



Alacrity : Journal Of Education

e-ISSN : 2775-4138

Volume 5 Issue 3 Oktober 2025

The Alacrity : Journal Of education is published 4 times a year in
(February, June, October)

Focus : Learning, Education, Including, Social, Curriculum,
Management Science, Educational Philosophy And Educational
Approaches.

LINK : <http://lpppipublishing.com/index.php/alacrity>

Menilai Mutu Instrumen Penelitian: Analisis Butir Soal Lingkup Kerja Akuntansi Fase E Berbasis Aplikasi Anates

**Rahma Ais Trilestari¹, Najwa Widia Ardana², Maulidia Tri Wulandari³, Monica
Desideria⁴, Ika Widyanti Putri⁵, Rizky Fathan Ramadhan⁶, Luqman Hakim⁷,
Vivi Pratiwi⁸**

^{1,2,3,4,5,6,7,8} Universitas Negeri Surabaya, Indonesia

ABSTRACT

Penelitian ini bertujuan untuk menganalisis mutu instrumen penilaian berupa 54 butir soal pilihan ganda pada materi Lingkup Kerja Akuntansi dan Keuangan Lembaga dengan menggunakan aplikasi Anates. Instrumen dianalisis berdasarkan lima aspek utama, yaitu validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan efektivitas pengecoh. Penelitian menggunakan pendekatan deskriptif kuantitatif dengan melibatkan 20 siswa SMK jurusan Akuntansi dari kelas X, XI, dan XII sebagai responden melalui pengumpulan data berbasis Google Form. Hasil analisis menunjukkan bahwa aspek validitas butir soal berada pada kategori sangat tinggi (17%), tinggi (28%), cukup (37%), rendah (15%), dan sangat rendah (4%). Reliabilitas instrumen memperoleh koefisien sebesar 0,97 yang termasuk kategori sangat tinggi, sehingga menunjukkan konsistensi instrumen yang sangat baik. Pada aspek daya pembeda, sebagian besar butir soal menunjukkan kemampuan diskriminatif yang memuaskan, meskipun beberapa butir soal seperti nomor 12, 15, 18, dan 40 perlu diperbaiki karena nilai diskriminasi yang rendah. Tingkat kesukaran soal didominasi kategori sedang sebesar 53,7%, diikuti kategori mudah (33,3%), sangat mudah (11,1%), dan satu butir sangat sukar (1,9%). Efektivitas pengecoh menunjukkan sebagian besar distraktor berfungsi baik, meskipun beberapa opsi tidak bekerja optimal dan perlu diperbaiki. Secara keseluruhan, hasil penelitian menegaskan bahwa instrumen penilaian yang dianalisis memiliki mutu yang baik dan layak digunakan, namun tetap memerlukan revisi pada butir-butir soal tertentu untuk meningkatkan kualitas evaluasi pembelajaran.

ARTICLE INFO

Article history:
Received
01 September 2025
Revised
10 Oktober 2025
Accepted
25 November 2025

Keywords

Analisis Butir Soal, Validitas, Reliabilitas, Aplikasi Anates, Akuntansi

Corresponding

Author :

24080304117@mhs.unesa.ac.id

PENDAHULUAN

Pendidikan memiliki peranan strategis dalam mengembangkan potensi manusia. Seperti dijelaskan oleh (Hawari et al., 2024), pendidikan merupakan proses sistematis untuk membantu peserta didik mencapai kematangan pribadi

serta kemampuan berpikir yang lebih matang. Lebih jauh, pendidikan tidak sekadar sebagai sarana transfer pengetahuan, tetapi juga sebagai instrumen penting proses pembentukan karakter dan sebagai persiapan bagi peserta didik agar mampu berperan aktif dan berkontribusi dalam kehidupan bermasyarakat. Melalui pendidikan, peserta didik dibimbing untuk berpikir kritis, beretika, dan mampu beradaptasi terhadap perubahan sosial dan teknologi yang terus berkembang. Salah satu unsur penting dalam pendidikan adalah kegiatan penilaian, karena untuk menilai sejauh mana tujuan pembelajaran dapat dicapai. Penilaian yang berkualitas tidak hanya mengukur hasil belajar, tetapi juga memberikan umpan balik bagi pendidik dan peserta didik untuk meningkatkan mutu proses pembelajaran, sebagaimana ditegaskan oleh (Ningsih et al., 2024) bahwa penilaian memberikan informasi penting bagi pendidik dan peserta didik untuk memperbaiki proses belajar. Penilaian menjadi bagian integral dari sistem pendidikan yang menentukan keberhasilan pencapaian kompetensi peserta didik.

