

## ANALISIS BUTIR SOAL DOKUMEN BERBASIS DIGITAL PADA SISWA SMK KELAS X MENGGUNAKAN PROGRAM ANATES

Anis Nurrahmah \*<sup>1</sup>  
Afifah Intan Febriyanti <sup>2</sup>  
Iklila Nur Muzayanah <sup>3</sup>  
Luqman Hakim <sup>4</sup>  
Vivi Pratiwi <sup>5</sup>

<sup>1,2,3,4,5</sup> Program Studi Pendidikan Akuntansi Universitas Negeri Surabaya

\*e-mail : [anis.23057@mhs.unesa.ac.id](mailto:anis.23057@mhs.unesa.ac.id), [afifah.23062@mhs.unesa.ac.id](mailto:afifah.23062@mhs.unesa.ac.id), [iklila.23118@mhs.unesa.ac.id](mailto:iklila.23118@mhs.unesa.ac.id),  
[luqmanhakim@unesa.ac.id](mailto:luqmanhakim@unesa.ac.id), [vivipratiwi@unesa.ac.id](mailto:vivipratiwi@unesa.ac.id).

### Abstrak

*Program anates digunakan untuk mengetahui ciri-ciri butir soal HOTS Dokumen berbasis digital pada siswa SMK kelas X. Pada penelitian ini pengumpulan data menggunakan Teknik dokumen dengan pendekatan kuantitatif. Alat tes berupa soal HOTS Dokumen berbasis digital berjumlah 10 soal berupa dokumen tes pilihan ganda. HOTS Dokumen berbasis digital direspon oleh 24 siswa dengan menggunakan aplikasi program anates versi 4.0.9 dengan jumlah 10 dokumen. Setelah dilakukan analisis, hasil yang didapat menunjukkan Berdasarkan hasil analisis soal HOTS Dokumen berbasis digital yang dilakukan pada siswa SMK kelas X yang diperoleh secara keseluruhan, maka dapat ditarik kesimpulan bahwa : Sebanyak 80% soal valid, menunjukkan soal mampu mengukur apa yang diharapkan. Namun, 20% butir soal perlu revisi. Reliabilitas tes sangat tinggi (0,80), sehingga konsisten dan dapat digunakan kembali. Mayoritas soal (60%) berada pada kategori sedang, yang sesuai dengan kriteria soal yang baik. Sebanyak 50% butir soal memiliki daya pembeda sangat baik, dan sisanya cukup hingga baik. Pengecoh hanya efektif pada 40% soal; perlu peningkatan kualitas opsi jawaban untuk 60% soal lainnya.*

**Kata Kunci:** Karakteristik Butir, Soal HOTS, Program Anates

### Abstract

*The anates program was used to determine the characteristics of digital-based HOTS Document question items for class X vocational school students. In this research, data collection used document techniques with a quantitative approach. The test tool is in the form of digital-based HOTS Document questions totaling 10 questions in the form of multiple choice test documents. HOTS Digital-based documents were responded to by 24 students using the Anates program application version 4.0.9 with a total of 10 documents. After the analysis was carried out, the results obtained showed that Based on the results of the digital-based HOTS Document question analysis carried out on class However, 20% of the questions need revision. The reliability of the test is very high (0.80), making it consistent and reusable. The majority of questions (60%) are in the medium category, which meets the criteria for good questions. As many as 50% of the questions have very good discriminating power, and the rest are fair to good. The distractor was only effective on 40% of the questions; need to improve the quality of answer options for the other 60% of questions.*

**Keywords:** Item Characteristics, HOTS Questions, Anates Program

## PENDAHULUAN

Terkait dengan isu perkembangan pendidikan di tingkat internasional, Kurikulum Merdeka dirancang dengan berbagai penyempurnaan. Penyempurnaan antara lain dilakukan pada standar isi yaitu mengurangi materi yang tidak relevan serta pendalaman dan perluasan materi yang relevan bagi peserta didik serta diperkaya dengan kebutuhan peserta didik untuk berpikir kritis dan analitis sesuai standar internasional. Penyempurnaan lainnya juga dilakukan pada standar penilaian, dengan mengadaptasi secara bertahap model-model penilaian standar internasional. Penilaian hasil belajar diharapkan dapat membantu peserta didik untuk meningkatkan kemampuan berpikir tingkat tinggi (Higher Order Thinking Skills/HOTS), karena berpikir tingkat tinggi dapat mendorong peserta didik untuk berpikir secara luas dan mendalam

