

**ANALISIS KELAYAKAN BUTIR SOAL KESEHATAN DAN KESELAMATAN
KERJA MENGGUNAKAN SOFTWARE ANATES**

Nadhirra Shaffa Aprillia Faatiha¹, Nayla Aurelia Putri², Nafaza Zakiyah Qolbi
Agustin³, Friska Putri Amelia⁴, Dian Merta Halatu⁵, Luqman Hakim⁶, Vivi Pratiwi⁷

Fakultas Ekonomika dan Bisnis, Universitas Negeri Surabaya

Alamat e-mail : 24080304053@mhs.unesa.ac.id¹,
24080304075@mhs.unesa.ac.id², 24080304124@mhs.unesa.ac.id³,
24080304127@mhs.unesa.ac.id⁴, 24080304159@mhs.unesa.ac.id⁵,
luqmanhakim@unesa.ac.id⁶, vivipratiwi@unesa.ac.id⁷

ABSTRACT

This study aims to analyze the quality of multiple-choice test items in the Occupational Safety and Health (K3) Teaching Module using a quantitative descriptive approach. A total of 45 items were administered to 25 respondents, consisting of vocational high school students and accounting students, through Google Forms. The Anates software was used to examine item validity, reliability, difficulty level, discrimination index, and distractor effectiveness. The results show that the test has a very high reliability coefficient of 0.96, indicating excellent internal consistency. The validity analysis reveals that 33.33% of the items fall into the very high category, 17.78% high, 15.56% moderate, 6.67% low, and 26.67% very low. In terms of difficulty, most items are categorized as easy to moderate, while several are very difficult, suggesting an imbalanced distribution. Analysis of the discrimination index indicates that most items effectively distinguish between high- and low-performing students, although some items require revision due to low or negative discrimination values. The distractor analysis also shows variation in effectiveness, with some options failing to function as intended. Overall, the test is considered reliable and generally representative, but revisions to items with low validity, extreme difficulty levels, weak discrimination power, and ineffective distractors are recommended to enhance the overall quality of the assessment instrument.

Keywords: *HOTS, K3LH, Anates, Validity, Reliability, Discrimination Power, Difficulty Level.*

ABSTRAK

Penelitian ini bertujuan untuk menganalisis kualitas butir soal pilihan ganda pada Modul Ajar Keselamatan dan Kesehatan Kerja (K3) dengan menggunakan pendekatan deskriptif kuantitatif. Sebanyak 45 butir soal diujikan kepada 25 responden yang terdiri atas siswa/i SMK sederajat dan mahasiswa/i jurusan Akuntansi melalui Google Form. Analisis butir soal dilakukan menggunakan software Anates untuk memperoleh informasi mengenai validitas, reliabilitas, tingkat kesukaran, daya pembeda, serta efektifitas pengecoh. Hasil penelitian menunjukkan bahwa instrumen

memiliki reliabilitas sangat tinggi sebesar 0,96 yang menandakan konsistensi tes sangat baik. Analisis validitas mengungkapkan bahwa 33,33% butir soal berkategori validitas sangat tinggi, 17,78% tinggi, 15,56% cukup, 6,67% rendah, dan 26,67% sangat rendah. Dari sisi tingkat kesukaran, sebagian besar soal tergolong mudah hingga sedang, namun beberapa soal termasuk sangat sukar sehingga distribusinya belum seimbang. Daya pembeda menunjukkan bahwa sebagian besar butir mampu membedakan peserta didik berkemampuan tinggi dan rendah, meskipun beberapa butir memerlukan revisi. Efektivitas pengecoh juga bervariasi, dengan sejumlah opsi tidak berfungsi sebagaimana mestinya. Secara keseluruhan, tes ini dinilai reliabel dan representatif, namun perbaikan diperlukan pada butir soal dengan validitas rendah, tingkat kesukaran ekstrem, daya pembeda lemah, serta pengecoh yang kurang efektif untuk meningkatkan kualitas instrumen evaluasi.

Kata Kunci: HOTS, K3LH, Anates, Validitas, Reliabilitas, Daya Pembeda, Tingkat Kesukaran.

A. Pendahuluan

Evaluasi hasil belajar merupakan komponen penting dalam proses pendidikan karena berfungsi untuk mengetahui tingkat pencapaian tujuan pembelajaran serta efektivitas strategi yang digunakan (Fauziah & Mustami, 2021). Melalui evaluasi, pendidik dapat memperoleh informasi akurat mengenai perkembangan peserta didik dan dasar dalam pengambilan keputusan, seperti pemberian remedial, pengayaan, maupun penentuan nilai akhir. Tes hasil belajar, khususnya tes pilihan ganda, masih menjadi instrumen evaluasi yang paling sering digunakan karena mampu menilai kemampuan kognitif secara sistematis dan efisien (Salsabila, Muslim, & Anwar, 2022). Namun, agar tes mencerminkan

kemampuan peserta didik secara objektif, instrumen harus memiliki kualitas teknis yang baik.

