

**ANALISIS BUTIR SOAL HOTS PILIHAN GANDA PADA MATA PELAJARAN  
AKUNTANSI FASE F MELALUI PROGRAM KOMPUTER ANATES V4**

Annara Nathania Sabella<sup>1</sup>, Linda Puspitasari<sup>2</sup>, Shafira Aulia Rahma<sup>3</sup>, Salsabila  
Nur Faidah<sup>4</sup>, Chelsie Nur Oktavia<sup>5</sup>, Vivi Pratiwi<sup>6</sup>, Luqman Hakim<sup>7</sup>

<sup>1234567</sup>Universitas Negeri Surabaya

Alamat e-mail : [124080304008@mhs.unesa.ac.id](mailto:124080304008@mhs.unesa.ac.id),

[224080304026@mhs.unesa.ac.id](mailto:224080304026@mhs.unesa.ac.id), [324080304069@mhs.unesa.ac.id](mailto:324080304069@mhs.unesa.ac.id),

[424080304078@mhs.unesa.ac.id](mailto:424080304078@mhs.unesa.ac.id), [524080304133@mhs.unesa.ac.id](mailto:524080304133@mhs.unesa.ac.id),

[6vivipratiwi@unesa.ac.id](mailto:6vivipratiwi@unesa.ac.id), [7luqmanhakim@unesa.ac.id](mailto:7luqmanhakim@unesa.ac.id).

**ABSTRACT**

*This study evaluates the quality of Higher Order Thinking Skills (HOTS) assessment instruments for Phase F Accounting. The core problem identified is the limited teacher capacity for conducting qualitative item analysis manually, which risks compromising assessment validity. Employing a quantitative descriptive method, this research analyzes response data from 45 exam questions to evaluate their psychometric characteristics, including validity, reliability, difficulty index, discriminating power, and distractor effectiveness using Anates V.4 software. The empirical findings confirm that 76.67% of the items meet validity criteria with a high overall reliability coefficient. The majority of the items were of moderate difficulty. However, an in-depth analysis of the distractors revealed the ineffectiveness of several options, indicating a need for systematic revision. A key practical implication of this study is the validation of Anates V.4 as a technological solution for enhancing the efficiency of item bank analysis. The original contribution of this research lies in its computational analysis of HOTS items for Phase F Accounting, with a specific focus on distractor effectiveness—an analytical dimension often overlooked in similar literature. This focused approach provides a critical and replicable framework for improving the quality of HOTS-based assessments.*

*Keywords: Item Analysis, HOTS, Accounting, Anates Program*

**ABSTRAK**

Keterampilan berpikir tingkat tinggi (HOTS) merupakan kompetensi krusial dalam Kurikulum Merdeka, namun implementasinya dalam instrumen penilaian, khususnya soal pilihan ganda Akuntansi Fase F, seringkali terkendala kesulitan guru dalam menganalisis kualitas butir soal secara manual. Penelitian ini bertujuan untuk menganalisis validitas, reliabilitas, tingkat kesukaran, daya beda, dan efektivitas pengecoh pada soal HOTS mata pelajaran Akuntansi menggunakan program komputer Anates V.4. Metode yang digunakan adalah deskriptif kuantitatif dengan menganalisis data dari 45 butir soal ujian siswa. Hasil penelitian mengungkapkan bahwa 45 butir soal (76,67%) valid dan memiliki reliabilitas keseluruhan yang tinggi, dengan dominasi tingkat kesukaran sedang. Namun, ditemukan bahwa beberapa pengecoh tidak berfungsi efektif, yang mengindikasikan perlunya revisi. Secara praktis, penelitian ini membuktikan

efisiensi Anates sebagai alat bantu guru untuk merevitalisasi bank soal, sehingga dapat menghasilkan instrumen penilaian HOTS yang lebih berkualitas. Kebaruan dari penelitian ini terletak pada integrasi analisis berbasis software terhadap soal HOTS Akuntansi Fase F secara spesifik, dengan fokus mendalam pada analisis efektivitas pengecoh yang masih jarang dieksplorasi dalam konteks serupa.

