

ANALISIS PEMAHAMAN SOAL HOTS SISWA XI AKUNTANSI SMKN 6 SURABAYA PADA FASE F KOMPUTER AKUNTANSI

Bintang Zada Aufa¹, Pramitha Rahmawati², Amanda Putri Salsabila³, Luqman Hakim⁴, Vivi Pratiwi⁵

^{1,2,3,4,5} Universitas Negeri Surabaya

Email: bintang.23102@mhs.unesa.ac.id¹, pramitha.23162@mhs.unesa.ac.id², amanda.23090@mhs.unesa.ac.id³, luqmanhakim@unesa.ac.id⁴, vivipratiwi@unesa.ac.id⁵.

ABSTRAK

Potensi siswa Indonesia dalam mengembangkan kemampuan berpikir kritis (HOTS) cukup besar, karena kemampuan ini sangat penting dalam menghadapi tantangan yang kompleks di era abad ke-21. Meski demikian, masih banyak instrumen penilaian yang belum bisa mengukur kemampuan berpikir tingkat tinggi secara tepat. Penelitian ini bertujuan untuk mengevaluasi kualitas soal pilihan ganda berbasis HOTS berdasarkan berbagai aspek seperti validitas, reliabilitas, daya pembeda, tingkat kesulitan, dan efektivitas pengecoh. Penelitian ini menggunakan metode kuantitatif deskriptif dengan melibatkan 30 siswa kelas XI di SMK Negeri 6 Surabaya. Sebanyak 15 soal diberikan melalui Google Form dan dianalisis dengan aplikasi Anates versi 4.0.2. Hasil menunjukkan bahwa 60% soal termasuk kategori sedang, 33% mudah, dan 7% sukar. Dalam hal daya pembeda, terdapat 27% soal sangat baik, 27% baik, dan 47% cukup. Untuk aspek efektivitas pengecoh, ditemukan beberapa opsi jawaban yang belum berfungsi optimal. Secara keseluruhan, instrumen penilaian berbasis HOTS cukup baik, tetapi masih perlu diperbaiki pada beberapa butir soal agar hasil pengukuran lebih akurat dan seimbang. Hasil penelitian ini membantu pengembangan ilmu evaluasi pendidikan, khususnya dalam meningkatkan kualitas instrumen asesmen berbasis HOTS yang bisa menjadi pedoman dalam proses pembelajaran Fase F di bidang Komputer Akuntansi di SMK, sehingga meningkatkan efektivitas serta produktivitas belajar untuk mencapai tujuan akademik.

Kata Kunci : *aplikasi anates; evaluasi pembelajaran; komputer akuntansi; pemahaman soal hots;*

ABSTRACT

Indonesian students have a strong potential to develop critical thinking skills, which are essential for dealing with the complex issues of the 21st century. However, many current tests are not fully effective in measuring these higher-order thinking skills. This study looks at the quality of multiple-choice questions that are meant to test higher-order thinking skills. It checks the questions based on several important factors like validity, reliability, how well they can tell apart different levels of understanding, difficulty level, and how effective the wrong answer choices are. The study used a quantitative approach involving 30 students from 11th grade at SMK Negeri 6 Surabaya. They were given 15 multiple-choice questions through Google Forms and analyzed using the Anates version 4.0.2 software. The results showed that 60% of the questions were of moderate difficulty, 33% were easy, and 7% were hard. When it came to how well the questions could differentiate between students, 27% were very good, 27% were good, and 47% were only fair. The analysis also found that some of the answer choices were not as effective as they should be. Overall, the HOTS-based test was considered good, but some parts still needed improvement to make sure the test measures students' abilities more accurately and fairly. These findings can help improve educational assessments, especially for HOTS-based tests used in the Computer Accounting field at vocational high schools, which can support better learning outcomes and help students reach their academic goals more effectively.

Keywords : *anates apps; learning evaluation; accounting computer; understanding hot*

PENDAHULUAN

Di era globalisasi, pendidikan memainkan peran penting dalam meningkatkan kualitas sumber daya manusia. Salah satu kompetensi penting yang harus dikembangkan dalam



pembelajaran adalah kemampuan berpikir tingkat tinggi atau *Higher Order Thinking Skills* (HOTS). Mengingat potensi pelajar Indonesia dalam menguasai skill berpikir kritis (HOTS) sangat besar mengingat pentingnya keterampilan tersebut dalam menghadapi tantangan abad 21 yang semakin kompleks (Rahmawati, 2023). HOTS tidak hanya menekankan pada kemampuan mengingat, tetapi juga dapat digunakan untuk mengukur kemampuan yang meliputi kemampuan menerapkan, menganalisis, mengevaluasi, dan mencipta apabila soal disusun dengan tepat (Warju et al., 2020). Dalam konteks Sekolah Menengah Kejuruan (SMK), penerapan HOTS menjadi sangat relevan karena dapat membantu siswa mempersiapkan diri menghadapi dunia kerja yang kompetitif maupun melanjutkan pendidikan ke jenjang yang lebih tinggi. Oleh karena itu, penerapan HOTS dalam pembelajaran perlu menitikberatkan pada keterlibatan aktif siswa dalam proses berpikir serta pada hasil belajar yang dicapai (Ananda et al., 2020).

Selain berfokus pada penerapan indikator HOTS, suatu tes yang diujikan kepada siswa juga harus memiliki kualitas yang baik. Suatu alat penilaian (tes) dikatakan mempunyai kualitas yang baik apabila alat tersebut memenuhi dua hal, yakni ketepatan atau validitas dan ketetapan atau keajegan atau reliabilitas (Rahayu & Djazari, 2016). Salah satu metode yang dapat digunakan untuk mengevaluasi kualitas instrumen penilaian adalah melalui analisis butir soal berbasis HOTS menggunakan aplikasi Anates. Aplikasi ini mampu memberikan informasi rinci mengenai validitas, reliabilitas, daya pembeda, tingkat kesukaran, serta efektivitas pengecoh suatu soal (Anggreini & Darmawan, 2017). Dengan melakukan analisis ini, guru dapat mengetahui sejauh mana suatu butir soal layak digunakan untuk mengukur kemampuan berpikir tingkat tinggi siswa.