Instrumen penilaian dalam evaluasi pembelajaran adalah alat atau perangkat yang digunakan untuk menilai sejauh mana peserta didik mencapai tujuan pembelajaran, baik dari aspek pengetahuan, keterampilan, maupun sikap. Instrumen ini mencakup berbagai bentuk, seperti tes tertulis, lembar observasi, rubrik penilaian, dan portofolio, yang disusun agar dapat menghasilkan data yang sah dan akurat mengenai kemampuan peserta didik (Zahroh et al., 2025). Mutu instrumen penilaian menjadi aspek yang sangat penting agar hasil evaluasi benar-benar mencerminkan kemampuan peserta didik. Instrumen yang kurang valid atau reliabel dapat menimbulkan interpretasi yang keliru mengenai kemampuan siswa, sehingga keputusan pembelajaran yang diambil menjadi kurang tepat. (Wilsa, Asrizal Wahdan, Rusilowati, Ani, Susilaningsih, Endang, Jaja, Nurpadillah, 2023) menekankan bahwa kualitas instrumen dapat dilihat dari aspek validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Pemenuhan kriteria tersebut memastikan setiap butir soal dapat menilai kemampuan peserta didik secara akurat dan adil. Analisis butir soal menjadi langkah penting untuk meningkatkan mutu penilaian dan menekan risiko ketidakakuratan hasil evaluasi.

Perkembangan teknologi pendidikan menyediakan alat bantu untuk menganalisis kualitas butir soal secara cepat dan komprehensif. Salah satu inovasi yang banyak digunakan adalah aplikasi Anates, yang memudahkan guru dan peneliti menghitung validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh secara akurat (Sabela et al., 2025). (Hervi & Afdarina, 2023) menemukan bahwa penggunaan Anates 4.0 mampu membantu

pendidik dalam menilai karakteristik butir soal dengan lebih efisien. Selain itu (Febriant et al., 2025), menyatakan bahwa Anates memberikan informasi objektif dan berbasis data mengenai kualitas soal dalam evaluasi, sehingga meningkatkan efektivitas dan keandalan evaluasi pembelajaran secara keseluruhan. Anates memungkinkan evaluasi yang lebih komprehensif terkait validitas, reliabilitas, tingkat kesukaran, daya pembeda soal, dan efektivitas pengecoh, dan membantu pendidik dalam perbaikan kualitas soal secara berkelanjutan.

Selain aspek teknis, penilaian juga perlu diarahkan untuk mengukur *Higher Order Thinking Skills* (HOTS). HOTS merupakan kemampuan berpikir tingkat tinggi yang mencakup keterampilan berpikir kritis, kreatif, analitis, dan evaluatif. HOTS menuntut peserta didik tidak hanya mengingat, tetapi juga mampu mengamati, menganalisis, menalar, serta mencipta solusi atas masalah yang kompleks (Tasrif, 2023). Implementasi soal berbasis HOTS dapat mengoptimalkan kapabilitas berpikir kritis, inovatif, dan kemampuan menyelesaikan masalah peserta didik, yang esensial dalam menghadapi realitas yang kompleks dan dunia pendidikan modern (Anayanti et al., 2024). Analisis butir soal menggunakan Anates memungkinkan guru menilai sejauh mana instrumen benar-benar mengukur kemampuan berpikir tingkat tinggi tersebut. Pendekatan ini memastikan bahwa evaluasi tidak hanya menguji hafalan, tetapi juga keterampilan berpikir kompleks yang relevan dengan tuntutan pendidikan saat ini.

