

## **Analisis Kualitas Butir Soal Pada Pembelajaran Praktikum Akuntansi Jasa, Dagang Dan Manufaktur Menggunakan Aplikasi Anates**

**Wahyu Titis Anggraini**  
Universitas Negeri Surabaya

**Luqman Hakim**  
Universitas Negeri Surabaya

**Vivi Pratiwi**  
Universitas Negeri Surabaya

**Faizatul Ummah**  
Universitas Negeri Surabaya

**Sevina Anggraini**  
Universitas Negeri Surabaya

**Dila Anggraini**  
Universitas Negeri Surabaya

**Araminta Fuadiyah**  
Universitas Negeri Surabaya

**Erlita Tri Febrianti**  
Universitas Negeri Surabaya

**Ervita Oktaviani**  
Universitas Negeri Surabaya

**Ayunda Martalina Chintiya Bella**  
Universitas Negeri Surabaya

**Alamat:**  
Program Studi Pendidikan Akuntansi, Fakultas Ekonomika dan Bisnis  
Universitas Negeri Surabaya  
Jl. Ketintang, Ketintang, Kec. Gayungan, Kota Surabaya, Jawa Timur 60231  
Korespondensi penulis: 24080304052@mhs.unesa.ac.id

**ABSTRACT.** *This study aims to assess the quality of Higher Order Thinking Skills (HOTS) items in the Accounting, Trade, and Manufacturing Phase F Practicum subject using the ANATES application. The research method employed is a quantitative descriptive approach, with data collected from item testing results on grade XII students*

*across various vocational high schools (SMK) specializing in accounting. The analysis includes evaluating test reliability, dividing students into superior and average groups, assessing discriminating power, difficulty level, the correlation between item scores and total scores, and the quality of distractors. The findings indicate that most questions are valid with high reliability. The questions' discriminating power ranges from fair to good, though some items still require revision. Overall, the HOTS items developed meet the criteria for instrument quality and are suitable for assessing students' higher-order thinking skills in accounting practical learning. However, they require further development before being disseminated broadly.*

**Keywords:** *Accounting, Anates, HOTS, Instrument Quality, Item Analysis*

**ABSTRAK.** Penelitian ini bertujuan untuk menganalisis kualitas butir soal Higher Order Thinking Skills (HOTS) pada mata pelajaran Praktikum Akuntansi Jasa, Dagang, dan Manufaktur Fase F dengan menggunakan aplikasi ANATES. Metode penelitian yang digunakan adalah deskriptif kuantitatif dengan data diperoleh dari hasil uji coba butir soal pada peserta didik kelas XII di beberapa Sekolah Menengah Kejuruan (SMK) bidang akuntansi. Analisis butir soal meliputi reliabilitas tes, pembagian kelompok unggul dan asor, daya pembeda, tingkat kesukaran, korelasi skor butir dengan skor total, serta kualitas pengecoh. Hasil penelitian menunjukkan bahwa sebagian besar soal berada dalam kategori valid dengan tingkat reliabilitas tinggi. Daya pembeda soal berada pada kategori cukup hingga baik, meskipun beberapa butir soal masih memerlukan revisi. Secara umum, butir soal HOTS yang dikembangkan telah memenuhi kriteria kualitas instrumen dan layak digunakan untuk mengukur kemampuan berpikir tingkat tinggi peserta didik dalam pembelajaran praktik akuntansi, namun perlu dikembangkan lagi sebelum disebarkan secara luas.

**Kata kunci:** Akuntansi, Analisis Butir Soal, Anates, HOTS, Kualitas Instrumen

## PENDAHULUAN

Evaluasi pembelajaran merupakan alat yang digunakan untuk mengukur ketercapaian tujuan pembelajaran, efektivitas pembelajaran, keberhasilan metode pembelajaran, dan kemampuan siswa dalam memahami proses belajar. Proses evaluasi ini dapat dilakukan dengan berbagai cara termasuk dengan mengikuti tes berbasis soal HOTS (Devina et al., 2024). Evaluasi tidak hanya berfungsi untuk mengukur pencapaian belajar siswa, tetapi juga sebagai umpan balik bagi guru untuk memperbaiki proses pembelajaran. Dengan adanya evaluasi pembelajaran, maka kita dapat mengetahui sejauh mana kemajuan dan perkembangan pendidikan (Riana et al., 2024).

Agar dapat mengukur kemampuan peserta didik, maka perlu dilakukan tes. Tes merupakan alat atau langkah yang sistematis. Tes biasanya terdiri atas pertanyaan atau tugas-tugas untuk dikerjakan peserta didik guna mengukur hasil belajar peserta didik (Syairana et al., 2024). Ada tes tertulis dan tidak tertulis, namun yang biasa digunakan di sekolah yaitu tes tertulis. Tes tertulis ada dua yaitu tes dalam bentuk soal esai dan pilihan ganda. Namun yang umum digunakan adalah soal pilihan ganda yang dapat mengacu ke soal HOTS (*Higher Order Thinking Skills*). Soal HOTS (*Higher Order Thinking Skills*) menjadi salah satu keterampilan utama yang harus dimiliki siswa dan guru dalam menghadapi tantangan abad ke-21. HOTS mencakup kemampuan analisis, evaluasi, dan sintesis yang penting untuk menyelesaikan permasalahan dunia nyata secara kritis (Susilowati et al., 2025).

