

# EVALUASI KUALITAS SOAL HOTS DALAM MENGIKUR PEMIKIRAN KRITIS SISWA SMK PADA MATA PELAJARAN AKUNTANSI DASAR MENGGUNAKAN APLIKASI ANATES

Ninis Cantika Zalianti<sup>1</sup>, Sinta Mei Agustin<sup>2</sup>, Mutia Az Zahra<sup>3</sup>, Najwa Asilah Safitri<sup>4</sup>, Elen Natasya Putri<sup>5</sup>,  
Luqman Hakim<sup>6</sup>, Vivi Pratiwi<sup>7</sup>  
Universitas Negeri Surabaya  
e-mail : [24080304102@mhs.unesa.ac.id](mailto:24080304102@mhs.unesa.ac.id), [24080304154@mhs.unesa.ac.id](mailto:24080304154@mhs.unesa.ac.id),  
[24080304034@mhs.unesa.ac.id](mailto:24080304034@mhs.unesa.ac.id), [24080304107@mhs.unesa.ac.id](mailto:24080304107@mhs.unesa.ac.id), [24080304029@mhs.unesa.ac.id](mailto:24080304029@mhs.unesa.ac.id),  
[luqmanhakim@unesa.ac.id](mailto:luqmanhakim@unesa.ac.id), [vivipratiwi@unesa.ac.id](mailto:vivipratiwi@unesa.ac.id)

## **Abstrak**

*Penelitian deskriptif kuantitatif ini bertujuan untuk mengevaluasi kualitas instrumen tes kemampuan Higher Order Thinking Skills (HOTS) pada materi Elemen Dasar Akuntansi dan Dasar Perbankan Fase E di SMK. Metode penelitian yang digunakan adalah analisis kuantitatif terhadap butir soal melalui uji coba instrumen. Instrumen berupa 45 soal pilihan ganda yang diuji cobakan kepada 20 siswa kompetensi keahlian Akuntansi. Pengumpulan data dilakukan melalui Google Form, kemudian dianalisis secara empiris menggunakan aplikasi ANATES versi 4.0. Analisis meliputi validitas butir (korelasi point biserial), reliabilitas instrumen (KR-20), daya pembeda, tingkat kesukaran, efektivitas pengecoh, serta hubungan skor butir dengan skor total. Hasil penelitian menunjukkan bahwa instrumen memiliki reliabilitas baik dengan koefisien 0,77. Pada tingkat butir, sebanyak 27 soal dinyatakan layak digunakan, 14 soal perlu direvisi karena kelemahan pada daya pembeda, tingkat kesukaran, atau validitas, dan 4 soal harus dibuang karena memiliki daya pembeda negatif atau korelasi negatif dengan skor total. Dengan demikian, meskipun kualitas instrumen secara umum baik, diperlukan revisi terarah agar instrumen semakin akurat dalam mengukur kompetensi berpikir tingkat tinggi siswa.*

*Kata kunci:* analisis butir soal, hots, evaluasi instrumen, anates, SMK

## **Abstract**

*This quantitative descriptive research aims to evaluate the quality of test instruments designed to measure Higher Order Thinking Skills (HOTS) in Basic Accounting Elements and Basic Banking for Phase E in Vocational High Schools. The study employed a quantitative item-analysis method based on instrument try-out data. The instrument consisted of 45 multiple-choice items administered to 20 Accounting students. Data were collected through Google Form and analyzed using ANATES version 4.0, focusing on item validity (point biserial correlation), instrument reliability (KR-20), discrimination power, difficulty level, distractor effectiveness, and the correlation between item score and total score. The results indicate that the instrument has good reliability, with a coefficient of 0.77. At the item level, 27 questions were categorized as acceptable, 14 items required revision due to issues in discrimination, difficulty, or validity, and 4 items were discarded due to negative discrimination or negative correlation with the total score. In conclusion, although the overall instrument meets good quality standards, targeted improvements are necessary to enhance its accuracy and consistency in measuring students' higher-order competencies.*

*Keywords:* item analysis, hots, instrument evaluation, anates, vocational education

## **1. PENDAHULUAN**

Penilaian merupakan tahap untuk merancang, mengumpulkan, dan menyajikan informasi yang diperlukan untuk membuat keputusan. Informasi yang diperoleh harus merupakan data yang relevan dan mendukung tujuan yang telah ditentukan. Proses penilaian ini dapat memberikan gambaran mengenai hasil belajar siswa yang dinilai dari pencapaian target pembelajaran di kelas.

