

Analisis Kualitas Butir Soal Prinsip-Prinsip dan Konsep Akuntansi serta Perbankan Dasar Menggunakan Software Anates

Erlinda Nur Fadillah^{1*}, Luqman Hakim², Vivi Pratiwi³, Nur Dwi Kholifatu Rohmah⁴, Nadya Aprillia Putri⁵, Btary Quinnina Azzahra⁶, Anggita Permadi Assufi Nur Najwa⁷, Ernezha Putri Pariarzha⁸, Zebina Fataturrohmah⁹,

^{1,2,3,4,5,6,7,8,9} Universitas Negeri surabaya, Surabaya, Indonesia

* E-mail Korespondensi: 24080304022@mhs.unesa.ac.id

Information Article

History Article

Submission: 20-11-2025

Revision: 29-11-2025

Published: 02-12-2025

DOI Article:

[10.24905/mlt.v6i2.289](https://doi.org/10.24905/mlt.v6i2.289)

A B S T R A K

Penelitian ini bertujuan untuk menganalisis kualitas butir soal berbasis Higher Order Thinking Skills (HOTS) pada mata pelajaran Prinsip-Prinsip dan konsep Akuntansi serta Perbankan Dasar menggunakan software Anates versi 4.0. Instrumen terdiri atas 63 butir soal yang dikembangkan oleh tujuh anggota kelompok dan diuji cobakan kepada 36 responden. Analisis dilakukan terhadap beberapa aspek, yaitu validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Hasil penelitian menunjukkan nilai reliabilitas sebesar 0,97, menandakan konsistensi tes yang sangat tinggi. Dari 63 butir soal, 46 soal dinyatakan valid dan 17 perlu revisi. Berdasarkan tingkat kesukaran, mayoritas soal tergolong mudah hingga sedang, sementara dari aspek daya pembeda, 68,26% soal berada pada kategori baik hingga sangat baik. Secara umum, terdapat 32 butir soal layak digunakan tanpa revisi, 23 soal layak dengan perbaikan, dan 8 butir soal perlu diperbaiki secara menyeluruh. Temuan ini menegaskan bahwa penggunaan software Anates efektif dalam membantu guru atau pengembang instrumen untuk mengevaluasi kualitas butir soal secara objektif dan sistematis, sehingga dapat menghasilkan alat ukur yang valid, reliabel, dan sesuai dengan karakteristik soal HOTS.

Kata Kunci: analisis butir soal; anates; HOTS

A B S T R A C T

This study aims to analyze the quality of Higher Order Thinking Skills (HOTS)-based items in the subjects of Accounting Principles and Concepts and Basic Banking using Anates software version 4.0. The instrument consists of 63 items developed by seven group members and tested on 36 respondents. The analysis was conducted on several aspects, namely validity, reliability, difficulty level, discriminating power, and distractor effectiveness. The results showed a reliability value of 0.97, indicating a very high test consistency. Of the 63 questions, 46 were declared valid and 17 needed revision. Based on the level of difficulty, the majority of questions were classified as easy to moderate, while in terms of discriminating power, 68.26% of the

Acknowledgment

questions were in the good to very good category. In general, there were 32 items suitable for use without revision, 23 items suitable with improvement, and 8 items that needed to be completely revised. These findings confirm that the use of Anates software is effective in helping teachers or instrument developers to evaluate the quality of items objectively and systematically, thereby producing valid, reliable measuring instruments that are in line with the characteristics of HOTS items.

Key word: item analysis; HOTS; software anates

© 2025 Published by multiplier. Selection and/or peer-review under responsibility of multiplier

PENDAHULUAN

Evaluasi hasil belajar memegang peranan penting dalam kegiatan pembelajaran karena membantu memastikan bahwa tujuan yang telah direncanakan benar-benar tercapai. Melalui proses evaluasi, guru dapat menilai sejauh mana siswa memahami materi yang telah dipelajari (Fuadiy, 2021). Secara umum, proses penilaian dibagi menjadi tiga aspek utama, yaitu aspek kognitif, aspek afektif, dan aspek psikomotor. Ketiga ranah ini digunakan untuk melihat kemampuan peserta didik dari sisi pengetahuan, sikap, dan keterampilan (Arif, 2014). Dalam mata pelajaran Prinsip-Prinsip dan Konsep Akuntansi Dasar serta Perbankan Dasar, evaluasi dilakukan melalui tes tertulis yang terdiri atas butir-butir soal pilihan ganda. Soal evaluasi digunakan untuk berbagai keperluan dalam bidang pendidikan. Instrumen ini berfungsi dalam proses seleksi dan penempatan peserta didik, memberikan motivasi serta arahan, menyediakan umpan balik bagi guru dan siswa, serta membantu mengidentifikasi kelemahan melalui diagnosis dan kegiatan remedial. Selain itu, hasil evaluasi juga dapat dimanfaatkan untuk meningkatkan proses pembelajaran, memperbaiki kurikulum dan program pendidikan, bahkan mendukung pengembangan ilmu pengetahuan (Ramadhan et al., 2023). Namun, agar evaluasi menghasilkan informasi yang maksimal, setiap butir tes perlu disusun berdasarkan standar kualitas yang jelas serta dapat diukur (Puteri et al., 2025). Dengan kata lain, kualitas soal harus memenuhi kriteria tertentu agar hasil penilaianya benar-benar akurat dan dapat dipercaya. Supaya hasil penilaian akurat, setiap butir soal perlu dianalisis kualitasnya sebelum digunakan secara luas. Analisis butir soal berfungsi untuk menilai kemampuan setiap item soal dalam membedakan tingkat kemampuan peserta didik. Melalui analisis ini, dapat diketahui apakah suatu soal benar-benar efektif dalam memisahkan siswa yang menguasai materi dan yang belum (N. H. Siregar et al., 2024). Soal yang berkualitas ialah soal yang dapat memberikan informasi secara akurat kepada pembaca mengenai materi atau topik yang telah dipelajari