Analisis butir soal perlu dilakukan untuk mengetahui sejauh mana butir soal tersebut dapat digunakan dalam pengujian tes sekaligus sebagai salah satu kontrol kompetensi peserta didik (Rudini, 2020). Dengan melakukan analisis terhadap soal, pendidik dapat mengetahui kualitas suatu butir soal berdasarkan tingkat kesukaran, daya pembeda, validitas, reliabilitas, serta efektivitas pengecoh (Karindi & Rufi'i, 2024). Adanya perangkat lunak seperti Anates dapat memudahkan guru dalam melakukan analisis kualitas terhadap butir soal, sehingga hasilnya dapat digunakan untuk memperbaiki mutu instrumen penilaian.

Beberapa penelitian sebelumnya telah membahas analisis kualitas soal berbasis HOTS dengan menggunakan aplikasi Anates pada berbagai mata pelajaran. Penelitian yang dilakukan oleh (Purwati et al., 2021) menunjukkan bahwa penggunaan Anates Windows versi 4 efektif dalam menganalisis soal berbasis HOTS, karena mampu menampilkan informasi yang rinci terkait validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Hasil serupa juga diperoleh dalam penelitian (Hervi et al., 2023) pada soal literasi sains diketahui bahwa Anates dapat memberikan hasil analisis yang akurat dalam menilai kualitas soal serta membantu guru dalam mengidentifikasi butir soal yang perlu direvisi. Selanjutnya, penelitian oleh (Sari Saraswati et al., 2021) pada mata kuliah Matematika Lanjut menunjukkan bahwa analisis butir soal berbasis HOTS sangat penting untuk memastikan kualitas instrumen penilaian. Melalui penggunaan aplikasi Anates, penelitian tersebut menemukan bahwa instrumen yang valid, reliabel, dan memiliki daya pembeda yang baik dapat menghasilkan alat evaluasi yang benar-benar mampu mengukur kemampuan berpikir tingkat tinggi secara akurat dan proporsional.

Namun demikian, dari berbagai penelitian terdahulu, belum banyak penelitian yang secara spesifik menganalisis kualitas soal berbasis HOTS pada mata pelajaran Komputer Akuntansi di SMK, khususnya pada Fase F. Padahal, mata pelajaran ini memiliki peran strategis dalam membentuk kompetensi siswa akuntansi berbasis teknologi, terutama dalam penguasaan perangkat lunak komputer akuntansi. Di sisi lain, pembelajaran pada era teknologi saat ini sudah sepatutnya berbasis digital agar mampu menyesuaikan diri dengan kebutuhan zaman. Dengan demikian, terdapat kesenjangan penelitian pada kajian ini, yaitu belum adanya evaluasi komprehensif terkait kualitas soal HOTS di bidang komputer akuntansi SMK.



Berdasarkan kondisi tersebut, diperlukan penelitian yang menganalisis kebutuhan dan kualitas soal berbasis HOTS yang relevan dengan perkembangan teknologi serta karakteristik peserta didik SMK. Analisis ini penting dilakukan agar instrumen penilaian yang dikembangkan tidak hanya memenuhi aspek validitas dan reliabilitas, tetapi juga mendukung pembelajaran berbasis digital. Oleh karena itu, penelitian ini bertujuan untuk menganalisis kualitas soal HOTS pilihan ganda pada mata pelajaran Komputer Akuntansi Fase F kelas XI Akuntansi SMKN 6 Surabaya, meliputi aspek validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan efektivitas pengecoh. Hasil penelitian ini diharapkan dapat memberikan kontribusi dalam penyusunan instrumen penilaian yang lebih baik sebagai bagian dari pengembangan kurikulum, sehingga pembelajaran menjadi lebih efektif dalam meningkatkan pemahaman dan kemampuan analisis siswa.

METODE PENELITIAN

Penelitian ini merupakan penelitian deskriptif dengan pendekatan kuantitatif. Menurut sugiyono dalam (Wardani & Listiadi, 2023), penelitian deskriptif merupakan pengumpulan data untuk menggambarkan atau menegaskan bagaimana sebuah konsep dan menjawab pertanyaan yang berhubungan dengan subjek suatu penelitian pada waktu tertentu. Menurut Sugiyono dalam (Jayusman & Shavab, 2020) pendekatan deskriptif digunakan untuk mengetahui nilai variabel mandiri, baik satu variabel atau lebih (independen) tanpa adanya perbandingan, atau menggabungkan dengan variabel yang lain. Penelitian ini dilaksanakan di SMK Negeri 6 Surabaya. Penyelidikan dimulai dari pengamatan materi bahan ajar yang diterapkan pada SMK Negeri 6 Surabaya hingga pembuatan soal HOTS dan distribusi angket pada bulan September – November 2024.

Subjek yang digunakan penelitian ini adalah seluruh siswa Kelas XI Akuntansi 2 SMK Negeri 6 Surabaya yang berjumlah 30 siswa. Pemilihan kelas tersebut didasarkan pada pertimbangan bahwa kelas XI Akuntansi 2 telah mempelajari materi Akuntansi Elemen Komputer, sehingga relevan dengan fokus penelitian. Objek yang digunakan penelitian ini adalah 15 soal pilihan ganda HOTS yang dikembangkan oleh peneliti untuk mengukur kemampuan berpikir kritis siswa pada mata pelajaran tersebut.

Instrumen Penelitian yang digunakan berupa kuesioner yang diakses melalui Google Form untuk mempermudah pengumpulan data. Menurut (Arsela, 2022), kuesioner atau angket merupakan suatu bentuk alat untuk mengumpulkan data yang bersifat fleksibel dan cukup mudah untuk digunakan. Angket tersebut digunakan untuk mengukur respon siswa terhadap soal HOTS yang diterapkan pada mata pelajaran Akuntansi Elemen Komputer, meliputi aspek kejelasan pertanyaan, tingkat kesulitan, dan daya tarik soal. Penelitian ini hanya menjadikan siswa Kelas XI Akuntansi 2 yang menjadi objek survei. Kuesioner yang dipakai peneliti bersifat tertutup. Angket digunakan sebagai pengukuran respon siswa terhadap alat penilaian yang diterapkan pada mata pelajaran Akuntansi Elemen Komputer.