Menurut Ayu Lestari, untuk memperoleh hasil pengukuran yang akurat, kualitas soal juga sangat perlu diperhatikan. Butir soal yang berkualitas harus memenuhi kriteria validitas, reliabilitas, pengecoh yang berfungsi secara efektif, daya pembeda yang baik, serta memiliki tingkat kesukaran yang proporsional. Tanpa adanya analisis butir soal, kita tidak dapat mengetahui apakah soal yang digunakan sudah memenuhi kriteria tersebut atau belum. Analisis butir soal perlu dilakukan guna menguji kelayakan mutu soal yang akan dikerjakan oleh peserta didik. Analisis butir soal penting untuk mengidentifikasi soal-soal mana saja yang kurang baik dan harus diperbaiki (Syairana et al., 2024). Di era teknologi saat ini telah tersedia software yang dapat digunakan oleh guru untuk analisis butir soal, salah satunya adalah program Anates versi 4.0. Program ini dapat menganalisis soal pilihan ganda maupun uraian dengan menghasilkan informasi tentang validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan aktifitas pengecoh (Syairana et al., 2024).

Penelitian ini dilakukan untuk menganalisis kualitas butir soal pilihan ganda pada mata pelajaran Praktikum Akuntansi Perusahaan Jasa dan Dagang Fase F yang dianalisis menggunakan program Anates versi 4.0. Hasil analisis ini diharapkan dapat memberikan informasi yang akurat dan dapat menjadi dasar untuk perbaikan dan dapat

memberikan kontribusi kepada guru untuk meningkatkan evaluasi pembelajaran di tingkat SMK (Sekolah Menengah Kejuruan) jurusan Akuntansi.

## **METODE PENELITIAN**

Penelitian ini menggunakan pendekatan deskriptif kuantitatif yang bertujuan untuk mengkaji dan menelaah kualitas setiap butir soal berbasis Higher Order Thinking Skills (HOTS). Data dianalisis menggunakan aplikasi software Anates untuk menentukan kualitas setiap butir soal melalui aspek validitas, reabilitas, daya pembeda, dan tingkat kesukaran. Dalam evaluasi pembelajaran, aspek-aspek tersebut berperan penting sebagai komponen penunjang. Subjek penelitian ini adalah peserta didik dari Sekolah Menengah Kejuruan (SMK) dengan jumlah partisipasi sebanyak 25 peserta didik. Instrumen penelitian meliputi 72 butir soal pilihan ganda, yang digunakan untuk mengukur kemampuan analitis, kreatif dan evaluatif peserta didik.

Pengumpulan data dilakukan dengan menyebarkan google form kepada peserta didik, kemudian diminta untuk menjawab seluruh soal secara daring. Tahapan selanjutnya adalah pengolahan data yang sudah dikumpulkan dari beberapa data, seperti butir soal yang digunakan, kunci jawaban, dan hasil jawaban dari responden penelitian. Data yang sudah dikumpulkan kemudian dianalisis menggunakan program Anates untuk mendapatkan hasil yang dibutuhkan, sehingga dapat diketahui butir soal yang memerlukan perbaikan, butir soal yang tidak perlu digunakan, serta butir soal yang telah memenuhi kriteria.

## **HASIL DAN PEMBAHASAN**

Data dalam penelitian ini diperoleh melalui uji coba instrumen yang dilakukan oleh mahasiswa Program Studi Pendidikan Akuntansi kepada 20 orang siswa sebagai responden. Soal berjumlah 72 butir dengan format pilihan ganda dan berbasis HOTS. Setiap butir soal mengandung stimulus dengan lima pilihan jawaban. Soal yang diujikan juga mengacu pada level kognitif taksonomi Bloom revisi pada kategori C4 sampai C6, yaitu kemampuan menganalisis (*analyze*), mengevaluasi (*evaluate*), dan mencipta

(create). Data hasil tes kemudian dianalisis menggunakan program ANATES untuk memperoleh indeks tingkat kesukaran, daya pembeda, efektivitas pengecoh, validitas butir, serta reliabilitas instrumen. Analisis ini dilakukan untuk menilai sejauh mana setiap butir soal bekerja secara optimal sebagai alat ukur. Analisis ini bertujuan untuk melihat apakah setiap butir soal telah berfungsi dengan baik sebagai alat ukur kemampuan berpikir tingkat tinggi siswa.

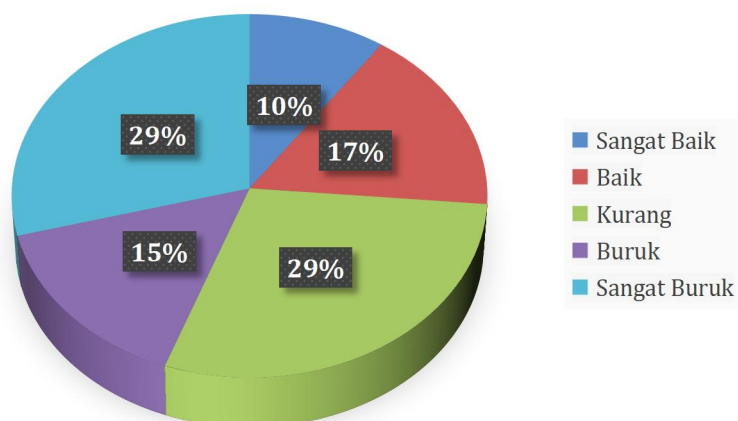
## Hasil

Adapun hasil dari analisis soal menggunakan aplikasi ANATES di bawah ini yang meliputi analisis kualitas pengecoh, tingkat kesukaran, daya pembeda, validitas, reabilitas, dan kualitas butir soal sebagai berikut.

### 1. Kualitas Pengecoh

Kualitas pengecoh, yaitu kemampuan pilihan jawaban salah untuk menarik perhatian siswa dengan pemahaman yang kurang, juga memerlukan perhatian khusus. Sebagian besar pengecoh dalam soal yang diuji tidak berfungsi dengan baik.