sebelumnya. Dengan kata lain, butir soal yang baik mampu mencerminkan isi pembelajaran secara tepat (Azmi & Fanani, 2024). Pendidikan kontemporer berfokus pada perkembangan kognitif, afektif, dan psikomotorik setiap siswa. Dalam konteks globalisasi dan revolusi industri 4.0, pendidikan tidak sekadar menjadi sarana penyampaian pengetahuan. Lebih dari itu, pendidikan juga berperan dalam membentuk karakter, menumbuhkan kreativitas, serta mengembangkan keterampilan yang relevan dengan tuntutan perkembangan zaman (Fitria et al., 2024).

Pada era digital saat ini, kebutuhan akan inovasi dalam dunia pendidikan semakin penting. Salah satu tantangan terbesar adalah mengembangkan instrumen evaluasi yang tidak hanya menilai penguasaan pengetahuan dasar, tetapi juga mampu mengukur keterampilan berpikir tingkat tinggi (Higher Order Thinking Skills/HOTS) (Sabela et al., 2025). Pertanyaan berbasis HOTS dirancang untuk mendorong siswa melakukan analisis dan evaluasi secara lebih mendalam. Melalui kemampuan tersebut, peserta didik diharapkan mampu menghadapi berbagai tantangan dalam kehidupan nyata dengan lebih baik, berdasarkan pemahaman yang mereka peroleh dari proses pembelajaran (Anandayu & Muslim, 2021). HOTS merupakan kemampuan berpikir tingkat tinggi yang diawali dengan penyusunan soal-soal pada ranah kognitif. Setelah itu, dilakukan pengembangan instrumen untuk memvalidasi kualitas soal, diikuti dengan proses pelaporan dan pemanfaatan hasilnya sebagai bagian dari penilaian kognitif (Markhamah, 2021). Penilaian menjadi bagian penting dari hasil belajar yang diharapkan baik oleh siswa maupun guru, karena melalui proses tersebut kemampuan berpikir tingkat tinggi (HOTS) dapat ditingkatkan (Syafiriya & Hakim, 2024). Indikator HOTS mencakup tingkat analisis (C4), evaluasi (C5), dan penciptaan atau kreasi (C6). Kemampuan berpikir kritis pada ranah ini diharapkan dapat membekali siswa agar mahir dalam menyelesaikan berbagai masalah secara efektif (Purwati et al., 2021).

Analisis soal merupakan langkah penting untuk menilai kualitas setiap butir soal dari segi aspek kuantitatif maupun kualitatif. Tujuannya untuk mendapatkan informasi mengenai karakteristik masing-masing soal, baik melalui kajian teoretis maupun analisis empiris. Software ANATES berperan sebagai alat yang dapat dipercaya untuk meningkatkan kualitas butir soal. Program ini melakukan analisis secara terstruktur dan menyeluruh sehingga mutu soal dapat dioptimalkan dengan efektif (Rachmadani et al., 2024). Software ini menyediakan berbagai indikator pengukuran, seperti skor data dibobot, reliabilitas tes, kelompok unggul dan asor, daya pembeda, tingkat kesukaran, korelasi skor butir dengan skor total, kualitas pengecoh, serta rekap analisis butir soal. Seluruh data tersebut dapat diperoleh dengan memasukkan hasil

penelitian yang dilakukan. Skor data dibobot digunakan untuk melihat distribusi nilai berdasarkan tingkat kesukaran. Reliabilitas tes berfungsi untuk menilai sejauh mana hasil pengukuran konsisten. Selain itu, konsep kelompok unggul dan inferior digunakan untuk membedakan kemampuan peserta didik yang tinggi dan yang rendah, sehingga performa mereka dapat dianalisis secara lebih tepat (Masruri et al., 2024). Daya pembeda suatu soal menunjukkan seberapa baik soal tersebut mampu membedakan peserta didik berdasarkan kemampuan mereka. Sementara itu, analisis tingkat kesukaran bertujuan untuk menilai seberapa sulit soal tersebut, sehingga soal dapat diklasifikasikan ke dalam kategori mudah, sedang, atau sulit (Nuranisa et al., 2025). Selain itu, korelasi skor butir dengan skor total sebagai indikator validitas empiris serta kualitas pengecoh digunakan untuk memastikan efektivitas setiap alternatif jawaban. Keseluruhan hasil olah data kemudian disajikan dalam bentuk rekap analisis butir soal yang dapat menjadi dasar bagi pendidik untuk memperbaiki atau mengganti butir soal yang kurang berkualitas. Melalui analisis ini, diharapkan butir-butir soal evaluasi pembelajaran memiliki validitas, reliabilitas, dan daya pembeda yang baik sehingga dapat mengukur kemampuan peserta didik secara tepat dan adil.

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode deskriptif. Tujuannya untuk menggambarkan dan mendeskripsikan secara sistematis kualitas soal HOTS yang diberikan kepada siswa SMK jurusan Akuntansi melalui Google Form, dengan landasan teori berupa pendekatan teori tes klasik yang menekankan pentingnya analisis terhadap karakteristik butir soal berdasarkan hasil empiris dan respon peserta tes. Sampel penelitian terdiri dari 36 siswa, sedangkan objek penelitian adalah 63 soal HOTS yang disusun oleh mahasiswa Prodi Pendidikan Akuntansi Universitas Negeri Surabaya. Data yang digunakan meliputi butir-butir soal HOTS dan jawaban siswa. Pengumpulan data dilakukan melalui Google Form.

