

ANALISIS KUALITAS BUTIR SOAL AKUNTANSI DAN PERBANKAN DASAR MENGUNAKAN APLIKASI ANATES

Dinda Putri Romiyanti¹, Luqman Hakim², Vivi Pratiwi³, Rahma Nur Indana⁴, Livia Dhea Intan Nuraini⁵, Yahya Ahmad Fatahillah⁶, Tya Rahma Sari⁷, Atika Mutiara Noer Kamari⁸, Aurel Feysa Resnandya⁹

^{1,2,3,4,5,6,7,8,9} Sarjana Pendidikan Akuntansi FEB Universitas Negeri Surabaya

¹24080304009@mhs.unesa.ac.id, ²luqmanhakim@unesa.ac.id,
³vivipratiwi@unesa.ac.id, ⁴24080304011@mhs.unesa.ac.id, ⁵24080304059@mhs.unesa.ac.id,
⁶24080304068@mhs.unesa.ac.id, ⁷24080304070@mhs.unesa.ac.id,
⁸24080304143@mhs.unesa.ac.id, ⁹24080304144@mhs.unesa.ac.id

ABSTRACT

This study aims to analyze the quality of test items developed by Accounting Education students at Universitas Negeri Surabaya for the topic of Principles and Concepts of Accounting and Basic Banking. The analysis was conducted using the Anates application to evaluate item validity, reliability, difficulty level, discrimination power, and distractor effectiveness. Twenty respondents, consisting of vocational students and undergraduate accounting education students, completed a set of 63 multiple-choice items. The results show that the test has a high reliability coefficient, indicating strong internal consistency. However, many items were categorized as very easy, had low discrimination power, or contained ineffective distractors, suggesting the need for revision. Only about half of the items met acceptable validity criteria. These findings emphasize the importance of empirical item analysis in producing accurate and fair evaluation instruments, and highlight the necessity of improving pre-service teachers' competence in test development and psychometric assessment.

Keywords: Anates, Item Analysis, Learning Evaluation.

ABSTRAK

Penelitian ini bertujuan menganalisis kualitas butir soal yang dikembangkan oleh mahasiswa Pendidikan Akuntansi Universitas Negeri Surabaya pada materi Prinsip-Prinsip dan Konsep Akuntansi serta Perbankan Dasar. Analisis dilakukan menggunakan aplikasi Anates untuk menilai validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Sebanyak 20 responden, yang terdiri atas siswa SMK dan mahasiswa pendidikan akuntansi, mengerjakan 63 soal pilihan ganda. Hasil penelitian menunjukkan bahwa instrumen memiliki koefisien reliabilitas tinggi, yang menandakan konsistensi internal yang baik. Namun, sebagian besar butir soal berada pada kategori sangat mudah, memiliki daya pembeda rendah, atau pengecoh yang tidak berfungsi, sehingga membutuhkan revisi. Hanya sekitar setengah dari keseluruhan butir yang memenuhi kriteria validitas yang dapat diterima. Temuan ini menegaskan pentingnya analisis empiris dalam menghasilkan

instrumen evaluasi yang akurat dan adil, serta perlunya peningkatan kompetensi calon guru dalam pengembangan tes dan asesmen psikometrik.

Kata Kunci: Anates, Analisis Butir Soal, Evaluasi Pembelajaran.

A. Pendahuluan

Evaluasi pembelajaran adalah proses yang dilakukan secara teratur untuk mengetahui sampai sejauh mana tujuan pendidikan telah tercapai. Evaluasi yang baik harus mampu menunjukkan hasil belajar siswa secara jelas dan tepat (Karengga et al., 2025). Ramadhani dkk. (2025) juga menekankan bahwa evaluasi penting dilakukan karena bisa membantu guru mengetahui kelebihan dan kekurangan siswa serta memancarkan cara mengajar yang digunakan. Dalam bidang pendidikan vokasi, khususnya pada jurusan akuntansi dan perbankan, alat evaluasi memiliki peran yang sangat penting. Selain mengukur pemahaman dasar, evaluasi juga digunakan untuk menilai kemampuan analitis, kesabaran, dan kemampuan berpikir yang dibutuhkan dalam dunia kerja. Oleh karena itu, kualitas soal dalam evaluasi menjadi hal yang sangat penting agar hasil penilaian menjadi obyektif dan tepat.

Analisis butir soal merupakan langkah penting agar soal dapat mengukur kemampuan siswa secara

tepat (Rahmawati & Rahman, 2025). Aplikasi Anates dikembangkan untuk membantu guru menganalisis kualitas soal berdasarkan beberapa indikator statistik, seperti validitas, reliabilitas, daya pembeda, dan tingkat kesukaran (Susilowati et al., 2025). Puteri dkk. (2025) menyatakan bahwa Anates mampu memberikan hasil analisis yang akurat untuk instrumen mata pelajaran akuntansi di SMK. Penelitian serupa juga dilakukan oleh Sudarto (2023), yang menunjukkan bahwa Anates membantu guru memancarkan karakteristik soal secara tujuan dan efisien.

