan aplikasi ANATES.

Perhitungan kualitas butir soal secara kuantitatif dilakukan menggunakan aplikasi ANATES versi 4.0.9. Lima aspek utama menjadi fokus analisis, meliputi reliabilitas, daya pembeda, tingkat kesukaran, validitas, dan kualitas pengecoh. Kategorisasi hasil pada setiap aspek mengacu pada kriteria: reliabilitas, daya pembeda, tingkat kesukaran (Widayanti et al., 2021). Reliabilitas tes dihitung dengan metode belah dua Spearman-Brown. Interpretasi koefisien reliabilitas mengacu pada kriteria, yaitu: $r \leq 0,20$ (sangat rendah), $0,20 < r \leq 0,40$ (rendah), $0,40 < r \leq 0,60$ (sedang), $0,60 < r \leq 0,80$ (tinggi), dan $0,80 < r \leq 1,00$ (sangat tinggi).

Penilaian daya pembeda soal dilakukan dengan membandingkan proporsi peserta yang menjawab benar antara kelompok atas dan kelompok bawah, yang kemudian ditafsirkan berdasarkan kriteria, yaitu: $D \leq 0,00$ (sangat jelek/dibuang), $0,00 < D \leq 0,20$ (jelek), $0,20 < D \leq 0,40$ (cukup), $0,40 < D \leq 0,70$ (baik), dan $0,70 < D \leq 1,00$ (sangat baik) (Nasution et al., 2023).

Tingkat kesukaran dihitung dari proporsi siswa yang menjawab benar (indeks kesukaran/P), dengan kategori interpretasi sebagai berikut: $P \leq 0,30$ (sukar), $0,30 < P \leq 0,70$ (sedang), dan $0,70 < P \leq 1,00$ (mudah) (Latifah et al., 2024). Validitas soal dianalisis menggunakan korelasi point biserial, sedangkan kualitas pengecoh ditentukan dari distribusi pilihan jawaban siswa.

Penilaian mutu pengecoh mengacu pada pola sebaran pilihan jawaban siswa terhadap setiap opsi yang salah (distraktor). Suatu pengecoh dikategorikan berfungsi efektif jika dipilih oleh sedikitnya 5% peserta tes dan lebih banyak dipilih kelompok bawah ketimbang kelompok atas; sementara itu, pengecoh dianggap tidak berfungsi apabila persentase pemilihnya kurang dari 5% atau bahkan tidak dipilih sama sekali oleh peserta.

HASIL DAN PEMBAHASAN

Deskripsi Hasil Tes

Berdasarkan analisis menggunakan aplikasi ANATES, diperoleh data skor hasil tes Pendidikan Pancasila pada Asesmen Sumatif Lingkup Materi II kelas 5 SDN Tritih Wetan 04 tahun pelajaran 2024/2025 sebagaimana disajikan pada Tabel 1.

Menurut Sugiyono (2022), penyajian data dalam bentuk table memudahkan peneliti untuk mengorganisasi dan membaca informasi secara sistematis:

Tabel 1. Skor Data Dibobot Hasil Tes Pendidikan Pancasila Kelas 5.

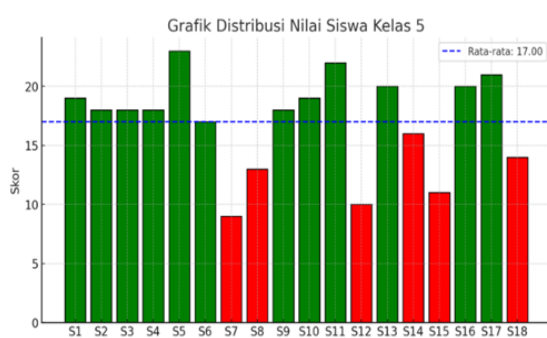
No	Kode Nama Siswa	Benar	Salah	Kosong	Skor Asli	Skor Bobot
1	S1	19	6	0	19	19
2	S2	18	7	0	18	18
3	S3	18	7	0	18	18
4	S4	18	7	0	18	18
5	S5	22	3	0	22	22
6	S6	17	8	0	17	17
7	S7	9	16	0	9	9

8	S8	13	12	0	13	13
9	S9	18	7	0	18	18
10	S10	19	6	0	19	19
11	S11	21	4	0	21	21
12	S12	15	10	0	15	15
13	S13	10	15	0	10	10
14	S14	20	5	0	20	20
15	S15	16	9	0	16	16
16	S16	11	14	0	11	11
17	S17	20	5	0	20	20
18	S18	14	11	0	14	14

Mengacu pada data Tabel 1, diperoleh skor tertinggi sebesar 22 dan skor terendah 9, dengan nilai rata-rata mencapai 16,56. Temuan ini mengindikasikan bahwa capaian hasil belajar mayoritas siswa berkumpul di sekitar nilai rata-rata tersebut. Sebagai ukuran pemusatan data, rata-rata mencerminkan kecenderungan umum dari hasil belajar yang diperoleh (Herbold, 2020).