Gambar 1. Kualitas Pengecoh



Sumber: Data Diolah Peneliti (2025)

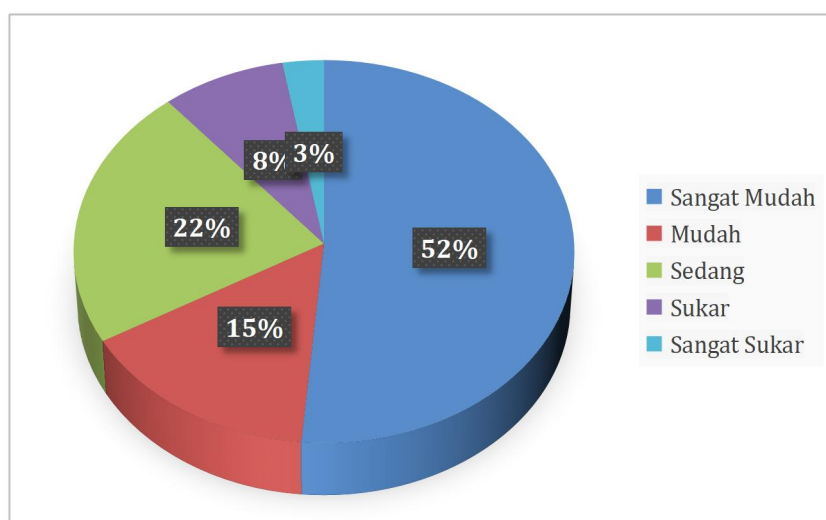
Dari grafik kualitas pengecoh, bisa dilihat kalau kondisi pengecoh seluruh soal masih belum merata. Pengecoh dengan kategori baik dan sangat buruk sama-sama punya persentase paling tinggi, yaitu 29%. Artinya, ada beberapa pengecoh yang

sebenarnya sudah baik dan berhasil membuat siswa berfikir dua kali, tapi ada juga yang benar-benar tidak berguna karena siswa langsung bisa melewatinya tanpa ragu. Kategori baik ada di 17%, dan buruk di 15%, yang menunjukkan bahwa masih banyak beberapa pengecoh yang kurang sehingga perlu diperbaiki agar lebih fungsional. Sementara itu, pengecoh sangat baik hanya 10%, yang berarti hanya sedikit opsi jawaban pengecoh yang benar-benar berhasil mengecoh siswa.

## 2. Tingkat Kesukaran

Tingkat Kesukaran digunakan untuk melihat sejauh mana sebuah soal dapat dijawab dengan benar oleh siswa, sehingga bisa mencerminkan tingkat pemahaman mereka terhadap materi. Tingkat kesukaran soal menggambarkan persentase siswa yang menjawab soal dengan benar. Semakin besar persentasenya, maka soal tersebut semakin mudah, dan semakin kecil persentasenya menunjukkan bahwa soal tersebut semakin sulit.

**Gambar 2. Tingkat Kesukaran**



Sumber: Data Diolah Peneliti (2025)

Berdasarkan grafik di atas, terlihat bahwa mayoritas soal termasuk dalam kategori sangat mudah, yakni mencapai 52%. Ini menunjukkan bahwa lebih dari setengah soal yang diujikan memiliki tingkat kesulitan yang belum cukup menantang

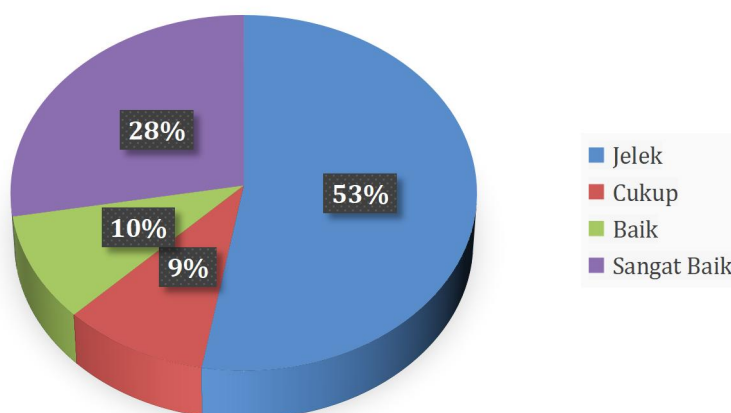
kemampuan siswa. Hal ini masih belum cukup ideal karena pertanyaan pada kategori tersebut belum bisa memberikan gambaran yang lebih valid tentang kemampuan berpikir siswa. Selanjutnya, ada 15% soal yang dikategorikan sebagai mudah dan 22% yang sedang. Persentase ini menunjukkan bahwa masih ada beberapa pertanyaan yang terlalu basic sehingga tidak memberikan tantangan bagi siswa dalam berpikir kritis. Pertanyaan dalam kategori sangat mudah dan mudah sebaiknya direvisi stimulusnya atau pilihan jawaban lebih mendorong analisis, bukan sekadar mengingat atau menebak.

Terdapat juga 8% soal yang termasuk dalam kategori sulit dan 3% yang sangat sulit. Soal dalam kategori ini berarti memiliki kompleksitas tinggi, seperti membutuhkan pemikiran yang lebih mendalam. Secara umum, distribusi tingkat kesulitan yang terlihat pada grafik ini cukup memuaskan, dengan banyaknya soal dari kategori sedang yang cocok untuk penilaian pembelajaran. Meskipun demikian, beberapa soal di kategori sangat mudah dan sangat sulit masih memerlukan perbaikan agar kualitas instrumen menjadi lebih seimbang dan benar-benar dapat dijadikan sebagai wadah untuk menilai kemampuan berpikir tingkat tinggi (HOTS) para siswa.