Penelitian mengenai analisis butir soal dalam bidang akuntansi telah banyak dilakukan. Sejumlah penelitian menunjukkan bahwa berbagai instrumen evaluasi di sekolah kejuruan masih belum memenuhi standar validitas dan reliabilitas yang memadai. Beberapa soal akuntansi juga dinilai kurang baik dalam hal daya pembeda dan tingkat kesulitan (Pengabdian et al., 2025; Adimas et al., 2025). Menurut penelitian yang dilakukan oleh Magdalena dan rekan-rekannya pada

tahun 2021, kualitas suatu instrumen pengukuran ditentukan oleh karakteristik psikometrik seperti validitas, reliabilitas, kemampuan membedakan, tingkat kesulitan, dan efektivitas pengecoh. Validitas merujuk pada sejauh mana alat pengukur mampu menilai apa yang seharusnya diukur, sementara reliabilitas menunjukkan konsistensi hasil yang didapatkan ketika alat tersebut digunakan berkali-kali (Himawan & Nurgiyantoro, 2022). Kemampuan membedakan menilai seberapa efektif soal dalam memisahkan peserta dengan tingkat kemampuan tinggi dan rendah. Tingkat kesulitan mengindikasikan persentase responden yang mampu memberikan jawaban yang benar, sedangkan pengecoh yang efektif dapat mendapatkan jawaban dari peserta yang tidak memiliki pemahaman yang baik tentang materi.

Permasalahan utama yang muncul dalam penelitian-penelitian terdahulu adalah belum konsistennya penggunaan analisis empiris dalam menilai kelayakan instrumen. Hakim dkk. (2024) menemukan bahwa sejumlah soal akuntansi SMK belum memenuhi standar validitas dan reliabilitas. Temuan serupa dilaporkan

oleh Indah dan Andriani (2021), terutama pada aspek daya pembeda dan tingkat kesukaran. Farida dan Musyarofah (2021) juga menegaskan bahwa sebagian besar instrumen evaluasi masih disusun tanpa dukungan analisis statistik yang memadai. Selain itu, sebagian besar penelitian hanya berfokus pada instrumen yang dibuat oleh guru, sementara penelitian yang melibatkan mahasiswa calon guru sebagai penyusun sekaligus analisis soal masih terbatas.

Dalam konteks pendidikan calon pendidik, kemampuan untuk menyusun dan mengevaluasi instrumen tes merupakan bagian dari kompetensi pedagogik dan psikometrik yang wajib dikuasai. Pemahaman mengenai validitas konstruk, reliabilitas internal, tingkat kesukaran, serta kualitas pengecoh menjadi dasar untuk menghasilkan instrumen evaluasi yang layak digunakan dalam pembelajaran (Saputra et al., 2022). Melalui penelitian ini, mahasiswa mendapatkan kesempatan untuk menerapkan teori evaluasi pembelajaran ke dalam praktik empiris, sekaligus memperkuat kemampuan mereka dalam

melakukan analisis butir soal berbasis data. Sejumlah penelitian menunjukkan bahwa profesionalisme guru berkontribusi pada meningkatnya efektivitas proses belajar (Barus & Darma, 2025). Efektivitas ini hanya dapat tercapai jika evaluasi pembelajaran dilakukan dengan instrumen yang akurat dan sesuai standar. Oleh karena itu, uji validitas menjadi tahap krusial dalam pengembangan instrumen evaluasi, termasuk butir soal, untuk memastikan bahwa setiap item relevan, konsisten, dan benar-benar mengukur tujuan pembelajaran yang ingin dicapai (Tania et al., 2025).

Penelitian ini bertujuan untuk mengevaluasi mutu butir soal yang dibuat oleh mahasiswa Program Studi Pendidikan Akuntansi di Universitas Negeri Surabaya, dengan fokus pada elemen Prinsip-Prinsip dan Konsep Akuntansi serta Perbankan Dasar. Metode analisis dilakukan menggunakan perangkat lunak Anates yang menyoroti lima aspek penting yaitu validitas, reliabilitas, tingkat kesulitan, kemampuan membedakan, dan efektivitas pilihan pengecoh. Hasil dari studi ini diharapkan dapat memberikan perspektif empiris tentang kelayakan

instrumen, menyarankan perbaikan dalam penyusunan soal, serta meningkatkan kapasitas mahasiswa dalam pengembangan evaluasi pembelajaran yang merujuk pada standar psikometrik.

B. Metode Penelitian

Penelitian ini menerapkan desain kuantitatif deskriptif dengan tujuan untuk menampilkan mutu butir soal berdasarkan analisis statistik (Jayusman & Shavab, 2020). Data yang dikumpulkan dalam bentuk kuantitatif deskriptif dimanfaatkan untuk menggambarkan keadaan nyata dan menjawab pertanyaan-pertanyaan yang berkaitan dengan kondisi atau keadaan subjek dalam penelitian tersebut (Isnawati et al., 2020). Target populasi dari penelitian ini adalah siswa yang mempelajari aspek Prinsip-Prinsip dan Konsep Akuntansi serta Perbankan Dasar. Sampelnya terdiri dari 20 responden, yang mencakup siswa SMK jurusan Akuntansi dan Perbankan di Sidoarjo serta mahasiswa di Program Studi Pendidikan Akuntansi Universitas Negeri Surabaya. Pemilihan sampel dilakukan dengan metode purposif karena semua responden memiliki latar belakang yang relevan dengan materi yang diuji.