Kata Kunci: Analisis Butir Soal, HOTS, Akuntansi, Program Anates

## **A. Pendahuluan**

Evaluasi hasil belajar merupakan bagian penting dalam proses pendidikan karena berfungsi untuk mengetahui sejauh mana peserta didik telah mencapai tujuan pembelajaran. Pemikiran ini juga sejalan dengan pandangan Bloom yang menempatkan evaluasi sebagai salah satu komponen penting dalam memahami kemampuan peserta didik pada berbagai ranah kognitif (Prasad, 2021)

Dalam praktik pembelajaran, tes pilihan ganda menjadi salah satu instrumen yang paling sering digunakan karena dianggap efisien, objektif, dan mampu mengukur berbagai level kemampuan secara lebih terstruktur (Quaigrain & Arhin, 2017). Meski demikian, efektivitas sebuah tes tidak hanya ditentukan oleh bentuknya saja, tetapi sangat bergantung pada kualitas setiap butir (Kumar et al., 2021). Arikunto menekankan bahwa analisis butir soal diperlukan untuk mengidentifikasi

soal-soal yang termasuk kategori baik dan kurang baik sehingga instrumen dapat diperbaiki sebelum digunakan. Pemikiran serupa juga dikemukakan oleh Purwanto yang menjelaskan bahwa instrumen penilaian harus memenuhi aspek-aspek seperti tingkat kesukaran, daya pembeda, serta konsistensi hasil pengukuran.

Hasil penelitian sebelumnya mendukung pentingnya proses analisis ini. (Anshari et al., 2024) menunjukkan bahwa uji validitas dan reliabilitas membantu memastikan bahwa instrumen benar-benar mengukur kompetensi yang dituju. Sementara itu, (Wahndianti & Sumarsih, 2018) menemukan bahwa analisis butir soal dapat memberikan informasi penting bagi guru dalam menentukan efektivitas evaluasi yang digunakan di kelas. Studi lain oleh (Nurjanah, 2017) juga mengungkapkan bahwa analisis karakteristik butir soal merupakan langkah penting dalam meningkatkan kualitas evaluasi pembelajaran sehingga hasil penilaian

lebih akurat dan dapat dipertanggungjawabkan.

Berdasarkan pemikiran tersebut, penelitian ini dilakukan untuk menganalisis kualitas soal pilihan ganda yang disusun oleh peneliti. Soal-soal tersebut dirancang berdasarkan Taksonomi Bloom revisi, dengan fokus pada kemampuan berpikir tingkat tinggi atau *Higher Order Thinking Skills (HOTS)*, terutama pada level C4 (menganalisis), C5 (mengevaluasi), dan C6 (mencipta) (Erdiana & Panjaitan, 2023). Pemilihan level ini didasari oleh kebutuhan dunia pendidikan modern yang menekankan pentingnya kemampuan berpikir kritis, analitis, dan kreatif pada peserta didik agar siap menghadapi tantangan abad ke-21.

Instrumen yang telah disusun kemudian disebarakan secara daring melalui Google Form kepada 25 responden yang terdiri atas siswa SMK Jurusan Akuntansi dan mahasiswa semester satu Program Studi Pendidikan Akuntansi. Pemilihan dua kelompok responden ini bertujuan untuk memperoleh gambaran yang lebih komprehensif mengenai variasi kemampuan dan

pemahaman antara peserta didik di tingkat menengah dan perguruan tinggi. Data yang diperoleh kemudian dianalisis menggunakan program ANATES untuk menilai kualitas setiap butir soal berdasarkan lima komponen utama, yaitu validitas, reliabilitas, tingkat kesukaran, daya pembeda, serta efektivitas pengecoh (Sagala & Andriani, 2019).

Melalui analisis ini, diharapkan dapat diperoleh informasi yang akurat mengenai kualitas setiap butir soal yang telah disusun. Hasil penelitian ini diharapkan dapat menjadi acuan dalam perbaikan instrumen tes selanjutnya, serta memberikan kontribusi bagi pendidik dalam menyusun alat evaluasi yang tidak hanya berfungsi menilai kemampuan dasar peserta didik, tetapi juga mendorong pengembangan kemampuan berpikir tingkat tinggi yang relevan dengan tuntutan pembelajaran masa kini (Utama et al., 2020).

## **B. Metode Penelitian**

Penelitian ini menerapkan penelitian kuantitatif dengan pendekatan deskriptif menggunakan alat analisis bernama Anates V4.

Penelitian deskriptif kuantitatif adalah pendekatan metodologis yang digunakan untuk mengamati dan menganalisis fenomena sebagaimana terjadi dalam keadaan alaminya. Pendekatan ini menggambarkan dan menjelaskan objek penelitian dengan memproses data secara numerik, memungkinkan peneliti untuk menarik kesimpulan didasarkan pada pengukuran gejala atau fenomena kuantitatif (Ratih et al., 2025). Terdapat beberapa tahapan dalam penelitian ini adalah: Mengumpulkan jawaban dari tes yang telah dikerjakan oleh peserta didik berupa pilihan ganda. Memasukan jawaban peserta didik pada aplikasi Anates V4, sehingga menghasilkan validitas, reliabilitas, daya pembeda, tingkat kesukaran, kualitas pengecoh, dan korelasi skor butir dengan skor total pada butir soal. Menganalisis hasil data sehingga data dapat dimaknai, kemudian menjadi dasar penyusunan hasil penelitian. Hasil penelitian dalam penelitian memberikan saran perbaikan terhadap soal - soal yang belum memenuhi kriteria, sehingga dapat membantu dalam pengembangan instrumen evaluasi yang valid.

