



## ANALISIS BUTIR SOAL HOTS ELEMEN PRINSIP DAN KONSEP DASAR AKUNTANSI DAN PERBANKAN MELALUI PROGRAM ANATES

**Gita Putri Rahayu**

Universitas Negeri Surabaya

**Luqman Hakim**

Universitas Negeri Surabaya

**Vivi Pratiwi**

Universitas Negeri Surabaya

**Fatma Nisa Nafilah**

Universitas Negeri Surabaya

**Naila Atika Sari**

Universitas Negeri Surabaya

**Sabrina Keashya Aurellia Putri**

Universitas Negeri Surabaya

**Nadia Zalfa Rahmatul Ulla**

Universitas Negeri Surabaya

**Suci Fadhila Rizti**

Universitas Negeri Surabaya

**Yesi Kristiani Putri**

Universitas Negeri Surabaya

Korespondensi penulis: [24080304085@mhs.unesa.ac.id](mailto:24080304085@mhs.unesa.ac.id)

**Abstract .** This study aims to analyze multiple-choice questions related to phase e elements, namely the Principles and Basic Concepts of Accounting and Basic Banking, using the Anates program. The study involved 20 participants, consisting of college students and vocational high school students. The Anates application program serves as a tool to process academic evaluation data accurately and efficiently. The method used in this study is quantitative and is described descriptively, resulting in analytical findings including weighted data scores, test reliability, discriminative power, difficulty level, validity level, and distractor quality. The results from the Anates program make it easier for teachers to evaluate the quality of learning and develop more focused, innovative, and student-centered teaching plans. This research is expected to provide benefits in improving students' accounting knowledge, and teachers are expected to be able to adjust their teaching methods and strategies dynamically, thereby supporting the optimal improvement of students' learning outcomes in this digital era.

**Keywords:** Anates, Question Items, Basic Accounting Principles and Concepts, Basic Banking Principles and Concepts, HOTS

**Abstrak.** Penelitian ini bertujuan untuk menganalisis butir soal bentuk pilihan ganda yang berkaitan dengan elemen fase e yaitu Prinsip dan Konsep Dasar Akuntansi dan Perbankan Dasar dengan menggunakan program anates. Penelitian ini melibatkan 20 partisipan yang mana terdiri dari mahasiswa dan siswa SMK. Program aplikasi anates berperan sebagai alat bantu dalam mengolah data evaluasi akademik secara akurat dan efisien. Metode yang digunakan dalam penelitian ini adalah kuantitatif dan dijabarkan secara deskriptif yang menghasilkan temuan analisis meliputi skor data berbobot, reliabilitas tes, daya pembeda, tingkat kesukaran, tingkat validitas, dan kualitas pengecoh. Hasil program anates ini memudahkan guru dalam mengevaluasi kualitas pembelajaran dan menyusun rencana pengajaran yang lebih terarah, inovatif, dan sesuai dengan kebutuhan peserta didik. Penelitian ini diharapkan dapat memberikan manfaat pada peningkatan pengetahuan akuntansi siswa serta guru diharapkan dapat menyesuaikan metode serta strategi pembelajaran secara dinamis sehingga mendukung peningkatan hasil belajar siswa secara optimal di era digital ini.

**Kata Kunci:** Anates, Butir Soal, Prinsip dan Konsep Akuntansi Dasar, Prinsip dan Konsep Perbankan Dasar, HOTS

## PENDAHULUAN

Pendidikan merupakan aspek penting dalam pembangunan bangsa. Dimana pendidikan yang efektif sangat bergantung pada kualitas evaluasi pembelajaran yang dilakukan. Evaluasi pembelajaran memiliki peran penting dalam menilai pencapaian hasil belajar peserta didik dan memberikan umpan balik untuk pengajaran. Dalam hal ini, butir soal berkualitas tinggi menjadi sangat penting untuk menilai pemahaman siswa dengan tepat. Salah satu metode evaluasi yang semakin mendapat perhatian adalah penggunaan butir soal HOTS (Higher Order Thinking Skills), yang dimana dirancang untuk menguji kemampuan berpikir kritis, analitis, dan kreatif siswa (Rahmawati et al., 2024).

Penilaian mutu soal adalah faktor penting dalam proses evaluasi pembelajaran. (Alfiani et al., 2024). Soal yang bermutu baik memiliki beberapa karakteristik. Pertama, mampu membedakan tingkat kemampuan yang dimiliki masing-masing siswa. Kedua, soal harus valid dan reliabel. Ketiga, untuk meningkatkan kevalidan dan keterandalan, analisis terhadap item soal tersebut sangat dibutuhkan. Untuk membantu menganalisis kualitas soal, aplikasi Anates menjadi alat bantu yang efektif (Anica et al., 2023). Hal ini penting karena berpikir kritis termasuk dalam kompetensi yang diperlukan di Abad-21, sehingga aspek tersebut bisa digunakan sebagai acuan dalam menganalisis soal-soal yang diberikan agar soal tersebut dapat mengukur kemampuan siswa dalam berpikir kritis dan analitis (Agustina et al., 2022).