### **3. Daya Pembeda**

Daya Pembeda merupakan salah satu parameter penting dalam analisis kualitas soal, terutama pada soal pilihan ganda. Parameter ini menunjukkan kemampuan suatu butir soal untuk membedakan siswa dengan kemampuan tinggi dan rendah.

#### **Gambar 3. Daya Pembeda**



Sumber: Data Diolah Peneliti (2025)

Berdasarkan grafik di atas, terlihat bahwa 53% dari jumlah seluruh soal berada dalam kategori jelek. Ini menunjukkan bahwa mayoritas soal belum dapat membedakan antara siswa yang berkemampuan tinggi dan rendah. Soal dalam kategori ini cenderung dijawab secara acak oleh siswa, baik benar maupun salah, sehingga kualitasnya sebagai alat ukur evaluasi dinilai kurang memuaskan, dan sebaiknya perlu direvisi atau diganti. Selanjutnya, ada 9% soal yang tergolong cukup. Soal di kategori ini memiliki kemampuan membedakan yang masih dapat diterima, tetapi belum mencapai tingkat optimal. Soal ini dapat dipertahankan, namun tetap perlu ada perbaikan kecil, seperti dalam bentuk stimulus, penyusunan pertanyaan, atau kualitas opsi jawaban yang disediakan. 10% dari seluruh soal termasuk dalam kategori baik. Soal di kategori ini sudah mampu memenuhi fungsinya dengan baik, yaitu dapat membedakan siswa berkemampuan tinggi dan rendah dengan jelas. Soal ini dapat dipertahankan tanpa perlu melakukan perubahan yang signifikan.

Terakhir, ada 28% soal yang berada dalam kategori sangat baik. Soal-soal ini sangat efektif menilai kemampuan siswa. Soal-soal ini mempunyai struktur yang logis, stimulus yang jelas, dan pilihan jawaban yang seimbang sehingga tidak dapat dengan mudah ditebak. Secara keseluruhan, hasil ini menunjukkan bahwa walaupun ada sebagian soal yang memiliki daya pembeda baik hingga sangat baik, namun proporsi soal dengan pembeda buruk masih lebih banyak. Ini menunjukkan perlunya perbaikan

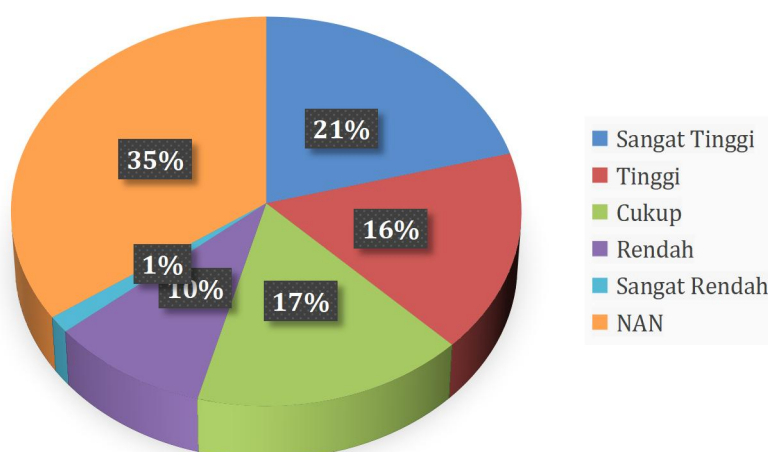


pada sebagian besar pertanyaan agar dapat mengukur kemampuan siswa dengan akurat dan selektif.

#### 4. Validitas

Perhitungan validitas dilakukan untuk mengetahui sejauh mana butir soal mampu mengukur apa yang seharusnya diukur dalam suatu instrumen penelitian. Validitas merupakan indikator penting dalam menentukan ketepatan instrumen dalam menggambarkan konstruk atau variabel yang diteliti (Sugiyono, 2022; Arikunto, 2021)..

**Gambar 4. Validitas**



Sumber: Data Diolah Peneliti (2025)

Berdasarkan grafik di atas, menunjukkan bahwa 21% soal termasuk dalam kategori sangat tinggi. Ini berarti butir butir soal sudah dengan tepat mengukur hal yang seharusnya diukur dalam pembelajaran Praktikum Akuntansi Jasa, Dagang, dan Manufaktur. Soal-soal dalam kategori ini bisa terus digunakan tanpa perlu dilakukan perubahan. Selanjutnya, terdapat 16% soal berada dalam kategori tinggi, dan 17% dalam kategori cukup. Kedua kategori ini menunjukkan bahwa soal-soal tersebut masih relevan, namun untuk soal dalam kategori cukup mungkin perlu ada sedikit penyesuaian pada penyampaian, kejelasan stimulus, atau pilihan jawaban untuk meningkatkan kualitas pengukurannya. Sementara itu, 10% dari soal tergolong dalam kategori rendah,

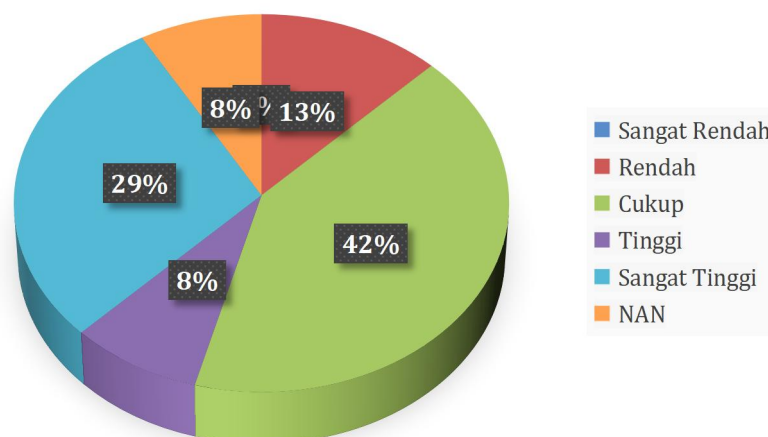
dan 1% berada dalam kategori sangat rendah. Soal-soal dalam kategori ini kurang efektif dalam mengukur kemampuan siswa, sehingga sebaiknya diperbaiki secara menyeluruh atau digantikan dengan soal baru yang lebih sesuai dengan indikator pembelajaran.