tentang materi pelajaran. Higher order thinking skills (HOTS) atau keterampilan berpikir tingkat tinggi merupakan bagian dari taksonomi Bloom hasil revisi yang berupa kata kerja operasional yang terdiri dari analyze (C4), evaluate (C5) dan create (C6) yang dapat digunakan dalam penyusunan soal. Guru harus memiliki pengetahuan dan keahlian untuk menunjang pekerjaannya, sehingga dapat mengembangkan keterampilan berpikir tingkat tinggi peserta didik Sudah diputuskan. Hasilnya digunakan untuk pengambilan keputusan di bidang pendidikan dan pendidikan. Selama pengukuran, hasil pengujian dibandingkan dengan standar yang ditetapkan. Pengukuran bersifat kuantitatif. Evaluasi erat kaitannya dengan evaluasi. Saat melakukan evaluasi, Anda perlu melakukan pengukuran terlebih dahulu. Oleh karena itu, pengukuran hasil belajar siswa menitik beratkan pada ruang lingkup penilaian. Gronlund & Linn (1990: 5), dalam Kuseri & Suprananto, mengartikan penilaian sebagai penentuan sejauh mana siswa atau kelompok belajar telah mencapai tujuan pembelajaran yang telah ditentukan dalam hal pengetahuan, sikap, dan keterampilan sebuah tujuan. Kegiatan tersebut meliputi pengumpulan, analisis, dan penjelasan informasi melalui evaluasi. Selama proses pembelajaran, tes berfungsi sebagai alat ukur untuk mengetahui hasil belajar siswa. Tes sering digunakan untuk mengukur kemampuan siswa ditinjau dari tingkat perkembangannya dan pencapaian tujuan pembelajaran yang dicapai selama belajar dalam jangka waktu tertentu. Tes digunakan sebagai penilaian oleh guru untuk memberikan informasi tentang kemajuan keterampilan siswa. Sebelum melakukan tes, harus dilakukan analisis kualitas tes dan analisis butir soal untuk menghasilkan tes yang berkualitas yang memenuhi syarat tes yaitu validitas, reliabilitas, objektivitas, kepraktisan, dan efektivitas biaya. Arikunto (2008).

Untuk menilai kualitas soal, perlu dilakukan analisis butir soal terlebih dahulu guna mendukung perbaikan dalam proses pembelajaran. Penilaian mutu soal dapat dilakukan melalui pendekatan kualitatif maupun kuantitatif dengan mengevaluasi sejumlah aspek. Analisis ini bertujuan untuk memperoleh informasi tentang karakteristik setiap butir soal yang digunakan, sehingga kualitas instrumen dan hasil belajar peserta didik dapat diketahui. Saat ini, berbagai aplikasi telah dikembangkan untuk mempermudah guru dalam menganalisis butir soal, salah satunya adalah aplikasi Anates versi 4.0.9.

Adapun tujuan dari penelitian ini adalah untuk mengetahui validitas, reliabilitas, tingkat kesukaran, daya beda dan efektifitas pengecoh butir-butir soal HOTS Dokumen berbasis digital pada siswa SMK kelas X dengan menggunakan Program aplikasi Anates. Pada penelitian ini dikategorikan penelitian deskriptif kuantitatif.

## **METODE**

Dalam penelitian ini menggunakan metode dengan teknik dokumen dengan pendekatan kuantitatif. Pengambilan data dalam peneliti ini dari siswa SMK kelas X. Penelitian ini dilakukan pada siswa SMK kelas X. Subjek ini dipilih karena memiliki ciri-ciri yang bisa menggambarkan kondisi dan situasi yang sesuai dengan harapan penelitian.

Subjek dalam penelitian ini adalah sumber data seluruh dokumen soal dan seluruh lembar jawaban penilaian soal HOTS Dokumen berbasis digital dengan jumlah soal 10 butir dengan tes yang dibuat dalam bentuk pilihan ganda dengan pilihan jawaban atau opsi berjumlah 4. Objek dalam penelitian ini adalah semua masalah yang nantinya akan dikaji, dibahas dan di teliti berdasarkan hasil analisis butir soal soal HOTS Dokumen berbasis digital yang berjumlah 10 butir soal dengan jumlah 24 siswa. Pengumpulan data pada penelitian ini menggunakan Teknik pengisian kuisioner soal pilihan ganda. Penelitian ini menggunakan program Anates versi 4.0.9. Butir soal dianalisis secara kuantitatif untuk informasi tentang mengetahui daya pembeda, validitas, reliabilitas, tingkat kesukaran, dan efektifitas distraktor pada butir soal.