Analisis butir soal adalah proses yang dilakukan guru untuk menilai kualitas butir soal yang digunakan. Melalui analisis ini, guru dapat menentukan soal mana yang berkualitas dan layak digunakan kembali serta soal mana yang perlu direvisi atau dihapus (Ida et al., 2021). Melalui Anates, para pendidik dapat mengidentifikasi

kekuatan dan kelemahan setiap butir soal berdasarkan data statistik, seperti indeks validitas, reliabilitas, daya pembeda, dan tingkat kesukaran (Fietri et al., 2021). Dengan demikian, penelitian ini dilakukan untuk menganalisis kualitas butir soal HOTS dalam pelajaran Prinsip dan Konsep Dasar Akuntansi dan Perbankan Dasar dengan menggunakan aplikasi Anates. Artikel ini akan menjelaskan hasil dari pemeriksaan soal HOTS di mata pelajaran tersebut, dengan tujuan untuk mengevaluasi kualitas alat evaluasi yang digunakan serta memberikan saran untuk meningkatkan penilaian pembelajaran di masa depan.

## KAJIAN TEORITIS

### 1. RELIABILITAS

Reliabilitas adalah istilah yang digunakan untuk menilai sejauh mana sebuah tes mampu memberikan hasil yang stabil dalam mengukur elemen yang seharusnya diukur. Tes yang mampu menghasilkan skor yang konsisten dari waktu ke waktu dianggap memiliki tingkat reliabilitas tinggi. Dengan kata lain, tes tersebut dapat dipercaya karena tidak menunjukkan perubahan yang signifikan dalam hasilnya. Tabel 1 menunjukkan klasifikasi reliabilitas, yang dijelaskan oleh (Akhmadi, 2021).

Tabel 1. Kriteria Koefisien Tingkat Reliabilitas

Koefisien Reliabilitas	Kategori
$0,00 \leq r_{11} \leq 0,20$	Sangat Rendah
$0,19 \leq DB \leq 0,39$	Rendah
$0,40 \leq DB \leq 0,69$	Cukup
$0,70 \leq DB \leq 0,89$	Tinggi
$0,90 \leq DB \leq 1,00$	Sangat Tinggi

### 2. DAYA PEMBEDA

Daya beda didefinisikan sebagai kemampuan seorang siswa untuk membedakan antara siswa yang memiliki tingkat kemampuan yang tinggi dan siswa yang memiliki tingkat kemampuan yang lebih rendah dalam sebuah soal (Akhmadi, 2021). Kualitas alat penilaian ditentukan oleh daya pembeda, karena alat tersebut dapat membedakan siswa yang memahami materi dengan siswa yang tidak. Namun, daya pembeda dengan indeks negatif (kurang dari 0), menunjukkan bahwa siswa dengan kemampuan rendah lebih sering menjawab pertanyaan dengan benar daripada siswa dengan kemampuan tinggi

(Mawardhi et al., 2023). Rincian klasifikasi daya pembeda dapat dilihat dalam Tabel 2 dibawah ini.

Tabel 2. Kriteria Koefisien Daya Pembeda

Besar Daya Beda	Keterangan
0,00 – 0,19	Tidak Baik/Jelek (soal di revisi)
0,20 - 0,39	Cukup Baik (soal digunakan dan di revisi)
0,40 – 0,69	Baik (soal digunakan)
0,70 – 1,00	Sangat Baik (soal digunakan)
Negatif	Soal dibuang

### 3. TINGKAT KESUKARAN

Menurut Hendrayadi et al. (2024) persentase siswa yang menjawab pertanyaan dengan benar menentukan tingkat kesukaran butir soal. Soal yang baik adalah yang memiliki tingkat kesulitan yang seimbang, artinya tidak terlalu mudah sehingga semua peserta dapat menjawabnya atau terlalu sulit sehingga hanya beberapa orang yang dapat menjawabnya (Bano et al., 2022). Soal yang dianggap sulit harus diubah atau diperbaiki. Begitu pula, masalah yang terlalu sederhana harus dipertimbangkan kembali. Jika semua siswa dapat menjawab soal dengan benar, ini menunjukkan soal tersebut terlalu sederhana dan tidak efektif untuk mengukur seberapa baik siswa memahaminya. Sementara itu, soal dengan tingkat kesulitan sedang dapat disimpan dalam bank soal untuk digunakan pada ujian berikutnya (Zihan Marcela Hapsari et al., 2024). Rincian klasifikasi tingkat kesukaran dapat dilihat dalam Tabel 3 dibawah ini.