Distribusi nilai siswa (Gambar 1) menunjukkan bahwa 12 siswa (66,67%) memperoleh skor di atas rata-rata, sementara 6 siswa (33,33%) berada di bawah rata-rata.

Distribusi nilai siswa dapat digambarkan dalam grafik berikut:



Gambar 1. Grafik Distribusi Nilai Siswa

Dari total 18 siswa peserta tes, tercatat 12 siswa (66,67%) meraih nilai di atas rata-rata, sementara 6 siswa (33,33%) berada di bawah rata-rata. Kondisi ini mengindikasikan bahwa pemahaman sebagian besar siswa terhadap materi

Pendidikan Pancasila yang diujikan tergolong cukup baik. Pola distribusi nilai tersebut sekaligus memberikan gambaran mengenai mutu instrumen tes serta tingkat capaian belajar siswa secara keseluruhan (Veronica *et al.*, 2023). Capaian hasil belajar yang mayoritas berada di atas rata-rata menunjukkan adanya penguasaan materi yang relatif baik oleh sebagian besar peserta didik (Silaban *et al.*, 2023).

Reliabilitas Tes

Pengujian reliabilitas tes menggunakan metode belah dua (split-half), yaitu dengan membagi butir soal ke dalam dua kelompok berdasarkan nomor ganjil dan nomor genap. Hasil pengujian reliabilitas tersebut, yang diperoleh melalui aplikasi ANATES, disajikan pada tabel berikut:

Tabel 2. Reliabilitas Tes Berdasarkan Skor Ganjil-Genap.

No	Aspek	Hasil
1	Jumlah Butir Soal	25
2	Jumlah Subjek	18
3	Korelasi XY	0,86
4	Reliabilitas Tes	0,92

Merujuk pada hasil yang tersaji dalam Tabel 2, koefisien reliabilitas tes Pendidikan Pancasila pada Asesmen Sumatif Lingkup Materi II kelas 5 SDN Tritih Wetan 04 tercatat sebesar 0,92. Merujuk pada kriteria interpretasi koefisien reliabilitas menurut

(Kalkbrenner, 2024), nilai reliabilitas 0,92 termasuk dalam kategori sangat tinggi (0,81-1,00). Temuan tersebut mengindikasikan bahwa instrumen tes tersebut memiliki tingkat konsistensi yang tinggi dalam mengukur kemampuan siswa.

Tingginya nilai reliabilitas ini mengindikasikan bahwa apabila tes yang sama diujikan kembali pada subjek dan kondisi yang serupa namun pada waktu berbeda, hasil yang diperoleh cenderung konsisten. Atas dasar tersebut, instrumen tes ini dapat diandalkan (reliable) sebagai alat ukur kemampuan siswa pada mata pelajaran Pendidikan Pancasila.

Analisis Daya Pembeda

Daya pembeda butir soal merepresentasikan sejauh mana suatu soal mampu membedakan siswa berkemampuan tinggi dari siswa berkemampuan rendah. Hasil pengukuran daya pembeda tersebut, yang diperoleh menggunakan aplikasi ANATES, disajikan pada tabel berikut:

Tabel 3. Hasil Analisis Daya Pembeda Butir Soal.

No. Butir Soal	Indeks Daya Pembeda	Interpretasi
1	0,40	Baik
2	0,60	Baik
3	0,80	Sangat Baik
4	0,40	Baik
5	0,20	Cukup

6	0,60	Baik
7	0,40	Baik
8	0,20	Cukup
9	0,60	Baik
10	0,40	Baik
11	0,60	Baik
12	0,20	Cukup
13	0,40	Baik
14	0,80	Sangat Baik
15	-0,20	Sangat Buruk
16	0,60	Baik
17	0,20	Cukup
18	0,40	Baik
19	0,60	Baik
20	0,00	Buruk
21	0,40	Baik
22	0,80	Sangat Baik
23	0,40	Baik
24	0,60	Baik
25	0,20	Cukup