### **C.Hasil Penelitian dan Pembahasan**

Berdasarkan hasil analisis melalui aplikasi Anates V4, penelitian ini memberikan identifikasi kualitas butir soal sebanyak 45 soal dengan jumlah responden sebanyak 25 peserta didik pada bangku SMA/SMK kelas dua belas dan mahasiswa pendidikan akuntansi atau akuntansi angkatan 2025. Diperoleh kesimpulan atas analisis yang telah dilakukan mulai dari uji reliabilitas, kelompok unggul dan asor, daya pembeda, tingkat kesukaran, kualitas pengecoh, dan kualitas butir soal (Andrianik Susilowati et al., 2025).

### **Uji Reliabilitas**

Analisis reliabilitas tes dilakukan untuk melihat apakah soal-soal yang digunakan dapat menghasilkan skor yang ajeg dan tidak dipengaruhi oleh faktor kebetulan. Reliabilitas penting karena menunjukkan seberapa dapat dipercaya hasil tes dalam menggambarkan kemampuan siswa yang sebenarnya. Jika suatu tes reliabel, maka hasilnya akan konsisten ketika diberikan pada kelompok siswa dengan tingkat kemampuan yang sama (Mamolo, 2021).

Berdasarkan hasil perhitungan menggunakan aplikasi ANATES, diperoleh nilai rata-rata skor peserta sebesar 32,60 dengan simpangan baku 10,97. Simpangan baku ini menunjukkan adanya tingkat variasi skor yang cukup lebar antar peserta tes. Variasi yang baik mencerminkan bahwa soal mampu mendistribusikan skor peserta didik secara proporsional, sehingga dapat membedakan siswa yang memiliki tingkat penguasaan materi tinggi, sedang, maupun rendah. Dengan kata lain, instrumen tidak bersifat terlalu mudah maupun terlalu homogen.

**Tabel 1 Kriteria Tingkat Reliabilitas**

| Koefisien Nilai Reliabilitas | Kriteria Tingkat Reliabilitas |
|------------------------------|-------------------------------|
| 0,00 – 0,20                  | Korelasi Sangat Rendah        |
| 0,21 – 0,40                  | Korelasi Rendah               |
| 0,41 – 0,70                  | Korelasi Cukup                |
| 0,71 – 0,90                  | Korelasi Tinggi               |
| 0,91 – 1,00                  | Korelasi Sangat Tinggi        |

Nilai Korelasi XY sebesar 0,94 menunjukkan hubungan yang sangat kuat antar butir tes. Berdasarkan tabel kriteria tingkat reliabilitas, nilai ini termasuk kategori “Korelasi Sangat Tinggi.” Selanjutnya, nilai reliabilitas tes yang diperoleh adalah 0,97, yang

juga berada dalam kategori “Reliabilitas Sangat Tinggi.” Tingginya nilai ini mengindikasikan bahwa keseluruhan butir soal bekerja secara konsisten dan memiliki kestabilan hasil yang sangat baik ketika diguna untuk mengukur kemampuan siswa.

Dengan mempertimbangkan tingginya reliabilitas yang diperoleh serta penyebaran skor yang memadai, dapat disimpulkan bahwa instrumen tes akuntansi perusahaan jasa, dagang, dan manufaktur yang disusun memiliki kualitas yang sangat baik dari sisi reliabilitas. Instrumen ini mampu mengukur kemampuan peserta didik secara konsisten, sehingga dapat dinyatakan layak digunakan sebagai alat evaluasi pembelajaran.

### Kelompok Unggul

**Tabel 2 Hasil Kelompok Unggul**

| No. Urut | No. Subyek | Kode/Nama Subyek | Skor |
|----------|------------|------------------|------|
| 1        | 17         | Jasmine          | 43   |
| 2        | 18         | Ayu Susanti      | 43   |
| 3        | 19         | Maharani         | 43   |
| 4        | 20         | Dimas Migi J     | 43   |
| 5        | 22         | Jelita Putri     | 43   |
| 6        | 23         | Lieza Puspa      | 43   |

|                                   |    |                  |    |   |    |                      |    |
|-----------------------------------|----|------------------|----|---|----|----------------------|----|
| 7                                 | 24 | Syakhis<br>Zahir | 43 | 2 | 11 | Rintan<br>Salsa      | 21 |
| <b>Sumber:</b> Data diolah (2025) |    |                  |    | 3 | 13 | Evanesia A.<br>S.    | 21 |
|                                   |    |                  |    | 4 | 2  | Shifadilla R.        | 18 |
|                                   |    |                  |    | 5 | 5  | Ramadhani<br>Syahnur | 13 |
|                                   |    |                  |    | 6 | 12 | Dina                 | 13 |
|                                   |    |                  |    | 7 | 3  | Fathur<br>Rozaki     | 13 |