Kualitas suatu tes tidak hanya ditentukan oleh kelayakan materi, tetapi juga oleh aspek psikometrik seperti validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh (Yuliasari & Sutrisno, 2020). Salah satu metode yang banyak digunakan dalam lima tahun terakhir untuk menganalisis butir soal adalah perangkat lunak ANATES, karena mampu memberikan gambaran terperinci mengenai performa setiap butir soal (Rohim & Puspitasari, 2021). Melalui analisis ini, pendidik dapat mengidentifikasi butir yang berfungsi baik maupun yang

memerlukan revisi agar instrumen semakin akurat dan adil.

Berdasarkan hasil analisis menggunakan ANATES, tes dengan 45 butir soal dan 25 peserta menunjukkan reliabilitas sangat tinggi sebesar 0,96, yang berarti instrumen konsisten dan layak digunakan dalam evaluasi pembelajaran (Fauziah & Mustami, 2021). Namun, karakteristik butir soal menunjukkan variasi, terutama pada tingkat kesukaran. Sebagian besar butir tergolong mudah hingga sedang, sementara beberapa berada pada kategori sangat sukar. Distribusi ini belum ideal, mengingat penelitian terbaru menyarankan proporsi soal yang seimbang—yakni sekitar 30% mudah, 40% sedang, dan 30% sukar untuk memperoleh pengukuran kemampuan peserta didik yang lebih komprehensif (Salsabila et al., 2022). Keberadaan soal yang terlalu mudah atau terlalu sulit berpotensi mengurangi kemampuan tes dalam membedakan peserta didik dengan tingkat kompetensi berbeda.

B. Metode Penelitian

Penelitian ini menggunakan metode deskriptif kuantitatif dengan

tujuan utama untuk menguji kualitas butir soal pilihan ganda pada Modul Ajar mata pelajaran Keselamatan dan Kesehatan Kerja. Pendekatan ini dipilih karena mampu menyajikan informasi numerik mengenai karakteristik soal, seperti validitas, reliabilitas, tingkat kesukaran, dan daya pembeda, yang menjadi indikator penting dalam memastikan efektivitas evaluasi pembelajaran. Seluruh proses analisis data ini dibantu oleh software Anates, sehingga penilaian terhadap setiap butir soal dapat dilakukan secara objektif dan terstandar.

Subjek dalam penelitian ini merupakan siswa/i SMK sederajat dan mahasiswa/i dari jurusan Akuntansi yang berdomisili di berbagai wilayah di Indonesia, terutama Surabaya dan Sidoarjo. Sebanyak 25 hasil tanggapan diperoleh melalui pengisian kuesioner secara sukarela tanpa ada pemilihan khusus. Jumlah ini dipertimbangkan cukup untuk memberikan gambaran awal mengenai kemampuan responden dalam memahami materi K3. Pengisian butir soal dilakukan secara daring melalui Google Form yang memungkinkan responden

mengerjakan soal dengan lebih fleksibel.

Instrumen pada penelitian ini berupa 45 butir soal pilihan ganda yang disusun berdasarkan kompetensi dasar mata pelajaran K3, yang dikembangkan sesuai dengan tingkat kognitif dalam Taksonomi Bloom. Dalam setiap soal memiliki lima alternatif pilihan jawaban, dengan satu jawaban benar.

Pengumpulan data dilakukan melalui beberapa langkah. Pertama, peneliti menyusun draft soal sesuai materi ajar kemudian meminta telaah dari ahli untuk memastikan kesesuaian isi dengan tujuan pembelajaran. Setelah revisi dilakukan berdasarkan masukan ahli, soal diunggah ke Google Form dan tautannya dibagikan kepada responden. Seluruh jawaban direkam otomatis oleh sistem sehingga memudahkan proses pengunduhan dan pengolahan data.

Tahap akhir penelitian adalah analisis data menggunakan Anates. Program ini memberikan berbagai informasi penting, seperti nilai validitas butir, reliabilitas tes, tingkat kesukaran, daya pembeda, hubungan

antara skor butir dengan skor total, serta efektivitas pengecoh. Anates juga menyajikan hasil dalam bentuk tabel dan ringkasan yang membantu peneliti maupun pembaca menilai kelayakan setiap butir soal dalam modul ajar K3.