Tabel 3. Kriteria Koefisien Tingkat Kesukaran

Indeks Kesukaran (P)	Penafsiran Butir Soal
$0,00 \leq TK \leq 0,30$	Sukar
$0,31 \leq TK \leq 0,70$	Sedang
$0,71 \leq TK \leq 1,00$	Mudah

### 4. VALIDITAS

Validitas adalah istilah yang digunakan untuk menggambarkan seberapa tepat suatu tes dalam menghasilkan data yang akurat (Puspasari & Puspita, 2022). Sebuah instrumen dianggap valid hanya jika dapat mengukur elemen yang dimaksud. Jenis soal yang memiliki korelasi yang signifikan dianggap sebagai jenis soal yang baik dan dapat

digunakan kembali pada ujian berikutnya (Wiguna, 2021). Rincian kategori validitas dapat dinilai dari Tabel 4.

Tabel 4. Kriteria Koefisien Tingkat Validitas

Koefisien Butir Soal	Interpretasi
0,80 – 1,00	Sangat Tinggi
0,60 - 0,79	Tinggi
0,40 – 0,59	Cukup
0,20 – 0,39	Rendah
0,00 – 0,19	Sangat Rendah

## 5. KUALITAS PENGECHOH

Menurut Leonardus et al. (2024), efektifitas pengecoh (distraktor) adalah pola yang menunjukkan bagaimana peserta tes memilih jawaban berdasarkan pilihan yang tersedia untuk setiap soal. Pola ini menunjukkan seberapa baik distraktor berfungsi untuk membuat peserta tes memilih jawaban yang tidak masuk akal tetapi masih salah. Rincian kategori kualitas pengecoh dapat dilihat dari Tabel 5.

Tabel 5. Kriteria Koefisien Tingkat Kualitas Pengecoh

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

## METODE PENELITIAN

Penelitian ini menggunakan metode pendekatan kuantitatif dan metode deskriptif. Penelitian kuantitatif ini dilaksanakan bukan untuk membuktikan atau menyangkal hipotesis. Namun, tujuannya adalah untuk mengeksplorasi dan menggambarkan keadaan nyata sesuai dengan fenomena yang sedang diteliti secara jelas dan berdasarkan fakta. Data dari responden yang berupa angka akan dianalisis dengan menggunakan program Anates V4. Setelah dilakukan analisis, data akan diinterpretasikan secara deskriptif dalam bentuk persentase (V et al., 2024).

Penelitian ini memiliki tujuan utama yaitu untuk menjelaskan dan memperlihatkan kualitas soal HOTS yang diberikan kepada siswa SMK jurusan Akuntansi dan mahasiswa Universitas Negeri Surabaya Fakultas Ekonomika dan Bisnis Semester 1 dengan menilai reliabilitas, daya beda, tingkat kesulitan, validitas, dan kualitas pengecoh dengan melalui Google Form yang telah dibuat berdasarkan standar teori pengujian klasik (Susilowati et al., 2025). Subjek dalam penelitian ini merupakan siswa SMK jurusan Akuntansi dan mahasiswa Universitas Negeri Surabaya Fakultas Ekonomika dan Bisnis Semester 1. Subjek penelitian diambil secara acak (random sampling).

Penyusunan soal berdasarkan pedoman Taksonomi Bloom yang dilakukan untuk memastikan soal dapat mengukur kemampuan berpikir tingkat tinggi, seperti C4 Menganalisis, C5 Mengevaluasi, dan C6 menciptakan. Proses ini sejalan dengan penelitian terdahulu yang menekankan pentingnya desain instrumen soal berbasis HOTS dalam mendukung pengembangan kemampuan berpikir kritis siswa (Hidayah et al., 2022). Kemudian ada 20 responden yang telah berpartisipasi secara sukarela mengisi kuesioner tersebut. Penelitian ini menggunakan alat penilaian berupa tes objektif yang terdiri dari pilihan ganda dengan jumlah 63 soal, setiap soal memiliki 5 pilihan jawaban dari A hingga E. Tes objektif adalah suatu metode penilaian yang memanfaatkan format pilihan ganda, dimana setiap pertanyaan disertai dengan beberapa pilihan jawaban yang memungkinkan peserta memilih jawaban yang dianggap tepat (Padang, 2021). Butir Soal HOTS yang dibuat oleh mahasiswa Program Studi Pendidikan Akuntansi di Universitas Negeri Surabaya. Data yang dikumpulkan meliputi soal HOTS dan jawaban dari siswa.