Teknik analisis data yang digunakan penelitian dalam analisis item ini meliputi sejumlah aspek yaitu validitas, keandalan, kemampuan diskriminatif, tingkat kesulitan dan efektivitas eliminasi masing-masing item. Bertanya. Data yang didapat dengan menyebarkan kuesioner kepada siswa Kelas XI program Akuntansi Keuangan Institusi dianalisis. Program Akuntansi Keuangan Institusi Kelas XI diajarkan dengan bantuan dianalisis menggunakan aplikasi Anates V4 untuk mengetahui mutu setiap soal dapat digunakan atau perlunya perbaikan pada elemen pertanyaan.

HASIL DAN PEMBAHASAN

Analisis kualitas butir soal pilihan ganda adalah alat yang dianggap penting untuk



mengidentifikasi butir soal yang dapat digunakan, diperbaiki, atau digantikan (Trianingsih, 2023). Analisis ini memungkinkan guru untuk mengambil langkah strategis guna memperbaiki soal yang memiliki kelemahan, baik dari segi validitas, reabilitas, daya beda, maupun tingkat kesukaran. Upaya ini dilakukan guna memastikan bahwa evaluasi pembelajaran mampu mencerminkan kemampuan siswa secara lebih akurat dan efektif. Soal yang dapat digunakan harus mempunyai angka validitas tes, reliabilitas tes, tingkat kesukaran, daya pembeda dan efektivitas distraktor (Ayu & Realita, 2024).

Berdasarkan data yang dianalisis, penelitian ini memilih penggunaan soal pilihan ganda yang berfokus pada materi Fase E Komputer Akuntansi untuk kelas XI Akuntansi di SMKN 6 Surabaya. Sampel yang terlibat dalam analisis berjumlah 30 peserta didik, dengan jumlah soal 15 butir dengan bentuk pilihan ganda. Hasil analisis butir soal mencakup sejumlah aspek perhitungan reliabilitas, uji validitas, analisis daya pembeda dan tingkat kesukaran soal. Hal tersebut digunakan untuk memberikan rekomendasi perbaikan atau penggunaan soal.

Validitas

Menurut (Arikunto, 2013), istilah validitas berawal dari kata yang mengacu pada akurasi dan kebenaran sebuah alat ukur dalam suatu tes. Validitas soal dari alat ukur dapat dimanfaatkan sebagai analisis untuk melakukan pengukuran yang bertujuan untuk menyempurnakan instrumen pengukuran. Pengukuran validitas soal tersebut memberikan informasi tentang soal mana yang layak diterapkan, soal yang tidak layak diterapkan, dan soal yang membutuhkan perbaikan. Menurut (Arikunto, 2013), suatu soal tes dianggap valid apabila memiliki korelasi yang tinggi pada skor keseluruhan. Sebab itu, soal yang memiliki korelasi tinggi dianggap lebih berarti untuk diterapkan pada tes berikutnya dan begitu juga sebaliknya. Kriteria penentuan validitas soal dalam penelitian ini mengacu pada klasifikasi nilai koefisien korelasi sebagaimana ditunjukkan pada tabel 1 berikut:

Tabel 1. Kriteria Validitas Butir Soal

Korelasi	Keterangan
0,00 - 0,200	Validitas soal sangat rendah
0,200 - 0,400	Validitas soal rendah
0,400 - 0,600	Validitas soal cukup
0,600 - 0,800	Validitas soal tinggi
0,800 - 1,00	Validitas soal sangat tinggi

Sumber: (Arikunto, 2013)

Suatu butir soal dapat dinyatakan mempunyai tingkat validitas tinggi jika hasil yang diperoleh dari butir soal mempunyai korelasi yang baik dengan nilai keseluruhan. Pada ANATES, perolehan korelasi antara nilai butir soal dan nilai keseluruhan dapat diterapkan guna penentuan validitas. (Khasanah et al., 2023). Berdasarkan tabel 1 kriteria validitas korelasi soal pertanyaan menurut (Arikunto, 2013), diketahui bahwa setiap butir soal yang mempunyai nilai korelasi 0,00 sampai dengan 0,200 dikatakan validitas soal sangat rendah, nilai korelasi 0,200 sampai dengan 0,400 dikatakan validitas soal rendah, nilai korelasi 0,400 sampai dengan 0,600 dikatakan validitas soal cukup, nilai korelasi 0,600 sampai dengan 0,800 dikatakan nilai validitas soal tinggi dan nilai korelasi 0,800 sampai dengan 1,00 dikatakan validitas soal sangat tinggi. Apabila nilai korelasi suatu soal memiliki nilai diatas 0,600 maka soal tersebut memiliki tingkatan soal yang tinggi dan dapat dikatakan layak digunakan. Jika soal tidak valid maka harus dibenahi dengan mengganti penyusunan



soal atau indikator pencapaiannya (Rahayu & Djazari, 2016). Aspek instrument penilaian siswa, dan jawaban yang diberikan siswa dapat menentukan valid atau tidaknya suatu soal (Zainal Ariin, 2012). Adapun hasil analisis validitas butir soal menggunakan aplikasi *Anates V4* disajikan pada tabel 2 berikut ini.

Tabel 2. Hasil Validitas

No. Butir Soal	Korelasi	Keterangan
1	0,532	Cukup
2	0,269	Rendah
3	0,657	Tinggi
4	0,528	Cukup
5	0,580	Cukup
6	0,357	Rendah
7	0,685	Tinggi
8	0,725	Tinggi
9	0,240	Rendah
10	0,589	Cukup
11	0,328	Rendah
12	0,326	Rendah
13	0,326	Rendah
14	0,541	Cukup
15	0,203	Rendah

Sumber: Hasil Olah Data Anates V4

Berdasarkan tabel 2 mengenai hasil validitas yang sudah diolah data nya oleh aplikasi Anates Versi 4, dapat diketahui bahwa soal HOTS pilihan ganda memiliki korelasi yang beragam. Soal valid memiliki arti bahwa kualitas soal yang telah disusun dapat menunjukkan kesesuaian antara suatu pengukuran dengan keefektifan dari suatu tes (Alpusari, 2014). Kategori rendah terdapat 7 butir soal yaitu pada soal nomor 2, 6, 9, 11, 12, 13, 15, kategori cukup terdapat 5 butir soal yaitu pada soal nomor 1, 4, 5, 10, 14, lalu, kategori tinggi terdapat 3 butir soal yaitu pada soal nomor 3, 7, 8. Butir soal dengan nilai validitas tinggi menunjukkan ketepatan dan kebenaran dalam mengukur kemampuan siswa. Sementara, butir soal dengan validitas sangat rendah perlu ditindak lanjuti terhadap butir soal yang sudah dibuat karna soal tersebut tidak valid. (Antari et al., 2021)