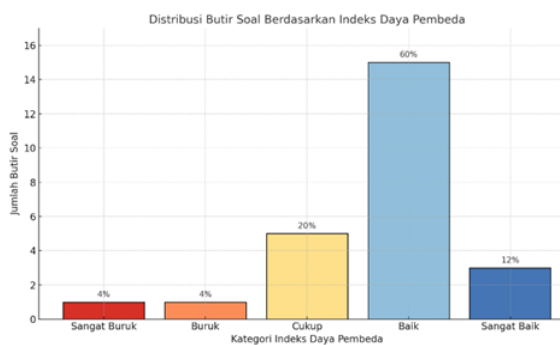
Daya pembeda butir soal menunjukkan bahwa butir soal dengan indeks daya pembeda 0,70–1,00 dikategorikan sangat baik, 0,40–0,69 dikategorikan baik, 0,20–0,39 dikategorikan cukup, 0,00–0,19 dikategorikan buruk, dan bernilai negatif dikategorikan sangat buruk, hal ini berdasarkan (Gusmizain, 2022). Indeks Daya Pembeda dapat dilihat pada tabel:

Tabel 4. Distribusi Butir Soal Berdasarkan Indeks Daya Pembeda.

No	Indeks Daya Pembeda	Interpretasi	Jumlah	Persentase	Nomor Butir Soal
1	< 0,00	Sangat Buruk	1	4%	15
2	0,00 - 0,19	Buruk	1	4%	20

3	0,20 - 0,39	Cukup	5	20%	5, 8, 12, 17, 25
4	0,40 - 0,69	Baik	15	60%	1, 2, 4, 6, 7, 9, 10, 11, 13, 16, 18, 19, 21, 23, 24
5	0,70 - 1,00	Sangat Baik	3	12%	3, 14, 22
		Total	25	100%	

Distribusi daya pembeda butir soal dapat digambarkan dalam grafik berikut:



Gambar 2. Distribusi Daya Pembeda

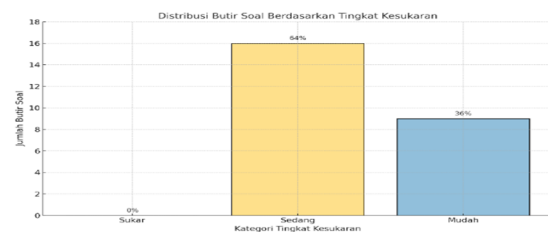
Merujuk pada Tabel 4 dan Gambar 2, dari total 25 butir soal yang dianalisis, diketahui sebanyak 1 butir (4%) berkategori daya pembeda sangat buruk, 1 butir (4%) berkategori buruk, 5 butir (20%) berkategori cukup, 15 butir (60%) berkategori baik, dan 3 butir (12%) berkategori sangat baik.

Mayoritas butir soal, yakni 72%, tergolong memiliki daya pembeda baik hingga sangat baik, yang berarti soal-soal tersebut efektif dalam memisahkan siswa berkemampuan tinggi dari siswa berkemampuan rendah. Di sisi lain, ditemukan 2 butir soal (8%) dengan daya pembeda buruk hingga sangat buruk. Khususnya butir soal nomor 15, yang menunjukkan nilai daya pembeda negatif — kondisi ini mengindikasikan soal tersebut justru bekerja terbalik, di mana siswa berkemampuan rendah lebih banyak menjawab benar dibandingkan siswa berkemampuan tinggi (Rishan & Sulaiman, 2023). Sementara itu, butir

soal nomor 20 tercatat memiliki daya pembeda sebesar 0, yang mengindikasikan bahwa soal tersebut sama sekali tidak mampu membedakan siswa berkemampuan tinggi dari siswa berkemampuan rendah.

Analisis Tingkat Kesukaran

Tingkat kesukaran butir soal mencerminkan derajat kemudahan atau kesulitan suatu soal bagi siswa. Sebaran butir soal menurut tingkat kesukarannya dapat dilihat pada gambar berikut:



Gambar 3. Distribusi Tingkat Kesukaran

Mengacu pada Gambar 3, dari keseluruhan 25 butir soal, tidak ditemukan satu pun soal (0%) berkategori sukar, sementara 16 butir soal (64%) berkategori sedang, dan 9 butir soal (36%) berkategori mudah. Dominasi butir soal berkategori sedang (64%) ini menunjukkan bahwa mayoritas soal berada pada tingkat kesulitan yang seimbang — tidak terlalu mudah maupun terlalu sukar bagi siswa. Soal dengan karakteristik semacam ini dinilai baik karena efektif dalam membedakan siswa berkemampuan tinggi dari siswa berkemampuan rendah (Kolin *et al.*, 2025).