Data dikumpulkan melalui penyelesaian 63 soal pilihan ganda yang disusun oleh mahasiswa. Instrumen ini dirancang berdasarkan kompetensi elemen materi dan kemudian diuji kepada para responden. Data dari jawaban tersebut digunakan sebagai dasar untuk menganalisis karakteristik dari setiap soal. Teknik analisis data dilakukan menggunakan aplikasi Anates untuk menghitung validitas materi, reliabilitas tes, tingkat kesulitan, daya pembeda, serta efektivitas pengecoh. Hasil dari analisis ditampilkan dalam bentuk tabel, grafik, gambar, dan interpretasi sesuai dengan kriteria evaluasi instrumen.

C. Hasil Penelitian dan Pembahasan

Bagian ini menyajikan hasil analisis kualitas butir soal yang diuji pada 20 subyek dengan total 63 butir soal menggunakan aplikasi Anates. Hasil analisis meliputi korelasi butir (validitas tes), reliabilitas tes, tingkat kesukaran, daya pembeda, dan kualitas pengecoh.

Validitas tes (Korelasi butir)

Tes evaluasi yang baik memiliki ciri dan sifat tertentu yang harus terpenuhi, yaitu tes tersebut harus

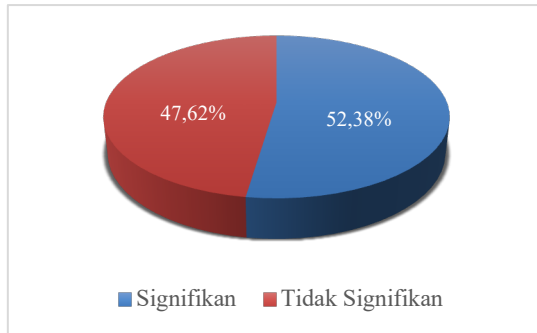
valid atau memiliki tingkat validitas yang cukup baik (Solichin, 2017). Korelasi skor butir dengan skor total digunakan untuk menilai validitas setiap butir soal, yaitu sejauh mana butir soal tersebut mampu mengukur hal yang sama dengan yang diukur oleh keseluruhan tes. Hasil analisis menunjukkan bahwa dari total 63 butir yang di mana 31 butir sangat signifikan dan 2 butir soal (butir 26 dan 58) signifikan, sehingga sebanyak 33 butir (52,38%) dikategorikan Valid, dengan perhitungan persentase sebagai berikut:

$$\frac{\text{Jumlah Butir Valid}}{\text{Total Butir Valid}} \times 100\% = \frac{33}{63} \times 100\% = 52,38\%$$

Sebaliknya, ditemukan sebanyak 14 butir tidak signifikan dan 16 butir menghasilkan nilai korelasi NAN, sehingga sebanyak 30 butir (47,62%) dikategorikan tidak valid atau membutuhkan perbaikan mendesak. Berikut perhitungan persentase korelasi yang tidak signifikan:

$$\frac{\text{Jumlah Butir Tidak Valid}}{\text{Total Butir Valid}} \times 100\% = \frac{30}{63} \times 100\% = 47,62\%$$

Persentase butir soal berdasarkan kriteria validitas disajikan pada Diagram 1.



Proporsi butir soal yang tidak valid hampir mencapai separuh dari total butir soal, yang menunjukkan adanya masalah besar terkait kualitas instrumen tes. Butir-butir tersebut tidak mampu membedakan tingkat pemahaman materi peserta didik dengan baik, atau tidak sesuai dengan tujuan pengukuran tes secara keseluruhan. Butir yang menghasilkan nilai korelasi NAN, terutama pada butir dengan TK 100%, adalah salah satu penyebab utama rendahnya validitas karena butir tersebut tidak berhasil mengukur perbedaan kemampuan antar peserta.

Reliabilitas Tes

Hera dkk. (2023) menyatakan bahwa reliabilitas adalah angka yang menggambarkan seberapa jauh suatu instrumen dapat diandalkan ketika dipakai berulang kali untuk mengevaluasi hal yang serupa, hasilnya akan tetap cukup stabil atau terjaga. Jika alat ukur yang sama dipakai berulang kali dan terus

menghasilkan hasil yang identik, artinya instrumen tersebut dapat dianggap dapat dipercaya. Dengan kata lain, alat ini menunjukkan konsistensi, kestabilan, dan kepastian. Berdasarkan hasil kalkulasi Anates, didapat hasil data seperti pada tabel berikut.