Reliabilitas

Menurut (Mulyani et al., 2024), keandalan sebuah tes soal adalah tingkat ketajaman atau keseimbangan hasil pengukuran. Alat ukur terpercaya akan memberikan hasil relatif konsisten ketika diaplikasikan berulang kali untuk mengukur hal yang serupa. Keandalan diperuntukkan



dalam mengetahui konsistensi suatu tes untuk menilai hal-hal yang perlu dinilai (Mulyani et al., 2024). Apabila semakin besar koefisien nilai keandalan atau mendekati 1, maka derajat konsistensi tes tersebut. Dalam penelitian ini pengujian reliabilitas soal dengan Anates V4 bertujuan mengukur sejauh mana perolehan dari analisis reabilitas soal HOTS elemen komputer dalam akuntansi. Untuk menafsirkan perhitungan nilai reliabilitas pada penelitian, digunakan perhitungannya yang didapat dengan tabel interpretasi nilai reliabilitas menurut (Arikunto, 2013). Kriteria interpretasi nilai reliabilitas menurut Arikunto (2013) dapat dilihat pada tabel 3 berikut:

Tabel 3. Interpretasi Nilai r

Besarnya nilai r	Interpretasi
0,00 sampai dengan 0,19	Sangat Rendah
0,20 sampai dengan 0,39	Rendah
0,40 sampai dengan 0,69	Cukup
0,70 sampai dengan 0,89	Tinggi
0,90 sampai dengan 1,00	Sangat Tinggi

Sumber: (Arikunto, 2013)

Berdasarkan tabel 3 interpretasi dari besarnya nilai r, diketahui apabila nilai r 0,00 – 0,19 memiliki interpretasi sangat rendah, nilai r 0,20 – 0,39 memiliki interpretasi rendah, nilai r 0,40 – 0,69 memiliki interpretasi cukup, nilai r 0,70 – 0,89 memiliki nilai interpretasi tinggi, dan nilai r 0,90 – 1,00 memiliki nilai interpretasi sangat tinggi. Soal atau tes yang valid dan reliabel maka soal yang dibuat sudah sesuai dengan tujuan pembelajaran dan kompetensi siswa (Fiska et al., 2021). Dengan demikian, semakin tinggi nilai r yang diperoleh, maka semakin tinggi pula tingkat hubungan antar-variabel yang diukur, serta semakin baik pula konsistensi instrumen evaluasi tersebut. Adapun hasil pengujian reliabilitas butir soal menggunakan aplikasi *Anates* versi 4 disajikan pada tabel 4 berikut ini:

Tabel 4. Hasil Reliabilitas

Rata-Rata	Simpang Baku	Korelasi XY	Reliabilitas
10,73	3,18	0,75	0,85

Sumber: Hasil Olah Data Anates V4

Berdasarkan tabel 4 dapat diketahui bahwa hasil analisis ini menunjukkan rata-rata sebesar 10,73; simpang baku sebesar 3,18; korelasi sebesar 0,75. Maka, hasil reliabilitas pada soal HOTS pilihan ganda ini menunjukkan angka 0,85 yang artinya kualitas soal ini masuk pada kriteria reliabilitas tinggi.

Daya Pembeda

(Elviana, 2020) mengatakan, Daya pembeda adalah indeks yang menunjukkan besaran siswa dalam melakukan tugas – tugas untuk memisahkan kelompok unggul dan kelompok asor. Kekuatan pembeda antar sebuah soal menunjukkan sejauh mana keunggulan soal tersebut dapat memisahkan antara kelompok soal unggul dan kelompok yang lemah. Ketika menafsirkan hasil perhitungan kekuatan pembeda, kriteria berikut ini dapat diterapkan. Untuk itu penafsiran perhitungan daya pembeda dengan tabel menurut (Arikunto, 2013). Sebagaimana ditunjukkan



pada tabel 5 berikut:

Tabel 5. Perhitungan Daya Pembeda

Besarnya nilai D	Hasil
0,00 - 0,20	Jelek (<i>Poor</i>)
0,21 - 0,40	Cukup (<i>Satisfactory</i>)
0,41 - 0,70	Baik (<i>Good</i>)
0,71 - 1,00	Sangat Baik (<i>Excellent</i>)
Negatif	Semuanya tidak baik.

Sumber: (Arikunto, 2013)

Berdasarkan Tabel 5, daya pembeda soal memperlihatkan sejauh mana butir soal dapat membedakan diantara peserta didik yang memiliki kemampuan tinggi dan rendah. Apabila nilai daya pembeda terdapat dalam rentang 0,00 sampai dengan 0,20, maka soal tersebut memiliki kategori jelek (*poor*) karena tidak mampu membedakan dengan baik diantara siswa yang berkemampuan tinggi dan rendah. Jika berada pada rentang 0,21 sampai dengan 0,40, maka soal dikategorikan cukup (*satisfactory*), artinya soal masih dapat membedakan meskipun kualitasnya belum optimal.