C. Hasil Penelitian dan Pembahasan

Penelitian ini mengkaji 45 butir soal pilihan ganda yang dikerjakan oleh 25 peserta didik melalui Google Form. Analisis terhadap soal dilakukan dengan bantuan software Anates untuk menilai aspek validitas, reliabilitas, daya pembeda, serta tingkat kesukaran setiap butir soal. Berikut hasil analisisnya:

Validitas

Menurut Sugiyono (2018) validitas merupakan tingkat kakuratan antara data yang terkumpul dengan realitas yang terjadi pada objek penelitian. Instrumen dikatakan valid apabila mampu menghasilkan data yang sesuai dengan tujuan pengukuran. Validitas dapat dibedakan menjadi beberapa jenis, antara lain validitas isi (content validity),

validitas konstruk (construct validity), dan validitas kriteria (criterion validity).

Validitas isi berfokus pada kesesuaian butir soal dengan materi atau kompetensi yang diujikan, sedangkan validitas konstruk berkaitan dengan sejauh mana tes mencerminkan konsep teoritis yang hendak diukur. Dalam penelitian, validitas sering diuji menggunakan korelasi antara skor tiap butir soal dengan skor total tes. Korelasi yang tinggi menunjukkan bahwa butir soal tersebut valid dan mampu mengukur kemampuan yang dimaksud.

Tingkat validitas biasanya diinterpretasikan melalui koefisien korelasi, dengan kriteria sebagai berikut:

Tingkat Korelasi Validitas	Interpretasi
0,80 - 1,00	Validasi Sangat Tinggi
0,60 - 0,80	Validasi Tinggi
0,40 - 0,60	Validasi Cukup
0,20 - 0,40	Validasi Rendah
0,00 - 0,20	Validasi Sangat Rendah

Gambar 1 Kriteria Korelasi Validitas

Tabel 1 Hasil Analisis Validitas Soal

No	Tingkat Validitas	Jumlah	%	No. Butir Soal
1.	Korelasi Sangat Tinggi	15	33,33%	1, 14, 15, 16, 20, 23, 26, 30, 33, 36, 39, 40, 42, 43, 44
2.	Korelasi Tinggi	8	17,78%	7, 10, 12, 13, 17, 24, 29, 38
3.	Korelasi Cukup	7	15,56%	2, 3, 4, 5, 19, 32, 45
4.	Korelasi Rendah	3	6,67%	28, 31, 37
5.	Korelasi Sangat Rendah	12	26,67%	6, 8, 9, 11, 18, 21, 22, 25, 27, 34, 35, 41

Sumber: data diolah (2025)

Berdasarkan hasil analisis validitas butir soal pada tabel 1, maka dapat disimpulkan bahwa sebesar 33,33% atau sebanyak 15 butir soal dengan validasi soal sangat tinggi, 17,78% atau 8 butir soal dengan

validasi soal tinggi, 15,56% atau 7 butir soal dengan validasi soal cukup, 6,67% atau 3 butir soal dengan validasi soal rendah, dan 26,67% atau 12 butir soal dengan validasi soal sangat rendah.

Reliabilitas

Reliabilitas merupakan tingkat ketetapan atau konsistensi suatu alat ukur dalam mengukur soal (Ary et al., 2023). Dengan kata lain, reliabilitas menunjukkan sejauh mana butir soal memberikan hasil yang sama ketika diuji berulang kali dalam kondisi serupa. Instrumen yang reliabel berarti dapat dipercaya karena memiliki hasil yang stabil dan konsisten. Nilai reliabilitas yang tinggi menandakan bahwa butir-butir soal saling berkaitan dan mampu mengukur kompetensi secara konsisten.

Nilai Realibilitas	Interpretasi
0,00 - 0,20	Sangat Rendah
0,20 - 0,40	Rendah
0,40 - 0,70	Cukup
0,70 - 0,90	Tinggi
0,90 - 1,00	Sangat Tinggi (Sempurna)

Gambar 2 Kriteria Korelasi Koefisien

RELIABILITAS TES

Ketaj= 23.84
 Simpang Baku= 8.04
 KorelasiXY= 0.92
 Reliabilitas Tes= 0.96
 Nama Sachas: BELUM ADA, NAMA_ASA