Analisis Korelasi Skor Butir dengan Skor Total

Korelasi antara skor butir dan skor total berfungsi sebagai indikator validitas butir soal, yang mencerminkan kuat-lemahnya hubungan antara skor tiap butir dengan skor keseluruhan. Hasil perhitungan korelasi tersebut, yang diperoleh melalui aplikasi ANATES, disajikan pada tabel berikut:

Tabel 5. Hasil Analisis Korelasi Skor Butir dengan Skor Total

No	No. Butir Soal	Koefisien Korelasi	Signifikan	Interpretasi
1	1	0,454	Sangat Signifikan	Valid
2	2	0,622	Sangat Signifikan	Valid
3	3	0,752	Sangat Signifikan	Valid
4	4	0,488	Sangat Signifikan	Valid
5	5	0,311	Signifikan	Valid
6	6	0,632	Sangat Signifikan	Valid
7	7	0,454	Sangat Signifikan	Valid
8	8	0,321	Signifikan	Valid
9	9	0,622	Sangat Signifikan	Valid
10	10	0,478	Sangat Signifikan	Valid
11	11	0,642	Sangat Signifikan	Valid
12	12	0,345	Signifikan	Valid
13	13	0,488	Sangat Signifikan	Valid
14	14	0,752	Sangat Signifikan	Valid
15	15	-0,189	-	Tidak Valid
16	16	0,632	Sangat Signifikan	Valid
17	17	0,311	Signifikan	Valid

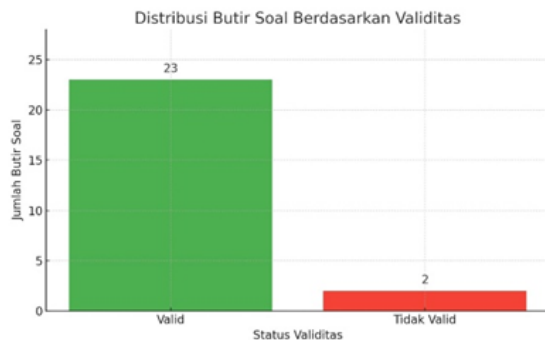
18	18	0,465	Sangat Signifikan	Valid
19	19	0,622	Sangat Signifikan	Valid
20	20	0,124	-	Tidak Valid
21	21	0,488	Sangat Signifikan	Valid
22	22	0,752	Sangat Signifikan	Valid
23	23	0,454	Sangat Signifikan	Valid
24	24	0,632	Sangat Signifikan	Valid
25	25	0,321	Signifikan	Valid

Mengacu pada hasil yang tercantum dalam Tabel 5, sebaran butir soal menurut kategori korelasi skor butir dengan skor total ditampilkan pada tabel berikut:

Tabel 6. Distribusi Butir Soal Berdasarkan Korelasi dengan Skor Total

No	Status	Jumlah	Persentase	Nomor Butir Soal
1	Valid	23	92%	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 17, 18, 19, 21, 22, 23, 24, 25
2	Tidak Valid	2	8%	15, 20
	Total	25	100%	

Distribusi kualitas butir soal berdasarkan korelasi dengan skor total dapat digambarkan dalam grafik berikut:

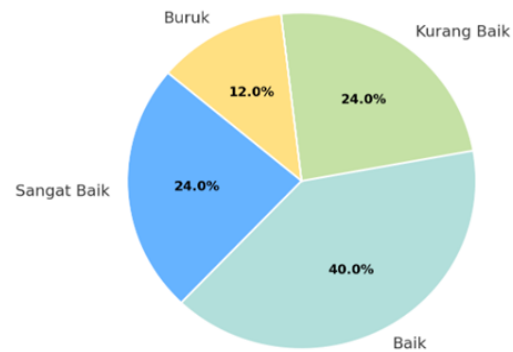


Gambar 4. Distribusi Butir Soal Berdasarkan Korelasi dengan Skor Total

Merujuk pada Tabel 6 dan Gambar 4, dari total 25 butir soal, sebanyak 23 butir (92%) dinyatakan valid, sementara 2 butir (8%) dinyatakan tidak valid. Status valid menandakan adanya korelasi yang signifikan antara skor butir dengan skor total, sedangkan status tidak valid menunjukkan sebaliknya — tidak terdapat korelasi yang signifikan di antara keduanya. Secara khusus, butir soal nomor 15 tercatat memiliki koefisien korelasi negatif (-0,189), yang berarti hubungannya dengan skor total bersifat berlawanan arah: siswa yang menjawab benar pada soal ini justru cenderung memperoleh skor total yang rendah, dan begitu pula sebaliknya (Kühbeck *et al.*, 2019).