Dalam proses penilaian, diperlukan informasi mengenai aspek yang sedang dinilai. Pada aktivitas pembelajaran, data yang dimaksud mencakup perilaku atau kinerja siswa selama mengikuti kegiatan belajar, serta nilai dari ulangan dan ujian

akhir semester. Untuk mengevaluasi hasil belajar, penting untuk menggunakan metode penilaian yang sesuai dengan tipe tes atau instrumen penilaian yang akan dipakai guna menilai hasil pengajaran, serta mengacu pada materi dan metode pengajaran yang diterapkan, serta tujuan pengajaran yang telah ditetapkan.

Dalam sektor pendidikan, evaluasi yang akurat bisa mendukung pengukuran efektivitas metode pengajaran yang diterapkan dan memberikan informasi yang diperlukan untuk memperbaiki proses pembelajaran di masa depan (Amir, 2015). Salah satu jenis evaluasi yang umum dilakukan di sekolah adalah tes, yang mencakup berbagai kategori soal seperti pilihan ganda, tulisan, atau praktik. Namun, untuk mencapai hasil evaluasi yang optimal, soal yang disusun harus memenuhi kriteria kualitas yang spesifik dan dapat diukur.

Tes adalah cara yang digunakan untuk menilai dan mengumpulkan data mengenai sifat suatu objek. Djemali menyatakan bahwa tes merupakan cara untuk mengevaluasi kemampuan seseorang yang terlihat secara tidak langsung melalui reaksi orang tersebut terhadap rangsangan dan pertanyaan. Tes dapat dipahami sebagai kumpulan pertanyaan yang dijawab dengan maksud untuk menilai sejauh mana kemampuan individu atau mengungkapkan aspek-aspek tertentu dari peserta tes.

Menurut Arikunto (2021), kualitas soal tes sangat penting untuk memastikan bahwa hasil evaluasi benar-benar mencerminkan kemampuan nyata siswa. Soal yang baik tidak hanya mampu menilai pengetahuan siswa, tetapi juga keterampilan serta pemahaman mendalam mengenai materi yang telah diajarkan. Oleh karena itu, sangat perlu untuk menganalisis butir-butir soal guna menentukan apakah soal tersebut valid, reliabel, dapat membedakan antara siswa dengan kemampuan tinggi dan rendah, serta memiliki tingkat kesulitan yang sesuai dengan pemahaman siswa. Hal ini diungkapkan oleh Arifin (2014), yang menyatakan bahwa validitas, reliabilitas, daya pembeda, dan tingkat kesulitan adalah empat aspek penting yang harus ada dalam setiap soal tes agar dapat dipercaya dalam menilai pembelajaran.

Salah satu materi yang diajarkan di Sekolah Menengah Kejuruan (SMK) pada program Akuntansi dan Keuangan adalah Dasar-Dasar Akuntansi. Materi ini bertujuan untuk memberikan pemahaman kepada siswa mengenai konsep-konsep fundamental dalam akuntansi. Dengan adanya pembelajaran ini, diharapkan siswa dapat mengerti cara melakukan pencatatan, pengelompokan, dan pelaporan transaksi keuangan. Mengingat pentingnya penguasaan materi ini untuk siswa SMK, terutama di kelas X dan XI, evaluasi melalui ujian berkualitas sangat dibutuhkan untuk menilai seberapa baik siswa memahami konsep serta penerapan materi tersebut. Penggunaan ujian pilihan ganda dalam evaluasi untuk Akuntansi Dasar dapat secara efektif mencakup berbagai aspek kompetensi, tetapi pembuatannya harus dilakukan dengan cermat agar bisa mengukur kemampuan berpikir tingkat tinggi (HOTS) dan mempertimbangkan perbedaan tingkat pemahaman dari siswa.