No. Urut	No. Soal	Kode/Nama Siswa	Skor Benar	Skor Salah	Skor Total
1	1	Galbi Agustini	12	14	26
2	2	Putri Maulida	14	15	29
3	3	Aefa Maulida	15	15	30
4	4	Nova Pelinda	12	14	26
5	5	Serina Hafiza	15	15	30
6	6	Dauca Ancha	10	15	25
7	7	Selma Shaleha	15	14	29
8	8	Randa Sembale	15	15	30
9	9	Adha Zahra	15	14	29
10	10	Hana Robiatul	11	12	23
11	11	Ananda	15	15	30
12	12	Serdian	14	14	28
13	13	Putri Nur	12	14	26
14	14	Cleo Kariliah	16	14	30
15	15	Agus Ramadani	7	6	13
16	16	Adinda Nur	14	13	27
17	17	Jessika Mariska	15	15	30
18	18	Dionisius Arya	1	4	5
19	19	Sissy Adam	8	8	16
20	20	Putuhammad Affan	4	1	5
21	21	Fadil Acan	4	2	6
22	22	Nurul Acan	18	4	22
23	23	Adha Faridha	13	13	26
24	24	Natalya Adhi	15	16	31
25	25	Intan Zahatus	15	16	31

Gambar 3 Hasil Analisis Reliabilitas Soal

Sumber: data diolah (2025)

Dari hasil perhitungan reliabilitas menggunakan Anates, diperoleh nilai reliabilitas tes sebesar 0,96. Berdasarkan kriteria korelasi reliabilitas, nilai tersebut termasuk dalam kategori sangat tinggi (sempurna) karena berada pada rentang 0,90 - 1,00. Hasil ini menunjukkan bahwa instrumen tes memiliki konsistensi yang sangat baik serta dapat dipercaya dalam mengukur kemampuan peserta didik secara tepat dan stabil.

Daya Pembeda

Daya pembeda merupakan karakteristik suatu butir soal yang menunjukkan kemampuannya dalam membedakan peserta didik berdasarkan tingkat kemampuannya.

Pengkajian daya pembeda dilakukan untuk melihat seberapa efektif suatu soal dalam mengelompokkan peserta didik yang berkemampuan tinggi dan peserta didik yang berkemampuan rendah, terutama dalam hal pencapaian prestasi belajar. Berikut adalah interpretasi dari koefisien daya beda:

Indeks Daya Beda	Interpretasi
0,00 - 0,20	Jelek (soal diperbaiki)
0,20 - 0,40	Cukup (sal diterima dan diperbaik)
0,40 - 0,70	Baik (soalditerima)
0,70 - 1,00	Sangat baik (soal diterima)
Negatif	Semuanya tidak bik (soal dibuang)

Gambar 4 Interpretasi Koeisien Daya Beda

Tabel 2 Hasil Analisis Daya Beda Butir Soal

No	Tingkat Daya Beda	Jum lah	%	No. Butir Soal
1.	Jelek (Soal Diperbaiki)	5	11,11%	8, 9, 21, 22, 25
2.	Cukup (Soal Diterima dan Diperbaiki)	1	2,22%	28

3.	Baik (Soal Diterima)	10	22,22%	1, 3, 4, 7, 13, 17, 24, 26, 31, 38
4.	Sangat baik (Soal Diterima)	20	44,44%	2, 5, 10, 12, 14, 15, 16,19, 20,23,2 9,30, 33,36,3 9,40,42 , 43,44,4 5
5.	Semuanya tidak baik (Soal Dibuang)	9	20%	6, 11, 18, 27, 32, 34, 35, 37, 41

Sumber: data diolah (2025)

Hasil analisis menunjukkan bahwa kualitas daya pembeda butir soal cukup beragam. Terdapat 5 soal (11,11%) yang termasuk kategori jelek dan 1 soal (2,22%) kategori cukup, yang berarti soal-soal tersebut belum bekerja optimal dalam membedakan peserta didik sehingga memerlukan revisi untuk meningkatkan fungsinya sebagai alat evaluasi pembelajaran. Sementara itu, terdapat 10 soal (22,22%) yang berada pada kategori

baik, menunjukkan bahwa soal-soal tersebut sudah mampu mengukur perbedaan kemampuan peserta didik secara efektif. Di sisi lain, kategori paling dominan adalah sangat baik dengan jumlah 20 soal (44,44%), yang menandakan bahwa sebagian besar soal berfungsi sangat optimal dalam membedakan peserta didik berdasarkan tingkat penguasaan materi. Meski demikian, masih ditemukan 9 soal (20%) pada kategori semuanya tidak baik, yang menunjukkan bahwa soal-soal tersebut tidak layak digunakan karena sama sekali tidak berfungsi sebagai pembeda. Oleh karena itu, soal dalam kategori ini harus dibuang dan diganti untuk menjaga kualitas instrumen evaluasi pembelajaran.