Proses penelitian mencakup penyebaran kuesioner menggunakan Google Form, yang mana pendekatan ini juga diterapkan dalam pembuatan dan penilaian soal yang berfokus pada HOTS dalam berbagai studi lainnya untuk tujuan efektivitas dan efisiensi. Program Anates ini berfungsi sebagai alat yang sangat bermanfaat untuk menilai mutu butir soal, yang mana sudah terbukti melalui sejumlah penelitian analisis butir soal yang memanfaatkan software ini (Mardiyah, 2024). Metode analisis terhadap soal ini dilakukan secara angka dengan pendekatan klasik yang termasuk penilaian mengenai keabsahan, ketepatan, tingkat kesulitan, kemampuan membedakan, serta kualitas pilihan soal. Sehingga hasilnya dapat digunakan sebagai acuan dalam meningkatkan kualitas instrumen evaluasi pembelajaran di masa mendatang.

## HASIL DAN PEMBAHASAN

Untuk mengevaluasi kualitas instrumen secara menyeluruh, program Anates telah digunakan untuk menganalisis 63 butir soal HOTS. Hasil pengolahan data ini mencakup informasi mengenai reliabilitas, daya pembeda, tingkat kesukaran, validitas, dan kualitas pengecoh pada setiap butir soal. Penjelasan berikut memaparkan hasil analisis yang digunakan untuk menentukan kelayakan instrumen yang digunakan.

### 1. RELIABILITAS

Dari data yang kita dapatkan dan sudah melewati hasil analisis menggunakan program anates yaitu sebagai berikut:

No Soal	No. Siswa	Nama/Warna Siswa	Rata-rata		Cukup
			Skor Rata	Skor Standar Deviasi	
1	1	Maryati,Azizah,Mardhi	33	28	62
2	2	Dian Fitriandini,Pemoneg Saputro	37	18	70
3	3	Tatina Hayuning Ayu Afifah	24	20	44
4	4	Aldila Candra Gunawan	36	26	63
5	5	Berry Vilandika	33	20	62
6	6	Umar Fathurrahman, Yudha	28	20	56
7	7	Priya Kristianti Andini Syahidah	31	18	61
8	8	Eka Asihulah,Tiyaputri	36	21	61
9	9	Chandika Hadi Prabangrah	19	18	38
10	10	Ria Herlina	23	21	54
11	11	Lida Oktavia,Virgoputra	28	18	59
12	12	Amilia Aritaya	33	20	63
13	13	Ayu Pia Riwuna	32	20	63
14	14	Azizahie Hafidz Syahrial	32	20	63
15	15	Fatimah Ambarwati	33	22	66
16	16	Elaiza Priscilia Septiana Putri	24	23	47
17	17	Zaidah Dwi Sugiharto	31	20	60
18	18	Eva Dewi Triandini	31	20	61
19	19	Putri Lestari Putriawati	30	20	59
20	20	Gita Puspita Elisa Putrianto	31	20	62

**Gambar 1. Hasil Anates Reliabilitas Bobot 1**

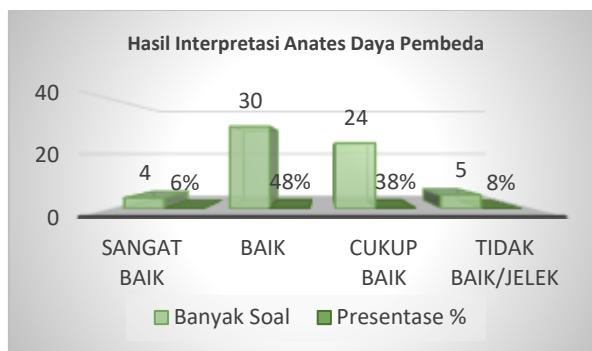
Sumber: Data Penelitian, 2025

Berdasarkan hasil perhitungan reliabilitas pada tampilan Anates, terlihat bahwa nilai reliabilitas tes adalah 0,97. Jika dibandingkan dengan *Tabel Kriteria Koefisien Tingkat Reliabilitas*, nilai tersebut berada pada rentang  $0,90 \leq r \leq 1,00$ , yang termasuk dalam kategori “Sangat Tinggi.” Hal ini menunjukkan bahwa instrumen tes memiliki konsistensi internal yang sangat baik, sehingga hasil pengukuran dapat dipercaya dan relatif stabil bila diberikan pada waktu atau situasi yang berbeda. Selain itu, nilai rata-rata skor sebesar 54,60 menggambarkan kecenderungan capaian peserta secara umum, sementara simpangan baku sebesar 9,55 menunjukkan bahwa variasi skor antar peserta cukup moderat—artinya terdapat perbedaan kemampuan, tetapi tidak terlalu ekstrem. Kombinasi antara reliabilitas yang sangat tinggi, rata-rata skor yang berada pada rentang menengah, serta simpangan baku yang masih dalam batas wajar mengindikasikan bahwa tes tidak hanya konsisten, tetapi juga mampu membedakan kemampuan peserta dengan

cukup baik. Dengan demikian, instrumen ini layak digunakan karena memberikan hasil yang stabil dan representatif terhadap kemampuan yang diukur.