Analisis Kualitas Pengecoh

Mutu pengecoh mencerminkan seberapa efektif pilihan jawaban selain kunci jawaban (distraktor) berperan dalam mengecoh siswa yang belum sepenuhnya memahami materi. Hasil analisis kualitas pengecoh menggunakan aplikasi ANATES disajikan pada gambar berikut:



Gambar 5. Hasil Analisis Kualitas Pengecoh

Mengacu pada hasil analisis pada Gambar 5, dari 25 butir soal yang diteliti, tercatat 6 butir (24%) berkategori kualitas pengecoh sangat baik, 10 butir (40%) berkategori baik, 6 butir (24%) berkategori kurang baik, dan 3 butir (12%) berkategori buruk.

Kategori kualitas pengecoh sangat baik diberikan pada butir soal yang seluruh opsi pengecohnya berfungsi secara efektif — yakni dipilih oleh sedikitnya 5% peserta tes dan lebih diminati oleh kelompok berkemampuan rendah dibandingkan kelompok berkemampuan tinggi — sehingga soal tersebut mampu membedakan siswa yang telah memahami materi dari yang belum (Rezigalla *et al.*, 2024).

Butir soal dengan kualitas pengecoh baik adalah butir soal yang sebagian besar pengecohnya berfungsi efektif, namun terdapat satu opsi pengecoh yang kurang diminati siswa sehingga fungsinya belum optimal. Butir soal dengan kualitas pengecoh kurang baik adalah butir soal yang memiliki lebih dari satu pengecoh tidak berfungsi, yaitu dipilih oleh kurang dari 5% peserta atau justru lebih banyak dipilih oleh kelompok atas dibandingkan kelompok bawah, sehingga pengecoh tersebut tidak mampu mengecoh siswa yang belum memahami materi.

Adapun butir soal dengan kualitas pengecoh buruk adalah butir soal yang sebagian besar atau seluruh opsi pengecohnya tidak dipilih sama sekali

oleh siswa (0%) atau menyimpang jauh dari pola sebaran yang diharapkan, sehingga opsi tersebut dianggap tidak efektif dan perlu direvisi total agar dapat berfungsi sebagai pengecoh yang baik

Berdasarkan hasil analisis butir soal

menggunakan aplikasi ANATES, rangkuman mutu butir soal tes Pendidikan Pancasila pada Asesmen Sumatif Lingkup Materi II kelas 5 SDN Tritih Wetan 04 tahun pelajaran 2024/2025 disajikan sebagai berikut:

Tabel 7. Rekapitulasi Hasil Analisis Butir Soal

No. Butir Soal	Valid	Daya Pembeda	Tingkat Kesukaran	Kualitas Pengecoh	Keputusan
1	Valid	Baik	Mudah	Kurang Baik	Direvisi
2	Valid	Baik	Sedang	Sangat Baik	Dipertahankan
3	Valid	Sangat Baik	Sedang	Sangat Baik	Dipertahankan
4	Valid	Baik	Mudah	Kurang Baik	Direvisi
5	Valid	Cukup	Sedang	Baik	Dipertahankan
6	Valid	Baik	Sedang	Sangat Baik	Dipertahankan
7	Valid	Baik	Mudah	Baik	Dipertahankan
8	Valid	Cukup	Sedang	Baik	Dipertahankan
9	Valid	Baik	Mudah	Kurang Baik	Direvisi
10	Valid	Baik	Sedang	Baik	Dipertahankan
11	Valid	Baik	Sedang	Baik	Dipertahankan
12	Valid	Cukup	Mudah	Buruk	Direvisi
13	Valid	Baik	Sedang	Baik	Dipertahankan
14	Valid	Sangat Baik	Sedang	Sangat Baik	Dipertahankan
15	Tidak Valid	Sangat Buruk	Sedang	Buruk	Dibuang
16	Valid	Baik	Sedang	Baik	Dipertahankan
17	Valid	Cukup	Mudah	Kurang Baik	Direvisi
18	Valid	Baik	Sedang	Sangat Baik	Dipertahankan
19	Valid	Baik	Mudah	Kurang Baik	Direvisi
20	Tidak Valid	Buruk	Sedang	Buruk	Dibuang
21	Valid	Baik	Sedang	Baik	Dipertahankan
22	Valid	Sangat Baik	Sedang	Sangat Baik	Dipertahankan
23	Valid	Baik	Mudah	Baik	Dipertahankan
24	Valid	Baik	Sedang	Baik	Dipertahankan
25	Valid	Cukup	Mudah	Kurang Baik	Direvisi