Terakhir, terdapat 35% item soal yang dicatat sebagai NAN (Not An Number). Persentase ini cukup tinggi dan muncul karena pilihan jawaban yang tidak merata (semua siswa menjawab pada pilihan sama yaitu benar atau salah), Hal ini menunjukkan bahwa item soal ini tidak dapat dianalisis keabsahannya, sehingga sangat penting untuk melakukan tinjauan kembali. Soal dalam kategori NAN memerlukan pemeriksaan lebih lanjut, revisi, atau pengujian ulang agar dapat diketahui kualitasnya yang sesungguhnya. Secara keseluruhan, hasil yang diperoleh menunjukkan bahwa meski ada sejumlah soal yang sudah valid dan dapat dipakai, masih banyak soal yang perlu disempurnakan agar instrumen evaluasi sesuai dengan tujuan pengukuran kemampuan berpikir tingkat tinggi (HOTS).

## 5. Reliabilitas

Reliabilitas soal digunakan untuk mengetahui tingkat konsistensi hasil pengukuran yang dilakukan dengan instrumen yang sama pada kondisi yang relatif serupa.

Gambar 4. Reliabilitas



Sumber: Data Diolah Peneliti (2025)

Berdasarkan grafik reliabilitas di atas, kualitas reliabilitas tes dapat dikatakan cukup baik. Hal ini dapat dilihat dari kategori cukup yang mendominasi sebesar 42%, diikuti oleh kategori sangat tinggi sebesar 29%. Artinya, sebagian besar butir soal sudah konsisten dalam mengukur kemampuan siswa, bahkan ada proporsi yang menunjukkan konsistensi sangat tinggi. Meski begitu, masih terdapat 13% soal dengan reliabilitas rendah dan 8% soal dengan reliabilitas sangat rendah, yang menandakan sebagian kecil butir soal belum stabil dan perlu diperbaiki.

Terakhir, sebanyak 8% soal adalah NAN, biasanya terjadi karena adanya distribusi skor yang tidak normal. Hal ini menunjukkan bahwa kualitas soal tersebut tidak dapat dianalisis secara reliabel sehingga memerlukan pemeriksaan ulang, baik dari sisi penyusunan, pemilihan stimulus, maupun pelaksanaan penilaiannya. Secara keseluruhan, hasil reliabilitas ini menunjukkan bahwa instrumen tes belum berada pada kategori sangat baik dan masih memerlukan revisi, terutama pada soal dengan reliabilitas rendah dan sangat rendah. Perbaikan yang dapat dilakukan adalah melalui penyempurnaan stimulus soal, penyesuaian tingkat kesukaran, serta peningkatan kualitas pengecoh agar tes menghasilkan skor yang lebih konsisten dan akurat dalam mengukur kemampuan siswa pada mata pelajaran Praktikum Akuntansi Jasa, Dagang, dan Manufaktur.

## **6. Kualitas Butir Soal**

Berdasarkan hasil analisis 72 butir soal HOTS pada mata pelajaran Praktikum Akuntansi Jasa, Dagang, dan Manufaktur menggunakan program ANATES, dapat disimpulkan bahwa kualitas soal secara keseluruhan masih belum optimal. Secara keseluruhan, hasil analisis menunjukkan bahwa kualitas butir soal masih perlu ditingkatkan. Tingkat kesukaran soal sebagian besar berada pada kategori sedang, namun masih terdapat butir yang terlalu mudah dan terlalu sulit sehingga perlu penyesuaian. Daya pembeda banyak yang berada pada kategori rendah, yang berarti sejumlah soal belum mampu membedakan siswa berkemampuan tinggi dan rendah. Kualitas pengecoh pada beberapa soal juga kurang optimal karena ada pilihan jawaban

yang tidak dipilih siswa, sehingga perlu diperbaiki agar lebih berfungsi. Dari segi validitas, hanya sebagian soal yang menunjukkan kesesuaian dengan kompetensi yang diukur, sedangkan lainnya perlu revisi. Selain itu, nilai reliabilitas tes masih tergolong rendah, sehingga konsistensi hasil pengukuran belum stabil. Hal ini menunjukkan bahwa instrumen tes belum sepenuhnya ideal dan memerlukan penyempurnaan sebelum digunakan secara lebih luas.

### **Pembahasan**

Penelitian ini bertujuan untuk menganalisis kualitas butir soal berbasis HOTS (*Higher Order Thinking Skills*) pada mata pelajaran Praktikum Akuntansi Jasa, Dagang, dan Manufaktur Fase F dengan memanfaatkan aplikasi ANATES sebagai alat bantu analisis. Berdasarkan hasil analisis terhadap 72 butir soal, diperoleh bahwa kualitas soal masih perlu ditingkatkan agar dapat berfungsi secara optimal dalam mengukur kemampuan berpikir tingkat tinggi siswa.