## 2. DAYA PEMBEDA

Dari data yang kita dapatkan dan sudah melewati hasil analisis menggunakan program anates yaitu sebagai berikut:



**Gambar 2. Hasil Interpretasi Anates Daya Pembeda**

Sumber: Data Penelitian, 2025

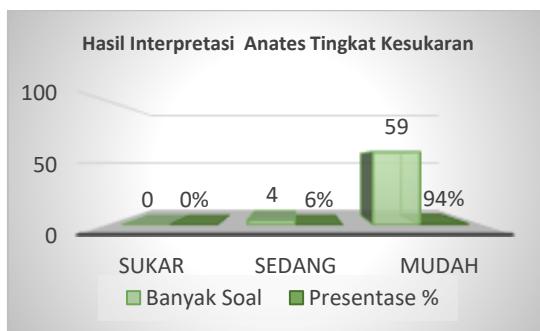
Berdasarkan hasil analisis butir soal menggunakan program Anates, diperoleh hasil daya pembeda dari 63 butir soal dengan jumlah subjek sebanyak 20 orang. Hasil perhitungan menunjukkan bahwa nilai indeks daya pembeda berkisar antara 0% hingga 80%, dengan interpretasi bahwa semakin tinggi nilai indeks daya pembeda maka semakin baik kemampuan soal dalam membedakan peserta didik berkemampuan tinggi dan rendah.

Hasil analisis daya pembeda butir soal menunjukkan bahwa kualitas butir soal secara keseluruhan berada pada kategori baik. Berdasarkan interpretasi Anates, diperoleh bahwa 4 butir soal (6%) termasuk kategori sangat baik, yang berarti hanya sebagian kecil soal memiliki kemampuan yang sangat optimal dalam membedakan peserta didik berkemampuan tinggi dan rendah. Sebagian besar butir soal, yaitu 30 soal (48%), berada pada kategori baik sehingga dapat dikatakan bahwa hampir setengah dari keseluruhan instrumen telah berfungsi efektif sebagai alat evaluasi pembelajaran. Selain itu, terdapat 24 butir soal (38%) yang tergolong cukup baik; meskipun masih dapat digunakan, butir-butir ini memerlukan perbaikan agar daya pembedanya dapat lebih optimal. Adapun 5 butir soal (8%) masuk kategori tidak baik atau jelek, sehingga tidak mampu berfungsi dalam membedakan kemampuan peserta dan sebaiknya direvisi atau dieliminasi dari instrumen. Dengan demikian, komposisi ini menunjukkan bahwa kualitas butir soal

memiliki kualitas yang memadai, meskipun masih diperlukan beberapa perbaikan untuk meningkatkan efektivitas keseluruhan butir soal.

### 3. TINGKAT KESUKARAN

Dari data yang kita dapatkan dan sudah melewati hasil analisis menggunakan program Anates yaitu sebagai berikut:



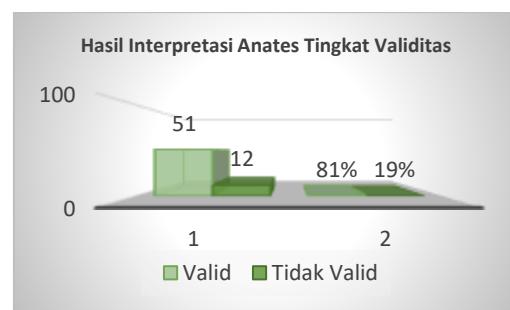
**Gambar 3. Hasil Interpretasi Anates Tingkat Kesukaran**

Sumber: Data Penelitian, 2025

Berdasarkan hasil pengolahan data menggunakan program Anates, terhadap 63 soal, diketahui bahwa tingkat kesulitan instrumen berada pada kategori yang kurang ideal untuk mengevaluasi kemampuan peserta didik secara menyeluruh. Menurut data, tidak ada satu butir soal (0%) yang termasuk dalam kategori sukar, yang berarti ujian tidak dapat menantang siswa dengan kemampuan tinggi. Selain itu, hanya 4 butir soal (6%) yang termasuk dalam kategori sedang, yang merupakan tingkat kesulitan ideal karena dapat memberikan informasi yang lebih akurat tentang variasi kemampuan peserta. Sebaliknya, sebanyak 94% atau 59 butir soal termasuk kategori mudah. Penggunaan soal yang dominan dalam kategori soal mudah ini menunjukkan bahwa sebagian besar siswa memiliki kemungkinan untuk menjawab soal dengan benar tanpa menunjukkan perbedaan kemampuan yang signifikan. Akibatnya, fungsi evaluasi tes berkurang untuk membedakan tingkat penguasaan materi. Secara keseluruhan, hasil ini menunjukkan bahwa instrumen harus direvisi dengan menambahkan butir soal yang lebih sukar dan sedang. Hal ini dilakukan agar tes menjadi lebih seimbang dan mampu mengukur kemampuan siswa secara optimal serta memberikan gambaran yang lebih akurat mengenai pencapaian kompetensi mereka.