Temuan pada Tabel 7 memperlihatkan bahwa mayoritas butir soal berada dalam kategori valid dengan daya pembeda baik hingga sangat baik, sehingga dinilai layak untuk

dipertahankan. Meski demikian, terdapat beberapa soal yang perlu direvisi karena pengecohnya kurang berfungsi, serta dua butir soal (nomor 15 dan 20) yang harus dibuang karena

tidak memenuhi syarat validitas. Kualitas soal tidak hanya ditentukan oleh validitas, tetapi juga harus diperkuat oleh daya pembeda dan efektivitas pengecoh. Mutu suatu soal tidak semata-mata bergantung pada validitas, melainkan juga perlu didukung oleh daya pembeda dan efektivitas pengecoh yang memadai (Ulfahyana, 2022).

Ditinjau dari aspek tingkat kesukaran, sebagian besar soal tergolong berkategori sedang, sebagian kecil lainnya berada pada kategori mudah, sedangkan sebagian kecil tergolong mudah. Komposisi ini mendekati distribusi ideal yakni soal sebaiknya terdiri dari proporsi mudah, sedang, dan sukar agar mampu mengukur kemampuan siswa secara menyeluruh (Jumini & Retnawati, 2022). Kendati masih terdapat sejumlah butir yang memerlukan revisi, instrumen ini secara keseluruhan dinilai telah memenuhi kriteria sebagai tes yang layak digunakan dalam evaluasi pembelajaran.

PENUTUP

Simpulan

Mengacu pada hasil analisis butir soal menggunakan aplikasi ANATES, dapat disimpulkan bahwa mutu soal tes Pendidikan Pancasila pada Asesmen Sumatif Lingkup Materi II kelas 5 SDN Tritih Wetan 04 tahun pelajaran 2024/2025 tergolong dalam kategori baik secara umum. Indikasi ini tercermin dari reliabilitas tes yang sangat tinggi (koefisien 0,92), 72% butir soal berdaya pembeda baik hingga sangat baik, 64% butir soal berkategori sedang dari segi tingkat kesukaran, serta 92% butir soal dinyatakan valid berdasarkan korelasi antara skor butir dan skor total. Dari sisi pengecoh, mayoritas juga berada pada kategori baik hingga sangat baik, kendati dua butir soal harus disingkirkan karena

dinyatakan tidak valid, dan sejumlah butir lainnya masih memerlukan revisi agar berfungsi lebih optimal.

Temuan penelitian ini menguatkan pandangan bahwa suatu instrumen tes dapat dikategorikan berkualitas apabila memenuhi aspek validitas, reliabilitas, daya pembeda, tingkat kesukaran, serta fungsi pengecoh yang berjalan optimal. (Permatasari, 2025). Secara praktis, temuan ini menegaskan bahwa guru perlu memperhatikan kualitas butir soal sebelum digunakan dalam asesmen sumatif agar hasil evaluasi lebih akurat dalam menggambarkan kemampuan peserta didik. Meskipun terdapat butir yang harus direvisi atau dibuang, instrumen yang dianalisis dapat dikatakan sudah cukup layak sebagai alat evaluasi pembelajaran.