Analisis dibuat secara kuantitatif menggunakan pendekatan klasik yang menilai aspek-aspek seperti validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan kualitas pengecoh. Seluruh data dianalisis dengan bantuan perangkat lunak Anates untuk mempermudah perhitungan dan memperoleh hasil yang akurat. Analisis dilakukan secara bertahap untuk menghasilkan penilaian kualitas soal yang lebih komprehensif dan mendalam. Proses ini memperhatikan beberapa aspek penting, yaitu validitas, daya pembeda, tingkat kesukaran, reliabilitas, serta kualitas pengecoh yang terdapat pada setiap soal (Saputra et al., 2025).

HASIL

Penelitian ini dilakukan untuk mengetahui apakah sebuah alat penilaian dapat dikatakan layak atau tidak untuk digunakan sebagai alat penilaian. Hal tersebut dapat diketahui dengan cara menganalisis butir-butir soal.

Skor Data Dibobot

SKOR DATA DIBOBOT											
No	Urt	No	Subyek	Kode/Nama	Benar	Salah	Kosong	Skr	Asli	Skr	Bobot
1	1	1	Meissy...	43	20	0	43	43	43		
2	2	2	Tiara ...	55	8	0	55	55	55		
3	3	3	Rafael...	39	24	0	39	39	39		
4	4	4	Dyah A...	27	36	0	27	27	27		
5	5	5	Eza Se...	59	4	0	59	59	59		
6	6	6	Zahrot...	26	37	0	26	26	26		
7	7	7	Cahya ...	55	8	0	55	55	55		
8	8	8	Gita C...	40	23	0	40	40	40		
9	9	9	Lijan ...	61	2	0	61	61	61		
10	10	10	Aura M...	49	14	0	49	49	49		
11	11	11	Muhamm...	56	7	0	56	56	56		
12	12	12	Muhamm...	62	1	0	62	62	62		
13	13	13	Muklis...	62	1	0	62	62	62		
14	14	14	Novael...	16	47	0	16	16	16		
15	15	15	Intani...	57	6	0	57	57	57		
16	16	16	Gilang...	57	6	0	57	57	57		
17	17	17	Gitali...	60	3	0	60	60	60		
18	18	18	Kayla ...	60	3	0	60	60	60		
19	19	19	Astri ...	60	3	0	60	60	60		
20	20	20	Septil...	52	11	0	52	52	52		
21	21	21	Prisci...	53	10	0	53	53	53		
22	22	22	Nurul ...	61	2	0	61	61	61		
23	23	23	Seira ...	49	14	0	49	49	49		
24	24	24	Retno ...	59	4	0	59	59	59		
25	25	25	Feby A...	62	1	0	62	62	62		
26	26	26	Intan ...	62	1	0	62	62	62		
27	27	27	Arum K...	22	41	0	22	22	22		
28	28	28	Putri ...	58	5	0	58	58	58		
29	29	29	Antare...	28	35	0	28	28	28		
30	30	30	Berlia...	57	6	0	57	57	57		
31	31	31	Kerina...	39	24	0	39	39	39		
32	32	32	Jagat ...	57	6	0	57	57	57		
33	33	33	Suci H...	56	7	0	56	56	56		
34	34	34	Asya M...	60	3	0	60	60	60		
35	35	35	Agnita...	59	4	0	59	59	59		
36	36	36	Maris ...	38	25	0	38	38	38		

Gambar 1. Skor data dibobot

Sumber: Data diolah (2025)

Dari gambar di atas diperoleh informasi bahwa bobot dari skor untuk soal dengan jawaban benar adalah 1 dan bobot skor untuk soal jawaban salah adalah 0. Dari gambar di atas

juga dapat diperoleh berapa perolehan skor yang diperoleh oleh masing-masing subjek.

Reliabilitas Tes

Hasil uji reliabilitas terhadap butir soal pilihan ganda dilakukan menggunakan software Anates untuk mengetahui tingkat konsistensi instrumen dalam mengukur kemampuan siswa. Berdasarkan hasil perhitungan, diperoleh rata-rata skor sebesar 50,44 dengan simpang baku 13,00. Nilai korelasi XY sebesar 0,94 menunjukkan adanya hubungan yang sangat kuat antara skor setiap butir dengan skor total. Dari hasil tersebut, diperoleh koefisien reliabilitas sebesar 0,97, yang berarti tes memiliki tingkat reliabilitas sangat tinggi.

Tabel 6. Hasil Reliabilitas Soal

Rata-Rata	50,44
Simpang Baku	13,00
Korelasi XY	0,94
Reliabilitas Tes	0,97

Sumber: Data diolah (2025)

Nilai ini mengindikasikan bahwa instrumen yang digunakan sudah cukup stabil dan konsisten dalam mengukur kemampuan siswa. Dengan kata lain, hasil yang diperoleh dari tes ini dapat dipercaya dan mencerminkan kemampuan peserta dengan baik. Dilihat pada tabel 6 menunjukkan koefisien reliabilitas soal uji coba adalah 0.97. menandakan bahwa pertanyaan-pertanyaan tersebut memiliki tingkat konsisten yang “sangat tinggi” dalam mengukur kemampuan siswa.

Kelompok Unggul dan Asor

Menunjukkan kelompok peserta yang menjadi bagian kelompok unggul di Kelompok atas dan bagian kelompok asor ada di Kelompok bawah.