Berdasarkan paparan tersebut, penelitian ini berfokus pada analisis mutu instrumen penilaian butir soal lingkup kerja akuntansi berbasis aplikasi Anates. Analisis dilakukan terhadap aspek validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh untuk memperoleh gambaran menyeluruh mengenai kualitas soal yang digunakan. Hasil penelitian diharapkan memberikan masukan praktis bagi pendidik dalam menyusun instrumen penilaian yang akurat, adil, dan menantang kemampuan berpikir tingkat tinggi peserta didik. Lebih jauh, penelitian ini bertujuan mendorong penerapan asesmen berbasis data yang akuntabel sehingga mutu instrumen penilaian tidak hanya sebagai alat pengukur, tetapi juga sebagai sarana pengembangan proses pembelajaran yang bermakna.

METODE PENELITIAN

Penelitian ini mengimplementasikan metode deskriptif kuantitatif dengan tujuan untuk mengevaluasi kualitas dari butir pertanyaan pilihan ganda yang terdapat dalam handout berbasis e-book mengenai Lingkup Kerja di bidang Akuntansi dan Keuangan Lembaga. Metode kuantitatif ini didasari oleh

paradigma positivistik yang fokus pada pengukuran variabel secara objektif. Variabel yang dianalisis meliputi validitas, reliabilitas, daya diferensiasi, dan tingkat kesulitan. Analisis data dilakukan menggunakan perangkat lunak Anates yang menawarkan cara penilaian instrumen secara sistematis dan objektif (*Metode Kuantitatif.Pdf*, n.d.).

Jumlah subjek penelitian adalah 20 siswa dari kelas X, XI, dan XII program keahlian akuntansi di beberapa Sekolah Menengah Kejuruan. Pemilihan subjek dilakukan dengan prinsip partisipasi sukarela tanpa adanya seleksi tertentu. Proses pengumpulan data menggunakan Google Form memungkinkan peserta untuk menjawab soal secara mandiri dengan perangkat digital mereka sendiri. Instrumen penelitian terdiri dari 54 butir soal pilihan ganda yang masing-masing memiliki lima pilihan jawaban. Penyusunannya disesuaikan dengan kompetensi dasar dari kurikulum akuntansi keuangan dan merujuk pada Taksonomi Bloom untuk meratakan tingkat kognitif. Proses pengumpulan data dimulai dengan penyusunan butir soal sesuai dengan tujuan pembelajaran, kemudian divalidasi oleh dosen yang ahli, dilanjutkan dengan distribusi soal melalui Google Form, hingga pengunduhan hasil jawaban dari peserta.

Metode pengumpulan data meliputi penyusunan butir soal yang sesuai dengan tujuan pembelajaran, validasi oleh dosen pengampu, distribusi soal menggunakan Google Form, serta pengunduhan hasil jawaban dari peserta. Format e-book memungkinkan akses terhadap materi dan soal secara fleksibel dalam periode waktu yang telah ditentukan. Semua jawaban peserta kemudian dianalisis melalui perangkat lunak Anates.

Anates adalah sebuah aplikasi yang dirancang untuk menganalisis kualitas alat ukur dalam format pilihan ganda dan essay. Beberapa fitur yang ditawarkan meliputi perhitungan nilai, koefisien keandalan, kemampuan membedakan, tingkat kesulitan, hubungan butir-total, serta analisis sejauh mana pengecoh berfungsi. Aplikasi ini dapat memberikan hasil analisis dengan cepat, tepat, dan objektif, sehingga membantu dalam menilai kualitas tes secara komprehensif (Zahir, 2018).

HASIL DAN PEMBAHASAN

Sebanyak 54 butir soal pilihan ganda telah dianalisis berdasarkan hasil jawaban 20 siswa yang diambil melalui Google Form dan diolah menggunakan software Anates.

Validitas

Validitas berasal dari kata *validity* yang memiliki arti sejauh mana ketepatan dan kecermatan suatu instrumen dalam melakukan fungsi ukurnya (Ramadhan et al., 2024). Artinya hasil ukur dari pengukuran tersebut

merupakan besaran yang mencerminkan secara tepat fakta atau keadaan sesungguhnya dari apa yang diukur.

Dalam penelitian ini menggunakan aplikasi AnatesV4, korelasi antara skor validitas dihitung dengan membandingkan hasil jawaban peserta didik dengan total skor pada pertanyaan yang telah dikerjakan. Hubungan ini memberikan pemahaman dan gambaran mengenai seberapa baik soal yang telah diberikan dapat membedakan peserta yang memiliki pemahaman mendalam mengenai materi dari mereka yang kurang memahami.