Saran

Mengacu pada temuan penelitian, berikut disampaikan beberapa saran dari penulis:

- a. Ditinjau dari sisi teoretis, temuan penelitian ini berpotensi menjadi acuan dalam memperkuat teori evaluasi pembelajaran, khususnya terkait urgensi validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan efektivitas pengecoh dalam penyusunan soal. Untuk itu, penelitian selanjutnya disarankan mengembangkan kajian sejenis dengan cakupan yang lebih luas guna memperkaya khazanah teori evaluasi pendidikan.
- b. Secara praktis, guru disarankan untuk secara rutin melakukan analisis butir soal guna mengetahui mutu instrumen evaluasi yang digunakan, serta melakukan revisi atau penggantian terhadap soal yang tidak valid atau memiliki daya pembeda rendah, sehingga

- efektivitas evaluasi pembelajaran dapat meningkat.
- c. Pihak sekolah perlu mendorong pemanfaatan aplikasi seperti ANATES sebagai alternatif praktis dan efisien dalam mengevaluasi kualitas butir soal, sekaligus memberikan dukungan berupa pelatihan atau workshop mengenai penyusunan soal yang baik serta penggunaan teknologi dalam analisis soal.
 - d. Penelitian ini memiliki keterbatasan karena hanya dilakukan pada satu sekolah dan satu lingkup materi, sehingga hasilnya belum dapat digeneralisasikan secara luas. Oleh karena itu, penelitian lanjutan disarankan melibatkan lebih banyak sekolah, berbagai mata pelajaran, serta pendekatan evaluasi yang berbeda agar diperoleh gambaran yang lebih komprehensif.

DAFTAR PUSTAKA

- Fajaruddin, S., Retnawati, H., Yusron, E., & Sofyaningsih, V. (2021). Exploring the final examination test item characteristics of Pancasila and civic education. *International Journal of Education and Learning*, 3(3), 232-240. <https://doi.org/10.31763/ijele.v3i3.435>
- Fuadiy, M. (2021). Evaluasi Pembelajaran Sebagai Sebuah Studi Literatur. *DIMAR: Jurnal Pendidikan Islam*, 3(1), 173-179. <https://doi.org/10.58577/dimar.v3i1.83>
- Gusmizain, A. (2022). Karakteristik Butir Soal Tes Mata Kuliah Matriks & Ruang Vektor Mahasiswa Matematika. *Jurnal Evaluasi Pendidikan*, 13(2), 127-130. <https://doi.org/10.21009/jep.v13i2.28328>
- Hasyim, A. A., Sucipto, H., Lazulfa, I., & Rizal, M. F. (2025). Rancang Bangun Aplikasi Penilaian Kinerja Guru Berbasis Website Menggunakan Metode Graphic Rating Scale (Studi Kasus: Mi Al-Hikmah Sidowarek). *Inovate : Jurnal Ilmiah Inovasi Teknologi Informasi* 9(2), 137-158. <https://doi.org/10.33752/inovate.v9i2.8882>
- Herbold, S. (2020). Autorank: A Python package for automated ranking of classifiers. *J. Open Source Softw.*, 5(48), 2173. <https://doi.org/10.21105/joss.02173>
- Hermaya, I., Helendra, Vauzia, & Arsih, F. (2024). Item Quality Analysis of Concept Understanding and Problem Solving in Environmental Change Materials Using ANATES. *Jurnal Ilmiah Pendidikan Profesi Guru*, 7(2), 371-381. <https://doi.org/10.23887/jippg.v7i2.84083>
- Hidayati, R. A., & Itidiyati, D. S. (2025). The Effect of Work Discipline, Work Motivation, And Work Environment on The Performance of Contract Employees In The Environmental Service of Gresik Regency. *Innovation Research Journal*, 6(1), 92-101. <https://doi.org/10.30587/innovation.v6i1.9353>
- Jumini, J., & Retnawati, H. (2022). Estimating Item Parameters and Student Abilities: An IRT 2PL Analysis of Mathematics Examination. *AL-ISHLAH: Jurnal Pendidikan*, 14(1), 385-398.