Penelitian ini memiliki tujuan untuk mengevaluasi kualitas item soal pilihan ganda yang terdapat dalam handout akuntansi dasar dan perbankan dasar yang berbasis e-book. Analisis yang dilakukan menggunakan perangkat lunak Anates, yang dirancang khusus untuk menilai validitas, reliabilitas, daya beda, dan tingkat kesulitan dari soal-soal ujian. Anates dipilih dalam studi ini karena kemampuannya untuk memberikan analisis yang komprehensif mengenai kualitas soal, sehingga peneliti dapat memperoleh hasil yang objektif tentang kualitas soal yang diteliti.

Tujuan jangka panjang dari penelitian ini adalah untuk memberikan kontribusi dalam pengembangan evaluasi pembelajaran yang menggunakan e-book di tingkat

SMK, khususnya pada materi akuntansi dan perbankan dasar. Dengan memanfaatkan perangkat lunak Anates untuk menilai kualitas soal, diharapkan hasil penelitian ini dapat membantu guru dalam merancang soal evaluasi yang lebih baik, yang tidak hanya memenuhi standar validitas dan reliabilitas tetapi juga mampu membedakan antara siswa yang memiliki kemampuan tinggi dan rendah. Selain itu, penelitian ini juga berupaya mendukung penggunaan teknologi dalam pendidikan secara lebih menyeluruh, terutama dalam pengembangan alat evaluasi berbasis digital. Dengan cara ini, para guru dapat lebih mudah dan efektif dalam mengelola proses evaluasi pembelajaran, yang pada akhirnya akan meningkatkan kualitas pendidikan di SMK dan mempersiapkan siswa untuk lebih siap menghadapi tantangan di dunia kerja.

## 2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan deskriptif kuantitatif untuk menganalisis kualitas instrumen tes pilihan ganda berorientasi Higher Order Thinking Skills (HOTS) pada mata pelajaran Elemen Akuntansi Dasar dan Perbankan Dasar Fase E. Pendekatan ini sesuai dengan praktik umum dalam analisis empiris butir soal berbasis Classical Test Theory (CTT) sebagaimana digunakan oleh beberapa penelitian terkini (Hermanto, Anggraeni, & Sutirman, 2023; Sarea & Ruslan, 2019). Populasi penelitian adalah seluruh peserta didik program keahlian Akuntansi di Sekolah Menengah Kejuruan (SMK), sedangkan sampel penelitian berjumlah 20 peserta didik yang dipilih menggunakan teknik purposive sampling dengan kriteria telah mempelajari materi yang diujikan. Subjek penelitian dianggap memadai untuk analisis kualitas butir soal karena fokus penelitian adalah pada evaluasi instrumen, bukan generalisasi populasi (Warju et al., 2020).

Instrumen penelitian berupa tes pilihan ganda sebanyak 45 butir soal dengan lima alternatif jawaban yang disusun berdasarkan kisi-kisi capaian pembelajaran dan karakteristik soal HOTS yang mencakup kemampuan menganalisis, mengevaluasi, dan mencipta. Penyusunan soal mengikuti prinsip konstruksi soal HOTS sebagaimana dijelaskan oleh Mulyani et al. (2021). Instrumen diadministrasikan melalui Google Form untuk memastikan konsistensi tampilan, kemudahan distribusi, serta meminimalkan kesalahan pencatatan data.