Tabel 1 Data Hasil Analisis Reliabilitas Tes

Statistik	Nilai	Posttest
Jumlah Subyek		
Butir Soal	63	-
Rata-Rata Skor	55,20	-
Simpang Baku	6,76	-
Reliabilitas Tes	0,90	Sangat Tinggi

Sumber: Data diolah tahun 2025

Koefisien reliabilitas dari tes mencapai 0,90, yang menandakan bahwa instrumen yang dipakai menunjukkan tingkat reliabilitas yang sangat tinggi. Angka ini mengindikasikan bahwa alat pengukur tersebut dapat menghasilkan hasil yang stabil saat digunakan dalam kondisi yang serupa. Dengan kata lain, pertanyaan dalam tes ini menunjukkan konsistensi yang baik, sehingga dapat diandalkan untuk menilai kemampuan peserta dengan akurat. Nilai reliabilitas 0,90 juga telah melebihi ambang batas minimum yang lazim digunakan untuk menentukan kelayakan sebuah instrumen ($\geq 0,70$), sehingga tes ini

bisa dikategorikan sebagai sangat reliabel dan pantas digunakan dalam evaluasi pembelajaran.

Tingkat Kesukaran

Jika diperlihatkan dalam format gambar, maka analisis mengenai tingkat kesulitan dari 63 butir soal terlihat seperti yang ada pada Gambar 1 dan Gambar 2.

TINGKAT KESUKARAN

Jumlah Subyek= 20
 Butir Soal= 63
 Nama berkas: C:\USERS\NADIA\DOWNLOADS\ANATES KELOMPOK.ANA

No Butir Baru	No Butir Asli	Jml Betul	Tkt. Kesukaran(%)	Tafsiran
1	1	20	100,00	Sangat Mudah
2	2	19	95,00	Sangat Mudah
3	3	19	95,00	Sangat Mudah
4	4	19	95,00	Sangat Mudah
5	5	20	100,00	Sangat Mudah
6	6	18	90,00	Sangat Mudah
7	7	20	100,00	Sangat Mudah
8	8	19	95,00	Sangat Mudah
9	9	19	95,00	Sangat Mudah
10	10	19	95,00	Sangat Mudah
11	11	20	100,00	Sangat Mudah
12	12	20	100,00	Sangat Mudah
13	13	20	100,00	Sangat Mudah
14	14	20	100,00	Sangat Mudah
15	15	20	100,00	Sangat Mudah
16	16	20	100,00	Sangat Mudah
17	17	20	100,00	Sangat Mudah
18	18	20	100,00	Sangat Mudah
19	19	18	90,00	Sangat Mudah
20	20	19	95,00	Sangat Mudah
21	21	19	95,00	Sangat Mudah
22	22	19	95,00	Sangat Mudah
23	23	19	95,00	Sangat Mudah
24	24	19	95,00	Sangat Mudah
25	25	20	100,00	Sangat Mudah
26	26	7	35,00	Sedang
27	27	10	50,00	Sedang
28	28	20	100,00	Sangat Mudah
29	29	19	95,00	Sangat Mudah
30	30	19	95,00	Sangat Mudah

Sumber: Data diolah tahun 2025
 Gambar 1 Data Hasil Analisis Tingkat Kesukaran

Gambar 1 di atas memaparkan hasil perhitungan tingkat kesukaran untuk 30 soal pertama. Guna melengkapi data tersebut, hasil analisis untuk nomor-nomor berikutnya disajikan dalam Gambar 2 berikut.

31	31	19	95,00	Sangat Mudah
32	32	19	95,00	Sangat Mudah
33	33	13	65,00	Sedang
34	34	19	95,00	Sangat Mudah
35	35	19	95,00	Sangat Mudah
36	36	13	65,00	Sedang
37	37	17	85,00	Mudah
38	38	20	100,00	Sangat Mudah
39	39	15	75,00	Mudah
40	40	19	95,00	Sangat Mudah
41	41	19	95,00	Sangat Mudah
42	42	19	95,00	Sangat Mudah
43	43	10	50,00	Sedang
44	44	18	90,00	Sangat Mudah
45	45	19	95,00	Sangat Mudah
46	46	19	95,00	Sangat Mudah
47	47	16	80,00	Mudah
48	48	18	90,00	Sangat Mudah
49	49	18	90,00	Sangat Mudah
50	50	19	95,00	Sangat Mudah
51	51	14	70,00	Sedang
52	52	20	100,00	Sangat Mudah
53	53	20	100,00	Sangat Mudah
54	54	13	65,00	Sedang
55	55	17	85,00	Mudah
56	56	17	85,00	Mudah
57	57	19	95,00	Sangat Mudah
58	58	8	40,00	Sedang
59	59	18	90,00	Sangat Mudah
60	60	2	10,00	Sangat Sukar
61	61	12	60,00	Sedang
62	62	19	95,00	Sangat Mudah
63	63	17	85,00	Mudah