Tingkat Kesukaran

Tingkat kesukaran merupakan indikator yang menggambarkan peluang peserta didik untuk dapat menjawab suatu butir soal dengan benar sesuai dengan tingkat kemampuan yang mereka miliki. Indikator ini biasanya dinyatakan melalui sebuah indeks yang membantu menentukan apakah suatu soal berada pada kategori mudah, sedang, atau sulit. Dalam konteks

evaluasi pembelajaran, suatu soal dikatakan berkualitas apabila tingkat kesukarannya berada pada kategori sedang, sehingga tidak terlalu tinggi maupun terlalu rendah (Sudijono & Pendidikan, 2007). Butir soal yang baik harus mampu menantang peserta didik tanpa membuat mereka kewalahan atau justru terlalu santai dalam menyelesaikannya.

Apabila soal yang diberikan terlalu sulit, siswa dapat mengalami tekanan, kehilangan motivasi, bahkan enggan melanjutkan pengerjaan karena merasa tidak memiliki kemampuan yang memadai. Sebaliknya, apabila soal terlalu mudah, peserta didik cenderung meremehkan tugas yang diberikan dan menganggap proses evaluasi tidak penting karena dapat diselesaikan tanpa usaha berarti. Kondisi tersebut dapat memengaruhi sikap dan keseriusan siswa dalam mengikuti proses penilaian. Oleh karena itu, penyusunan butir soal dengan tingkat kesukaran yang seimbang sangat diperlukan agar instrumen evaluasi dapat berfungsi optimal dalam mengukur kemampuan peserta didik secara objektif dan adil (Halik et al., 2019). Kriteria tingkat

kesukaran yang digunakan adalah sebagai berikut:

43, 44

Tingkat Kesukaran	Interpretasi
0,00 - 0,30	Sukar
0,31 - 0,70	Sedang
0,71 - 0,80	Mudah
0,81 - 1,00	Sangat Mudah

Gambar 5 Kriteria Tingkat Kesukaran

Tabel 3 Hasil Analisis Tingkat Kesukaran Butir Soal

No	Tingkat Kesukaran	Jumlah	%	No. Butir Soal
1.	Sukar	14	31,11 %	6, 8, 9, 11, 18, 21, 22, 25, 27, 28, 32, 34, 37, 41
2.	Sedang	7	15,55 %	2, 5, 12, 16, 31, 35, 45
3.	Mudah	22	48,88 %	1, 4, 7, 10, 13, 14, 15, 17, 19, 20, 23, 26, 29, 30, 33, 36, 38, 39, 40, 42,

4. Sangat Mudah 2 4,44 % 3, 24

Sumber: data diolah (2025)

Berdasarkan hasil analisis tingkat kesukaran butir soal, diperoleh bahwa sebagian besar soal berada pada kategori sedang dan mudah. Soal kategori sedang berjumlah 22 soal (48,88%), menunjukkan bahwa hampir separuh soal memiliki tingkat kesukaran yang ideal untuk mengukur kemampuan peserta didik. Sedangkan kategori mudah berjumlah 12 soal (26,66%), yang menunjukkan bahwa soal-soal tersebut relatif mudah dijawab oleh siswa, namun masih dapat digunakan selama tidak mendominasi keseluruhan instrumen evaluasi.

Di sisi lain, ditemukan pula soal dengan tingkat kesukaran sukar sebanyak 7 soal (15,55%), yang berarti soal tersebut cenderung sulit dijawab dan hanya dapat dikerjakan oleh sebagian kecil peserta didik. Selain itu terdapat 4 soal (8,88%) yang termasuk kategori sangat mudah, sehingga soal tersebut kurang efektif dalam mengukur kemampuan

karena terlalu mudah ditebak jawabannya. Dengan demikian, perbaikan diperlukan pada soal-soal yang terlalu sukar maupun terlalu mudah agar distribusi tingkat kesukaran lebih seimbang, sehingga instrumen evaluasi dapat mengukur kemampuan peserta didik secara lebih akurat.

Efektivitas Pengecoh

Pengecoh (distractor) merupakan sesuatu yang membuat orang terkecoh atau salah dalam memilih atau menilai sesuatu. Dalam konteks tes, pengecoh yang efektif adalah pilihan jawaban yang salah atau terlihat mirip dan dirancang untuk mengecoh atau mengalihkan perhatian peserta tes dari jawaban yang benar, sehingga beberapa peserta tes mungkin salah memilihnya. Efektivitas pengecoh menggambarkan sejauh mana pilihan jawaban yang bukan kunci dapat bekerja dengan baik dalam menarik peserta tes untuk memilihnya. Dengan demikian, kualitas pengecoh turut menentukan keberhasilan suatu butir soal dalam mengevaluasi kemampuan peserta didik secara akurat.