Pengumpulan data dilakukan melalui beberapa tahap: penyusunan kisi-kisi dan butir soal, konversi instrumen ke dalam format Google Form, penyebaran tautan kepada peserta dengan waktu penggerjaan terkontrol, serta pengunduhan respons dalam format spreadsheet untuk dianalisis. Teknik analisis data dilakukan menggunakan aplikasi ANATES versi 4.0, yang banyak digunakan dalam penelitian analisis butir soal berbasis CTT (Hermanto et al., 2023). Aspek yang dianalisis meliputi: validitas butir menggunakan korelasi point biserial, reliabilitas instrumen menggunakan rumus KR-20, daya pembeda berdasarkan perbandingan proporsi jawaban benar kelompok atas-bawah, tingkat kesukaran berdasarkan proporsi peserta menjawab benar, efektivitas pengecoh berdasarkan distribusi pilihan, serta korelasi skor butir terhadap total skor untuk menilai kontribusi setiap butir terhadap keseluruhan tes. Kriteria interpretasi merujuk pada standar analisis butir soal menurut literatur terbaru dan relevan (Sarea & Ruslan, 2019; Hermanto et al., 2023).

Alur penelitian terdiri dari lima tahap utama: penyusunan kisi-kisi dan butir soal, validasi teoritis instrumen, pelaksanaan tes dengan Google Form, serta analisis butir soal dan penarikan kesimpulan menggunakan ANATES.

### 3. HASIL DAN PEMBAHASAN

Hasil dari analisis ini dimanfaatkan untuk mengevaluasi kualitas soal berdasarkan data yang diperoleh dari hasil ujian. Proses analisis ini memfasilitasi guru untuk melakukan langkah-langkah strategis yang didasarkan pada data guna memperbaiki soal-soal yang memiliki kelemahan dalam hal validitas, daya beda, serta tingkat kesulitan. Tindakan perbaikan ini tidak hanya bertujuan untuk meningkatkan mutu soal, tetapi juga untuk memastikan bahwa evaluasi pembelajaran dapat mencerminkan kemampuan siswa dengan lebih akurat, efektif, dan adil.

Analisis ini diharapkan menjadi bagian yang tidak terpisahkan dari siklus pembelajaran, sehingga kualitas pendidikan dapat terus ditingkatkan secara berkesinambungan oleh karena itu, penerapan metode ini tidak hanya berkontribusi pada perbaikan kualitas soal, tetapi juga pada peningkatan kemampuan guru dalam merancang dan menilai alat pembelajaran secara profesional.

Berdasarkan data yang telah dianalisis, penelitian ini memanfaatkan soal pilihan ganda yang berfokus pada topik Prinsip-Prinsip dan Konsep Dasar Akuntansi untuk siswa SMKN jurusan akuntansi. Dalam analisis ini, terdapat 20 peserta didik yang terlibat, dengan total 45 butir soal dalam format pilihan ganda. Hasil analisis soal mencakup beberapa aspek penting, termasuk penghitungan reliabilitas tes, kriteria validitas, pengelompokan peserta didik ke dalam kelompok unggul dan kelompok bawah, analisis daya pembeda, tingkat kesulitan soal, korelasi antara skor butir dan skor total, serta penilaian kualitas pengecoh, yang kemudian digunakan untuk memberikan rekomendasi perbaikan atau penggunaan soal.

Reliabilitas merujuk pada konsistensi dan kestabilan hasil yang dihasilkan oleh suatu alat ukur atau tes, yang esensinya adalah menilai sejauh mana instrumen tersebut mampu memberikan hasil yang sama atau tidak jauh berbeda ketika diterapkan secara berulang kali pada subjek yang sama dan di berbagai kondisi, tingkat reliabilitas yang baik menunjukkan bahwa instrumen pengukuran dapat dipercaya dan menghasilkan data yang akurat serta stabil.