**Sumber:** Data diolah (2025)

Berdasarkan tampilan hasil untuk kelompok unggul, terlihat bahwa seluruh subjek dalam kelompok ini konsisten menjawab benar pada hampir seluruh nomor soal, mulai dari nomor 1 hingga nomor 45. Setiap subjek yang memiliki skor tinggi (skor 43) menunjukkan pola jawaban benar yang sama. Hal ini menunjukkan bahwa soal-soal tersebut dengan baik dapat dijawab oleh peserta berkemampuan tinggi, sehingga mencerminkan bahwa instrumen tes berada dalam tingkat kesesuaian yang tepat terhadap kemampuan kelompok unggul. Kondisi ini mengindikasikan bahwa sebagian besar soal memiliki tingkat kesukaran rendah bagi kelompok berkemampuan tinggi, atau dengan kata lain, soal-soal tersebut berada pada kategori mudah hingga sedang untuk kelompok unggul.

### Kelompok Asor

**Tabel 3 Hasil Kelompok Asor**

| No.<br>Urut | No.<br>Subyek | Kode>Nama<br>Subyek | Skor |
|-------------|---------------|---------------------|------|
| 1           | 14            | Azzahra<br>Keisha   | 23   |

Kelompok asor terdiri dari tujuh peserta dengan skor terendah, yang masing-masing memperoleh skor jauh di bawah rata-rata kelas. Pola jawaban kelompok ini menunjukkan bahwa mereka mengalami kesulitan signifikan dalam menjawab sebagian besar butir soal, terutama pada butir yang memiliki tingkat kesukaran sedang hingga sukar. Banyak jawaban mereka jatuh pada distraktor yang kurang tepat, dan terdapat pula beberapa nomor yang tidak dijawab sama sekali. Hal ini menunjukkan bahwa penguasaan materi pada kelompok asor masih rendah, serta kemungkinan kurang memahami konsep-konsep dasar yang diukur oleh tes. Dari sudut pandang analisis butir, kinerja kelompok asor membantu memperlihatkan fungsi diskriminatif dari soal yang diberikan. Pada sebagian besar butir, kelompok

unggul menjawab benar sementara kelompok asor menjawab salah, menandakan bahwa butir tersebut memiliki daya pembeda yang baik. Secara keseluruhan, analisis kelompok asor memberikan gambaran penting mengenai kelemahan pemahaman siswa sekaligus membantu pihak guru mengevaluasi kualitas soal yang digunakan.

### **Daya Pembeda Soal**

Daya Pembeda adalah indeks yang mengukur sejauh mana sebuah soal dapat membedakan antara peserta didik yang pandai (kelompok tinggi) dengan peserta didik yang kurang pandai (kelompok rendah) dalam kemampuan yang diukur. Dalam konteks soal HOTS (Higher Order Thinking Skills) Akuntansi, daya pembeda tidak hanya mengukur kemampuan menghafal, tetapi lebih pada kemampuan analisis, evaluasi, dan mencipta dalam konteks akuntansi (Hawk, 2024).

ANATES V4 umumnya menggunakan metode korelasi point-biserial atau pembagian kelompok atas-bawah (27% kelompok tinggi vs 27% kelompok rendah). Program ini secara otomatis mengelompokkan

responden berdasarkan skor total, kemudian menganalisis setiap butir soalnya.

**Tabel 4 Interpretasi Koefisien Daya Beda**

| Indeks Daya | Kriteria                             |
|-------------|--------------------------------------|
| 0,00 – 0,20 | Jelek (Soal Direvisi)                |
| 0,20 – 0,40 | Cukup (Soal Diterima dan Diperbaiki) |
| 0,40 – 0,70 | Baik (Soal Diterima)                 |
| 0,70 – 1,00 | Sangat Baik (Soal Diterima)          |
| Negatif     | Semuanya Tidak Baik (Soal Dibuang)   |

Berdasarkan hasil analisis daya beda butir soal, dapat disimpulkan bahwa kualitas instrumen evaluasi secara keseluruhan berada dalam kategori sangat baik. Sebanyak 84,3% butir soal, yang terdiri dari 17 soal berkategori baik dan 21 soal berkategori sangat baik, telah berhasil membedakan dengan jelas antara peserta didik yang berkemampuan tinggi dan rendah. Meskipun demikian, masih terdapat ruang untuk perbaikan, yaitu pada 5 butir soal (3 soal berkategori cukup dan 2 soal berkategori jelek) yang perlu direvisi, serta 2 butir soal yang disarankan untuk dibuang karena dinilai tidak efektif sama sekali. Berdasarkan data pada tabel 5 dapat diketahui bahwa mayoritas soal sudah layak

dipertahankan, sementara sebagian kecil lainnya memerlukan penyempurnaan untuk meningkatkan validitas tes secara keseluruhan pada pelaksanaan di masa yang akan datang.