Sumber: Data diolah tahun 2025
 Gambar 2 Data Hasil Analisis Tingkat Kesukaran

untuk menggambarkan seberapa gampang atau sulitnya seorang peserta ujian untuk menjawab sebuah pertanyaan. Pertanyaan yang berkualitas adalah yang memiliki tingkat kesulitan yang sedang, yaitu tidak terlalu gampang dan tidak terlalu sulit (Sumanti & Susanti, 2019). Melalui analisis terhadap tingkat kesulitan, ditemukan bahwa 42 butir soal atau sekitar 66,67% termasuk dalam kategori sangat mudah (TK > 85%). Ini menunjukkan bahwa sebagian besar peserta didik dapat menjawab dengan tepat, sehingga soal-soal tersebut belum cukup menantang untuk menguji variasi keterampilan peserta. Selanjutnya, terdapat 5 butir soal (7,94%) yang dikategorikan mudah dengan indeks

kesulitan antara 70% hingga 85%, yang menyiratkan bahwa hanya sedikit dari soal tersebut yang memiliki tingkat kesulitan yang sesuai untuk membedakan kemampuan peserta. Terdapat 15 butir soal (23,81%) yang masuk dalam kategori sedang dengan indeks kesulitan di antara 30% hingga 70%. Di sisi lain, tidak ada butir soal yang dianggap sulit (0,00%) dengan indeks kesulitan antara 15% hingga 30%, dan hanya 1 butir soal (1,58%) yang termasuk dalam kategori sangat sulit ($TK < 15\%$). Secara keseluruhan, total butir soal yang dianalisis mencapai 63 butir dengan total persentase 100%. Distribusi tingkat kesulitan menandakan bahwa komposisi soal belum merata, karena banyaknya butir soal yang sangat mudah dapat mengurangi efektivitas tes dalam menilai kemampuan peserta dengan optimal.

Daya Pembeda

Daya pembeda pertanyaan merujuk pada kemampuan suatu soal dalam membedakan antara siswa yang benar-benar menguasai materi dan mereka yang belum menguasainya (Qomariyah, 2022). Memahami potensi setiap soal menjadi hal yang sangat krusial,

karena salah satu dasar dalam penyusunan soal adalah keyakinan bahwa kemampuan antara siswa bervariasi. Soal yang tidak memiliki daya pembeda yang baik atau kurang memadai tidak mampu menunjukkan perbedaan kemampuan antara siswa yang berprestasi dengan siswa yang kurang berprestasi (Nafs et al., 2023). Berdasarkan nilai indeks DP yang didapatkan, butir soal kemudian dikelompokkan pada Tabel 2.

Tabel 2 Data Hasil Analisis Daya Pembeda

Kategori Daya Pembeda	Rentang DP (%)	Jumlah Butir	Persen
Sangat Baik	70-100	5	7,49%
Baik	40-69	11	17,46%
Sedang	20-39	23	36,51%
Kurang	0-19	20	31,75%
Negatif	< 0	4	6,34%
Total	-	63	100%

Sumber: Data diolah tahun 2025

Berdasarkan analisis yang dilakukan, terlihat bahwa mayoritas soal, yaitu 36,51%, menunjukkan daya bedanya dalam kategori cukup. Ini berarti, soal-soal ini dapat sedikit membedakan antara siswa dengan kemampuan tinggi dan rendah. Selain itu, ada 17,46% soal yang tergolong baik dan 7,94% dalam kategori sangat baik, yang menandakan adanya beberapa soal berkualitas tinggi dalam membedakan kemampuan

peserta. Namun, terdapat pula 31,75% soal yang berada dalam kategori buruk dan 6,34% yang memiliki daya pembeda negatif. Ini menunjukkan bahwa soal-soal ini gagal dalam membedakan siswa yang berkemampuan tinggi dan rendah, bahkan sering kali salah. Soal dengan daya pembeda negatif harus diperbaiki atau dihapus karena cenderung dijawab benar oleh siswa dengan kemampuan rendah dan salah oleh siswa yang memiliki kemampuan lebih baik. Secara keseluruhan, ada kebutuhan untuk meningkatkan kualitas daya pembeda soal. Hal ini disebabkan karena proporsi soal yang berada dalam kategori baik dan sangat baik belum cukup mendominasi. Idealnya, sebagian besar soal harus berada pada kategori baik hingga sangat baik agar ujian benar-benar efektif dalam menilai perbedaan kemampuan siswa.

Kualitas Pengecoh

Analisis tentang kualitas pengecoh dilakukan untuk mengevaluasi seberapa efektif pilihan jawaban yang salah dalam mengalihkan perhatian siswa, terutama untuk kelompok Asor (Warju

et al., 2020). Jika diperlihatkan dalam format diagram, maka persentase kualitas pengecoh terlihat seperti yang ada pada tabel berikut.