Selanjutnya, nilai daya pembeda 0,41 sampai dengan 0,70 dikategorikan baik (*good*), yang menunjukkan soal sudah dapat memisahkan dengan jelas kelompok siswa berkemampuan tinggi dan rendah. Sementara itu, nilai daya pembeda pada rentang 0,71 sampai dengan 1,00 tergolong sangat baik (*excellent*) karena soal benar-benar efektif dalam membedakan kemampuan peserta didik. Selanjutnya, hasil perhitungan daya pembeda untuk setiap butir soal dapat dilihat pada tabel 6. hasil daya pembeda berikut ini :

Tabel 6. Hasil Daya Pembeda

No.Butir Soal	Kel. Atas	Kel. Bawah	Beda	Indeks DP (%)	Kategori Soal
1	8	5	3	37,50	Cukup
2	8	6	2	25,00	Cukup
3	7	0	7	87,50	Sangat Baik
4	8	4	4	50,00	Baik
5	8	3	5	62,50	Baik
6	8	5	3	37,50	Cukup
7	8	1	7	87,50	Sangat Baik
8	8	1	7	87,50	Sangat Baik
9	8	5	3	37,50	Cukup
10	8	2	6	75,00	Sangat Baik



11	8	6	2	25,00	Cukup
12	7	3	4	50,00	Baik
13	7	4	3	37,50	Cukup
14	8	3	5	62,50	Baik
15	6	4	2	25,00	Cukup

Sumber: Olah Data Anates V4

Berdasarkan Tabel 6, hasil analisis daya pembeda menunjukkan bahwa dari 15 butir soal, terdapat 7 soal berkategori cukup (nomor 1, 2, 6, 9, 11, 13, 15), 4 soal berkategori baik (nomor 4, 5, 12, 14), dan 4 soal berkategori sangat baik (nomor 3, 7, 8, 10). Hasil ini menunjukkan bahwa sebagian besar soal mampu membedakan siswa berkemampuan tinggi dan rendah. Sesuai dengan pendapat (Zahiroh & Ritonga, 2021), kategori daya pembeda dalam analisis butir soal ujian akhir semester dapat dibagi menjadi beberapa tingkatan, yaitu jelek, cukup, baik, dan sangat baik. Sehingga, disarankan untuk butir soal yang memiliki daya pembeda sedang, baik, dan sangat baik disimpan dalam bank soal dan digunakan kembali pada tes berikutnya, lalu untuk soal dengan daya pembeda yang buruk dapat diperbaiki sehingga dapat digunakan kembali pada tes berikutnya, sedangkan soal dengan daya pembeda negatif atau sangat buruk lebih baik dibuang atau tidak usah digunakan. (Rahayu & Djazari, 2016).

Tingkat Kesukaran

Menurut (Radja et al., 2023), menganalisis tingkat kesukaran soal berarti membedakan soal-soal mana yang termasuk golongan mudah, sedang, dan sulit. Tingkat kesulitan diindikasikan dengan indeks kesukaran, yaitu angka yang memperlihatkan perbandingan siswa yang menjawab pertanyaan dengan benar (Arikunto, 2013). Jika indeks kesulitannya, semakin tinggi, maka soal tersebut dikatakan mudah untuk dikerjakan. Sebaliknya, jika indeks kesukaran semakin kecil, maka soal tersebut dikatakan sulit dalam pengerjaannya. Dalam menafsirkan hasil perhitungan tingkat kesulitan soal, dapat digunakan interpretasi sebagaimana ditunjukkan pada tabel 7 berikut:

Tabel 7. Tingkat Kesukaran

Soal dengan P 0,00 sampai 0,30	Soal sukar
Soal dengan P 0,31 sampai 0,70	Soal sedang
Soal dengan P 0,71 sampai 1,00	Soal mudah

Sumber: (Arikunto, 2013)

Berdasarkan Tabel 7, tingkat kesulitan soal ditentukan oleh nilai proporsi (P) dari murid yang bisa menjawab soal dengan tepat. Jika suatu soal memiliki nilai P yaitu 0,00 hingga 0,30, dengan begitu soal tersebut termasuk kategori sukar karena hanya beberapa saja peserta yang mampu menjawab dengan benar. Apabila nilai P berada dalam rentang 0,31 hingga 0,70, soal termasuk kategori sedang, artinya tingkat kesukarannya seimbang karena dapat dijawab dengan benar oleh sebagian peserta dan salah oleh sebagian lainnya. Sementara itu, soal yang mempunyai nilai P yaitu 0,71 hingga 1,00 dikategorikan sebagai mudah, karena setengah dari peserta didik dapat menjawab dengan tepat. Soal dikatakan baik adalah soal yang tidak terlalu gampang serta tidak terlalu sulit.



Siswa tidak akan dapat meningkatkan kemampuannya jika hanya bisa memecahkan soal yang terlalu mudah. Sehingga, sebaiknya pertanyaannya dibuat menjadi terlalu sukar, jadi siswa akan berusaha mencoba lebih keras lagi karena soal yang diselesaikan tampak tidak mudah untuk dituntaskan (Ambarwati & Ismiyati, 2021). Dengan demikian, penentuan tingkat kesukaran soal ini sangat penting untuk memastikan kualitas butir soal, agar tidak terlalu gampang maupun terlalu sulit, melainkan bervariasi dan sesuai dengan tujuan evaluasi pembelajaran. Hasil perhitungan tingkat kesukaran soal secara lengkap disajikan pada tabel 8 berikut:

Tabel 8. Hasil Tingkat Kesukaran

No. Butir Soal	Jumlah Betul	Tingkat kesukaran	Tafsiran
1	27	90,00	Sangat Mudah
2	25	83,33	Mudah
3	19	63,33	Sedang
4	23	76,67	Mudah
5	23	76,67	Mudah
6	25	83,33	Mudah
7	21	70,00	Sedang
8	19	63,33	Sedang
9	20	66,67	Sedang
10	19	63,33	Sedang
11	25	83,33	Mudah
12	18	60,00	Sedang
13	18	60,00	Sedang
14	21	70,00	Sedang
15	19	63,33	Sedang

Sumber: Hasil Olah Data Anates V4

Berdasarkan tabel 8, tingkat kesukaran butir soal HOTS pilihan ganda ini memiliki kategori sangat mudah terdapat 1 butir soal yaitu nomor 1 dengan persentase 90% , kategori mudah terdapat 5 butir soal yaitu pada nomor 2, 4, 5, 6, 11, dengan presentase 76,67%-83,33%, lalu, kategori sedang terdapat 9 butir soal yaitu pada nomor 3, 7, 8, 9, 10, 12, 13, 14, 15 dengan persentase 60,00%-70,00%.