## **HASIL DAN PEMBAHASAN**

Proses analisis butir soal bertujuan untuk mendapatkan informasi yang berguna dalam memperbaiki atau menyempurnakan soal atau tes. Analisis ini digunakan sebagai umpan balik untuk meningkatkan kualitas butir soal yang akan dipakai dalam kegiatan evaluasi. Soal yang

disusun oleh guru harus mampu berfungsi dengan baik sebagai alat pengukur hasil belajar yang akurat. Fungsi suatu soal dapat dievaluasi melalui analisis yang dilakukan secara menyeluruh.

Analisis butir soal dapat dilakukan menggunakan metode kualitatif atau kuantitatif. Dalam penelitian ini, pendekatan kuantitatif digunakan dengan cara mengumpulkan dokumen hasil tes sehingga datanya dapat dibuktikan secara empiris. Analisis berdasarkan bukti empiris ini mampu menunjukkan kemampuan butir soal dalam membedakan antara siswa yang berkemampuan tinggi dan rendah.

Salah satu aplikasi yang membantu proses analisis ini adalah Anates versi 4.0.9, sebuah perangkat lunak yang dirancang untuk memudahkan analisis soal pilihan ganda maupun tes uraian secara cepat dan efisien. Anates memberikan kemudahan karena menggunakan Bahasa Indonesia, sehingga mudah diakses oleh guru. Aplikasi ini memungkinkan pengguna menganalisis butir soal secara otomatis, mencakup aspek daya pembeda, reliabilitas, tingkat kesukaran, korelasi skor dengan skor total, hingga kualitas pengecoh.

Dalam penelitian ini, analisis butir soal HOTS Dokumen berbasis digital yang dilakukan pada siswa SMK kelas X menggunakan Anates menghasilkan data sebagai berikut:

### Daya Pembeda

Kemampuan kuesioner untuk membedakan hasil tes yang benar dan salah merupakan fungsi dari kekuatan diskriminatifnya. Arti lain dari kekhasan adalah kemampuan suatu soal untuk membedakan siswa berkemampuan tinggi dan rendah. Pernyataan tersebut didukung oleh pendapat Aliqunto bahwa ciri suatu pertanyaan adalah apakah item-item kuesioner dapat membedakan responden yang berkemampuan tinggi dan rendah. Sudijono (2007) menyatakan bahwa

pencarian daya pembeda bertujuan untuk membedakan apakah suatu butir pertanyaan mampu membedakan kelompok siswa yang diukur dengan perbedaan kemampuan dalam kelompok tersebut. Koefisien korelasi biserial digunakan dalam penelitian ini untuk menunjukkan hubungan antara dua skor: skor item pertanyaan dan skor total peserta ujian yang sama. Berbagai faktor kinerja berkisar dari -1,00 hingga +1,00. Kekuatan selisih +1,00 berarti kelompok teratas menjawab semua pertanyaan dengan benar dan kelompok terbawah menjawab salah satu pertanyaan dengan benar. Sebaliknya, kekuatan selisihnya adalah -1,00, artinya kelompok teratas menjawab semua pertanyaan dengan salah, dan kelompok terbawah menjawab satu item pertanyaan dengan benar.

Nilai koefisien korelasi biserial digunakan dalam penelitian ini untuk menunjukkan hubungan antara dua skor: skor item pertanyaan dan skor total peserta ujian yang sama. Nilai koefisien daya diferensial berkisar antara -1,00 hingga +1,00. Perbedaan kinerja sebesar +1,00 berarti kelompok teratas menjawab semua pertanyaan dengan benar dan kelompok terbawah menjawab setiap item pertanyaan dengan benar. Sebaliknya, kekuatan selisihnya adalah -1,00, artinya kelompok teratas menjawab semua pertanyaan dengan salah, dan kelompok terbawah menjawab satu item pertanyaan dengan benar.