Tabel 7. Kelompok Unggul

No. Urut	No. Subjek	Kode/Nama Subjek	Skor
1	12	Muhammad Maki	62
2	13	Mukhlison Abid	62
3	25	Feby Ayu Puspita	62
4	26	Intan Nurazizah	62
5	9	Jihan Rahma	61

No. Urut	No. Subjek	Kode/Nama Subjek	Skor
6	22	Nurul Khumairoh	60
7	17	Gitalis Dwi	60
8	18	Kayla aulia	60
9	19	Astri Ayu	60
10	34	Asya Mutiara	60

Sumber: Data diolah (2025)

Tabel 8. Kelompok Asor

No. Urut	No. Subjek	Kode/Nama Subjek	Skor
1	1	Meissy Dwi	43
2	8	Gita Candra	40
3	3	Rafael Adivia	39
4	31	Kerina Utami	39
5	36	Maris ayu	38
6	29	Antaresti dwi	28
7	4	Dyah Ayu Safitri	27
8	6	Zahrotus Shita	26
9	27	Arum Kusuma	22
10	14	Novaello	16

Sumber: Data diolah (2025)

Berdasarkan hasil pengolahan data menggunakan *software* Anates, analisis terhadap kelompok unggul dan kelompok asor menunjukkan adanya perbedaan tingkat kemampuan antar peserta. Dari total 36 peserta uji coba soal, sebanyak 27,77% atau sepuluh orang siswa termasuk ke dalam kelompok unggul (kelompok atas) karena memperoleh jumlah jawaban benar paling tinggi. Adapun nama-nama yang termasuk dalam kelompok ini yaitu Muhammad Maki, Mukhlison Abid, Feby Ayu Puspita, Intan Nurazizah, Jihan Rahma, Nurul Khumairoh, Gitalis Dwi, Kayla Aulia, Astri Ayu, dan Asya Mutiara.

Sementara itu, 27,77% lainnya dari jumlah peserta yang sama dikategorikan sebagai kelompok asor (kelompok bawah) karena memiliki jumlah jawaban benar paling rendah. Peserta yang masuk kelompok ini antara lain Meissy Dwi, Gita Candra, Rafael Adivia, Kerina Utami, Maris Ayu, Antaresti Dwi, Dyah Ayu Safitri, Zahrotus, Arum Kusuma, dan Novaello. Hasil pengelompokan ini menggambarkan distribusi kemampuan siswa dalam menjawab soal

uji coba, yang kemudian menjadi dasar untuk menganalisis daya pembeda dan validitas butir soal secara lebih mendalam.

Daya Pembeda

Software Anates digunakan sebagai alat bantu untuk menganalisis kualitas setiap butir soal melalui perhitungan daya pembeda. Hasil analisis mengelompokkan nilai daya pembeda menjadi lima kategori, yaitu sangat baik, baik, cukup baik, kurang baik, dan negatif (Ikhsaniyah et al., 2024). Nilai daya pembeda dapat bernilai positif maupun negatif, yang berfungsi untuk menunjukkan sejauh mana suatu soal mampu membedakan siswa yang benar-benar memahami materi dari mereka yang belum menguasai materi tersebut. Dengan demikian, semakin tinggi nilai positif daya pembeda, semakin efektif butir soal dalam mengidentifikasi perbedaan tingkat kemampuan siswa. Berdasarkan data yang kami dapatkan dan analisis yang dilakukan dengan menggunakan aplikasi anates, hasilnya adalah sebagai berikut.

Tabel 9. Hasil Daya Pembeda

Kategori Soal	Jumlah	Presentase (%)	Nomor Soal
Sangat Baik	15	23,81	5, 12, 33, 39, 42, 43, 46, 49, 51, 55, 56, 57, 59, 61, 63
Baik	28	44,45	3, 8, 10, 13, 16, 17, 19, 21, 23, 24, 25, 29, 30, 31, 32, 36, 37, 38, 41, 44, 45, 47, 48, 50, 52, 54, 60, 62
Cukup Baik	17	26,98	1, 2, 4, 6, 7, 9, 14, 15, 22, 26, 27, 28, 34, 35, 40, 53, 58
Tidak Baik	2	3,17	11, 18
Negatif	1	1,59	20
Jumlah	63	100	

Sumber: Data diolah (2025)

Berdasarkan hasil analisis daya pembeda butir soal menggunakan software Anates, diperoleh bahwa dari keseluruhan soal yang dianalisis, terdapat 15 butir soal yang tergolong dalam kategori sangat baik, yaitu pada nomor 5, 12, 33, 39, 42, 43, 46, 49, 51, 55, 56, 57, 59, 61, dan 63. Butir-butir soal tersebut memiliki nilai daya pembeda berada pada rentang 0,70 hingga 1,00, dengan persentase sebesar 23,81%, yang menandakan bahwa soal-soal tersebut mampu secara optimal membedakan peserta didik berkemampuan tinggi dan rendah.

Selanjutnya, ditemukan 28 butir soal yang termasuk dalam kategori baik, yakni pada nomor 3, 8, 10, 13, 16, 17, 19, 21, 23, 24, 25, 29, 30, 31, 32, 36, 37, 38, 41, 44, 45, 47, 48, 50, 52, 54, 60, dan 62. Nilai daya pembeda untuk kelompok ini berada pada kisaran 0,40 hingga 0,70, dengan persentase 44,45%, yang menunjukkan bahwa sebagian besar butir soal telah berfungsi dengan baik dalam mengukur perbedaan tingkat pemahaman siswa.

Kemudian, terdapat 17 butir soal yang dikategorikan cukup baik, mencakup nomor 1, 2, 4, 6, 7, 9, 14, 15, 22, 26, 27, 28, 34, 35, 40, 53, dan 58. Soal-soal tersebut memiliki nilai daya pembeda pada rentang 0,20 hingga 0,40, dengan persentase 26,98%. Hal ini mengindikasikan bahwa soal-soal tersebut masih mampu membedakan peserta didik, meskipun dengan tingkat efektivitas yang sedang.