Tabel 1.

Tabel Korelasi Validitas

Tingkat Korelasi Validitas	Interpretasi
0,80 - 1,00	Validasi soal sangat tinggi
0,60 - 0,80	Validasi soal tinggi
0,40 - 0,60	Validasi soal cukup
0,20 - 0,40	Validasi soal rendah
0,00 - 0,20	Validasi soal sangat rendah

Sumber: data diolah (2025)

Tabel 2.

Hasil Analisis Validitas

No	Tingkat Daya Beda	Jumlah	Persentase	Nomor Butir Soal
1.	Validasi Soal Sangat Tinggi	9	17%	22, 28, 32, 37, 39, 41, 45, 49, 52
2.	Validasi Soal Tinggi	15	28%	4,9,11,13,14,16,19,23,29, 33,35,44,46,47,53
3.	Validasi Soal Cukup	20	37%	1,2,3,5,7,8,10,17,20,25,26,27, 31,34,36,38,48,50,51,54
4.	Validasi Soal Rendah	8	15%	6,12,15,21,24,30,42,43
5.	Validasi Soal Sangat Rendah	2	4%	18,40

Sumber: data diolah (2025)

Berdasarkan analisis validitas yang telah dilakukan, dapat disimpulkan bahwa kategori validitas soal dengan tingkat sangat tinggi mencapai 17% yang terdiri dari 9 butir soal. Kategori validitas soal yang tergolong tinggi adalah 28% dengan 15 butir soal. Sedangkan kategori validitas soal yang dinilai cukup berada pada angka 37% dengan 20 butir soal. Kategori validitas soal dengan tingkat rendah adalah 15% yang mencakup 8 butir soal. Terakhir, terdapat kategori validitas soal yang sangat rendah sebesar 4% dengan 2 butir soal.

Reliabilitas

Reliabilitas dalam penelitian ini diartikan sebagai tingkat kesesuaian skor yang diperoleh dari suatu tes ketika dilakukan di situasi serupa. Dengan kata lain, reliabilitas mengukur sejauh mana sebuah alat dapat diandalkan dalam menghasilkan data yang konsisten dan minim dari kesalahan pengukuran (Sekolah & Kejuruan, 2020).

Tabel 3.

Tabel Kriteria Penafsiran Reabilitas Soal	
Nilai Reabilitas	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)

Sumber: data diolah (2025)

Analisis yang dilakukan menunjukkan bahwa koefisien reliabilitas untuk item soal terkait laporan keuangan pada perusahaan jasa, dagang, dan manufaktur mencapai 0,97. Berdasarkan kriteria untuk menafsirkan reliabilitas, angka ini termasuk dalam kategori sangat tinggi atau sempurna dengan rentang 0,90–1,00. Ini menunjukkan bahwa konsistensi internal yang sangat baik dan dapat diandalkan sebagai sarana untuk menilai pembelajaran. Koefisien 0,97 juga menunjukkan bahwa sebagian besar variasi skor dari peserta didik disebabkan oleh perbedaan kemampuan yang nyata, bukan karena kesalahan dalam pengukuran, sehingga alat ini dapat diandalkan dalam pengukuran, sehingga alat ini dapat diandalkan untuk menilai kompetensi siswa.

Daya Pembeda

Daya pembeda adalah ukuran seberapa efektif suatu item tes dapat membedakan antara pelajar yang memiliki kemampuan tinggi dan yang memiliki kemampuan rendah. Makin meningkatnya nilai daya pembeda, semakin tinggi pula kualitas butir tes dalam mengidentifikasi perbedaan tingkat kemampuan peserta. Analisis ini dilakukan menggunakan perangkat lunak Anates dengan sekitar 54 butir soal dan 20 peserta. Nilai indeks daya pembeda bervariasi antara 0% hingga 100%. Daya pembeda berada dalam kategori baik hingga sangat baik. Berikut adalah hasil yang mencerminkan variasi keterampilan untuk setiap pembeda.

Tabel 4.

Tabel Daya Pembeda Soal

Indeks DP	Interpretasi	Nomor Soal
100%	Sangat Baik	28, 39, 41, 45, 52, 53
80%	Sangat Baik	2, 9, 22, 25, 32, 33, 35, 37, 46, 47, 49

Tabel 5.