Kriteria indeks kekhasan dijelaskan pada tabel berikut:

**Tabel.1** Kriteria Indeks Daya Pembeda Soal

Indeks Daya Pembeda	Kriteria Daya Pembeda
Negatif - 9%	Sangat buruk
10% - 19%	Buruk
20% - 29%	Cukup
30% - 49%	Baik
50% ke atas	Sangat Baik

Koefisien daya pembeda memiliki rentang nilai antara -1,00 hingga +1,00. Nilai positif

menunjukkan semakin tinggi kemampuan butir soal untuk membedakan kelompok atas dan bawah. Butir soal dianggap memiliki daya beda yang baik jika nilai koefisiennya mendekati 1, yang berarti soal tersebut efektif membedakan antara kelompok dengan kemampuan tinggi dan rendah. > 0,5.

Dari hasil Anates diperoleh data sebagai berikut.

Kategori Daya Pembeda	Jumlah Soal	Persentase	Nomor Soal
Sangat Baik	5	50%	3, 5, 6, 7, 8
Baik	3	30%	1, 2, 4
Cukup	2	20%	9, 10

Dari hasil analisis di atas, jika nilai tes siswa tinggi berarti menjawab soal dengan benar, dan jika nilai tes siswa rendah berarti menjawab soal salah. Pertanyaan-pertanyaan tersebut menunjukkan bahwa terdapat perbedaan antara siswa berkemampuan tinggi yang memahami isi dan siswa berkemampuan rendah yang tidak memahami isi. Butir soal "Sangat puas" dan "Sangat tidak puas" merupakan soal yang tidak membedakan siswa berkemampuan tinggi dan rendah. Soal dengan nilai negatif sebaiknya tidak digunakan karena siswa pada kelompok yang lebih tinggi kemungkinan besar akan menjawab salah karena kualitasnya sangat rendah.

**Tingkat Kesukaran**

Tingkat kesukaran soal adalah peluang untuk menjawab benar suatu soal pada tingkat kemampuan tertentu yang biasanya dinyatakan dalam bentuk indeks. Indeks tingkat kesukaran ini pada umumnya dinyatakan dalam bentuk proporsi yang besarnya berkisar 0,00 - 1,00 (Bungin 2008). Semakin besar indeks tingkat kesukaran yang diperoleh dari hasil hitungan, berarti semakin mudah soal itu. Suatu soal memiliki TK= 0,00 artinya bahwa tidak ada siswa yang menjawab benar dan bila memiliki TK= 1,00 artinya bahwa siswa menjawab benar. Perhitungan indeks tingkat kesukaran ini dilakukan untuk setiap nomor soal. Pada prinsipnya, skor rata-rata yang diperoleh peserta didik pada butir soal yang bersangkutan dinamakan tingkat kesukaran butir soal itu. Soal yang terlalu mudah tidak merangsang siswa untuk mempertinggi usaha memecahkannya. Sebaliknya soal yang terlalu sukar akan menyebabkan siswa menjadi putus asa dan tidak mempunyai semangat untuk mencoba lagi karena di luar jangkauannya (Arikunto 2008).

Banyaknya responden yang mampu menjawab suatu soal dengan benar pada instrumen tergantung pada tingkat kesukaran soal berdasarkan tingkat sedang, sukar atau mudah dan dapat dilihat pada tabel berikut :

**Tabel 2.** Kriteria Tingkat Kesukaran Butir Soal

Kriteria	Kategori
$P \geq 0,76$	Mudah
$0,25 \leq p \leq 0,75$	Sedang
$P \leq 0,24$	Sukar

Dari hasil analisis Tingkat kesukaran menunjukkan proporsi siswa yang menjawab soal dengan benar. Distribusi tingkat kesukaran:

Tingkat Kesukaran	Jumlah Soal	Persentase	Nomor Soal
Mudah	2	20%	1, 2
Sedang	6	60%	3, 4, 5, 6, 7, 8
Sukar	2	20%	9, 10

Interpretasi: Mayoritas soal (60%) berada pada kategori sedang, yang sesuai dengan prinsip soal yang baik karena memberikan tantangan moderat bagi siswa. Namun, butir soal yang terlalu mudah dan terlalu sulit memerlukan evaluasi.