### 4. VALIDITAS

Dari data yang kita dapatkan dan sudah melewati hasil analisis menggunakan program anates yaitu sebagai berikut:



Gambar 4. Hasil Interpretasi Anates Tingkat Validitas

Sumber: Data Penelitian, 2025

Berdasarkan hasil interpretasi Anates terhadap 63 butir soal menunjukkan bahwa sebagian besar butir soal memenuhi kriteria validitas yang memadai. Dari keseluruhan butir soal, sebanyak 51 butir soal atau sekitar 81%, dinyatakan valid. Ini menunjukkan bahwa butir-butir tersebut memiliki kemampuan untuk mengukur apa yang seharusnya diukur secara tepat dan konsisten. Jumlah soal valid yang tinggi menunjukkan bahwa instrumen tersebut memiliki kualitas konstruksi yang cukup baik dan sesuai dengan kompetensi yang diujikan. Namun, sekitar 19% atau 12 butir soal masih dianggap tidak valid. Jenis soal ini tidak dapat memberikan informasi yang akurat tentang kemampuan siswa karena tidak berkorelasi dengan skor total. Soal yang tidak valid dapat mengurangi efektivitas dan reliabilitas instrumen secara keseluruhan sehingga harus diperhatikan. Oleh karena itu, bagian-bagian ini harus direvisi atau dieliminasi untuk meningkatkan kualitas tes dan memberikan gambaran yang lebih akurat tentang pencapaian siswa.

## 5. KUALITAS PENGECHOH

Dari data yang kita dapatkan dan sudah melewati hasil analisis menggunakan program anates yaitu sebagai berikut:



Gambar 5. Hasil Interpretasi Anates Kualitas Pengecoh

Sumber: Data Penelitian, 2025

Berdasarkan hasil analisis atas 63 butir soal dengan menggunakan program Anates terhadap kualitas pengecoh menunjukkan variasi yang cukup beragam dalam efektivitas. Dari keseluruhan butir soal, terdapat 12 soal (19%) yang memiliki kualitas pengecoh yang sangat baik, dimana setiap opsi berfungsi dengan baik untuk menarik peserta yang belum menguasai materi. Selain itu, sebagian besar soal berada dalam kategori baik yaitu sebanyak 18 soal (29%) menunjukkan bahwa sebagian besar pengecoh melakukannya dengan baik dan mampu mengurangi kemungkinan peserta menebak jawaban. Sementara itu, kategori terbanyak berada pada kualitas kurang baik, yaitu sebanyak 20 soal (32%). Pada kategori ini, hanya satu pengecoh yang berfungsi, sehingga efektivitas soal dalam mengelompokkan peserta berkemampuan rendah masih tergolong kurang optimal.

Terdapat 10 soal (16%) yang termasuk kategori buruk, karena pengecoh pada butir tersebut tidak berfungsi sama sekali sehingga seluruh peserta cenderung mengarahkan pilihan pada satu jawaban, yang mengurangi kualitas penilaian. Selain itu, ada 3 soal (5%) yang termasuk kategori sangat buruk, di mana pengecoh yang dipilih lebih dominan daripada kunci jawaban. Hal itu, menunjukkan bahwa butir tersebut sangat bermasalah dan memerlukan perbaikan. Secara keseluruhan, hasil menunjukkan bahwa meskipun beberapa butir memiliki pengecoh yang baik, ada beberapa butir yang memerlukan perbaikan untuk membuat fungsi pengecohnya bekerja dengan lebih baik dan instrumen evaluasi memberikan hasil penilaian yang lebih akurat.

## KESIMPULAN DAN SARAN

Secara keseluruhan, instrumen tes, yang terdiri dari 63 soal HOTS dan dianalisis melalui program Anates, menunjukkan kualitas yang cukup baik, meskipun masih ada beberapa perbaikan yang diperlukan. Tes memiliki konsistensi internal yang kuat dan hasil pengukurnya dapat dipercaya, seperti yang ditunjukkan oleh nilai reliabilitas yang sangat tinggi (0,97). Dilihat dari daya pembeda, sebagian besar soal masuk dalam kategori baik atau cukup. Oleh karena itu, instrumen ini relatif mampu membedakan peserta dengan kemampuan rendah dan tinggi. Meskipun demikian, distribusi tingkat kesulitan yang didominasi oleh soal mudah (94%) menunjukkan bahwa elemen tersebut belum sepenuhnya mencerminkan ciri-ciri soal HOTS yang seharusnya sulit dan menuntut penalaran berpikir yang tinggi. Dari aspek validitas, sebanyak 81% soal dinyatakan valid, yang berarti memenuhi kompetensi yang dimaksud. Sebaliknya, ada perbedaan dalam