Tabel Daya Pembeda Soal

60%	Baik	3, 4, 11, 13, 14, 17, 19, 20, 21, 23, 26, 27, 29, 34, 36, 38, 44, 48, 50, 51, 54
40%	Cukup	6, 7, 8, 10, 16, 24, 30, 31, 42, 43
20%	Rendah	1, 5, 12, 15, 40
0%	Sangat Rendah	18

Sumber: data diolah (2025)

Mengacu pada data diatas, sebagian besar butir soal mempunyai daya pembeda tinggi dengan skor minimal 40%. Item soal dengan daya pembeda 100% (nomor 28, 39, 41, 45, 52, dan 53) dikategorikan sangat baik karena dapat menggolongkan siswa dengan tingkat keterampilan tinggi dan rendah secara efektif. Sebaliknya, butir soal dengan daya pembeda rendah atau 0% (nomor 18) tidak efektif dalam membedakan kemampuan. Hasil penelitian membuktikan bahwa umumnya butir soal berkualitas baik. Namun, item soal nomor 12, 15, 18, dan 40 perlu ditinjau ulang karena skor daya diskriminasinya rendah. Skor daya diskriminasi rendah dapat disebabkan oleh soal yang terlalu mudah atau sulit, kesalahan jawaban, atau interferensi malfungsi.

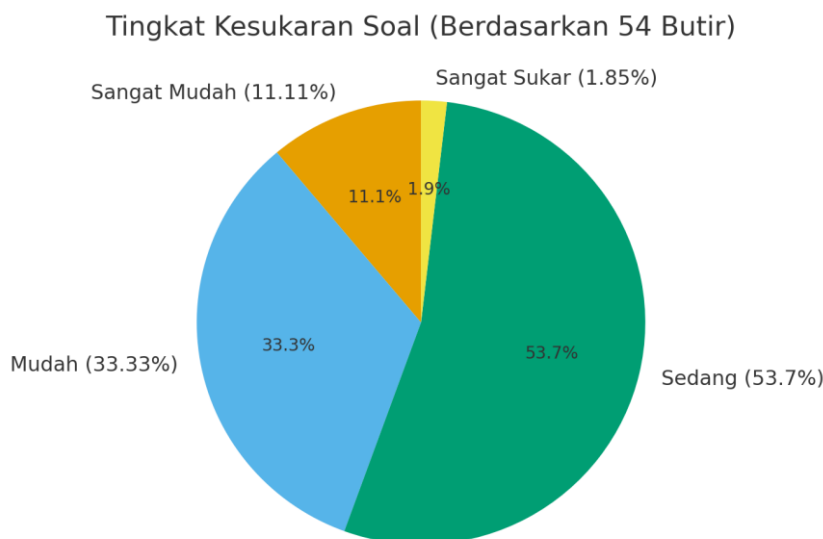
Dapat disimpulkan bahwa alat evaluasi yang diterapkan menunjukkan kemampuan diskriminatif secara keseluruhan yang memuaskan. Butir-butir soal dengan kemampuan diskriminatif tinggi perlu dipertahankan untuk digunakan dalam penilaian selanjutnya, sementara butir-butir soal dengan kemampuan diskriminatif rendah perlu ditingkatkan untuk meningkatkan kualitas alat ukur.

Tingkat Kesukaran

Tingkat kesukaran merupakan indeks yang menunjukkan sejauh mana soal itu benar-benar mengukur apa yang ingin diukur Sugiyono (2005). Dari hasil analisis terhadap 54 butir soal, diketahui bahwa tingkat kesukaran berfungsi menunjukkan apakah sebuah soal cenderung mudah atau sulit dijawab dengan benar oleh peserta tes, yang dinyatakan dalam bentuk persentase. Semakin besar nilai persentasenya, maka semakin mudah soal tersebut dikerjakan.