Selain itu, terdapat 2 butir soal yang masuk dalam kategori tidak baik, yaitu nomor 11 dan 18, dengan nilai daya pembeda berada pada batasan 0,00 hingga 0,20 serta persentase 3,17%. Butir soal dalam kategori ini dianggap kurang efektif dalam membedakan kemampuan siswa.

Adapun 1 butir soal, yaitu nomor 20, memiliki nilai daya pembeda negatif dengan persentase 1,59%, yang menunjukkan bahwa siswa dengan kemampuan rendah justru cenderung menjawab benar, sementara siswa berkemampuan tinggi menjawab salah. Kondisi ini mengindikasikan adanya masalah pada konstruksi soal atau kunci jawaban, sehingga butir soal tersebut perlu direvisi atau diganti.

Tingkat Kesukaran

Hasil analisis dengan menggunakan software Anates menunjukkan bahwa butir-butir soal memiliki tingkat kesukaran yang beragam. Soal-soal tersebut dapat diklasifikasikan ke dalam tiga kategori utama, yaitu sangat mudah, sedang, dan sangat sulit (Syafiriya & Hakim, 2024). Pengelompokan ini dilakukan untuk melihat sejauh mana butir soal tersebut dapat membedakan kemampuan siswa. Hasil analisis tingkat kesukaran dapat dilihat sebagai berikut.

Tabel 10. Hasil Tingkat Kesukaran

Kategori Soal	Jumlah	Persentase (%)	Nomor Soal
Sangat Mudah	23	36	1, 2, 3, 4, 6, 8, 9, 10, 11, 14, 16, 17, 18, 21, 22, 23, 28, 34, 40, 45, 47, 48, 50
Mudah	34	54	7, 12, 13, 15, 19, 24, 25, 26, 27,

Kategori Soal	Jumlah	Presentase (%)	Nomor Soal
			29, 30, 31, 32, 35, 36, 37, 38, 39,
			41, 42, 44, 46, 51, 52, 53, 54, 55,
			56, 59, 60, 61, 62, 63
Sedang	5	7	5, 33, 43, 49, 57
Sangat Sukar	1	3	20
Jumlah	63	100	

Sumber: Data diolah (2025)

Berdasarkan data diatas, dapat disimpulkan bahwa sebagian besar butir soal tergolong mudah dan sangat mudah, sehingga keseluruhan tes masih terlalu sederhana bagi peserta. soal-soal tersebut ternyata perlu diperbaiki agar memiliki tingkat kesulitan yang lebih seimbang, sehingga soal tes akan lebih efektif dalam mengukur kemampuan siswa secara menyeluruh. Soal yang terlalu mudah tidak mampu menstimulasi kemampuan motivasi belajar siswa, sedangkan soal yang terlalu sukar justru menurunkan motivasi belajar siswa.

Korelasi Skor Butir dengan Skor Total

Yang dimaksud dengan korelasi skor butir soal dengan skor total adalah suatu indikator penting dalam melakukan analisis tes. Korelasi ini menunjukkan sejauh mana jawaban pada suatu butir soal berhubungan dengan pencapaian keseluruhan peserta. Dengan kata lain, hal ini mencerminkan seberapa erat keterkaitan antara jawaban individu dan kinerja umum siswa (Syafiriya & Hakim, 2024).

Tabel 11. Hasil Korelasi Skor Butir dengan Skor Total

Kategori Soal	Jumlah	Presentase (%)	Nomor Soal
Validitas Sangat Tinggi	2	3,17	61 dan 62
Validitas Tinggi	23	36,5	3, 5, 12, 13, 16, 21, 23, 31, 39, 41, 42, 43, 46, 47, 48, 49, 51, 52, 54, 56, 57, 59, 63
Validitas Cukup	21	33,33	1, 4, 8, 9, 10, 14, 17, 19, 22, 24, 28, 29, 30, 32, 34, 36, 38, 45, 50, 55, 60
Validitas Rendah	13	20,63	2, 6, 7, 11, 18, 25, 27, 33, 35, 37, 44, 53, 58

Kategori Soal	Jumlah	Presentase (%)	Nomor Soal
Validitas Sangat Rendah	4	6,35	15, 20, 26, 40
Jumlah	63	100	

Sumber: Data diolah (2025)

Berdasarkan hasil analisis data yang telah diperoleh, dapat disimpulkan bahwa dari keseluruhan butir soal yang dianalisis, terdapat 46 soal yang dinyatakan valid, sedangkan 17 soal lainnya memerlukan revisi atau perbaikan. Hal ini menunjukkan bahwa apabila suatu butir soal tidak memenuhi kriteria validitas, maka hasil penilaian yang dihasilkan tidak akan menggambarkan kemampuan peserta didik secara akurat. Dengan kata lain, soal yang tidak valid tidak mampu mengukur aspek kompetensi yang seharusnya diukur.

Dalam evaluasi pendidikan, validitas dan reliabilitas merupakan dua aspek yang saling terkait. Validitas menilai sejauh mana sebuah tes mampu mengukur hal yang memang seharusnya diukur, sementara reliabilitas menunjukkan seberapa konsisten hasil pengukuran tersebut jika dilakukan berulang kali (Halim & Wahyuningsih, 2025). Sebuah instrumen tes yang reliabel umumnya juga memiliki kecenderungan menghasilkan data yang valid. Oleh karena itu, suatu butir soal dapat dikategorikan baik apabila telah memenuhi kedua aspek tersebut.

Butir soal yang terbukti valid dan reliabel layak untuk disimpan dan digunakan kembali pada kegiatan evaluasi berikutnya, karena telah teruji secara kualitas dan mampu memberikan hasil penilaian yang objektif. Sebaliknya, butir soal yang tidak valid sebaiknya diperbaiki atau diganti agar sesuai dengan standar penyusunan soal yang baik dan memenuhi kriteria instrumen penilaian yang efektif.