- <https://doi.org/10.35445/alishlah.v14i1.926>
- Kalkbrenner, M. (2024). Choosing Between Cronbach's Coefficient Alpha, McDonald's Coefficient Omega, and Coefficient H: Confidence Intervals and the Advantages and Drawbacks of Interpretive Guidelines. *Measurement and Evaluation in Counseling and Development*, 57, 93–105.
<https://doi.org/10.1080/07481756.2023.2283637>
- Kolin, A. A., Rosliyah, Y., Setiawati, A., & Oesman, A. M. (2025). Analisis Butir Soal Ulangan Akhir Semester Bahasa Jepang Kelas XI Di SMA Negeri 1 Semarang. *Jurnal Pendidikan Bahasa Jepang Undiksha*, 11(1), 81-90.
<https://doi.org/10.23887/jpbj.v11i1.78399>
- Kühbeck, F., Berberat, P., Engelhardt, S., & Sarikas, A. (2019). Correlation of online assessment parameters with summative exam performance in undergraduate medical education of pharmacology: a prospective cohort study. *BMC Medical Education*, 19, Article 412.
<https://doi.org/10.1186/s12909-019-1814-5>
- Latifah, I., Ardiyaningrum, M., Putri, M. A., & Arifin, A. S. (2024). Analisis Hasil Ujian Sekolah Berstandar Nasional di Sekolah Dasar Kecamatan Bulu pada Muatan Matematika. *Indonesian Journal of Elementary Education and Teaching Innovation*, 3(1), 52-63.
[https://doi.org/10.21927/ijeeti.2024.3\(1\).52-63](https://doi.org/10.21927/ijeeti.2024.3(1).52-63)
- Luciano, R. (2025). Innovative test item analysis using optical mark recognition technology: An evaluation. *International Journal of Advanced And Applied Sciences*, 12(4), 1-11.
<https://doi.org/10.21833/ijaas.2025.04.001>
- Nasution, A. F., Rahmawati, B. L., Salma, M. F., Qomari, N., & Syarifuddin, M. (2023). Tahlilu Bunud Asilah Muta'ddidah Ikhtiyarat Al-Lughah Al-Arabiyah. *Lahjah Arabiyah: Jurnal Bahasa Arab Dan Pendidikan Bahasa Arab*.
<https://doi.org/10.35316/lahjah.v4i1.34-43>
- Paskalin, G., & Susanti, M. (2020). Analisis Butir Soal untuk Materi Gaya di Sekolah Dasar. 3, 23–32.
<https://doi.org/10.21831/DIDAKT IKA.V3I1.31431>
- Permatasari, N. (2025). Analisis kualitas instrumen penilaian materi keanekaragaman hayati melalui tes klasik dan Rasch model. *Bio-Pedagogi*, 14(1), 10-26.
<https://doi.org/10.20961/bio-pedagogi.v14i1.88470>
- Reinita, & Syahroh, K. (2024). Quizizz-Assisted Cognitive Diagnostic Assessment in Pancasila Education Learning in Elementary Schools. *Jurnal Imiah Pendidikan Dan Pembelajaran*, 8(3), 1476-1490.
<https://doi.org/10.23887/jipp.v8i3.83527>
- Rezigalla, A., Eleragi, A., Elhussein, A., Alfaifi, J., Alghamdi, M., Ameer, A. Al, Yahia, A., Mohammed, O., & Adam, M. (2024). Item analysis: the impact of distractor efficiency on the difficulty index and discrimination power of multiple-choice items. *BMC Medical Education*, 24.
<https://doi.org/10.1186/s12909-024-05433-y>

- Rishan, M., & Sulaiman, S. (2023). Analisis Butir Soal Tipe Multiple Choice Ujian Akhir Semester Ganjil Kelas X Tahun Ajaran 2022/2023 Mata Pelajaran Pendidikan Agama Islam di SMA Negeri 1 Padang. *Islamika*, 5(3), 981-997. <https://doi.org/10.36088/islamika.v5i3.3517>
- Silaban, M. S. R., Silaban, P., & Sitepu, A. (2023). The Effect Of The Mastery Learning Model On The Students' Learning Outcomes At Class V Of Public Elementary School 097795 Perasmian. *Jurnal Pajar (Pendidikan Dan Pengajaran)*, 7(3), 640-657. <https://doi.org/10.33578/pjr.v7i3.9446>
- Ulfahyana, H. (2022). Analisis Kualitas Butir Soal Ujian Sekolah Kelas Ix Mata Pelajaran Matematika. *Jurnal Penalaran Dan Riset Matematika*, 1(2), 70-78. <https://doi.org/10.62388/prisma.v1i2.212>
- Veronica, S. F., Pandra, V., & Rosalina, E. (2023). Pengembangan Instrumen Tes untuk Mengukur Hasil Belajar Materi Pecahan Siswa Kelas IV SD. *Journal of Elementary School (JOES)*, 6(2), 301-307. <https://doi.org/10.31539/joes.v6i2.6777>
- Widayanti, W., Bistari, B., & Suparjan, S. (2021). Analisis Butir Soal Pilihan Ganda Penilaian Tengah Semester Pada Pembelajaran Tematik Kelas V Sekolah Dasar Negeri 39 Pontianak Kota. *Jurnal DIDIKA: Wahana Ilmiah Pendidikan Dasar*, 7(2), 279-296. <https://doi.org/10.29408/didika.v7i2.4370>