Efektivitas Pengecoh

Pola yang menunjukkan bagaimana seorang penguji membandingkan setiap pilihan jawaban dengan alternatif yang tersedia pada tiap butir soal disebut distraktor (Interdiana et al., 2014). Sebuah pengecoh dinyatakan efektif jika dipilih oleh minimal lima persen peserta tes



(Fatimah, Laela Umi dan Alfath, 2019). Fungsi utama pengecoh adalah untuk memikat perhatian peserta didik yang belum mencerna materi dengan baik. Oleh karena itu, distraktor sebaiknya mampu menarik perhatian siswa yang belum menguasai konsep tanpa terlihat terlalu jelas sebagai jawaban yang salah. Pada aplikasi Anates, kualitas distraktor dapat diukur menggunakan Indeks Fungsi Distraktor, yang menunjukkan sejauh mana distraktor berperan dalam mengecoh peserta. Distraktor yang baik akan ditunjuk oleh peserta yang menjawab kurang tepat atau salah, sedangkan opsi yang tidak akan dipilih dianggap tidak berfungsi. Berdasarkan (Arikunto, 2013), suatu strategi pengecualian atau pengalihan dapat dikategorikan berhasil dan sukses jika pengecualian itu ditunjuk oleh setidaknya 5% dari peserta didik ujian. Dalam menilai efektivitas ekstrak setiap item, peneliti menerapkan kriteria yang diambil dari penggunaan Anates V4. Adapun kriteria penilaian efektivitas pengecoh disajikan pada tabel 9 berikut.

Tabel 9. Kriteria Penilaian Efektivitas Pengecoh

Pengecoh yang berfungsi	Kriteria
---	Sangat Buruk
--	Buruk
-	Kurang Baik
+	Baik
++	Sangat baik

Sumber: Anates V4

Berdasarkan Tabel 9, efektivitas pengecoh dapat dinilai dari simbol yang diberikan pada setiap alternatif jawaban. Jika pengecoh diberi tanda (---) maka efektivitasnya dikategorikan sangat buruk, artinya pengecoh sama sekali tidak berfungsi karena tidak dipilih oleh peserta. Apabila diberi tanda (--), maka efektivitasnya dianggap buruk, dan jika diberi tanda (-) berarti pengecoh masih memiliki fungsi tetapi hanya pada kategori kurang baik.

Sebaliknya, pengecoh yang mendapat tanda (+) dikategorikan baik, yang berarti pengecoh cukup efektif digunakan karena berhasil menarik perhatian sebagian peserta. Adapun pengecoh yang mendapat tanda (++) termasuk kategori sangat baik, karena berhasil berfungsi optimal dalam mengelabui peserta ujian. Dengan demikian, semakin tinggi kategori yang ditunjukkan, maka semakin efektif pula pengecoh dalam menjalankan perannya. Hasil penilaian kualitas pengecoh serta distribusi efektivitas pengecoh secara keseluruhan disajikan pada tabel 10 dan tabel 11 berikut:

Tabel 10. Hasil Kualitas Pengecoh

No. Butir Soal	A	B	C	D	E
1	0--	2--	0--	27**	1+
2	0--	1++	3---	25**	1++
3	1-	19**	3++	4+	3++
4	3-	3-	0--	1+	23**



5	23**	3-	0--	0--	4---
6	2--	2-	25**	1++	0--
7	0--	1-	3+	21**	5---
8	19**	7---	1-	3++	0--
9	5--	20**	3++	0--	2++
10	4+	5--	2+	19**	0--
11	1++	4---	25**	0--	0--
12	4+	18**	6--	2+	0--
13	3++	2+	4+	3++	18**
14	5---	2++	1-	21**	1-
15	2+	19**	3++	3++	3++

Sumber: Hasil Olah Data Anates V4

Keterangan:

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

Tabel 11. Hasil Analisis Distribusi Efektivitas Pengecoh

No. Butir Soal	A	B	C	D	E
1	Buruk	Buruk	Buruk	Kunci Jawaban	Baik
2	Buruk	Sangat Baik	Sangat Buruk	Kunci Jawaban	Sangat Baik
3	Kurang Baik	Kunci Jawaban	Sangat Baik	Baik	Sangat Baik
4	Kurang Baik	Kurang Baik	Buruk	Baik	Kunci Jawaban
5	Kunci jawaban	Kurang Baik	Buruk	Buruk	Sangat Buruk
6	Buruk	Kurang Baik	Kunci jawaban	Sangat Baik	Buruk
7	Buruk	Kurang Baik	Baik	Kunci Jawaban	Sangat Buruk
8	Kunci Jawaban	Sangat Buruk	Kurang Baik	Sangat Baik	Buruk
9	Buruk	Kunci Jawaban	Sangat Baik	Buruk	Sangat Baik
10	Baik	Buruk	Baik	Kunci Jawaban	Buruk
11	Sangat Baik	Sangat Buruk	Kunci Jawaban	Buruk	Buruk
12	Baik	Kunci Jawaban	Buruk	Baik	Buruk



13	Sangat Baik	Baik	Baik	Sangat Baik	Kunci Jawaban
14	Sangat Buruk	Sangat Baik	Kurang Baik	Kunci Jawaban	Kurang Baik
15	Baik	Kunci Jawaban	Sangat Baik	Sangat Baik	Sangat Baik

Sumber: Hasil Olah Data Peneliti

Berdasarkan hasil analisis pengecoh soal HOTS pilihan ganda pada tabel 10 dan 11, penelitian ini menunjukkan bahwa efektivitas pengecoh memiliki hubungan sebab akibat yang logis terhadap kualitas butir soal. Jika pengecoh disusun dengan tepat dan menarik bagi siswa yang kurang memahami materi, maka soal bisa membedakan kemampuan siswa secara lebih akurat. Sebaliknya, jika pengecoh tidak efektif, maka daya pembeda soal akan menurun dan soal menjadi kurang menantang. Temuan utama dalam penelitian ini adalah variasi efektivitas pengecoh, di mana hanya beberapa soal yang memenuhi kriteria baik hingga sangat baik. Beberapa faktor yang memengaruhi hal ini adalah ketidakkonsistenan variasi opsi jawaban, ketidaktepatan dalam merumuskan pertanyaan, serta ketidakrelevanan opsi jawaban terhadap miskonsepsi peserta didik. Kelebihan penelitian ini adalah analisis yang mendalam mengenai efektivitas pengecoh dan hubungannya dengan tingkat kesukaran serta indeks diskriminasi soal, sementara kekurangannya adalah keterbatasan dalam generalisasi hasil karena tidak melibatkan data kualitatif dalam analisis penyebabnya. Penelitian ini sesuai dengan hasil penelitian (Mochammad Noor Akhmadi, 2021) yang menekankan bahwa efektivitas pengecoh berpengaruh terhadap indeks diskriminasi dan tingkat kesukaran soal, serta mendukung pendapat (Rahayu & Djazari, 2016) bahwa pengecoh tidak efektif apabila tidak dipilih oleh siswa yang kurang memahami materi. Kesamaan ini menunjukkan bahwa penelitian ini memperkuat temuan sebelumnya mengenai pentingnya peran pengecoh dalam menentukan kualitas instrumen evaluasi. Implikasi dari penelitian ini adalah perlunya peningkatan kemampuan para penyusun soal dalam membuat pengecoh yang relevan dan menarik, sehingga butir soal yang dihasilkan dapat secara optimal mengukur keterampilan berpikir kritis tingkat tinggi siswa serta meningkatkan validitas instrumen evaluasi pembelajaran.