Kualitas Pengecoh

Hasil analisis Kualitas pengecoh terbagi menjadi 5, yaitu: Sangat Buruk, Buruk, Kurang Baik, Baik, Sangat Baik yang bertujuan mengetahui kemampuan siswa. Hasil analisis kualitas pengecoh butir soal dapat dilihat sebagai berikut.

Tabel 12. Hasil Kualitas Pengecoh

Kategori Soal	Jumlah	Presentase (%)	Nomor Soal
Sangat Buruk	6	9,5	50, 51, 52, 55, 56, 57
Buruk	9	14,3	49 – 57
Kurang Baik	10	15,9	2, 6, 11, 16, 18, 21, 30, 34, 45, 47

Kategori Soal	Jumlah	Presentase (%)	Nomor Soal
Baik	20	31,7	4, 9, 12, 14, 17, 19, 22, 24, 26, 27, 29, 32, 33, 35, 36, 38, 39, 41, 44, 46, 48
Sangat Baik	18	28,6	1, 3, 5, 7, 8, 10, 13, 15, 20, 23, 25, 28, 31, 37, 40, 42, 58
Jumlah	63	100	

Sumber: Data diolah (2025)

Berdasarkan hasil analisis data yang telah diperoleh melalui *software* Anates, dapat disimpulkan bahwa kualitas pengecoh (distraktor) pada butir soal menunjukkan variasi tingkat efektivitas. Dari keseluruhan butir soal yang dianalisis, terdapat 6 soal dengan nomor 50, 51, 52, 55, 56, dan 57 yang termasuk dalam kategori sangat buruk dengan persentase sebesar 9,5%. Hal ini menunjukkan bahwa distraktor pada butir tersebut tidak berfungsi optimal dalam mengalihkan perhatian peserta didik yang belum menguasai materi.

Selanjutnya, terdapat 9 butir soal (nomor 49 hingga 57) yang tergolong dalam kategori buruk dengan persentase 14,3%. Soal-soal pada kategori ini masih perlu dilakukan perbaikan agar pengecohnya mampu berfungsi sebagaimana mestinya. Sementara itu, 10 butir soal (nomor 2, 6, 11, 16, 18, 21, 30, 34, 45, dan 47) memiliki efektivitas pengecoh kurang baik dengan persentase 15,9%, yang berarti pengecoh pada soal tersebut hanya berfungsi sebagian dan belum sepenuhnya efektif.

Adapun 20 butir soal (nomor 4, 9, 12, 14, 17, 19, 22, 24, 26, 27, 29, 32, 33, 35, 36, 38, 39, 41, 44, 46, dan 48) berada dalam kategori baik dengan persentase 31,7%, menandakan bahwa sebagian besar pengecoh telah berfungsi sesuai dengan tujuannya, yaitu menarik perhatian siswa yang kurang memahami materi. Selain itu, terdapat 18 butir soal (nomor 1, 3, 5, 7, 8, 10, 13, 15, 20, 23, 25, 28, 31, 37, 40, 42, dan 58) yang termasuk dalam kategori sangat baik dengan persentase 28,6%. Hal ini menunjukkan bahwa pengecoh pada soal-soal tersebut mampu bekerja dengan sangat efektif dalam membedakan peserta didik yang memahami materi dengan yang belum menguasainya.

Rekap Hasil Analisis butir Soal

Tabel 13. Hasil Rekap Analisis Butir Soal

Kategori Soal	Jumlah	Persentase (%)	Nomor Soal
Soal Diperbaiki	8	12,7	2, 6, 7, 11, 14, 18, 20, 35
Soal Digunakan dan Diperbaiki	23	36,5	1, 3, 4, 8, 9, 10, 15, 16, 17, 19, 22, 23, 25, 26, 27, 28, 34, 38, 40, 44, 50, 53, 58
Soal Digunakan	32	50,8	5, 12, 13, 21, 24, 29, 30, 31, 32, 33, 36, 37, 39, 41, 42, 43, 45, 46, 47, 48, 49, 51, 52, 54, 55, 56, 57, 59, 60, 61, 62, 63
Jumlah	63	100	

Sumber: Data diolah (2025)

Berdasarkan hasil rekapitulasi analisis butir soal yang ditampilkan pada Tabel 11, dapat diketahui bahwa terdapat 8 butir soal dengan nomor 2, 6, 7, 11, 14, 18, 20, dan 35 yang perlu diperbaiki secara menyeluruh. Hal ini disebabkan oleh rendahnya daya pembeda, tingkat kesukaran yang kurang ideal, serta kualitas pengecoh yang tidak berfungsi secara efektif. Persentase soal yang termasuk kategori ini mencapai 12,7% dari keseluruhan butir yang dianalisis.

Selanjutnya, terdapat 23 butir soal (nomor 1, 3, 4, 8, 9, 10, 15, 16, 17, 19, 22, 23, 25, 26, 27, 28, 34, 38, 40, 44, 50, 53, dan 58) yang masih dapat digunakan, namun memerlukan revisi sebagian. Perbaikan ini perlu dilakukan karena masih terdapat beberapa aspek yang belum memenuhi kriteria ideal, baik dari segi daya pembeda, tingkat kesukaran, maupun efektivitas pengecoh. Dengan adanya penyempurnaan pada bagian-bagian tersebut, soal dapat menjadi instrumen evaluasi yang lebih representatif dalam mengukur kemampuan peserta didik.