Simbol	Keterangan
**	Kunci Jawaban
++	Sangat Baik
+	Baik
-	Kurang Baik
--	Buruk
---	Sangat Buruk

Gambar 6 Simbol Kualitas Butir Soal

Tabel 4 Hasil Kualitas Pilihan Jawaban

Nomor	Kualitas Pilihan Ganda				
Soal	A	B	C	D	E
1	--	**	++	--	++
2	---	--	**	--	-
3	--	--	---	**	+
4	-	**	-	--	++
5	++	-	**	-	---
6	---	**	--	--	--
7	**	++	---	--	--
8	---	--	--	**	+
9	--	**	-	-	---
10	-	**	++	-	--
11	-	**	---	-	--
12	+	**	-	++	+
13	--	-	**	-	++
14	--	**	--	+	+
15	++	**	---	++	--

16	+	--	**	++	+
17	--	**	++	--	++
18	--	**	---	-	--
19	-	++	**	++	--
20	++	---	+	--	**
21	---	--	--	**	--
22	**	--	-	---	--
23	++	--	**	++	---
24	+	**	+	+	--
25	-	---	**	--	--
26	**	--	--	---	++
27	-	**	---	--	--
28	-	+	---	--	**
29	---	-	**	--	--
30	+	**	--	+	+
31	---	++	--	**	+
32	--	---	**	--	--
33	---	-	**	--	--
34	**	---	--	--	--
35	**	-	++	---	-
36	+	--	--	+	**
37	--	---	-	--	**
38	--	**	--	++	++
39	**	++	--	---	++

40	-	**	--	-	++
41	--	---	**	-	--
42	**	--	--	-	---
43	--	**	--	+	+
44	**	---	--	--	-
45	++	**	--	---	++

Sumber: data diolah (2025)

Berdasarkan hasil analisis, setiap butir soal menunjukkan variasi kualitas pengecoh. Pengecoh dengan simbol “+” dan “++” berfungsi dengan baik karena mampu menarik perhatian peserta didik untuk memilihnya, sehingga membantu membedakan antara peserta yang memahami konsep dan yang belum. Sebaliknya, pengecoh dengan simbol “-”, “--”, atau yang tidak dipilih sama sekali termasuk kategori rendah karena terlalu jelas salah dan tidak mampu mengalihkan peserta dari kunci jawaban. Dengan demikian, efektivitas pengecoh sangat menentukan kontribusi soal terhadap kemampuan evaluasi pemahaman peserta didik. Berdasarkan temuan tersebut:

- Pengecoh efektif disimbolkan dengan “+” / “++” yang berarti

sering dipilih peserta yang menjawab salah, berfungsi optimal.

- Pengecoh tidak efektif disimbolkan dengan “-” / “--” yang berarti tidak dipilih, tidak menarik peserta, kualitas rendah.

Secara keseluruhan, beberapa butir soal masih memerlukan revisi karena pengecohnya belum bekerja secara optimal, terutama soal yang memiliki banyak pengecoh bertanda “-” atau “-”. Sementara itu, soal dengan pengecoh bertanda “+” dan “++” sudah memenuhi kriteria pengecoh yang baik dan layak dipertahankan. Perbaikan terhadap pengecoh yang kurang efektif menjadi langkah penting untuk meningkatkan mutu instrumen evaluasi secara keseluruhan.

D. KESIMPULAN

Berdasarkan hasil analisis butir soal menggunakan program Anates, dapat disimpulkan bahwa instrumen tes yang dianalisis memiliki kualitas yang cukup baik dan reliabilitas sangat tinggi, dengan nilai reliabilitas sebesar 0,96. Hal ini menunjukkan bahwa tes memiliki tingkat konsistensi

yang sangat kuat dalam mengukur kemampuan peserta didik. Dari hasil analisis tingkat kesukaran, sebagian besar butir soal berada dalam kategori mudah hingga sedang, sementara beberapa soal termasuk kategori sangat sukar, sehingga perlu dilakukan penyesuaian agar distribusi soal menjadi lebih seimbang.

Dari segi daya pembeda, mayoritas butir soal menunjukkan nilai yang baik dan mampu membedakan peserta didik berkemampuan tinggi dan rendah, meskipun masih terdapat beberapa butir dengan daya pembeda negatif yang perlu direvisi. Sementara itu, hasil korelasi antara skor butir dan skor total menunjukkan bahwa sebagian besar butir memiliki hubungan signifikan, yang berarti butir-butir tersebut mengukur konstruk yang sama dengan keseluruhan tes.

Secara keseluruhan, tes ini dapat dikategorikan reliabel dan representatif, namun masih memerlukan revisi terhadap beberapa butir soal agar kualitas instrumen semakin meningkat. Dengan perbaikan pada soal yang terlalu sulit, daya pembeda rendah, serta pengecoh yang tidak berfungsi optimal, diharapkan tes dapat

memberikan hasil evaluasi yang lebih valid, objektif, dan menggambarkan kemampuan peserta didik secara menyeluruh.