KESIMPULAN DAN SARAN

Berdasarkan hasil analisis butir soal HOTS pada mata pelajaran Komputer Akuntansi Fase F di kelas XI SMKN 6 Surabaya menggunakan aplikasi Anates versi 4, dapat disimpulkan bahwa dari total 15 butir soal yang dianalisis, sebanyak 53% soal dianggap valid, sedangkan 47% lainnya masih perlu diperbaiki agar memenuhi standar kualitas yang baik. Hasil reliabilitas menunjukkan nilai koefisien sebesar 0,85, yang termasuk dalam kategori tinggi, artinya tes tersebut memiliki tingkat konsistensi yang bagus.

Analisis tingkat kesukaran menunjukkan bahwa 60% soal termasuk kategori sedang, 33% kategori mudah, dan 7% kategori sukar. Dalam hal daya pembeda, terdapat 27% soal dengan daya pembeda sangat baik, 27% baik, dan 47% cukup. Dalam aspek efektivitas pengecoh, beberapa opsi jawaban belum berfungsi dengan baik. Secara umum, hasil penelitian menunjukkan bahwa instrumen penilaian berbasis HOTS sudah cukup baik, tetapi masih perlu ada perbaikan pada beberapa butir soal agar pengukuran kemampuan berpikir tingkat tinggi siswa lebih tepat dan seimbang.

Berdasarkan hasil penelitian dan kesimpulan yang diberikan, peneliti menyarankan kepada guru dan pengembang soal untuk melakukan revisi pada butir soal yang tidak valid serta memperbaiki pengecoh yang kurang efektif agar setiap item bisa mengukur kemampuan berpikir tingkat tinggi siswa secara akurat. Selain itu, penelitian berikutnya disarankan untuk menggunakan sampel dan materi yang lebih banyak serta beragam agar hasil analisis lebih lengkap dan mewakili.



Peneliti juga dapat membandingkan hasil analisis menggunakan aplikasi lain seperti TAP atau Iteman untuk memastikan konsistensi dan ketepatan hasil. Penelitian lanjutan juga bisa fokus pada pengembangan instrumen HOTS yang lebih sesuai dengan karakteristik siswa SMK, sehingga tes tidak hanya mengukur pengetahuan tetapi juga kemampuan berpikir kritis dan pemecahan masalah. Diharapkan hasil penelitian ini dapat berkontribusi dalam pengembangan ilmu evaluasi pendidikan serta menjadi acuan dalam meningkatkan kualitas asesmen berbasis HOTS di lingkungan pendidikan kejuruan.

DAFTAR PUSTAKA

- Alpusari, M. (2014). Analisis Butir Soal, Program Komputer Anates Versi 4.0 For Windows. *Jurnal Primary Program Studi Pendidikan Guru Sekolah Dasar Universitas Riau*, 3(2), 107. <https://pdfs.semanticscholar.org/0cb3/b3e65e93596f7007710b2dd229f1ea2a5205.pdf>
<https://doi.org/10.33578/jpkip.v3i2.2501>
- Ambarwati, Y. F., & Ismiyati, I. (2021). Analisis butir soal pilihan ganda ulangan akhir semester genap mata pelajaran kearsipan. *Measurement In Educational Research*, 1(2), 64–75. <https://doi.org/10.33292/meter.v1i2.144>
- Ananda, D., Muhyani, M., & Suhandi, T. (2020). Systematic Literature Review Implementasi Higher Order Thinking Skills (HOTS) terhadap Hasil Belajar Siswa. *Al-Adzka: Jurnal Ilmiah Pendidikan Guru Madrasah Ibtidaiyah*, 10(2), 106. <https://doi.org/10.18592/aladzkapgmi.v10i2.4005>
- Anggreini, D., & Darmawan, C. A. (2017). Analisis Kualitas Soal Try Out Ujian Nasional Dengan Menggunakan Aplikasi Program Anates. *JP2M (Jurnal Pendidikan Dan Pembelajaran Matematika)*, 2(1), 20. <https://doi.org/10.29100/jp2m.v2i1.213>
- Antari, B., Susanta, A., & Siagian, T. A. (2021). Analisis Soal Penilaian Akhir Semester Matematika SMA Negeri 8 Kota Bengkulu Tahun Ajaran 2020/2021. *JEMS (Jurnal Edukasi Matematika Dan Sains)*, 9(2), 299–312. <https://doi.org/10.25273/jems.v9i2.10421>
- Arikunto, S. (2013). *Prosedur penelitian: Suatu pendekatan praktik*. Rineka Cipta.
- Arsela, Z. (2022). *Pengembangan Modul Pembelajaran Bahasa Indonesia Materi Teks Ulasan Pada Siswa Kela VIII DI SMP Negeri 4 Kota Bengkulu*. 2(2), 163.
- Ayu, D., & Realita, R. (2024). Analisis Butir Soal pada Mata Pelajaran Pendidikan Agama Islam di Kelas V SDN 3 Tanjung Pura. *ALACRITY: Journal of Education*, 4(3), 140–151. <https://doi.org/10.52121/alacrity.v4i3.418>
- Elviana. (2020). Analisis Butir Soal Evaluasi Pembelajaran Pendidikan Agama Islam Menggunakan Program Anates. *Mudarrisuna*, 10(2), 58–74.
- Fatimah, Laela Umi dan Alfath, K. (2019). *ANALISIS KESUKARAN SOAL, DAYA PEMBEDA DAN FUNGSI DISTRAKTOR*. 17, 302. <https://doi.org/10.36668/jal.v8i2.115>
- Fiska, J. M., Hidayati, Y., Qomaria, N., & Hadi, W. P. (2021). Analisis Butir Soal Ulangan Harian Ipa Menggunakan Software Anates Pada Pendekatan Teori Tes Klasik. *Natural Science Education Research*, 4(1), 65–76. <https://doi.org/10.21107/nser.v4i1.8133>
- Hervi, F., Syamsurizal, S., Arsih, F., Advinda, L., & Sari, M. (2023). Analysis of the Quality of Scientific Literacy Question Items on Biodiversity Material Using the ANATES Application. *Jurnal Penelitian Pendidikan IPA*, 9(SpecialIssue), 1062–1069. <https://doi.org/10.29303/jppipa.v9ispecialissue.6304>
- Interdiana, A., Sari, C., & Herawati, M. (2014). Aplikasi Anates Versi 4 Dalam Menganalisis Butir Soal. *Faktor Jurnal Ilmiah Kependidikan*, 1(2), 203–214.
- Jayusman, I., & Shavab, O. A. K. (2020). Aktivitas Belajar Mahasiswa Dengan Menggunakan Media Pembelajaran Learning Management System (Lms) Berbasis Edmodo Dalam Pembelajaran Sejarah. *Jurnal Artefak*, 7(1), 13. <https://doi.org/10.25157/ja.v7i1.3180>
- Karindi, D. R., & Rufi'i. (2024). Penggunaan ANATES 4.0.9 dalam menguji kelayakan soal pilihan ganda sumatif matematika. *LINEAR: Journal of Mathematics Education*, 5(2), 181–196. <https://doi.org/10.32332/2zg4f753>
- Khasanah, I., Fuady, A., & Sunismi, S. (2023). Analisis Soal Ulangan Harian Matematika Bentuk Pilihan Ganda. *Mathema: Jurnal Pendidikan Matematika*, 5(2), 110.