**Tabel 5 Hasil Analisis Daya Pembeda**

| No | Tingkat Daya Beda                     | Jumlah | Prese-ntase | No. Butir Soal                                                                |
|----|---------------------------------------|--------|-------------|-------------------------------------------------------------------------------|
| 1  | Jelek (Soal Direvisi)                 | 2      | 4,4%        | 15, 22                                                                        |
| 2  | Cukup (Soal Diterima dan Diperbaiki ) | 3      | 6,6%        | 3, 40, 43                                                                     |
| 3  | Baik (Soal Diterima)                  | 17     | 37,7%       | 1, 2, 5, 10, 11, 12, 13, 14, 16, 17, 18, 27, 28, 9, 32, 37, 44, 45            |
| 4  | Sangat Baik (Soal Diterima)           | 21     | 46,6%       | 4, 6, 7, 8, 9, 10, 19, 17, 21, 23, 24, 25, 26, 30, 31, 33, 34, 35, 36, 28, 42 |
| 5  | Semuanya Tidak Baik (Soal Dibuang)    | 2      | 4,4%        | 39, 41                                                                        |

**Sumber:** Data diolah (2025)

### Tingkat Kesukaran

Di dalam analisis butir soal, tingkat kesukaran menjadi instrumen penting untuk menunjukkan kualitas menggambarkan seberapa mudah dan sulit suatu butir soal yang dipahami oleh peserta didik . Soal yang memiliki tingkat kesukaran sedang dianggap paling ideal karena mampu membedakan kemampuan peserta didik dengan jelas (Sabrina et al., 2024). Sebaliknya, soal yang terlalu mudah atau terlalu sulit kurang tepat untuk menggambarkan kemampuan peserta didik secara akurat (Karindi & Ruffi'i, 2024). Pada aplikasi Anates V4 mengklasifikasikan tingkat kesukaran dari butir soal ke dalam tiga kriteria sebagai berikut:

**Tabel 6 Indeks Tingkat Kesukaran**

| Indeks Kesukaran   | Kriteria |
|--------------------|----------|
| $0,00 < TK < 0,30$ | Sukar    |
| $0,31 < TK < 0,70$ | Sedang   |
| $0,71 < TK < 1,00$ | Mudah    |

Berikut adalah hasil analisis butir soal menggunakan aplikasi Anates V4:

**Tabel 7 Hasil Tingkat Kesukaran Butir Soal**

| No | Kate-<br>gori<br>Su-<br>kar | Nomor<br>Soal | Banyak<br>Soal | Pre-<br>Senta-<br>se |
|----|-----------------------------|---------------|----------------|----------------------|
|----|-----------------------------|---------------|----------------|----------------------|

|   |              |                                                                                           |    |         |
|---|--------------|-------------------------------------------------------------------------------------------|----|---------|
| 1 | Sangat Mudah | 1, 2, 3, 15, 22, 29, 40, 43                                                               | 8  | 17,78 % |
| 2 | Mudah        | 5, 10, 11, 12, 13, 14, 16, 17, 21, 23, 25, 27, 28, 30, 31, 32, 33, 34, 36, 37, 42, 44, 45 | 23 | 51,11 % |
| 3 | Sedang       | 4, 6, 7, 8, 9, 18, 19, 20, 24, 26, 35, 38, 41                                             | 13 | 28,89 % |
| 4 | Sangat Sukar | 39                                                                                        | 1  | 2,22%   |

**Sumber:** Data diolah (2025)

Dari analisis butir soal dengan elemen Akuntansi Perusahaan Jasa, Dagang, dan Manufaktur Fase F menggunakan aplikasi Anates V4, menunjukkan tingkat kesukaran mayoritas dalam tingkat mudah (Augustia et al., 2025). Butir soal dianggap baik jika tidak terlalu sukar atau terlalu mudah (Karindi & Ruffi, 2024). Pada hasil analisis tingkat kesukaran pada soal Akuntansi Perusahaan Jasa, Dagang, dan Manufaktur berbentuk pilihan ganda menunjukkan soal dominan dengan tingkat kesukaran mudah sebanyak

51,11%, yaitu nomor 5, 10, 11, 12, 13, 14, 15, 16, 17, 21, 23, 25, 27, 28, 30, 31, 32, 33, 34, 36, 37, 42, 44, 45, maka masih dianggap kurang baik, tetapi terdapat 28,89% soal menunjukkan tingkat kesukaran sedang, dianggap memiliki kualitas yang baik yaitu nomor 4, 6, 7, 8, 9, 18, 19, 20, 24, 26, 35, 38, 41. Dibandingkan dengan penelitian terdahulu "Analisis Butir Soal Pilihan Ganda pada Siswa SMK Menggunakan Anates, hasilnya masih terlalu mudah dan perlu direvisi supaya memiliki variasi tingkat kesukaran" (Ikawaty et al., 2024). Pada hasil analisis kami, hasil menunjukkan bahwasannya terdapat variasi tingkat kesukaran yang menunjukkan soal tidak terlalu mudah dan tidak terlalu sukar sehingga dapat dikatakan baik (Puteri et al., 2025).