DAFTAR PUSTAKA

- Ramadhan, M. F., Siroj, R. A., & Afgani, M. W. (2024). Validitas and reliabilitas. *Journal on Education*. [http://jonedu.org/index.php/joe]
- Fadli, R., Hidayati, S., Cholifah, M., Siroj, R. A., & Afgani, M. W. (2023). Validitas dan reliabilitas pada penelitian motivasi belajar pendidikan agama Islam menggunakan Product Moment. *JlIP: Jurnal Ilmiah Ilmu Pendidikan*. Radja, S. P., Bano, V. O., & Ina, A. T. Analisis kualitas butir soal hasil belajar peserta didik berdasarkan tingkat kesukaran, daya beda, dan efektivitas pengecoh di SMAN 1 Pandawai. *Jurnal Pendidikan Biologi, FKIP, Universitas Kristen Wira Wacana Sumba*.
- Fauziah, R., & Mustami, M. K. (2021). Analisis kualitas butir soal dalam evaluasi pembelajaran. *Jurnal Pendidikan*, 6(2), 123–132.
- Rohim, A., & Puspitasari, N. (2021). Pemanfaatan ANATES dalam analisis kualitas soal pilihan ganda. *Jurnal Evaluasi Pendidikan*, 9(1), 45–56.
- Salsabila, A., Muslim, S., & Anwar, K. (2022). Analisis kualitas butir soal berbasis teori tes klasik. *Jurnal Pendidikan Dasar*, 14(3), 210–220.
- Yuliasari, R., & Sutrisno, M. (2020). Validitas dan reliabilitas instrumen penilaian hasil belajar. *Jurnal Penilaian Pendidikan*, 8(2), 98–108.
- Nazliati, N (2018). Penggunaan Software Anates Dalam Pembelajaran Evaluasi Pendidikan Pada Mahasiswa Non Matematika Ftik lain Langsa. *Al Khawarizmi: Jurnal Pendidikan dan ...*, jurnal.ar-raniry.ac.id, <https://jurnal.ar-raniry.ac.id/index.php/alkhawarizmi/article/view/4503>
- Puteri, ND, Putricia, ND, Arsyanti, K, & ... (2025). Analisis Kelayakan Butir Soal Pada Handout Akuntansi Keuangan Dana Kas Kecil Berbasis E-Book Menggunakan Software Anates. *JIBEMA: Jurnal Ilmu ...*, jibema.murisedu.id, <http://jibema.murisedu.id/index.php/JIBEMA/article/view/108>
- Cahyani, DA, Aurelia, AH, Hakim, L, & ... (2024). Analisis Butir Soal HOTS

Akuntansi Perusahaan Jasa Menggunakan Software Anates pada Siswa SMK Kelas 11. ... *Pendidikan, Sosial dan ...*, ulilalbabinstitute.co.id, <https://ulilalbabinstitute.co.id/index.php/PESHUM/article/view/7108>

Augustia, AD, Agustia, CN, Azzahra, D, & ... (2025). Analisis Validitas dan Reliabilitas Soal Pilihan Ganda dengan Menggunakan Software Anates pada Mata Pelajaran Perpajakan. ... *Mahasiswa Ekonomi & ...*, ojs.pseb.or.id, <https://www.ojs.pseb.or.id/index.php/jme/article/view/1165>

Asyaroh, SP, Febriyanti, AC, & ... (2025). Analisis Butir Soal Hots Pada Elemen Perpajakan Fase F SMK Akuntansi Menggunakan Software Anates. *PESHUM: Jurnal ...*, ulilalbabinstitute.co.id, <https://ulilalbabinstitute.co.id/index.php/PESHUM/article/view/7176>

Hidayat, T, Biantoro, RNUR, & ... (2019). PELATIHAN ANALISIS SOAL DENGAN SOFTWARE ANATES UNTUK MENINGKATKAN KEMAMPUAN ASESMEN MAHASISWA PENDIDIKAN MATEMATIKA STKIP *PROSIDING* ...

repository.stkippacitan.ac.id, <https://repository.stkippacitan.ac.id/id/eprint/1269/>

Fiska, JM, Hidayati, Y, Qomaria, N, & ... (2021). Analisis butir soal ulangan harian IPA menggunakan software Anates pada pendekatan teori tes klasik. *Natural Science ...*, journal.trunojoyo.ac.id, <https://journal.trunojoyo.ac.id/nser/article/view/8133>

Syafiriya, DM, Rahmawati, PA, & ... (2024). Analisis Butir Soal HOTS untuk Menguji Pemahaman Siswa SMK Jurusan Akuntansi Menggunakan Software Anates. *Jurnal Nirta ...*, ejournal.nlc-education.or.id, <https://ejournal.nlc-education.or.id/index.php/JNSI/article/view/107>

Putra, RS, & Furqan, M (2024). Quality Analysis Of Islamic Religious Education Learning Questions At Smkn 1 Meulaboh Through Anates Software. *Tadrib: Jurnal Pendidikan Agama Islam*

Mutmainah, S, & Rofi, I (2024). Assessing of Biology Test Question Quality Using Anates Software.