- <https://doi.org/10.33365/jm.v5i2.2818>
- Mochammad Noor Akhmadi. (2021). Analisis Butir Soal Evaluasi Tema 1 Kelas 4 Sdn Plumbungan Menggunakan Program Anates. *Ed-Humanistics: Jurnal Ilmu Pendidikan*, 6(1), 799–806. <https://doi.org/10.33752/ed-humanistics.v6i1.1464>
- Mulyani, A. P., Septyanti, E., & Mustika, T. P. (2024). Analisis Butir Tes Pilihan Ganda Berbantuan Software Anates Mata Kuliah Menyimak Mahasiswa PBSI Universitas Riau. *Jurnal Multidisiplin Teknologi Dan Arsitektur*, 2(2), 694–704. <https://doi.org/10.57235/motekar.v2i2.3824>
- Purwati, L. M., Arianty, R., Syakilah, D. M., Ridlo, S., & Susilaningsih, E. (2021). Analisis Soal Tes Pilihan Ganda Berbasis Higher Order Thinking Skill menggunakan Aplikasi Anates Windows Versi 4.0.9 For Windows. *Jurnal Pendidikan UNIGA*, 15(2), 460. <https://doi.org/10.52434/jp.v15i2.1287>
- Radja, S. P., Bano, V. O., & Ina, A. T. (2023). Analisis Kualitas Butir Soal Hasil Belajar Peserta Didik Berdasarkan Tingkat Kesukaran, Daya Beda Dan Efektivitas Pengecoh Di Sman 1 Pandawai. *Jurnal Edusavana*, 1(1), 30–41.
- Rahayu, R., & Djazari, M. (2016). Analisis Kualitas Soal Pra Ujian Nasional Mata Pelajaran Ekonomi Akuntansi. *Jurnal Pendidikan Akuntansi Indonesia*, 14(1). <https://doi.org/10.21831/jpai.v14i1.11370>
- Rahmawati, Z. A. & B. (2023). Penerapan Model Pembelajaran Problem Based Learning dalam Meningkatkan Keterampilan Berpikir Tingkat Tinggi (HOTS). *Journal of Classroom Action*, 5(4), 449–453. <http://jppipa.unram.ac.id/index.php/jcar/indexAILATUL FITRIANI LENGKAP TUGAS AKHIR - Lailatul Fitriani%281%29.pdf>
- Rudini, M. (2020). Efektivitas Analisis Butir Soal Mata Pelajaran Matematika pada Siswa Kelas IV dalam Meningkatkan Kualitas Guru di SDN Sabang. *Tolis Ilmiah: Jurnal Penelitian*, 2(1), 17–27. https://ojs.umada.ac.id/index.php/Tolis_Ilmiyah/article/view/90
- Sari Saraswati, Iesyah Rodliyah, & Novia Dwi Rahmawati. (2021). Analisis Instrumen Penilaian Berbasis Higher Order Thinking Skills pada Mata Kuliah Matematika Lanjut. *Inomatika*, 3(2), 138–151. <https://doi.org/10.35438/inomatika.v3i2.275>
- Trianingsih, L. (2023). Analisis Kualitas Tes dan Butir Soal Integral pada Evaluasi Formatif Matematika Teknik. *Indonesian Journal Of Civil Engineering Education*, 9(2), 44. <https://doi.org/10.20961/ijcee.v9i2.84711>
- Wardani, E. Y. S., & Listiadi, A. (2023). Efektivitas Penggunaan Media Pembelajaran Administrasi Pajak Berbasis Articulate Storyline Terhadap Hasil Belajar Siswa Kelas Xi Akuntansi Smkn 1 Kemlagi. *EDUNOMIA: Jurnal Ilmiah Pendidikan Ekonomi*, 4(1), 25–34. <https://doi.org/10.24127/edunomia.v4i1.4864>
- Warju, W., Ariyanto, S. R., Soeryanto, S., & Trisna, R. A. (2020). Analisis Kualitas Butir Soal Tipe Hots Pada Kompetensi Sistem Rem Di Sekolah Menengah Kejuruan. *Jurnal Pendidikan Teknologi Dan Kejuruan*, 17(1), 95. <https://doi.org/10.23887/jptk-undiksha.v17i1.22914>
- Zainal Ariin. (2012). *Evaluasi Pembelajaran*. www.diktis.kemenag.go.id