Berdasarkan hasil analisis reliabilitas, diperoleh bahwa rata-rata skor total peserta didik adalah 19,90 dengan simpangan baku 6,19. Nilai rata-rata tersebut menunjukkan bahwa kemampuan siswa berada pada kategori sedang, sementara besarnya simpangan baku mengindikasikan adanya keberagaman kemampuan di antara peserta tes. Koefisien korelasi antara skor ganjil dan genap (Korelasi XY) sebesar 0,63 menunjukkan hubungan yang cukup kuat dan konsistensi antara dua bagian tes, sehingga mencerminkan adanya konsistensi internal pada butir soal. Selain itu, nilai reliabilitas sebesar 0,77 menandakan bahwa tes memiliki reliabilitas yang baik, karena nilai di atas 0,70 umumnya dianggap memadai untuk keperluan evaluasi pembelajaran. Dengan demikian, instrumen ini dapat dinyatakan layak dan cukup konsisten untuk digunakan. Meskipun demikian, simpangan baku yang relatif besar tetap perlu diperhatikan untuk menilai apakah tingkat kesulitan soal sudah sesuai dengan karakteristik kemampuan siswa.

Dalam menilai kualitas suatu tes, kriteria validitas menjadi aspek penting yang perlu diperhatikan. Korelasi antara skor setiap butir dengan skor total berperan sebagai indikator konsistensi internal yang menunjukkan sejauh mana butir soal mampu membedakan tingkat penguasaan peserta dan daya diskriminasi terhadap keseluruhan tes. Dengan demikian, hasil tes dapat dipastikan memiliki validitas dan reliabilitas yang baik.

Tabel 1. Korelasi Skor Butir dengan Skor Total

Kategori	Jumlah
Sangat Signifikan	18
Signifikan	4
Tidak Signifikan	23

*Sumber: Data diolah (2025)*

Berdasarkan analisis terhadap 45 butir soal dengan jumlah subjek N=20, diperoleh koefisien reliabilitas tes sebesar 0,77 yang menunjukkan bahwa instrumen memiliki tingkat konsistensi yang baik. Hasil analisis korelasi skor butir dengan skor total menunjukkan bahwa dari seluruh butir, terdapat 22 soal yang memiliki validitas baik (Signifikan maupun Sangat Signifikan) sesuai kategorisasi program ANATES. Sebanyak 18 butir termasuk kategori Sangat Signifikan karena memiliki korelasi  $r$  yang tinggi dan konsisten terhadap skor total, yaitu butir bernomor 3, 5, 14, 15, 17, 18, 19, 20, 21, 22, 23, 24, 29, 30, 31, 33, 40, dan 44. Sementara itu, terdapat 4 butir lainnya, yaitu 8, 11, 25, dan 27, yang dikategorikan sebagai Signifikan karena nilai korelasinya berada di atas batas  $r$  tabel ( $r \geq 0,444$ ) sehingga layak digunakan sebagai bagian dari instrumen. Meskipun ANATES mengelompokkan butir-butir tersebut ke dalam kategori tertentu, berdasarkan standar statistik baku ( $r \geq 0,561$  pada taraf signifikansi 0,01), hanya sebagian dari butir kategori Sangat Signifikan yang benar-benar memenuhi kriteria tersebut. Dengan demikian, terdapat perbedaan antara kategorisasi program dan persyaratan statistik yang sesungguhnya, sehingga interpretasi validitas harus memperhatikan kedua dasar analisis tersebut.

Dari keseluruhan soal, terdapat 23 butir yang memiliki korelasi di bawah  $r$  tabel dan dinyatakan tidak valid karena tidak menunjukkan hubungan yang signifikan dengan skor total. Butir-butir ini perlu direvisi atau dibuang karena kurang efektif dalam mengukur kemampuan siswa. Khusus untuk butir dengan korelasi negatif, yaitu nomor 2, 7, 10, 37, 41, dan 43, butir tersebut harus dihapus sebab dijawab benar oleh siswa dengan kemampuan rendah dan dijawab salah oleh siswa dengan kemampuan tinggi, sehingga arah pengukurannya berlawanan dengan tujuan tes.

Secara keseluruhan, 22 butir soal yang valid terbukti berfungsi baik sebagai instrumen pengukur yang akurat dan efektif, sedangkan butir-butir tidak valid memerlukan perhatian lebih lanjut melalui proses revisi atau penggantian. Perbaikan pada butir tidak valid diharapkan dapat meningkatkan kualitas instrumen secara keseluruhan sehingga menghasilkan data yang lebih konsisten, reliabel, dan dapat dipercaya.