Pada aspek efektivitas pengecoh, hasil analisis menunjukkan bahwa distribusi fungsi pengecoh belum merata. Sebanyak 29% pengecoh berada pada kategori baik, namun persentase yang sama (29%) juga termasuk kategori sangat buruk. Sementara itu, 17% berada pada kategori baik, 15% kurang, dan hanya 10% tergolong sangat baik. Kondisi ini menunjukkan perlunya perbaikan opsi jawaban agar lebih logis dan mampu mengecoh siswa, sehingga setiap pilihan dapat berfungsi secara optimal dalam mengukur kemampuan analisis siswa. Selanjutnya, Hasil uji tingkat kesukaran memperlihatkan bahwa sebagian besar butir soal, yakni 52%, berada pada kategori sangat mudah. Persentase ini menunjukkan bahwa tingkat kesukaran masih belum sesuai karena belum merepresentasikan kemampuan siswa. Namun demikian, terdapat 15% soal mudah, 22% sedang, 8% sulit, dan 3% sangat sulit. Distribusi tersebut belum sesuai dengan proporsi ideal, yaitu 25% mudah, 50% sedang, dan 25% sukar. Oleh karena itu, beberapa soal masih perlu disesuaikan agar dapat mencerminkan variasi kemampuan siswa secara proporsional dan tetap selaras dengan tuntutan soal HOTS.

Pada uji daya pembeda, diperoleh hasil bahwa 53% butir soal termasuk kategori jelek, 9% cukup, 10% baik, dan 28% sangat baik. Persentase tinggi pada kategori jelek

menunjukkan bahwa sebagian besar soal belum mampu membedakan siswa berkemampuan tinggi dan rendah. Soal dengan daya pembeda rendah cenderung tidak informatif karena dapat dijawab secara acak. Sebaliknya, soal dengan daya pembeda baik dan sangat baik dapat dipertahankan karena telah mampu mengukur kemampuan siswa secara efektif. Soal dengan daya pembeda jelek dan cukup perlu direvisi, terutama pada aspek kejelasan stimulus, konteks soal, dan kualitas opsi jawaban. Uji validitas menunjukkan bahwa 21% soal tergolong sangat tinggi, 16% tinggi, 17% cukup, 10% rendah, dan 1% sangat rendah. Selain itu, 35% soal tidak dapat dianalisis karena masuk kategori NAN (*Not a Number*), yang umumnya disebabkan oleh pola jawaban yang seragam atau distribusi respons yang tidak seimbang. Temuan ini mengindikasikan bahwa sebagian besar soal masih perlu ditinjau ulang agar benar-benar mengukur kemampuan berpikir tingkat tinggi. Soal dengan validitas tinggi dapat digunakan kembali, sementara soal dengan validitas rendah hingga sangat rendah perlu direvisi atau diganti. Validitas tinggi mencerminkan kesesuaian antara indikator pembelajaran dan hasil pengukuran.

Pada uji reliabilitas, diperoleh bahwa 42% soal memiliki reliabilitas cukup, 13% rendah, 8% sangat rendah, 29% sangat tinggi, dan 8% NAN. Hasil ini menunjukkan bahwa konsistensi instrumen masih belum stabil. Butir soal dengan reliabilitas rendah tidak mampu menghasilkan pengukuran yang konsisten ketika digunakan pada waktu atau kelompok yang berbeda. Meskipun terdapat beberapa soal dengan reliabilitas tinggi, sebagian besar instrumen tetap memerlukan perbaikan. Upaya peningkatan reliabilitas dapat dilakukan melalui peningkatan validitas dan perbaikan pengecoh, karena kualitas tiap butir soal sangat berpengaruh terhadap keandalan tes secara keseluruhan. Secara keseluruhan, hasil analisis menunjukkan bahwa soal berbasis HOTS pada mata pelajaran Praktikum Akuntansi Jasa, Dagang, dan Manufaktur memiliki potensi sebagai alat ukur kemampuan berpikir tingkat tinggi, namun belum mencapai kualitas yang ideal. Aspek tingkat kesukaran relatif seimbang, tetapi daya pembeda dan reliabilitas masih rendah, serta beberapa pengecoh tidak berfungsi dengan baik. Hal ini menunjukkan bahwa guru perlu melakukan revisi dan analisis berkelanjutan terhadap instrumen tes. Penggunaan aplikasi ANATES terbukti efektif dalam mengidentifikasi

kelemahan soal, sehingga dapat menjadi dasar dalam perbaikan kualitas instrumen evaluasi. Penelitian ini juga konsisten dengan hasil penelitian dari berbagai sumber yang menunjukkan bahwa sebagian besar soal buatan guru masih belum memenuhi kriteria validitas dan reliabilitas yang memadai. Oleh karena itu, analisis butir soal merupakan langkah penting dalam meningkatkan mutu evaluasi pembelajaran.

## KESIMPULAN DAN SARAN

Analisis kualitas butir soal pilihan ganda berbasis Higher Order Thinking Skills (HOTS) pada mata pelajaran Praktikum Akuntansi Jasa, Dagang, dan Manufaktur Fase F menggunakan aplikasi ANATES menunjukkan bahwa secara keseluruhan kualitas soal masih perlu ditingkatkan. Sebagian besar soal berada pada tingkat kesukaran sedang yang sesuai dengan kemampuan siswa, namun terdapat soal yang terlalu mudah atau sulit sehingga perlu penyesuaian. Daya pembeda banyak yang rendah sehingga soal belum optimal dalam membedakan kemampuan siswa yang tinggi dan rendah. Kualitas pengecoh masih kurang maksimal karena ada beberapa pilihan jawaban yang tidak efektif dalam mengecoh siswa. Dari segi validitas, hanya sebagian soal yang sudah layak digunakan, sementara yang lain memerlukan revisi atau penggantian. Reliabilitas soal juga masih rendah, menunjukkan hasil pengukuran yang belum konsisten. Oleh karena itu, perbaikan dan revisi soal sangat diperlukan agar instrumen evaluasi dapat mengukur kemampuan berpikir tingkat tinggi siswa secara akurat dan andal. Penggunaan aplikasi ANATES sangat membantu dalam mengidentifikasi kelemahan butir soal untuk peningkatan mutu tes di masa mendatang.