Di dalam dunia pendidikan, menilai tingkat kesulitan secara menyeluruh sangatlah penting (Buana et al., 2024). Soal yang terlalu sederhana berisiko menurunkan minat belajar peserta didik dan menghambat pengembangan kemampuan berpikir kritis serta analitis. Ini menjadi sangat penting terutama dalam pembelajaran akuntansi, yang menuntut kecakapan

berpikir yang kompleks dan kemampuan beradaptasi tinggi (Ikawaty et al., 2024).

### **Kualitas Pengecoh**

Efektivitas distraktor menggambarkan pola bagaimana peserta tes memilih opsi jawaban pada setiap butir soal. Pola ini menunjukkan sejauh mana pilihan jawaban salah berfungsi menarik peserta untuk memilih opsi yang keliru namun tetap masuk akal. Distraktor yang efektif berperan penting dalam membedakan peserta didik yang benar-benar memahami materi dan yang belum menguasainya. Ketika pengecoh bekerja dengan baik, daya pembeda soal meningkat dan tingkat kesukaran soal cenderung lebih tinggi, sedangkan pengecoh yang tidak berfungsi menyebabkan daya pembeda rendah dan soal menjadi terlalu mudah. Oleh karena itu, analisis efektivitas pengecoh dilakukan untuk mengetahui apakah opsi salah sudah cukup mirip dengan jawaban benar sehingga dapat menjerumuskan peserta didik berkemampuan rendah.

Setiap distractor idealnya dipilih oleh sebagian peserta, terutama oleh peserta berkemampuan rendah,

sedangkan peserta berkemampuan tinggi seharusnya memilih distractor dalam jumlah lebih sedikit. Jika suatu distractor hanya dipilih satu orang, maka pemilih tersebut seharusnya berasal dari kelompok peserta berkemampuan rendah. Distraktor yang tidak dipilih sama sekali dikategorikan tidak efektif dan perlu diperbaiki atau dihapus (Putri & Rosliyah, 2020). Pengecoh yang kurang efektif biasanya muncul pada soal dengan opsi jawaban yang terlalu mirip atau justru tidak relevan dengan materi sehingga peserta didik yang memahami konsep dapat langsung mengabaikannya. Untuk meningkatkan kualitas pengecoh, opsi salah harus dirancang sedemikian rupa agar tetap logis, relevan, dan secara konseptual mendekati jawaban benar tanpa kehilangan sifat keliru. Dalam perangkat lunak Anates, kualitas pengecoh dinilai melalui simbol tertentu, yaitu;

\*\* : kunci jawaban

++ : sangat baik

+ : baik

- : kurang baik

- - : buruk

- - - : sangat buruk

Berikut ini merupakan hasil dari analisis pada soal pilihan ganda menggunakan Anates:

**Tabel 8 Hasil Kualitas Pengecoh**

| No | Kategori Kualitas Pengecoh | Nomor Soal                                                                                                         | Banyak Soal | Pre-sentase |
|----|----------------------------|--------------------------------------------------------------------------------------------------------------------|-------------|-------------|
| 1  | Sangat Baik                | 5, 7, 8, 9, 10, 11, 12, 13, 16, 17, 18, 20, 21, 23, 24, 26, 27, 28, 30, 32, 32, 33, 34, 35, 36, 37, 38, 41, 42, 44 | 2           | 64,44%      |
| 2  | Baik                       | 1, 3, 4, 6, 14, 15, 19, 22, 25, 29, 31, 39, 40, 43, 45                                                             | 1           | 33,33%      |
| 3  | Kurang Baik                | 2                                                                                                                  | 1           | 2,22%       |
| 4  | Buruk                      | 0                                                                                                                  | 0           | 0%          |
| 5  | Sangat Buruk               | 0                                                                                                                  | 0           | 0%          |