Daya pembeda soal merujuk pada sejauh mana sebuah butir soal bisa membedakan antara siswa yang memiliki kemampuan tinggi dan yang memiliki kemampuan rendah. Ukuran yang menunjukkan kapasitas pembeda tersebut disebut sebagai indeks diskriminasi, dengan rentang nilai antara 0,00 sampai 1,00. Indeks ini memberikan gambaran mengenai kualitas soal dalam mengevaluasi perbedaan pemahaman siswa terhadap materi yang telah dipelajari.

Soal yang memiliki daya pembeda yang baik dapat mengidentifikasi sejauh mana siswa memahami materi. Di sisi lain, apabila soal tidak dapat membedakan siswa berdasarkan tingkat pemahaman mereka, hal ini mungkin disebabkan oleh beberapa

hal, seperti kunci jawaban yang keliru atau kehadiran lebih dari satu jawaban yang benar untuk soal tersebut.

Tabel 2. Hasil Analisis Daya Pembeda

Kriteria Daya Beda Soal	Indeks	Jumlah Soal
Tidak Baik	Negatif	4
Buruk	0,00 – 0,20	14
Cukup	0,20 – 0,40	
Baik	0,40 – 0,70	22
Sangat Baik	0,70 – 1,00	5

*Sumber: Data diolah (2025)*

Dari analisis terhadap 45 soal, 27 soal (60,00%) termasuk dalam kategori baik dan sangat baik, menandakan adanya kemampuan diskriminasi yang efektif. Namun, terdapat 14 soal (31,11%) yang perlu dilakukan perbaikan atau revisi, dan 4 soal (8,89%) yang harus dibuang karena menunjukkan daya pembeda negatif. Ini menandakan bahwa meskipun sebagian besar soal memiliki kualitas yang memadai dalam membedakan kemampuan peserta, masih ada beberapa soal yang perlu diperbaiki untuk meningkatkan keseluruhan kualitas tes.

Tingkat kesulitan adalah perbandingan antara persentase jawaban benar dari peserta uji dengan item soal yang ada. Analisis mengenai tingkat kesulitan bertujuan untuk memastikan soal yang diberikan dapat mencakup berbagai tingkat kemampuan siswa. Soal dikategorikan baik apabila tidak terlalu mudah dan juga tidak terlalu sulit. Sebaliknya, jika soal terlalu sulit, hal ini dapat membuat siswa merasa putus asa dan kurang bersemangat untuk mengerjakannya. Demikian pula, jika soal terlalu mudah, siswa bisa menjadi terlalu percaya diri dan cenderung meremehkan soal yang ada.

Tingkat kesulitan soal umumnya dinyatakan dalam bentuk indeks yang berkisar antara 0,00 hingga 1,00. Nilai indeks yang lebih tinggi menunjukkan bahwa soal tersebut lebih mudah. Indeks kesulitan ini digunakan untuk menunjukkan seberapa banyak peserta didik yang mampu menjawab soal dengan benar.