**Validitas**

Untuk mengetahui kualitas soal dalam penelitian ini, kami melakukan analisis kuantitatif berdasarkan data empiris, khususnya untuk soal pilihan ganda. Dalam penelitian ini digunakan program aplikasi Anates versi 4.0.9 untuk memudahkan analisis dan menghitung sifat-sifat aspek yang akan dihitung. Analisis kuantitatif soal meliputi aspek validitas butir soal, aspek reliabilitas butir soal, aspek tingkat kesukaran, aspek kinerja butir soal, dan aspek fungsi pengecoh atau pengecoh. Suatu instrumen dikatakan baik apabila mempunyai tingkat validitas yang tinggi atau validitas (Arikunto, 2008). Saat mengumpulkan data, perhatian harus diberikan pada prinsip reliabilitas instrumen untuk memastikan bahwa pengukuran dan observasi menghasilkan data yang valid. Menurut (Arikunto 2008), hasil pengalaman dapat menentukan validitas tes. Ada dua kelompok validitas uji yaitu validitas logis dan validitas empiris. Validitas logika meliputi validitas isi dan validitas konstruk, sedangkan validitas empiris meliputi validitas konkordansi dan validitas proksimat.

Menurut Sukardi (2008), suatu pengujian dikatakan valid jika koefisiennya 0,5, dan dapat diterima jika pengujian dilakukan satu kali, asumsi lain tetap sama, dimana nilai koefisiennya lebih tinggi dari satu koefisien. 0,5 tidak diterima. Oleh karena itu, tes dianggap valid jika koefisiennya paling sedikit 0,5. Menurut Azwar (2008), jika item kuesioner dianggap valid maka nilai koefisien yang diperoleh antara 0,30 sampai dengan 0,50. Pengujian yang valid berarti pengujian tersebut mencapai ketelitian dan ketelitian dalam menjalankan fungsi pengukurannya atau mencapai tujuannya. Semakin akurat dan teliti pengujiannya, semakin tinggi nilai koefisiennya. Validitas menunjukkan kemampuan soal untuk mengukur apa yang seharusnya diukur. Hasil analisis menggunakan aplikasi Anates versi 4.0.9 menunjukkan:

-	Nilai	Validitas	Total:	0,80	(kategori	sangat	tinggi).
---	-------	-----------	--------	------	-----------	--------	----------

**Rincian Validitas per Butir Soal:**

Nomor Soal: 1-8 (Valid)	9-10 (Tidak Valid).
-------------------------	---------------------

Interpretasi: Sebanyak 80% butir soal telah memenuhi kriteria validitas yang sangat baik, sedangkan 20% lainnya memerlukan perbaikan pada redaksi atau materi.

**Reliabilitas**

Menurut Ruffi (2020), reliabilitas mengacu pada konsistensi mengukur dan mengamati apa yang diukur dalam suatu tes. Suatu tes dikatakan reliabel atau reliabel apabila jawaban siswa terhadap soal-soal tersebut konsisten dan stabil sepanjang waktu. Akurasi, prediktabilitas, konsistensi, dan stabilitas merupakan hasil reliabilitas tes. Jika suatu pengukuran dianggap andal, maka pengukuran tersebut menunjukkan atau menghasilkan keandalan yang signifikan atau tinggi. Informasi yang diperoleh dari suatu tes dianggap dapat diandalkan jika tes tersebut secara konsisten mengukur apa yang ingin diukur. Suatu soal mempunyai indeks reliabilitas yang tinggi jika hasil pengukurannya kurang lebih sama setelah dilakukan pengukuran berulang kali. Sebaliknya jika selisihnya relatif besar, maka dapat dikatakan reliabilitas soal tersebut rendah. Reliabilitas menunjukkan seberapa andal hasil pengukuran pengujian yang dilakukan. Tujuan dari reliabilitas adalah untuk mengetahui seberapa besar variasi yang disebabkan oleh kesalahan pengukuran dan seberapa besar variasi yang sebenarnya dimiliki suatu tes (Supranata 2004).

Reliabilitas tes dapat ditentukan sebagai berikut: (1) Metode pengujian dua kali, yaitu memberikan tes kepada siswa yang mengikuti tes. Korelasi ditentukan dari hasil dua pengujian. (2) Metode pengujian. Artinya, Anda melaksanakan dua tes kepada siswa pada waktu yang berbeda dan menentukan korelasi berdasarkan hasil kedua tes tersebut. (3) Metode "split-half" (tes). Artinya membagi tes menjadi dua bagian berdasarkan kesamaan kesulitan, isi, dan format. Korelasi ditentukan berdasarkan hasil setiap setengah pengujian. (Purwan 2010).