**Sumber:** Data diolah (2025)

Berdasarkan hasil analisis terhadap 45 butir soal menggunakan Anates, ditemukan bahwa sebagian besar pengecoh telah berfungsi secara efektif. Sebanyak 29 butir soal (64,44%) berada pada kategori sangat baik, menunjukkan bahwa distraktor mampu bekerja optimal dalam mengarahkan peserta berkemampuan rendah untuk memilih jawaban yang salah. Sebanyak 15 butir soal (33,33%) berada pada kategori baik, menandakan bahwa pengecohnya tetap efektif meskipun tidak sekuat kategori sebelumnya. Hanya satu soal (2,22%) yang termasuk kategori cukup baik dan memerlukan perbaikan. Tidak ditemukan soal dengan pengecoh berkategori kurang baik maupun tidak baik, sehingga seluruh soal memiliki distraktor yang berfungsi sebagaimana mestinya. Temuan ini menunjukkan bahwa kualitas pengecoh pada instrumen tes yang dianalisis tergolong sangat baik dan layak digunakan sebagai alat evaluasi pembelajaran.

### **E. Kesimpulan**

Berdasarkan hasil analisis butir soal menggunakan ANATES, dapat disimpulkan bahwa tes memiliki

kualitas yang sangat baik dengan tingkat reliabilitas sebesar 0,97, yang menunjukkan bahwa instrumen tersebut konsisten dan dapat dipercaya dalam mengukur kemampuan peserta didik. Rata-rata skor peserta adalah 32,60 dari 45 butir soal, dengan simpang baku 10,97 yang mengindikasikan adanya variasi kemampuan antar siswa. Pembagian kelompok unggul dan asor juga memperlihatkan perbedaan yang jelas antara peserta berkemampuan tinggi dan rendah, sehingga memungkinkan identifikasi butir soal yang mampu membedakan kedua kelompok tersebut dengan baik.

Dari segi karakteristik butir, sebagian besar soal berada pada kategori mudah hingga sangat mudah, sedangkan beberapa soal berada pada kategori sedang, dan hanya satu soal yang masuk kategori sangat sukar (butir 39). Daya pembeda mayoritas soal menunjukkan kualitas yang baik, dengan banyak butir memiliki indeks cukup hingga sangat baik, meskipun terdapat dua soal yang memiliki daya pembeda negatif (butir 39 dan 41) sehingga tidak berfungsi dengan baik dalam membedakan kemampuan siswa. Korelasi butir

dengan skor total sebagian besar signifikan, menandakan bahwa butir-butir tersebut relevan dengan kompetensi yang diukur. Namun terdapat beberapa butir dengan korelasi rendah atau tidak signifikan sehingga perlu direvisi. Secara keseluruhan, tes dapat dinyatakan valid dan layak digunakan meskipun beberapa butir perlu diperbaiki atau dibuang untuk meningkatkan kualitas instrumen secara keseluruhan.

#### **DAFTAR PUSTAKA**

- Andrianik Susilowati, D., Saputri, N., Ni, K., Hakim, L., & Pratiwi, V. (2025). Penerapan Aplikasi Anates Untuk Analisis Soal HOTS Dalam Evaluasi Pembelajaran Spreadsheet Di SMKN 1 Kendal. *Jurnal Pendidikan, Sosial dan Humaniora*, 4(2).
- Anshari, M. I., Nasution, R., Irsyad, M., Alifa, A. Z., & Zuhriyah, I. A. (2024). Analisis Validitas dan Reliabilitas Butir Soal Sumatif Akhir Semester Ganjil Mata Pelajaran PAI. *Edukatif: Jurnal Ilmu Pendidikan*, 6(1), 964–975. <https://doi.org/10.31004/edukatif.v6i1.5931>
- 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/jmeh.v5i1.1165>
- Buana, A. C., Azhari, C. A., Hakim, L., & Pratiwi, V. (2024). Analisis Hasil Uji Coba Soal Pilihan Ganda Pada Pelajaran Akuntansi dan Keuangan Lembaga Menggunakan Anates. *Jurnal Pendidikan Tambusai*, 8(3), 48111–48118.
- Dewi Ratih, P., Aulia Pristanti, V., Hakim, L., Pratiwi, V., Akuntansi, P., & Negeri Surabaya, U. (2025). Analisis Soal Pilihan Ganda HOTS Menggunakan Anates 4.0 Pada Etika Profesi Akuntansi SMK Kelas X. *Jurnal Pendidikan Tambusai*, 9.
- Dwi Rahmadhani, A., Andriani Ma, N., Hakim, L., & Pratiwi, V. (2024). Penerapan Aplikasi Anates Dalam Menganalisis Kelayakan Butir Soal Pilihan Ganda Pada Elemen Etika Profesi Akuntansi. *Jurnal Nirta : Studi Inovasi*, 4(1), 477–497. <https://ejournal.nlc-education.or.id/>
- Erdiana, N., & Panjaitan, S. (2023). How is HOTS Integrated into the Indonesian High School English Textbook? *Studies in English Language and Education*, 10(1), 60–77.  
<https://doi.org/10.24815/siele.v10i1.26052>
- Hawk, H. (2024). Unlocking the potential: enhancing higher-order thinking skills in accounting education. *Accounting Education*, 1–28.  
<https://doi.org/10.1080/09639284.2024.2375600>
- Ikawaty, D. S., Agustin, E. W., Hakim, L., & Pratiwi, V. (2024). Analisis Butir Soal Pilihan Ganda Menggunakan Anates. *Edukatif: Jurnal Ilmu Pendidikan*, 6(6), 6773–6781.
- Jatmoko, D., Pambudi, A., Alief Setiyono, K., & Nugroho, B. (2025). Analisis Butir Soal Teknik Dasar Otomotif Berbasis Diagnostik Kognitif dengan Anates Versi 4.02. *Jurnal Riset Rumpun Ilmu Pendidikan*, 4(1), 461–471.  
<https://doi.org/10.55606/jurripen.v4i1.4812>
- Karindi, D. R., & Rofi'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>
- Kumar, D., Jaipurkar, R., Shekhar, A., Sikri, G., & Srinivas, V. (2021). Item analysis of multiple choice questions: A quality assurance test for an assessment tool. *Medical Journal Armed Forces India*, 77, S85–S89.  
<https://doi.org/10.1016/j.mjafi.2020.11.007>