Diagram 1. Hasil Analisis Tingkat Kesukaran

*Sumber: Data diolah (2025)*

Tingkat kesulitan soal dalam ujian ini terbagi menjadi lima kelompok berdasarkan persentase jawaban benar yang diberikan oleh peserta. Dalam kategori Sangat Mudah, terdapat soal nomor 37 dan 43 dengan persentase jawaban benar mencapai 90%, yang menandakan bahwa hampir semua siswa dapat menjawab tanpa kesulitan, dan oleh karena itu kurang efektif dalam membedakan kemampuan peserta. Kategori Mudah mencakup soal nomor 1, 3, 13, dan 40 dengan persentase antara 75% dan 85%, yang menunjukkan bahwa mayoritas siswa mampu menjawab dengan tepat. Pada kategori Sedang, terdapat 22 soal, yaitu nomor 2, 4, 5, 7, 8, 10, 11, 14, 15, 19, 20, 21, 22, 23, 24, 26, 27, 32, 34, 39, 42 dan 44, dengan tingkat kesulitan antara 35% hingga 70%, sehingga cocok untuk merepresentasikan variasi kemampuan siswa dengan lebih akurat. Sementara itu, kategori Sukar terdiri dari 17 soal, yakni nomor 9, 12, 17, 18, 25, 28, 29, 31, 33, 35, 36, 38, 41, dan 45, dengan persentase jawaban benar berada pada kisaran 20% hingga 30%, menunjukkan bahwa hanya sebagian kecil siswa yang memahami materi pada level ini. Di sisi lain, kategori Sangat Sukar mencakup dua soal, yaitu nomor 6 dan 30, dengan persentase jawaban benar yang rendah, yaitu 10% dan 15%, menunjukkan bahwa hampir semua peserta mengalami kesulitan menjawabnya oleh karena itu, soal-soal ini perlu direvisi untuk meningkatkan kejelasan atau kesesuaian dengan materi yang diuji.

#### 4. KESIMPULAN DAN SARAN

Evaluasi dalam proses pembelajaran berfungsi sebagai alat ukur untuk mengetahui sejauh mana kemampuan peserta didik. Instrumen evaluasi dapat berbentuk tes pilihan ganda, menjodohkan, maupun isian singkat. Dalam penelitian ini digunakan soal pilihan ganda berbasis HOTS yang diberikan kepada peserta didik. Untuk memastikan soal tersebut memiliki kualitas yang baik serta memenuhi kriteria tertentu, dilakukan analisis menggunakan aplikasi Anates. Perangkat lunak ini dirancang untuk membantu guru dalam menganalisis mutu soal, baik pilihan ganda maupun uraian. Melalui Anates, guru dapat dengan mudah memasukkan data nilai siswa beserta kunci jawaban, kemudian sistem akan mengolahnya dan menampilkan hasil berupa tingkat reliabilitas, identifikasi kelompok atas dan bawah, tingkat kesukaran soal, efektivitas pengecoh, daya pembeda, serta hubungan antara skor butir dengan skor total.

Berdasarkan hasil analisis, instrumen tes menunjukkan reliabilitas tinggi dengan nilai 0,77 yang menandakan konsistensi yang baik. Dari total 45 butir soal yang dianalisis, terdapat 22 butir yang memiliki korelasi signifikan dengan skor total dan tergolong valid, sedangkan 23 butir memiliki korelasi di bawah r tabel yang artinya rendah dan dinyatakan tidak valid sehingga perlu direvisi atau dibuang karena tidak efektif dalam mengukur kemampuan siswa. Daya pembeda soal juga masih bervariasi, dengan 27 butir layak digunakan, 14 butir perlu perbaikan, dan 4 butir harus dibuang karena daya pembedanya kurang baik. Dilihat dari tingkat kesukarannya, sebagian besar soal berada pada kategori sedang (22 butir), yang menunjukkan distribusi cukup seimbang, meskipun masih terdapat beberapa butir yang terlalu mudah atau terlalu sulit. Dengan demikian, tes ini dapat digunakan setelah dilakukan revisi pada 14 butir soal agar pengukuran kemampuan berpikir kritis peserta didik menjadi lebih optimal.

#### Saran

Berdasarkan hasil analisis instrumen tes menggunakan aplikasi Anates, peneliti menyarankan agar penelitian selanjutnya melakukan penyempurnaan terhadap butir-butir soal yang tidak valid maupun yang memiliki daya pembeda rendah. Penelitian

berikutnya juga perlu mempertimbangkan proses penyusunan soal secara lebih cermat. Selain itu, uji coba instrumen sebaiknya dilakukan lebih dari satu tahap sehingga revisi terhadap butir soal dapat dilakukan secara bertahap hingga diperoleh tingkat validitas, reliabilitas, daya pembeda, dan tingkat kesukaran yang optimal. Peneliti selanjutnya diharapkan juga dapat menggunakan sampel yang lebih luas agar hasil analisis lebih representatif dan dapat digunakan sebagai acuan dalam pengembangan instrumen berbasis HOTS pada konteks yang lebih beragam.