Tabel 3 Rangkuman Hasil Analisis Kualitas Pengecoh

Kualitas Pengecoh	Jumlah Pengecoh	Perhitungan	Persentase Pengecoh
Sangat Baik (++)	10	$\frac{10}{252} \times 100\%$	3,97%
Baik (+)	12	$\frac{12}{252} \times 100\%$	4,76%
Kurang Baik (-)	8	$\frac{8}{252} \times 100\%$	3,17%
Buruk (--)	40	$\frac{40}{252} \times 100\%$	15,87%
Sangat Buruk (---)	182	$\frac{182}{252} \times 100\%$	72,22%
TOTAL	252		100,00%

Sumber: Data diolah tahun 2025

Hasil evaluasi menunjukkan bahwa instrumen pengujian ini memiliki kekurangan yang signifikan terkait kualitas pilihan salah. Secara keseluruhan, 91,27% (230 pilihan salah) dianggap tidak efisien (termasuk kategori Kurang Baik, Buruk, dan Sangat Buruk). Proporsi ini dipegang oleh kategori Sangat Buruk (---), yang mencakup 72,22% (182 pilihan salah), diikuti oleh kategori Buruk (--) dengan persentase 15,87%. Sebaliknya, pilihan yang berfungsi baik (Sangat Baik dan Baik) hanya memberikan kontribusi 8,73% dari total keseluruhan. Tingginya angka

ketidakefektifan dari pilihan ini disebabkan oleh dominasi butir soal yang memiliki Tingkat Kesulitan Sangat Mudah, di mana pilihan tersebut tidak dipilih sama sekali oleh responden (yang membuatnya dikategorikan sebagai Sangat Buruk). Ketidakmampuan pilihan untuk berfungsi secara efektif ini mengurangi kemampuan tes dalam mengidentifikasi kemampuan peserta yang sebenarnya, serta menjadi indikasi jelas bahwa jawaban yang salah tersebut terlalu mudah untuk dihapus atau tidak logis, sehingga perlu dilakukan perumusan kembali secara komprehensif.

Berdasarkan analisis tentang kualitas instrumen pengujian dengan 63 pertanyaan ini menunjukkan adanya kontradiksi antara konsistensi keseluruhan tes dan kualitas setiap pertanyaan secara individu. Di satu sisi, instrumen ini menunjukkan Reliabilitas yang Sangat Tinggi (0,90), yang menandakan adanya konsistensi dalam hasil pengukuran. Namun, pemeriksaan yang lebih rinci terhadap setiap pertanyaan mengungkapkan berbagai masalah serius: (1) Rasio pertanyaan didominasi oleh kategori Sangat Mudah (66,67%), menunjukkan bahwa tes ini tidak

dapat menguji kemampuan dengan optimal; (2) Akibatnya, sebagian besar pertanyaan (66,66%) memiliki Daya Pembeda yang rendah atau negatif, tidak mampu membedakan antara peserta dengan kemampuan tinggi dan rendah; (3) Situasi ini diperburuk oleh temuan validitas item yang buruk, di mana hampir setengah dari pertanyaan (47,62%) dinyatakan Tidak Valid; dan (4) Masalah utama terletak pada Kualitas Pengecoh, di mana sebagian besar (91,27%) dianggap tidak efektif (terutama 72,22% sangat buruk). Secara keseluruhan, meskipun tes ini dapat diandalkan, tingginya proporsi pertanyaan yang bermasalah pada Tingkat Kesulitan, Daya Pembeda, dan Pengecoh menunjukkan bahwa instrumen tes ini belum efektif dalam mengukur tingkat penguasaan kompetensi para peserta didik.

D. Kesimpulan

Penelitian ini bertujuan untuk mengevaluasi kualitas 63 butir soal pada materi Prinsip-Prinsip dan Konsep Akuntansi serta Perbankan Dasar yang dikembangkan oleh mahasiswa Pendidikan Akuntansi Universitas Negeri Surabaya dengan menggunakan aplikasi Anates. Analisis dilakukan pada lima indikator

utama, yaitu validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh.

Secara umum, hasil penelitian menunjukkan adanya ketidakseimbangan antara konsistensi keseluruhan tes dan kualitas setiap butir soal secara individu. Instrumen tes memiliki reliabilitas sangat tinggi (0,90), yang menandakan konsistensi internal yang kuat. Namun, hampir separuh butir soal tidak memenuhi kriteria validitas, terutama disebabkan oleh nilai korelasi yang tidak signifikan serta kemunculan nilai NAN pada butir-butir dengan tingkat kesukaran 100%. Hal ini menunjukkan bahwa sejumlah soal gagal merepresentasikan konstruk yang hendak diukur dan tidak mampu membedakan kemampuan peserta.

Dari aspek tingkat kesukaran, komposisi butir soal didominasi oleh kategori sangat mudah. Kondisi ini menyebabkan tes kurang mampu menguji variasi kemampuan peserta dan berdampak langsung pada rendahnya daya pembeda. Analisis daya pembeda menunjukkan bahwa sebagian besar butir berada pada kategori cukup, buruk, atau bahkan negatif. Hal ini menandakan bahwa banyak soal yang tidak efektif dalam

membedakan peserta berkemampuan tinggi dan rendah. Selain itu, kualitas pengecoh berada pada kategori kurang baik hingga sangat buruk, dengan 91,27% pengecoh tidak berfungsi. Proporsi pengecoh yang tidak efektif yang sangat tinggi mengindikasikan bahwa opsi jawaban salah belum bekerja sebagaimana mestinya, karena tidak dipilih oleh peserta dan tidak membantu soal dalam memisahkan peserta berkemampuan tinggi dan rendah.