Hasil analisis memperlihatkan bahwa sebagian besar soal termasuk dalam kategori sedang, yaitu sebanyak 29 butir soal (53,7%) yang meliputi nomor 2, 6,

9, 17, 18, 21, 24, 25, 27, 28, 30, 32, 33, 34, 35, 36, 39, 40, 41, 42, 45, 48, 50, 51, 52, 53, dan 54. Selanjutnya, terdapat 18 butir soal (33,3%) yang tergolong mudah, yaitu nomor 3, 4, 7, 10, 11, 13, 14, 19, 20, 22, 23, 26, 29, 37, 38, 43, 44, 46, 47, dan 49, Sementara itu, 6 butir soal (11,1%) termasuk kategori sangat mudah, yaitu nomor 1, 5, 8, 15, dan 16. Di sisi lain, terdapat 1 butir soal (1,9%) yang tergolong sangat sukar, yaitu nomor 12.



Gambar 1.
Diagram Tingkat Kesukaran Soal

Merujuk pada hasil telaah butir soal diperoleh bahwa komposisi soal dalam tes ini tergolong baik, karena sebagian besar butir berada dalam kategori sedang, yang berarti soal dapat menjadi tolok ukur kemampuan peserta secara optimal. Soal yang sangat mudah dan sangat sukar tetap perlu direvisi agar distribusi tingkat kesukaran lebih seimbang dan kualitas tes dapat meningkat.

Efektivitas Pengecoh

Dalam jenis soal pilihan ganda, selain terdapat satu jawaban yang benar, juga ada beberapa jawaban yang keliru yang dikenal sebagai pengecoh atau distraktor. Pengecoh atau distraktor adalah jawaban yang tidak benar dalam soal pilihan ganda yang dirancang untuk menarik perhatian peserta ujian yang belum menguasai materi dengan baik. Dalam soal yang mengukur kemampuan berpikir tingkat tinggi, pilihan yang mengalihkan perhatian tidak hanya sekedar salah, tetapi juga harus tampak masuk akal, relevan, dan terkait dengan konteks soal, sehingga peserta diharuskan untuk menganalisis dan mempertimbangkan setiap opsi sebelum membuat pilihan. Pilihan jawaban yang digunakan adalah sebagai berikut.

Tabel 6.
Tabel Efektivitas Pengecoh
Kualitas Pilihan Jawaban

Nomor Soal	A	B	C	D	E
1	--	**	---	--	--

Tabel 7.
Tabel Efektivitas Pengecoh
Kualitas Pilihan Jawaban

Nomor Soal	A	B	C	D	E
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	**	+	---	+	--

Tabel 8.
Tabel Efektivitas Pengecoh
Kualitas Pilihan Jawaban

Nomor Soal	A	B	C	D	E
41	+	**	+	+	+
42	+	++	-	**	+
43	--	++	-	-	**
44	**	++	--	--	+
45	+	**	---	--	-
46	--	**	++	--	++
47	--	**	++	--	++
48	-	---	**	--	--
49	++	**	-	--	-
50	+	**	+	--	+
51	**	--	-	++	-
52	-	**	+	+	+
53	+	**	--	+	--
54	+	---		++	--

Sumber: data diolah (2025)

Kualitas pilihan salah dalam hasil analisis tes yang ada pada tabel menunjukkan bahwa setiap soal disertai dengan pilihan salah yang memiliki berbagai tingkat kualitas. Pilihan salah yang efektif adalah jawaban yang dipilih oleh peserta yang tidak sepenuhnya mengerti konsep yang benar, sedangkan pilihan salah yang kurang efektif tidak dapat menarik perhatian peserta. Berdasarkan informasi yang didapatkan:

- a. Pengecoh yang menggunakan simbol “++” dianggap sangat baik karena distraktor tersebut dipilih dengan proporsi seimbang dan efektif untuk mengelabui peserta berkemampuan rendah.
- b. Pengecoh yang menggunakan simbol “+” dianggap baik karena distraktor tersebut masih berfungsi namun intensitas pemilih tidak terlalu banyak.
- c. Pengecoh yang menggunakan simbol “-” dianggap kurang baik karena distraktor tersebut jarang dipilih sehingga daya tariknya rendah.
- d. Pengecoh yang menggunakan simbol “--” dianggap tidak baik karena distraktor tersebut tidak dipilih sama sekali sehingga tidak berfungsi.