aplikasi anates dihasilkan indeks reliabilitas Klasifikasi hasil koefisien reliabilitas dapat dilihat pada tabel dibawah ini:

**Tabel 3.** Klasifikasi Tingkat Reliabilitas

Tes Kategori Reliabilitas	Koefisien Korelasi
Sangat Tinggi	0,800 – 1,000
Tinggi	0,600 – 0,799
Cukup	0,400 – 0,599
Rendah	0,200 – 0,399
Sangat Rendah	0,000- 0,199

Reliabilitas menggambarkan konsistensi alat ukur dalam menghasilkan hasil yang stabil. Hasil dari Anates menunjukkan:

- Nilai Reliabilitas Total: 0,80 (kategori sangat tinggi).
- Metode Uji: Split-half (ganjil-genap).

Interpretasi: Reliabilitas yang tinggi mengindikasikan soal dapat digunakan kembali dengan hasil yang konsisten dan terpercaya.

### Efektifitas Pengecoh

Jawaban yang mungkin dikombinasikan dengan setiap item jawaban dalam pertanyaan pilihan ganda digunakan untuk menggambarkan pola yang disebut pengecoh. Distraktor dianggap telah menjalankan fungsinya dengan benar jika siswa memilih minimal 5% pilihan jawaban untuk setiap item. Pedoman untuk menentukan efektifitas suatu distraksi adalah dengan mengalikan  $5D44 \times$  jumlah siswa. Menurut (Supranata 2004), bagi siswa berbakat, menyontek pada kemungkinan jawaban akan mengganggu. Pertanyaan pilihan ganda memiliki jawaban alternatif untuk membantu membodohi Anda dengan jawaban yang benar. Fungsi distraktor ini sendiri adalah untuk menipu jawaban yang benar. Pilihan jawaban alternatif berkisar antara 3 hingga 5 dan disertakan dalam setiap bagian tanya jawab.

Ada beberapa faktor yang mempengaruhi memilih untuk tidak terganggu. Artinya unsur-unsur pertanyaannya sulit dibedakan, dan unsur-unsur pertanyaannya biasanya sama. Penyebab lainnya bisa jadi karena konten belum sepenuhnya dipahami siswa dan pilihan jawaban yang tersedia kurang menarik. Purwanti (2014) berpendapat bahwa konten yang tidak dapat dipahami siswa akan mengganggu dan tidak menarik bagi siswa. Distraksi yang kurang berfungsi dengan baik digantikan dengan distraksi yang lain oleh guru, dan distraktor yang berfungsi dengan baik digunakan guru pada saat melakukan kegiatan penilaian, beserta instrumen tes lainnya dalam kurun waktu tertentu yang didokumentasikan sebagai bank soal. Itu adalah 5 tahun.

Berdasarkan hasil analisis soal disimpulkan bahwa soal evaluasi memberikan pengaruh distraksi pada kriteria sangat baik, baik, buruk, dan sangat buruk. Sebuah item pertanyaan dianggap sebagai pengalih perhatian yang baik jika dipilih oleh 5% responden (Supranata 2004). Jika suatu pengecoh dipilih secara merata oleh responden, maka dapat dikatakan bahwa pengecoh tersebut sangat baik. Tidak mudah memberikan jawaban yang mengganggu terhadap pertanyaan penilaian. Hal ini karena mempengaruhi karakteristik dan kesulitan elemen itu sendiri. Dengan kata lain, ketika efektifitas pengecoh rendah, maka indeks diskriminabilitasnya juga rendah, dan tingkat kesukaran soal juga rendah. Begitu pula jika jawaban suatu unsur pertanyaan mempunyai efek mengganggu, maka sebaliknya itu baik maka indeks daya pembeda akan baik dan tingkat kesukaran soal menjadi tinggi. Rincian Butir Soal dengan Pengecoh Efektif: Nomor 3, 4, 5, 6, 7. Interpretasi: Pengecoh pada 60% soal perlu direvisi agar lebih menarik perhatian siswa yang belum memahami materi.

### KESIMPULAN

Berdasarkan hasil analisis soal HOTS Dokumen berbasis digital yang dilakukan pada siswa SMK kelas X yang diperoleh secara keseluruhan, maka dapat ditarik kesimpulan bahwa : Sebanyak 80% soal valid, menunjukkan soal mampu mengukur apa yang diharapkan. Namun, 20% butir soal perlu revisi. Reliabilitas tes sangat tinggi (0,80), sehingga konsisten dan dapat digunakan kembali.