Secara keseluruhan, meskipun alat ukur yang digunakan menunjukkan tingkat keandalan yang cukup tinggi, kualitas soal dalam hal validitas, kesulitan, kemampuan membedakan, dan efektivitas pilihan salah masih rendah. Hasil penelitian ini menunjukkan bahwa tes belum siap dipakai sebagai sarana penilaian tanpa dilakukan perbaikan mendalam. Perbaikan sangat dibutuhkan, terutama dalam merumuskan indikator, variasi tingkat kesulitan, meningkatkan kemampuan membedakan, serta menyusun pilihan jawaban yang lebih efektif. Penelitian ini juga menekankan pentingnya penguasaan keterampilan psikometrik bagi calon guru agar mereka dapat merancang alat evaluasi yang tepat,

adil, dan sesuai dengan kriteria evaluasi pembelajaran.

DAFTAR PUSTAKA

- Adimas, I., Saputra, 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*. 6(3), 1288–1297.
- Barus, R., & Darma, J. (2025). Pengaruh Kompetensi Profesional Guru, Manajemen Kelas Dan Sarana Prasarana Terhadap Efektivitas Pembelajaran Siswa Konsentrasi Keahlianakuntansi Smk Negeri 13 Medan. *Jurnal Review Pendidikan Dan Pengajaran*, 8(2), 6770–6779. <https://doi.org/10.31004/jrpp.v8i2.48730>
- Hakim, L., Pratiwi, V., Ketintang, A. J., Gayungan, K., Timur, J., Negeri, U., & Kampus, S. (2024). *PROSES BISNIS DI BIDANG AKUNTANSI KEUANGAN LEMBAGA FASE E KELAS 10 SMK MELALUI APLIKASI ANATES* Miftachul Jannah Universitas Negeri Surabaya Evie Dwi Erwanti Universitas Negeri Surabaya Silvira Bela Agustin Universitas Negeri Surabaya Pendidikan akuntansi . 9(2).
- Hera, A. S., Zulhijrah, Nabila, J. L., & Shaleh. (2023). 2268-Article Text-9893-1-10-20231224. *Jurnal Ilmiah PGSD FKIP Universitas Mandiri*, 9, 2990–2991.
- Himawan, R., & Nurgiyantoro, B. (2022). Analisis Butir Soal Latihan Penilaian Akhir Semester Ganjil Mata Pelajaran Bahasa Indonesia Kelas VIII SMPN 1 Bambanglipuro Bantul Menggunakan program ITEMAN (Analysis of exercise items for odd semester end of semester Indonesian language subjects class. *Kembara: Jurnal Keilmuan Bahasa, Sastra, Dan Pengajarannya*, 8(1), 160–180.
- Ida, F. F., & Musyarofah, A. (2021). Validitas dan reliabilitas dalam analisis butir soal. *Al-muarrib journal of arabic education*, 1(1), 34-44.
- Indah, D., & Andriani, A. (2021). *Item Analysis of Basic Accounting Subjects for Class X Accounting*. 5.
- Isnawati, I., Jalinus, N., & Risfendra, R. (2020). Analisis Kemampuan Pedagogi Guru SMK yang