Hasil analisis menunjukkan bahwa butir soal yang memiliki banyak pengecoh dengan simbol “-” atau “--” kurang berfungsi dan tidak berfungsi. Butir soal yang memiliki pengecoh tersebut memerlukan revisi, baik dari segi tingkat kesulitan soal, kesesuaian konteks maupun kejelasan bahasa. Sedangkan butir soal yang memiliki banyak pengecoh dengan simbol “+” atau “++” sudah cukup memenuhi kriteria sebagai pengecoh yang baik. Pengecoh yang baik masih relevan dengan materi, tampak logis dan mampu menarik peserta didik sehingga banyak dipilih oleh siswa yang berkemampuan rendah, sementara siswa yang berkemampuan tinggi tetap dapat membedakan mana jawaban yang benar.

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa butir soal pilihan ganda dalam *handout* berbasis *e-book* Lingkup Kerja pada Bidang Akuntansi dan Keuangan Lembaga menggunakan aplikasi AnatesV4 dikategorikan sangat baik atau memuaskan pada uji validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan efektivitas pengecoh yang berfungsi dengan baik. Hasil analisis tersebut bisa dilihat berdasarkan hasil validitas 17% sangat tinggi, 28% tinggi, 37% cukup, 15% rendah, dan 4% sangat rendah. Untuk uji reliabilitas soal tergolong sangat tinggi dengan nilai sebesar 0,97. Dari aspek daya pembeda, pada butir soal nomor 12, 15, 18, dan 40 perlu ditinjau ulang karena skor daya diskriminasinya rendah. Dari tingkat kesukaran sebagian besar soal termasuk dalam kategori sedang sebesar 53,7%, kategori yang tergolong mudah sebesar 33,3%, kategori sangat mudah sebesar 11,1%, dan terdapat 1 butir soal yang tergolong sangat sukar sebesar 1,9%.

DAFTAR PUSTAKA

- Amalia, A. N., & Widayati, A. (2012). *Analisis Butir Soal Tes Kendali Mutu Kelas XII SMA Mata Pelajaran Ekonomi Akuntansi di Kota Yogyakarta Tahun 2012*. X(1), 1-26.

- Amanda, C., Ran, M., Avrista, S., Hakim, L., & Pratiwi, V. (2024). *Analisis Butir Soal Komputer Akuntansi Kelas XI Melalui Aplikasi Anates*. 4(1), 1-11.
- Anayanti, D. K., Dini, F. S., Putri, S. D., & Sari, L. E. (2024). *Implementasi pembelajaran berbasis High Order Thinking Skills (HOTS) Mahasiswa Kelas 4D PGSD Universitas Muhammadiyah Surakarta*. 1(1), 12-21.
- Arif, M. (2014). *Penerapan Aplikasi Anates Bentuk Soal Pilihan Ganda*. 1(1), 1-9.
- Arlindianie, N. K., A'isyah, E. T., Rahmah, E. A., Hakim, L., & Pratiwi, V. (2024). *Analisis Kualitas Butir Soal Akuntansi Perusahaan Jasa, Dagang, dan Manufaktur Menggunakan Aplikasi Anates Terhadap Siswa-Siswi SMK/MAK*. 4(1), 1-18.
- Az-Zahra, I. B., & Rochmawati. (2024). *Pengembangan Asesmen Berbasis Kurikulum Merdeka Mata Pelajaran Dasar Akuntansi Berbantuan Liveworksheet di SMK PGRI 2 Sidoarjo*. 12(2), 317-330.
- Dianova, F. R., & Anwar, N. (2024). *Analisis Butir Uji Validitas , Reliabilitas , Tingkat Kesukaran , dan Daya Pembeda Soal Sumatif Bahasa Arab SD Islam*. 1(3), 1-13.
- Fatimah, L. U., & Alfath, K. (2019). *Analisis Kesukaran Soal, Daya Pembeda dan Fungsi Distraktor*. 8(2), 37-64.
- Febriant, S., Khoerunnisa, Z., Chotimah, K. A., Dewi, L. P., Aliandra, A. L., & Putri, H. E. (2025). *Pemanfaatan ANATES untuk Menganalisis Validitas dan Reliabilitas Butir Soal Uraian IPAS Materi Perkembangbiakan Tumbuhan*. 8(1), 232-242. <https://doi.org/10.56338/jks.v8i1.6688>
- Hakim, R. Al, Mustika, I., & Yuliani, W. (2021). *Validitas dan Reliabilitas Angket Motivasi Berprestasi*. 4(4), 263-268.
- Hawari, M. F. A., Istiqomah, T. I., & Bakar, M. Y. A. (2024b). *Tujuan Pendidikan dalam Perspektif Islam*.
- Hervi, F., Afdarina, O., Syamsurizah, & Lutfri. (2023). *Analysis of Science Subject Questions in Class IX Junior High School using Anates 4 . 0 Program*. 14, 638-648.
- Ismail Suardi Wekke, D. (2019). *Metode Penelitian Sosial* (pp. 1-181).
- Journal, H., Multidisiplin, O., Amalia, R., Yoko, Y. U., Fauziyah, L. U., Pratiwi, V., Hakim, L., & Surabaya, U. N. (2024). *Analisis Soal Menggunakan Anates Mata Pelajaran Spreadsheet Dikelas Smk Akuntansi Kelas 10*. 2(1), 1-8.
- Jusrianto, Zahir, A., & Megawati. (2018). *Analisis Kualitas Tes Ujian Akhir Semester Mata Kuliah Pengetahuan Komputer*. 1(1), 1-9.
- Kadir, A. (2015). *Menyusun dan Menganalisis Tes Hasil Belajar*. 8(2), 70-81.
- Mulyati, B. (2021). *Analisis Butir Soal Uraian Mata Kuliah Pengantar Akuntansi 1*. 4(2), 178-187.
- Ningsih, N. R., Rosidah, N. A., & Pradana, D. A. (2024). *The Role of Formative*