Berdasarkan hasil penelitian ini, Peneliti menyarankan untuk merevisi soal - soal yang tingkat kesulitannya sangat mudah atau sangat sulit agar hasil lebih sesuai dengan kemampuan peserta didik. Untuk daya pembeda pada soal harus ditingkatkan agar soal lebih efektif dalam membedakan kemampuan peserta didik dari tingkatan rendah sampai tingkatan tertinggi. Selain itu, perlu diadakan perbaikan pada kualitas pengecoh soal dengan mengganti opsi jawaban yang kurang efektif agar dapat mengecoh peserta didik hal ini juga berpengaruh dalam hasil analisis. Soal yang validitasnya rendah harus di revisi/diganti sehingga instrumen evaluasi dapat mengukur kemampuan peserta didik secara jelas. Penggunaan software ANATES sangat membantu dalam mengidentifikasi

kelemahan butir soal dan juga sebaiknya dimanfaatkan dengan baik dalam pembuatan, pengembangan dan juga perbaikan soal agar sesuai dengan kriteria.

## 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. <https://ulilalbabainstitute.co.id/index.php/PESHUM/article/view/6870>
- Aini, U., Lathifa, N., & Jelita, M. (2025). *Anates : Analysis of Objective Test Item Quality*. 1(1), 45–59. <https://jeici.kips.or.id/index.php/jeici/article/view/3>
- Asyaroh, 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. <https://ulilalbabainstitute.co.id/index.php/PESHUM/article/view/7176>
- Augustia, A. D., & Augustia, C. N. (2025). *Analisis Validitas dan Reliabilitas Soal Pilihan Ganda dengan Menggunakan Software Anates pada Mata Pelajaran Perpajakan*. 5(1), 250–265. <https://www.ojs.pseb.or.id/index.php/jmeh/article/view/1165>
- 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. <https://ulilalbabainstitute.co.id/index.php/PESHUM/article/view/6792>
- Hakim, L. (2025). *ANALISIS BUTIR SOAL ELEMEN EKONOMI BISNIS DAN ADMINISTRASI UMUM MENGGUNAKAN APLIKASI ANATES*. 5(1). <https://felifa.net/index.php/DIDACTIC/article/view/83>
- Hartanto, M. A., Yon, A. E., & Sukmana, U. (n.d.). *ANALISIS SOAL PADA LKS MATA PELAJARAN EKONOMI KELAS X SEMESTER 1*. 2(2), 334–341. <https://samudrapublisher.com/index.php/jpgenus/article/view/167>
- Hidayat, R., & Rohmawati, B. (2025). *Pengembangan Butir Soal HOTS : Tantangan*



- dan Solusi Untuk Meningkatkan Kemampuan Analisis Siswa Pada Pembelajaran PAI. 4(1), 1–20.  
<https://jurnal.catimoredansahabat.my.id/index.php/jumper/article/view/255>
- Ikhsaniyah, S. N., Kurnia, A. D., Zuroida, M., Pratiwi, V., & Hakim, L. (2024). ANALISIS BUTIR SOAL PERPAJAKAN PPH PASAL 21 MENGGUNAKAN SOFTWARE ANATES PADA PENDEKATAN TEORI TES KLASIK. 12(2).  
<https://journal.uir.ac.id/index.php/Peka/article/view/19917>
- Jl, T., No, G., Rejo, T., & Penulis, K. (2024). Pengembangan analisis butir soal uraian berbasis ict. 1(4), 405–415.  
<https://www.ejurnal.kampusakademik.co.id/index.php/jmia/article/view/2204>
- Manusia, H., & Ix, K. (2021). Junior High School Students In Jember Regency Instrumen Four Tier Diagnostic Test Untuk Mengidentifikasi Miskonsepsi Materi. 5(2), 111–121. [https://repository.unmuhjember.ac.id/12957/1/artikel\\_UNP\\_2021.pdf](https://repository.unmuhjember.ac.id/12957/1/artikel_UNP_2021.pdf)
- Masullah, B. D., Zuhry, L. H., & Usman, L. H. (2024). ANALISIS BUTIR SOAL UJIAN AKHIR SEMESTER MATA PELAJARAN MATEMATIKA SMP NEGERI 6 PRAYA TIMUR. 5(September), 152–161.  
<https://journal.unpacti.ac.id/index.php/ELIPS/article/view/1219>
- Okyanida, I. Y., Mayanty, S., & Widiyatun, F. (2024). Analisis Butir Soal Kemampuan Berpikir Kritis Siswa SMAIT Nururrohmah Depok. 15(1), 73–79.  
<https://doi.org/10.26877/jp2f.v15i1>
- Parisu, C. Z. L., Ekadayanti, W., Sisi, L., & Juwairiyah, A. (2024). Analisis Butir Soal Pengetahuan Dasar Matematika Menggunakan Pendekatan Rasch. 36–45.  
<https://jurnal.ustjogja.ac.id/index.php/sciencetech/article/view/17112>
- Permadhani, R. M., & Rohayati, S. (2025). Pengembangan E-LKPD Berbasis Google Sites Terintegrasi Artificial Intelligence pada Elemen Perpajakan Siswa Kelas XII AKL. 13(2), 61–69. <https://ejournal.unesa.ac.id/index.php/jpak/article/view/72265>
- Purwati, L. M., Arianty, R., Syakilah, D. M., Ridlo, S., & Susilaningih, E. (2021). Analisis Soal Tes Pilihan Ganda Berbasis Higher Order Thinking Skill menggunakan Aplikasi Anates Windows Versi. 15, 460–473.  
<https://journal.uniga.ac.id/index.php/JP/article/view/1287>
- Puteri, N. D., Arsyanti, K., Putricia, N. D., Hakim, L., & Pratiwi, V. (2025). Analisis Kelayakan Butir Soal Pada Handout Akuntansi Keuangan Dana Kas Kecil Berbasis E-Book Menggunakan Software Anates. 2(3), 205–214.  
<https://jibema.murisedu.id/index.php/JIBEMA/article/view/108>
- Putri, D. R., & Hakim, L. (n.d.). SMK MENGGUNAKAN APLIKASI ANATES. 2047–2060. <https://jutepe-joln.net/index.php/JURPERU/article/view/767>
- Rahayu, L. P. (2024). Kualitas Soal Bahasa Indonesia Kelas XI SMAN 2 Bangko Pusako : Analisis Butir Soal. 10(4), 3755–3762. <https://ejournal.my.id/onoma/article/view/4481>
- Rahmawati, N. S., & Rahman, M. F. (2025). Analisis Butir Soal Objektif Penilaian Sumatif Akhir Semester Gasal Mata Pelajaran Ekonomi. 8, 8863–8873.  
<https://jiip.stkipyapisdompu.ac.id/jiip/index.php/JIIP/article/view/8837>
- Rem, S., Di, S., & Menengah, S. (2020). Analisis kualitas butir soal tipe hots pada kompetensi sistem rem siswa di sekolah menengah kejuruan. 17(1), 95–104.  
<https://ejournal.undiksha.ac.id/index.php/JPTK/article/view/22914>