- sedang Mengambil Pendidikan Profesi Guru dengan Metode Deskriptif Kuantitatif dan Metode Kualitatif. *INVOTEK: Jurnal Inovasi Vokasional Dan Teknologi*, 20(1), 37–44. <https://doi.org/10.24036/invotek.v20i1.652>
- Jayusman, I., & Shavab, O. A. K. (2020). Aktivitas Belajar Mahasiswa Dengan Menggunakan Media Pembelajaran Learning Management System (Lms) Berbasis Edmodo Dalam Pembelajaran Sejarah. *Jurnal Artefak*, 7(1), 13. <https://doi.org/10.25157/ja.v7i1.3180>
- Karengga, F. I., Rizko, U., & Bashith, A. (2025). Analisis Problematika Pelaksanaan Evaluasi Pembelajaran IPA dalam Mencapai Tujuan Pendidikan pada Kurikulum Merdeka SD/MI. *Al-Madrasah Jurnal Pendidikan Madrasah Ibtidaiyah*, 9(2), 533. <https://doi.org/10.35931/am.v9i2.4401>
- Magdalena, I., Fauziah, S. N., Faziah, S. N., & Nupus, F. S. (2021). Analisis validitas, reliabilitas, tingkat kesulitan dan daya beda butir soal ujian akhir semester tema 7 kelas III SDN Karet 1 Sepatan.
- Mu, A.-, Pendidikan, J., & Arab, B. (2021). *Validitas dan Reliabilitas dalam Analisis Butir Soal*. ليحتب. قنيسلا وديج يرغ. 1(1), 34–44.
- Nafs, H., Sridana, N., Hikmah, N., & Soeprianto, H. (2023). Analisis Kualitas Butir Soal Ulangan Akhir Semester Genap Mata Pelajaran Matematika Kelas VII SMPN 6 Mataram Tahun Ajaran 2022/2023. *Jurnal Ilmiah Profesi Pendidikan*, 8(4), 2324–2331. <https://doi.org/10.29303/jipp.v8i4.1785>
- Pengabdian, J., Ilmu, M., Pendidikan, M. A., Lampung, U., Rohman, F., Fisika, P., Lampung, U., Farha, S. A., Pendidikan, M. A., Lampung, U., Putri, H. S., Pendidikan, M. A., Lampung, U., Sucitra, D. A., Pendidikan, M. A., Lampung, U., & History, A. (2025). *Optimalisasi Kualitas Butir Soal melalui Pelatihan Penyusunan Indikator dengan Uji Validitas dan Reliabilitas*. 4(2). <https://doi.org/10.23960/jpm-ip.vol.4i.2.1083>
- Puteri, N. D., Putricia, 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. <https://doi.org/10.62421/jibema.v2i3.108>
- Qomariyah, L. (2022). Analisis Tingkat Kesukaran dan Daya Pembeda Butir Soal TOAFL Universitas Hasyim Asy'ari Tebuireng Jombang. *Lisanan Arabiya: Jurnal Pendidikan Bahasa Arab*, 6(1), 1–18. <https://doi.org/10.32699/liar.v6i1.2549>
- Rahmawati, N. S., & Rahman, M. F. (2025). Analisis Butir Soal Objektif Penilaian Sumatif Akhir Semester Gasal Mata Pelajaran Ekonomi. *JlIP - Jurnal Ilmiah Ilmu Pendidikan*, 8(8), 8863–8873. <https://doi.org/10.54371/jiip.v8i8.8837>
- Ramadhani, T., Aulia, T. H., & Anastasya, S. D. (2025). *PGSD Universitas Pendidikan Indonesia*. 10, 467–481.
- Saputra, H. D., Purwanto, W., Setiawan, D., Fernandez, D., & Putra, R. (2022). HASIL BELAJAR MAHASISWA : ANALISIS BUTIR SOAL TES pembelajaran sebagai upaya meningkatkan hasil belajar mahasiswa . Efektivitas. *Edukasi: Jurnal Pendidikan*, 20, 15–27.
- Solichin, M. (2017). Analisis Daya Beda Soal, Taraf Kesukaran, Validitas Butir Tes, Interpretasi Hasil Tes dan Validitas Ramalan dalam Evaluasi Pendidikan. *Dirāsāt: Jurnal Manajemen & Pendidikan Islam*, 2(2), 192–213. www.depdiknas.go.id/evaluasi-proses-
- Sudarto, S., Abd. Hafid, & Abd. Kadir. (2023). Tingkat Pencapaian Kemampuan Guru Sd Dalam Menggunakan Aplikasi Anates Melalui Pelatihan. *J-ABDI: Jurnal Pengabdian Kepada Masyarakat*, 3(5), 793–800. <https://doi.org/10.53625/jabdi.v3i5.6595>
- Sumanti, E., & Susanti, D. (2019). ANALISIS TINGKAT KESUKARAN, DAYA BEDA DAN EFEKTIVITAS PENGECOHpSOALpUJIAN AKHIRISEMESTERIMATAIPELAJARAN EKONOMI SISWA SMA NEGERI 2 GUNUNG TALANG TAHUN AJARAN 2018/2019.

Jurnal Ecogen, 2(4), 678.

[https://doi.org/10.24036/jmpe.v2i](https://doi.org/10.24036/jmpe.v2i4.7845)

4.7845

Susilowati, D. A., Saputri, N., Ni, K., & Hakim, L. (2025). Penerapan Aplikasi Anates Untuk Analisis Soal HOTS Dalam Evaluasi Pembelajaran Spreadsheet Di SMKN 1 Kendal. *PESHUM: Jurnal Pendidikan, Sosial Dan Humaniora*, 4(2), 2466–2477.

Tania, I., Setyawan, F., & Adiwibowo, Y. (2025). Jurnal Review Pendidikan dan Pengajaran, Volume 8 Nomor 2, 2025 | 5771. *Jurnal Review Pendidikan Dan Pengajaran*, 8, 5771–5777. <http://journal.universitaspahlawan.ac.id/index.php/jrpp>

Warju, W., Ariyanto, S. R., Soeryanto, S., & Trisna, R. A. (2020). Analisis Kualitas Butir Soal Tipe Hots Pada Kompetensi Sistem Rem Di Sekolah Menengah Kejuruan. *Jurnal Pendidikan Teknologi Dan Kejuruan*, 17(1), 95. <https://doi.org/10.23887/jptk-undiksha.v17i1.22914>