Sementara itu, terdapat 32 butir soal (nomor 5, 12, 13, 21, 24, 29, 30, 31, 32, 33, 36, 37, 39, 41, 42, 43, 45, 46, 47, 48, 49, 51, 52, 54, 55, 56, 57, 59, 60, 61, 62, dan 63) yang dinyatakan layak digunakan tanpa perbaikan tambahan. Butir soal pada kategori ini telah memenuhi seluruh kriteria kualitas, baik dari aspek validitas, reliabilitas, tingkat kesukaran, daya pembeda, maupun kualitas pengecoh, sehingga dapat langsung diimplementasikan dalam

proses evaluasi pembelajaran.

Hasil rekapitulasi ini menunjukkan bahwa sebagian besar soal telah memenuhi kriteria sebagai alat evaluasi yang baik, namun masih terdapat beberapa soal yang perlu diperbaiki agar hasil pengukuran dapat menggambarkan kemampuan peserta didik secara objektif dan akurat. Melalui proses revisi berbasis analisis Anates, pendidik dapat meningkatkan kualitas instrumen penilaian, sehingga proses evaluasi menjadi lebih valid, reliabel, dan adil bagi seluruh siswa.

SIMPULAN

Berdasarkan hasil analisis butir soal menggunakan software Anates versi 4.0, dapat disimpulkan bahwa instrumen tes yang terdiri dari 63 butir soal HOTS pada mata pelajaran Prinsip-Prinsip dan Konsep Akuntansi serta Perbankan Dasar memiliki kualitas yang baik dan layak digunakan sebagai alat evaluasi pembelajaran. Nilai reliabilitas sebesar 0,97 menunjukkan bahwa tes ini memiliki konsistensi yang sangat tinggi. Dari hasil analisis daya pembeda, sebanyak 68,26% butir soal tergolong dalam kategori baik hingga sangat baik, yang berarti sebagian besar soal mampu membedakan peserta didik berdasarkan tingkat kemampuannya.

Dilihat dari tingkat kesukaran, mayoritas butir soal termasuk kategori mudah dan sangat mudah, sehingga perlu dilakukan penyesuaian agar tingkat kesulitan lebih seimbang. Semenara itu, berdasarkan uji validitas, terdapat 46 soal yang valid dan 17 soal yang perlu diperbaiki, menunjukkan bahwa sebagian besar instrumen telah mengukur kompetensi yang sesuai. Analisis kualitas pengecoh juga memperlihatkan bahwa sebagian besar distraktor berfungsi dengan baik, meskipun beberapa butir masih memerlukan revisi agar lebih efektif.

Secara keseluruhan, hasil analisis menunjukkan bahwa 32 butir soal layak digunakan tanpa revisi, 23 soal dapat digunakan setelah perbaikan, dan 8 soal perlu diperbaiki secara menyeluruh. Oleh karena itu, penelitian ini menegaskan pentingnya analisis butir soal menggunakan software Anates dalam memastikan bahwa instrumen evaluasi tidak hanya valid dan reliabel, tetapi juga mampu menilai kemampuan berpikir tingkat tinggi peserta didik secara objektif dan komprehensif.

DAFTAR PUSTAKA

Alpusari, M. (2014). *Analisis Butir Soal Konsep Dasar IPA 1 Melalui Penggunaan Program Komputer Anates Versi 4.0 For Windows*. 3, 106–115.

Anandayu, R., & Muslim, A. B. (2021). Pengembangan Kemampuan Berpikir Tingkat Tinggi

- dalam Pembelajaran Daring Bahasa Inggris Promoting Higher-Order Thinking Skills in English Online Learning. *Jurnal Penelitian Pendidikan*, 21(3), 110–121.
- Arif, M. (2014). Penerapan aplikasi anates bentuk soal pilihan ganda. *Jurnal Ilmiah Edutic*, 1(1), 1–9.
- Augustia, A. D., Agustia, C. N., Azzahra, D., Hakim, L., & Pratiwi, V. (2025). Analisis Validitas dan Reliabilitas Soal Pilihan Ganda dengan Menggunakan Software Anates pada Mata Pelajaran Perpajakan. *Jurnal Mahasiswa Ekonomi & Bisnis*, 5(1), 250–265.
- Azmi, Q. L., & Fanani, U. Z. (2024). *Analisis Tingkat Kesulitan Level Kognitif Pada Soal Penilaian Sumatif Bahasa Mandarin Kelas XI SMA SEMEN GRESIK Tahun Ajaran 2023/2024. XX(Xx)*.
- Fatimah, L. U., & Alfath, K. (2019). Analisis Kesukaran Soal, Daya Pembeda dan fungsi Distaktor. *Al-Manar*, 8, 37–64.
- Firandhika, A., Apriliyanti, K. D., Saputri, A. K. A., Hakim, L., & Pratiwi, V. (2024). Analisis Butir Soal HOTS Pilihan Ganda Pada Elemen Komputer Akuntansi Di SMKN 10 Surabaya Menggunakan Aplikasi Anates. *PESHUM : Jurnal Pendidikan, Sosial Dan Humaniora*, 4(1), 727–738.
- Fitria, R. A. S., Sari, C. A., Pratiwi, V., & Hakim, L. (2024). *Analisi Butir Soal Pilihan Ganda Elemen Profil Pekerjaan / Profesi dan Peluang Usaha di Bidang Akuntansi dan Keuangan Lembaga Menggunakan Aplikasi Anates*. 8, 50966–50973.
- Fuadiy, M. R. (2021). Evaluasi Pembelajaran Sebuah Studi Literatur. *Jurnal Pendidikan Islam*, 3.
- Halim, F. Y., & Wahyuningsih, R. (2025). Pengaruh Bar Model Terhadap Penggunaan Konsep , Berpikir Logis , dan Keterampilan Memecahkan Masalah Matematika Siswa Kelas III SD Sekolah XYZ di Jakarta Utara. *JIIP (Jurnal Ilmu Ilmiah Pendidikan)*, 8, 5067–5074.
- Hidayat, R., & Sesmiarni, Z. (2025). *Evaluasi Pengembangan Kurikulum PAI di Sekolah dan Madrasah*. 9(20), 18430–18435.
- Ikhsaniyah, S. N., Kurnia, A. D., Zuroida, M., Pratiwi, V., & Hakim, L. (2024). Analisis Butir Soal Perpajakan PPH Pasal 21 Menggunakan Software Anates pada Pendekatan Teori Tes Klasik. *PeKA: Jurnal Pendidikan Ekonomi Akuntansi*, 12(2).
- Khumaira, R., Ramadan, R. R., Khairunnisa, S., Jannah, M., & Marhadi, H. (2024). Analisis Daya Beda pada Soal Tes Mata Pelajaran Matematika di Kelas IV SDN 136 Pekanbaru. *Jurnal Penelitian Ilmiah Multidisiplin*, 8(5), 533–540.
- Listyowati, I. (2025). Analisis Butir Soal Multiple Choice Mata Pelajaran Bahasa Indonesia Siswa Kelas 4 di SDN Bareng 5 Jombang. *Jurnal Review Pendidikan Dan Pengajaran*, 8, 3968–3976.
- Markhamah, N. (2021). Pengembangan Soal Berbasis HOTS (Higher Order Thinking Skills) pada Kurikulum 2013. *Nusantara: Jurnal Pendidikan Indonesia*, 1(2), 385–418.
- Masruri, F. B., Farila, F. N., Putri, M. A., & Pratiwi, V. (2024). *Artikel Optimasi Evaluasi*