- Riani, D., Almujab, S., Dina, A., & Budiarto, R. (2020). *ANALISIS BUTIR SOAL DAN KEMAMPUAN SISWA DALAM MENJAWAB SOAL UJIAN NASIONAL PADA MATA PELAJARAN EKONOMI. IV*, 70–79.  
<https://journal.unpas.ac.id/index.php/oikos/article/view/2425>
- Sabela, O. R., Krisdayanty, D., Taqqiyah, A. Z., Hakim, L., Pratiwi, V., & Surabaya, U. N. (2025). *Education Achievmnt: Journal of Science and Research*. 6(1), 251–262.  
<https://pusdikra-publishing.com/index.php/jsr>
- Siregar, S. R. N., Luckytasari, A., Ariani, A., Hakim, L., & Pratiwi, V. (2024). *Analisis Soal Hots Pada Isu Serta Perkembangan Terkini Akuntansi Keuangan Fase E SMK Menggunakan Aplikasi Anates*. 4(1), 1386–1395.  
<https://ulilalbabainstitute.co.id/index.php/PESHUM/article/view/7157>
- Sudarto, Hafid.A, K. . (2023). *Tingkat Pencapaian Kemampuan Guru SD Dalam Menggunakan Aplikasi Anates Melalui Pelatihan*. 3(5), 793–800.  
<https://bajangjournal.com/index.php/J-ABDI/article/view/6595>
- 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.  
<https://ulilalbabainstitute.co.id/index.php/PESHUM/article/view/7291>
- Syafiriya, D. M., & Hakim, L. (2024). *Analisis Butir Soal HOTS untuk Menguji Pemahaman Siswa SMK Jurusan Akuntansi Menggunakan Software Anates*. 4(1).  
<https://ejournal.nlc-education.or.id/index.php/JNSI/article/view/107>
- Tilaar, A. L. F., Sulangi, V. R., & Pelealu, E. M. C. (2020). *ANALISIS KUALITAS BUTIR SOAL TES BUATAN GURU MATA PELAJARAN MATEMATIKA KELAS X SMA NEGERI 1 AIRMADIDI*. 8(April).  
[https://download.garuda.kemdikbud.go.id/article.php?article=1527833&val=1458&title=ANALISIS KUALITAS BUTIR SOAL TES BUATAN GURU MATA PELAJARAN MATEMATIKA KELAS X SMA NEGERI 1 AIRMADIDI TAHUN AJARAN 20192020](https://download.garuda.kemdikbud.go.id/article.php?article=1527833&val=1458&title=ANALISIS%20KUALITAS%20BUTIR%20SOAL%20TES%20BUATAN%20GURU%20MATA%20PELAJARAN%20MATEMATIKA%20KELAS%20X%20SMA%20NEGERI%201%20AIRMADIDI%20TAHUN%20AJARAN%2020192020)
- Ubaidurrahman, A., & Kartini, T. (2025). *Inflasi : Jurnal Ekonomi , Manajemen dan Perbankan Analisis Butir Soal Penilaian Akhir Semester Akuntansi Kelas XI SMK Negeri 4 Jember Inflasi : Jurnal Ekonomi , Manajemen dan Perbankan*. 2, 32–37.  
<https://ejurnal.faaslibsmmedia.com/index.php/inflasi/article/view/58>
- V, M. A., Hakim, L., & Pratiwi, V. N. : S. I. A. B. S. P. G. P. E. A. K. G. M. E. C. D. L. U. N. S. (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.  
<https://ejournal.nlc-education.or.id/index.php/JNSI/article/view/95>
- Wahyu, G., Pidekso, S., Pandu, G., & Ramadhan, S. (2025). *AKUNTANSI FASE F MENGGUNAKAN SOFTWARE ANATES ANALYSIS OF ITEM CHARACTERISTICS OF PHASE F ACCOUNTING*. 84–95.  
<https://jurnal.uny.ac.id/index.php/jpakun/article/view/91595>
- Wahyuni, Fauzia, Y. (2024). *ANALISIS BUTIR SOAL ASESMEN SUMATIF BIOLOGI MATERI PERUBAHAN LINGKUNGAN DENGAN MENGGUNAKAN ANATES PADA KELAS X SMA*. 4(1), 126–135.  
<https://jurnalp4i.com/index.php/learning/article/view/2777>