- Magdalena, I., Fauzi, H. N., & Putri, R. (2020). PENTINGNYA EVALUASI DALAM PEMBELAJARAN DAN AKIBAT MEMANIPULASINYA. *Jurnal Pendidikan Dan Sains*, 2(2), 244–257.  
<https://ejournal.stitpn.ac.id/index.php/bintang>
- Mamolo, L. A. (2021). Development of an Achievement Test to Measure Students' Competency in General Mathematics. *Anatolian Journal of Education*, 6(1), 79–90.  
<https://doi.org/10.29333/aje.2021.616a>
- Nurjanah, N. (2017). Analisis Butir Soal Pilihan Ganda Dari Aspek Kebahasaan. *Faktor Jurnal Ilmiah Kependidikan*, 2(1), 69–78.
- Prasad, G. N. R. (2021). Evaluating student performance based on bloom's taxonomy levels. *Journal of Physics: Conference Series*, 1797(1).  
<https://doi.org/10.1088/1742-6596/1797/1/012063>
- 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>
- Putri, R. A., & Rosliyah, Y. (2020). Analisis Distraktor Butir Soal Bunpou Shokyuu Kouhan Semester 2 Tahun 2018/2019 Prodi Pendidikan Bahasa Jepang UNNES. *Kiryoku*, 4(1), 18–25.  
<https://doi.org/10.14710/kiryoku.v4i1.18-25>
- Quaigrain, K., & Arhin, A. K. (2017). Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. *Cogent Education*, 4(1), 1301013.  
<https://doi.org/10.1080/2331186X.2017.1301013>
- Sabrina, M., Hemi, D., Hildayanti, S., Hakim, L., & Pratiwi, V. (2024). PENERAPAN APLIKASI ANATES DALAM ANALISIS SOAL PILIHAN GANDA DI SMKN 2 TUBAN. *Jurnal Pendidikan Ilmiah*.
- Sagala, P. N., & Andriani, A. (2019). Development of Higher-Order Thinking Skills (HOTS) Questions of Probability Theory Subject Based on Bloom's Taxonomy. *Journal of Physics: Conference Series*, 1188(1).  
<https://doi.org/10.1088/1742-6596/1188/1/012025>
- Syafiriya, D. M., Rahmawati, P. A., Qolbi, Z., Hakim, L., & Pratiwi, V. (2024). Analisis Butir Soal HOTS untuk Menguji Pemahaman Siswa SMK Jurusan Akuntansi Menggunakan Software Anates. *Jurnal Nirta : Studi Inovasi*, 4(1).

- Tasrif. (2022). Higher Order Thinking Skills (HOTS) dalam pembelajaran social studies di sekolah menengah atas. *Jurnal Pembangunan Pendidikan: Fondasi Dan Aplikasi*, 10(1).  
<https://doi.org/10.21831/jppfa.v10i1.29347>
- Utama, C., Sajidan, Nurkamto, J., & Wiranto. (2020). The instrument development to measure higher-order thinking skills for pre-service biology teacher. *International Journal of Instruction*, 13(4), 833–848.  
<https://doi.org/10.29333/iji.2020.13451a>
- Wahndianti, S., & Sumarsih. (2018). Analisis Butir Soal Ujian Akhir Semester Gasal Mata Pelajaran Hitung Dagang. *Kajian Pendidikan Akuntansi Indonesia*, 11(4), 1–14.