Pembelajaran : Analisis Butir Soal Pilihan Ganda dengan Aplikasi Anates. 8, 49618–49628.

Nuranisa, W., Lailiyah, Z. N., Daud, N., Henery, C., Hakim, L., Arif, A., & Surabaya, U. N. (2025). *Analisis Butir Soal Elemen Prinsip-Prinsip dan E Kelas X AKL SMK Menggunakan Aplikasi Anates Versi 4.0.9.* 07(1), 291–309.

Nurhalimah, S., Hidayati, Y., Rosidi, I., & Hadi, W. P. (2022). Hubungan antara Validitas Item dengan Daya Pembeda dan Tingkat Kesukaran Soal Pilihan Ganda PAS. *Jurnal Natural Science Educational Research*, 4(3), 249–257.

Purwati, L. M., Arianty, R., Syakilah, D. M., Ridlo, S., & Susilaningsih, E. (2021). Analisis Soal Tes Pilihan Ganda Berbasis Higher Order Thinking Skill menggunakanAplikasi Anates Windows Versi 4.0.9 For Windows. *Jurnal Pendidikan Universitas Garut*, 15, 460–473.

Puteri, N. D., Arsyanti, K., Putricia, N. D., Hakim, L., & Pratiwi, V. (2025). Analisis Kelayakan Butir Soal Pada Handout Akuntansi Keuangan Dana Kas Kecil Berbasis E-Book Menggunakan Software Anates. *JIBEMA: Jurnal Ilmu Bisnis, Ekonomi, Manajemen, Dan Akuntansi*, 2(3), 205–214.

Rachmadani, F. C., Lestari, C. D., Hakim, L., & Pratiwi, V. (2024). Analisis Butir Soal Pilihan Ganda pada Elemen Akuntansi Keuangan Guna Mengoptimasi Evaluasi Menggunakan Anates V4.0. *Jurnal Nirta : Studi Inovasi*, 4, 1–18.

Ramadhan, W., Malahati, F., Romadhon, K., & Ramadhan, S. (2023). Analisis Butir Soal Tipe Multiple Choice Questions pada Penilaian Harian Sekolah Dasar. *Tarbiyah Wa Ta'lim: Jurnal Penelitian Pendidikan Dan Pembelajaran*, 10(2), 93–105.

Sabela, O. R., Krisdayanty, D., Taqiqiyah, A. Z., Hakim, L., & Pratiwi, V. (2025). Analisis Butir Soal HOTS Elemen Dokumen Berbasis Digital (FASE E) Menggunakan Program Anates. *Education Achievement: Journal of Science and Research*, 6(1), 251–262.

Saputra, I. A. G., Hakim, L., Pratiwi, V., Rohayati, S., Wahjudi, E., & Farid, M. M. (2025). Pelatihan Analisis Butir Soal Berbasis ANATES untuk Meningkatkan Kualitas Evaluasi Pembelajaran Guru MGMP Akuntansi. *Madaniya*, 6(3), 1288–1297.

Siregar, N. H., Remiswal, & Khadijah. (2024). Analisis Butir Soal Ujian Tengah Semester Pada Mata Pelajaran Pendidikan Agama Islam. *Jurnal Studi Kependidikan Dan Keislaman*, 13(2), 179–189.

Siregar, S. R. N., Luckytasari, A., Ariani, A., Hakim, L., & Pratiwi, V. (2024). Analisis Soal Hots Pada Isu Serta Perkembangan Terkini Akuntansi Keuangan Fase E SMK Menggunakan Aplikasi Anates. *PESHUM : Jurnal Pendidikan, Sosial Dan Humaniora*, 4(1), 1386–1395.

Subhaktiyasa, P. G. (2024). Evaluasi Validitas dan Reliabilitas Instrumen Penelitian Kuantitatif : Sebuah Studi Pustaka. *Journal of Education Research*, 5(4), 5599–5609.

Syafiriya, D. M., & Hakim, L. (2024). *Analisis Butir Soal HOTS untuk Menguji Pemahaman Siswa SMK Jurusan Akuntansi Menggunakan Software Anates*. 4(1).