kualitas pengecoh, yang dimana terdapat beberapa komponen berfungsi dengan baik, tetapi yang lain masih kurang efisien, sehingga diperlukan perbaikan. Dengan demikian, meskipun instrumen ini memiliki reliabilitas tinggi dan mayoritas butir valid, revisi tetap diperlukan terutama pada tingkat kesukaran, daya pembeda rendah, dan pengecoh yang tidak bekerja optimal agar tes HOTS ini benar-benar mampu mengevaluasi kemampuan peserta secara komprehensif dan akurat.

## DAFTAR REFERENSI

- Akhmadi, M. N. (2021). Analisis Butir Soal Evaluasi Tema 1 Kelas 4 SD NPlumbungan menggunakan Program Anates. *Ed-Humanistics : Jurnal Ilmu Pendidikan*, 6(1), 799–806.
- Alfiani, A. R., Mania, S., & Hasrianti, A. (2024). Analisis Kualitas Butir Soal Ulangan Akhir Semester Ganjil Pada Mata Pelajaran Matematika Siswa Kelas X Sma Negeri. *Jurnal THEOREMS (The Original Research of Mathematics)*, 8(2), 284-296.
- Anica, A., Idi, A., & Ismail, F. (2023). EVALUASI HASIL BELAJAR SISWA PADA SISTEM PEMBELAJARAN DARING SELAMA MASA PANDEMI. *TAFANI: Jurnal Pengabdian Masyarakat*, 1(1), 21-34.
- Asyarah, S. P., Febriyani, A. C., Sitanggang, C. J. C., & Hakim, L. (2025). Analisis Butir Soal Hots Pada Elemen Perpajakan Fase F SMK Akuntansi Menggunakan Software Anates. *PESHUM: Jurnal Pendidikan, Sosial Dan Humaniora*, 4(2), 1594–1602.
- 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. *AKADEMIK: Jurnal Mahasiswa Ekonomi & Bisnis*, 5(1), 250–265. <https://doi.org/10.37481/jmeb.v5i1.1165>
- Bano, V. O., Marambaawang, D. N., & Njoeroemana, Y. (2022). Analisis Kriteria Butir Soal Ujian Sekolah Mata Pelajaran IPA di SMP Negeri 1 Waingapu. *Ideas: Jurnal Pendidikan, Sosial, Dan Budaya*, 8(1), 145. <https://doi.org/10.32884/ideas.v8i1.660>
- Fatimah, S., Badarudin, K., Islam, U., Raden, N., & Sumatra, S. (2023). *Jurnal Pengabdian Kepada Masyarakat Volume 1, No. 1, Juni 2023 Open Access*: <http://jurnal.radenfatah.ac.id/index.php/TAFANI/>. 1(1), 11–20.
- Fietri, W. A. Zulyusri, & Violita.(2021). Analisis butir soal biologi kelas xi madrasah aliyah sakinah kerinci menggunakan program komputer anates 4.0 for windows. *Natural Science: Jurnal Penelitian Bidang IPA Dan Pendidikan IPA*, 7(1), 28-35.
- 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.
- Hamidah, M. H., & Wulandari, S. S. (2021). Pengembangan Instrumen Penilaian Berbasis Hots Menggunakan Aplikasi “Quizizz.” *Efisiensi : Kajian Ilmu Administrasi*, 18(1), 105–124. <https://doi.org/10.21831/efisiensi.v18i1.36997>
- Hendrayadi, H., Kustati, M., & Amelia, R. (2024). Analisis Ulangan Harian Mata Pelajaran PAI di SMA Negeri 10 Padang Tahun Pelajaran 2023/2024 (Telaah Terhadap Reliabilitas, Daya Beda Dan Tingkat Kesukaran Menggunakan Software Anates). *Jurnal Review Pendidikan dan Pengajaran (JRPP)*, 7(3), 6954-6961.
- Hidayah, N., Tureni, D., Shamdas, G. B. N., Studi, P., Biologi, P., & Tadulako, U. (2022). *Analisis Soal Ujian Biologi Berbasis HOTS (Higher Order Thinking Skill) di SMAN 9 Palu HOTS ( Higher Order Thinking Skill ) Based Biologi Examp Analysis at SMAN 9 Palu*. 10(1), 59–65.
- Ikhsaniyah, S. N., Kurnia, A. D., Zuroida, M., Pratiwi, V., & Hakim, L. (2024). *maharani*.23089. 12(2).
- Ismafitri, R., Alfan, M., & Kusumaningrum, S. R. (2022). Characteristics of HOTS (High Order Thinking Skills) and its relationship with numeracy literacy skills in primary schools. *Jurnal Riset Intervensi Pendidikan*, 4(1), 49–55.
- Karindi, D. R., & Rufi'i. (2024). Penggunaan ANATES 4.0.9 dalam menguji kelayakan soal pilihan ganda sumatif matematika. *LINEAR: Journal of Mathematics Education*, 5(2), 181–196. <https://doi.org/10.32332/2zg4f753>
- Leonardus Kaka, Vidriana Oktoviana Bano & Yohana Njoeroemana. 2024. Efektivitas Analisis Butir Soal Pilihan Ganda Menggunakan Aplikasi Anates di SMPN 2 Kanatang. *Jurnal Inovasi Pendidikan*, 4(9), 1441-1449, h
- Listiani, W., & Rachmawati. (2022). Transformasi Taksonomi Bloom dalam Evaluasi. *Jurnal Jendela Pendidikan*, 2(03), 397–402.
- Maimun dan Tzahira. (2022). Prinsip Dasar Perbankan. *Sharia Economic Law*, 1(1), 125–142.
- Mardiyah, H. (2024). *PENGGUNAAN APLIKASI ANATES PADA ANALISIS*. 11(2).
- Mawardi, M.S., Fuady A., & Sunismi 2023. Analisis Butir Soal Pilihan Ganda Menggunakan Anates pada Penilaian Tengah Semester Kelas VII D SMP Negeri 1 Ngajum Kabupaten Malang. Wahana: Tridarma Perguruan Tinggi, 75(1), 31-