- Assessment in Developing English Language Curriculum and Learning*. 1(2), 70–78.
- Putri, D. R., Anugrah, P., Lestari, A. S., Hakim, L., & Pratiwi, V. (n.d.). *Analisis Butir Soal Perpajakan Kelas XI pada Siswa SMK Menggunakan Aplikasi Anates*. 2047–2060.
- Putri, N. D., Arsyanti, K., Patricia, N. D., Hakim, L., & Pratiwi, V. (2025). *Analisis Kelayakan Butir Soal pada Handout Akuntansi Keuangan Dana Kas Kecil Berbasis E-Book Menggunakan Software Anates* (pp. 1–10).
- Rahma, A., Meivia, A., Hakim, L., & Pratiwi, V. (2025). *Analisis Butir Soal K3LH pada Tingkat SMK Menggunakan Aplikasi Anates*. 02(1), 1–8.
- Ramadhan, M. F., Siroj, R. A., & Afgani, M. W. (2024). *Validitas and Reliabilitas*. 06(02), 10967–10975.
- Sabela, O. R., Krisdayanty, D., Taqqiyah, A. Z., Hakim, L., Pratiwi, V., & Surabaya, U. N. (2025). *Analisis Butir Soal HOTS Elemen Dokumen Berbasis Digital (FASE E) Menggunakan Program Anates*. 6(1), 251–262.
- Syafiriya, D. M., Rahmawati, P. A., Qolbi, Z., Hakim, L., & Pratiwi, V. (2024). *Analisis Butir Soal HOTS untuk Menguji Pemahaman Siswa SMK Jurusan Akuntansi Menggunakan Software Anates*. 4(1), 1–23.
- Tasrif. (2023). *Higher Order Thinking Skills (HOTS) dalam Pembelajaran Social Studies di Sekolah Menengah Atas*. 10(1), 50–61.
- Warju, Ariyanto, S. R., Soeryanto, & Trisna, R. A. (2020). *Analisis Kualitas Butir Soal Tipe HOTS pada Kompetensi Sistem Rem di Sekolah Menengah Kejuruan*. 17(1), 1–11. <https://doi.org/10.23887/jptk-undiksha.v17i1.22914>
- Wilsa, Asrizal Wahdan, Rusilowati, Ani, Susilaningsih, Endang, Jaja, Nurpadillah, V. (2023). *Validity, Reliability, and Item Characteristics of Cell Material Science Literacy Assessment Instruments*. 27(2). <https://doi.org/10.21831/pep.v27i2.61577>
- Zahroh, F. L., Muflih, M., & Hilmiyati, F. (2025a). *Pembuatan Instrumen Evaluasi Pembelajaran; Analisis Literatur Review*. 5, 319–328.