Mayoritas soal (60%) berada pada kategori sedang, yang sesuai dengan kriteria soal yang baik. Sebanyak 50% butir soal memiliki daya pembeda sangat baik, dan sisanya cukup hingga baik. Pengecoh hanya efektif pada 40% soal; perlu peningkatan kualitas opsi jawaban untuk 60% soal lainnya.

#### DAFTAR PUSTAKA

- Arifin, Zainal. 2011. *Evaluasi Pembelajaran*. Bandung: Remaja Rosdakarya.
- Arikunto, Suharsimi. 2008. *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Azwar, Saifuddin. 2008. *Reliabilitas Dan Validitas*. Yogyakarta: Pustaka Pelajar. Bungin, Burhan. 2008. *TiMetodologi Penelitian Kuantitatif*. Jakarta: Kencana.
- Purwanti, Mustika. 2014. "Analisis Butir Soal Ujian Akhir Mata Pelajaran Akuntansi Keuangan Menggunakan Microsoft Excel 2010." *Jurnal Pendidikan Ajuntansi Indonesia, XII (1)*. 2014.
- Purwanto. 2010. *Prinsip-Prinsip Dan Teknik Evaluasi Pengajaran*. Bandung: Rosdakarya.
- Rufi'i, noor achmad. 2020. *Evaluasi Dan Pengukuran*. Surabaya: Adi Buana University Press.
- Sudijono, Anas. 2007. *Pengantar Evaluasi Pendidikan*. Jakarta: Raja Grafindo.
- Sukardi. 2008. *Evaluasi Pendidikan Prinsip & Operasionalnya*. Jakarta: Bumi Aksara.
- Supranata, Sumarna. 2004. *Analisis, Validitas, Reliabilitas Dan Interpretasi Hasil Tes*. Bandung: Remaja Rosdakarya.
- Setiawan, Mochamad Fajar. 2020. "Efektivitas Penggunaan Anates dalam Meningkatkan Kualitas Soal Evaluasi Pembelajaran." *Jurnal Teknologi Pendidikan*, 15(2): 89-97.
- Sugiyono. 2019. *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Sumintono, Bambang, dan Widhiarso, Wahyu. 2015. *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. Jakarta: Trim Komunikata.
- Yamin, Martinis, dan Bansu I. Maisah. 2010. *Standar Proses Pendidikan*. Jakarta: Gaung Persada Press.
- Zainuddin, Arif. 2012. *Evaluasi Pembelajaran Berbasis Kompetensi*. Jakarta: Kencana Prenada Media Group.
- Zuraida, Fitri. 2021. "Pengaruh Penggunaan Aplikasi Anates terhadap Validitas Soal HOTS." *Jurnal Pendidikan Indonesia*, 12(1): 34-41.
- Ahmad, Sofyan. 2016. *Statistik Pendidikan dan Evaluasi Pembelajaran*. Jakarta: Raja Grafindo Persada.
- Basuki, Imam, dan Hariyanto. 2014. *Asesmen Pembelajaran*. Bandung: PT Remaja Rosdakarya.
- Dimiyati dan Mudjiono. 2006. *Belajar dan Pembelajaran*. Jakarta: Rineka Cipta. Haryati, Tini. 2019. *Penerapan Analisis Butir Soal dalam Evaluasi Pendidikan*. Bandung: Alfabeta.
- Kusaeri, dan Suprananto. 2012. *Pengukuran dan Penilaian Hasil Belajar*. Yogyakarta: Graha Ilmu.
- Majid, Abdul. 2014. *Penilaian Autentik Proses dan Hasil Belajar*. Bandung: PT Remaja Rosdakarya.
- Mulyasa, E. 2008. *Kurikulum Tingkat Satuan Pendidikan*. Bandung: PT Remaja Rosdakarya.
- Nurgiyantoro, Burhan. 2009. *Penilaian dalam Pengajaran Bahasa dan Sastra*. Yogyakarta: BPFE.
- Rusman. 2015. *Belajar dan Pembelajaran Berbasis Komputer*. Jakarta: PT Raja Grafindo Persada.
- Sani, Ridwan Abdullah. 2014. *Pembelajaran Sainifik untuk Implementasi Kurikulum 2013*. Jakarta: Bumi Aksara.