EDUTEC: Journal of Education And Technology

Magdalena, I, Fauziah, SN, Faziah, SN, & Nupus, FS (2021). *Analisis validitas, reliabilitas, tingkat kesulitan dan daya beda butir soal ujian akhir semester tema 7 kelas III SDN Karet 1 Sepatan.*, ejournal.stitpn.ac.id,
<https://ejournal.stitpn.ac.id/index.php/bintang/article/view/1291>

Lestari, CA (2020). *Kajian Tingkat Kesulitan Dan Daya Beda Pada Soal Ujian Mata Pelajaran Bahasa Indonesia Kelas VIII SMP Negeri 1 Ujung Batu Kabupaten Rokan Hulu Tahun*, repository.uir.ac.id,
<https://repository.uir.ac.id/13577/>

Khasanah, U (2024). *Validitas, reliabilitas, daya beda, dan tingkat kesulitan soal penilaian akhir tahun bahasa Arab kelas vii tahun ajaran 2023/2024 di MTSN 2 Kota Malang.*, repository.um.ac.id,
<https://repository.um.ac.id/408058/>

Sudjana, N (2022). *Analisis validitas, reliabilitas, tingkat kesulitan, dan daya beda butir soal.*

Pradipta, AW, & Kurniawan, R (2023). *Tingkat Kesulitan dan Daya Beda Butir Soal Ujian Akhir Semester*

Matakuliah Penelitian Pendidikan. Jurnal Pendidikan

Rahma, A, Meivia, A, Hakim, L, & ... (2025). *Analisis Butir Soal K3LH pada Tingkat SMK Menggunakan Aplikasi ANATES. Journal of Learning and ...*, ejournal.reysha.org,
<https://ejournal.reysha.org/index.php/jt/article/view/39>

Rahmawati, N, Qoiri, FNP, Cahyo, FT, & ... (2024). *Analisis Butir Soal Hots Elemen Kesehatan Keselamatan Kerja Dan Lingkungan Hidup Untuk Menguji Pemahaman Siswa Smk Jurusan Akuntansi Keuangan Dan Hikamatzu| Journal ...*, yasyahikamatzu.com,
<https://yasyahikamatzu.com/index.php/hjm/article/view/160>

Ainiyah, NN, NP, AN, Hakim, L, & ... (2024). *Analisis Butir Soal Pilihan Ganda Materi Lingkup Kerja Akuntansi dan Keuangan Lembaga Menggunakan Anates. ... : Jurnal Penelitian dalam ...*, journal.upgris.ac.id,
<http://journal.upgris.ac.id/index.php/mediapenelitianpendidikan/article/download/21071/9245>

Siregar, SRN, Luckytasari, A, & ... (2024). Analisis Soal Hots Pada Isu Serta Perkembangan Terkini Akuntansi Keuangan Fase E SMK Menggunakan Aplikasi Anates. ... *Pendidikan, Sosial dan ...*, ulilalbabainstitute.co.id, <https://ulilalbabainstitute.co.id/index.php/PESHUM/article/view/7157>

Rachmadani, FC, Lestari, CD, & ... (2024). Analisis Butir Soal Pilihan Ganda pada Elemen Akuntansi Keuangan Guna Mengoptimasi Evaluasi Menggunakan Anates V4. 0. *Jurnal Nirta: Studi ...*, ejournal.nlc-education.or.id, <https://ejournal.nlc-education.or.id/index.php/JNSI/article/view/95>

Syafiriya, DM, Rahmawati, PA, & ... (2024). Analisis Butir Soal HOTS untuk Menguji Pemahaman Siswa SMK Jurusan Akuntansi Menggunakan Software Anates. *Jurnal Nirta ...*, ejournal.nlc-education.or.id, <https://ejournal.nlc-education.or.id/index.php/JNSI/article/view/107>

Cahyani, DA, Aurelia, AH, Hakim, L, & ... (2024). Analisis Butir Soal HOTS Akuntansi Perusahaan Jasa Menggunakan Software Anates pada

Siswa SMK Kelas 11. ... *Pendidikan, Sosial dan ...*, ulilalbabainstitute.co.id, <https://ulilalbabainstitute.co.id/index.php/PESHUM/article/view/7108>