- 41Padang, U. N. (2021). Natural Science: Jurnal Penelitian Bidang IPA dan Pendidikan IPA. 7(1), 28–35.
- Muhammad Afif Marta, Dimas Purnomo, & Gusmameli Gusmameli. (2024). Konsep Taksonomi Bloom dalam Desain Pembelajaran. *Lencana: Jurnal Inovasi Ilmu Pendidikan*, 3(1), 227–246. <https://doi.org/10.55606/lencana.v3i1.4572>
- Padang, U. N. (2021). *Natural Science: Jurnal Penelitian Bidang IPA dan Pendidikan IPA*. 7(1), 28–35.
- Puspasari, H., & Puspita, W. (2022). Validity Test and Reliability Instrument Research Level Knowledge and Attitude of Students Towards Elections Health Supplements in Facing Covid-19. *Jurnal Kesehatan*, 13(1), 65–71. <http://ejurnal.poltekkes-tjk.ac.id/index.php/JK>
- Putri, F. Z., Widi, M., & Sugiarto, A. (2024). Penerapan Prinsip Dasar Akuntansi Dalam Pembelajaran Operasional Perbankan Sederhana Melalui Lkpd. *Pionera| Journal Of Multi Research And Devotion*, 1(1), 36–40.
- Rahmawati, N., Qoiri, F. N. P., Cahyo, F. T., & ... (2024). ... Keselamatan Kerja Dan Lingkungan Hidup Untuk Menguji Pemahaman Siswa Smk Jurusan Akuntansi Keuangan Dan Lembaga Menggunakan Aplikasi Anates. *Hikamatzu| Journal* ..., 2(1), 418–427. <https://yasyahikamatzu.com/index.php/hjm/article/view/160%0Ahttps://yasyahikamatzu.com/index.php/hjm/article/download/160/160>
- Rianda, C. N. (2024). Prinsip dan Konsep Dasar Bank. *KIRANA : Social Science Journal*, 1(2), 100–106. <https://doi.org/10.61579/kirana.v1i2.152>
- Sabela, O. R., Krisdayanty, D., Taqqiyah, 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. <https://doi.org/10.51178/jsr.v6i1.2328>
- 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*. 4(2), 2466–2477.
- V, M. A., Hakim, L., & Pratiwi, V. (2024). *Jurnal Nirta : Studi Inovasi ANALISIS BUTIR SOAL PILIHAN GANDA PADA ELEMEN AKUNTANSI KEUANGAN GUNA MENGOPTIMASI EVALUASI* Cahyani Dwi Lestari Universitas Negeri Surabaya. 4, 1–18.

- Wiguna, S. (2021). Aplikasi anates dalam evaluasi pembelajaran.
- Zihan Marcela Hapsari, Dedi Muhtadi, & Sukirwan. (2024). Analisis Kesulitan Peserta Didik Dan Faktor-Faktor Penyebabnya Pada Materi Luas Permukaan Kerucut Dan Tabung. *UJMES (Uninus Journal of Mathematics Education and Science)*, 9(2), 055–067. <https://doi.org/10.30999/ujmes.v9i2.3203>