Dengan demikian, penelitian lanjutan diharapkan mampu menghasilkan instrumen evaluasi yang lebih komprehensif dan akurat dalam mengukur kemampuan berpikir tingkat tinggi peserta didik.

## DAFTAR PUSTAKA

- Agustin, L., Zahra, G., Amraow, W., & Hakim, L. (2024). *Analisis butir soal pilihan ganda pada elemen akuntansi lembaga atau instansi pemerintah menggunakan program komputer Anates*, 4(1), 826–835.
- Alpusari, M. (2014). *Analisis butir soal, program komputer Anates versi 4.0 for Windows*. 3, 106–115.
- Arif, M., Madura, U. T., & Wibisono, Y. (2014). *Penerapan aplikasi Anates bentuk soal pilihan ganda*. 1(1), 1–9.
- Arifin, Z. (2014). *Evaluasi pembelajaran: Prinsip, teknik, prosedur*.
- Arikunto, S. (2021). *Penelitian tindakan kelas*.
- Asyarah, S. P., Febriyanti, A. C., & Sitanggang, C. J. C. (2025). *Analisis butir soal HOTS pada elemen perpajakan fase F SMK Akuntansi menggunakan software Anates*. 4(2), 1594–1602.
- Butir, A., Evaluasi, S., Kelas, T., & Plumbungan, S. D. N. (1990). *Menggunakan program Anates*. 799–806.
- Butir, A., Pilihan, S., Pada, G., Job, E., Dan, P., & Usaha, P. (2024). *Journal of Learning and Teaching*, 01(2014), 66–74.
- Cahyani, D. A., Aurelia, A. H., Hakim, L., & Pratiwi, V. (2024). *Analisis butir soal HOTS akuntansi perusahaan jasa menggunakan software Anates pada siswa SMK kelas 11*, 4(1), 1334–1346.
- Darmawan, C. A. (2016). *Analisis kualitas soal try out ujian nasional*. 2(1), 20–34.
- Sabrina, N. M., Nayla, K., Hemi, D., Hidayanti, S., Hakim, L., & Pratiwi, V. (n.d.). *Jurnal Pendidikan Ilmiah Transformatif*, 8(12), 68–76.
- Febriant, S., Khoerunnisa, Z., Chotimah, K. A., Dewi, L. P., Aliandra, A. L., Putri, H. E., Purwakarta, U. P. I. K., & Kunci, K. (2025). Pemanfaatan ANATES untuk menganalisis validitas dan reliabilitas butir soal uraian IPAS materi perkembangbiakan tumbuhan. *Jurnal Kreatif Sains*, 8(1), 232–242. <https://doi.org/10.56338/jks.v8i1.6688>
- Firandhika, A., Apriliyanti, K. D., Kasih, A., & Saputri, A. (2024). *Analisis butir soal HOTS pilihan ganda pada elemen komputer akuntansi di SMKN 10 Surabaya menggunakan aplikasi Anates*, 4(1), 727–738.
- Hakim, L., Akuntansi, P., Ekonomi, F., & Surabaya, K. K. (n.d.). *Pelajaran ekonomi materi akuntansi kelas XI di SMA Negeri 1 Driyorejo Gresik*.
- Hermanto, A., Anggraeni, A. D., & Sutirman, S. (2023). Efisiensi: Kajian Ilmu Administrasi. *Jurnal Penelitian dan Pengembangan Pendidikan*, 20(2), 159–170. <https://journal.uny.ac.id/index.php/jpms/article/view/57576>
- Ikawati, D. S., Agustin, E. W., Hakim, L., Pratiwi, V., & Surabaya, U. N. (2024). Analisis butir soal pilihan ganda pada siswa SMK menggunakan Anates. *Edukatif*: