



ANALISI BUTIR SOAL PILIHAN GANDA PADA ELEMEN APLIKASI PENGOLAH ANGKA (SPREADSHEET) FASE E MELALUI PENGGUNAAN PROGRAM KOMPUTER ANATES

Salsa Cintiya Putri Utami
Universitas Negeri Surabaya

Catur Pamungkas Hari Subekti
Universitas Negeri Surabaya

Lailatul Mukarromah
Universitas Negeri Surabaya

Luqman Hakim
Universitas Negeri Surabaya

Vivi Pratiwi
Universitas Negeri Surabaya

Alamat: Jl. Ketintang, Ketintang, Kec. Gayungan, Surabaya, Jawa Timur 60231
Universitas Negeri Surabaya - Kampus 1

Korespondensi penulis: salsa.23016@mhs.unesa.ac.id, catur.23024@mhs.unesa.ac.id,
lailatul.23094@mhs.unesa.ac.id, luqmanhakim@unesa.ac.id, vivipratiwi@unesa.ac.id.

Abstract. This research aims to analyze the quality of Higher Order Thinking Skills (HOTS) based questions using the Anates application for class X Accounting students at SMK 1 Taman. The questions tested were 15 multiple choice questions which were used to train students' analysis, evaluation and synthesis skills, with data obtained from 21 students. The analysis focused on reliability, level of difficulty, discriminating power, distractor effectiveness and validity. The results of the analysis show that the test reliability is 0.55, which indicates fairly good instrument consistency. However, only 3 of the 15 questions met the criteria for significant validity, while the rest had a low correlation with the total score. The majority of the questions are in the medium category at 53.33% and the rest are in the difficult and easy categories, so the questions are in the sufficient category to challenge students to think critically. As many as 40% of the questions had good discriminating power and the rest were in the poor category, indicating a lack of ability in the questions to differentiate students with high and low ability. Additionally, most distractors did not function optimally, with 53.33% of distractors not being selected by students. Based on these findings, most of the questions need to be revised in order to meet the criteria for good HOTS questions. This research is quantitative research using descriptive methods. Utilizing the Anates application has proven to help identify weaknesses in questions systematically and provide guidance for making improvements. It is hoped that this research can contribute to improving the quality of learning evaluation instruments, especially in the application of HOTS-based questions at the vocational high school level.

Keywords: HOTS questions, Anates, Validity, Reliability.

Abstrak. Penelitian ini bertujuan untuk menganalisis kualitas soal berbasis Higher Order Thinking Skills (HOTS) menggunakan aplikasi Anates pada siswa kelas X Akuntansi di SMK 1 Taman. Soal yang diuji berupa 15 soal pilihan ganda yang digunakan untuk melatih kemampuan analisis, evaluasi, dan sintesis siswa, dengan data yang diperoleh dari 21 siswa. Analisis difokuskan pada reliabilitas, tingkat kesukaran,

daya pembeda, efektivitas pengecoh dan validitas. Hasil analisis menunjukkan bahwa reliabilitas tes sebesar 0,55, yang mengindikasikan konsistensi instrumen yang cukup baik. Namun, hanya 3 dari 15 soal yang memenuhi kriteria validitas signifikan, sementara sisanya memiliki korelasi rendah dengan skor total. Mayoritas soal berada dalam kategori sedang sebesar 53,33% dan sisanya dalam kategori sukar dan mudah, sehingga soal masuk kategori yang cukup untuk menantang siswa untuk berpikir kritis. Sebanyak 40% soal memiliki daya pembeda baik serta sisanya dalam kategori jelek, menunjukkan kurangnya kemampuan soal dalam membedakan siswa dengan kemampuan tinggi dan rendah. Selain itu, sebagian besar pengecoh tidak berfungsi secara optimal, dengan 53,33% pengecoh tidak dipilih oleh siswa. Berdasarkan temuan ini, sebagian besar soal perlu direvisi agar dapat memenuhi kriteria soal HOTS yang baik. Penelitian ini merupakan penelitian kuantitatif dengan menggunakan metode deskriptif. Pemanfaatan aplikasi Anates terbukti membantu mengidentifikasi kelemahan soal secara sistematis dan memberikan panduan untuk melakukan perbaikan. Penelitian ini diharapkan dapat memberikan kontribusi dalam meningkatkan kualitas instrumen evaluasi pembelajaran, khususnya dalam penerapan soal berbasis HOTS di tingkat sekolah menengah kejuruan.

Kata kunci: Soal Hots, Anates, Validitas, Reliabilitas.

LATAR BELAKANG

Dalam pendidikan abad ke-21, kemampuan berpikir kritis dan kemampuan untuk menyelesaikan masalah kompleks menjadi kompetensi penting yang harus dimiliki oleh siswa. Salah satu pendekatan yang digunakan untuk mengembangkan keterampilan ini adalah melalui soal berbasis Higher Order Thinking Skills (HOTS). HOTS mencakup kemampuan berpikir tingkat tinggi seperti analisis, sintesis, evaluasi, dan pemecahan masalah yang lebih mendalam, berbeda dengan soal-soal yang hanya menguji kemampuan dasar (lower-order thinking skills) seperti mengingat dan memahami. Penerapan HOTS di tingkat sekolah menengah kejuruan (SMK) sangat penting untuk mempersiapkan siswa menghadapi tantangan di dunia kerja yang semakin kompleks (Anderson & Krathwohl, 2001).

Meskipun penerapan soal HOTS diharapkan dapat meningkatkan kualitas pembelajaran, kenyataannya tidak semua soal yang disusun mampu mengukur kemampuan berpikir tingkat tinggi siswa dengan efektif. Oleh karena itu, diperlukan analisis mendalam terhadap kualitas soal HOTS yang digunakan dalam pembelajaran untuk memastikan bahwa soal-soal tersebut dapat memenuhi tujuan yang diinginkan, yaitu mendorong siswa untuk berpikir kritis dan analitis. Salah satu alat yang dapat digunakan untuk menganalisis kualitas soal adalah aplikasi Anates. Aplikasi ini dirancang untuk membantu pengujian atau pendidik dalam menganalisis berbagai aspek soal ujian, termasuk validitas, reliabilitas, tingkat kesukaran, daya pembeda, serta efektivitas pengecoh. Dengan menggunakan Anates, pendidik dapat mengetahui sejauh mana soal-soal yang digunakan dalam ujian dapat diandalkan (reliabel) dan valid dalam mengukur

kompetensi yang dimaksud (Widodo, 2019). Selain itu, aplikasi ini juga membantu untuk mengidentifikasi kelemahan soal dan memberikan panduan untuk perbaikan, sehingga kualitas soal dapat lebih baik dan dapat diandalkan dalam evaluasi pembelajaran.

Penelitian ini bertujuan untuk mengeksplorasi dan menganalisis kualitas soal dengan menggunakan aplikasi Anates, dengan fokus pada validitas dan reliabilitas soal dalam konteks pendidikan. Dengan menggunakan Anates, diharapkan hasil evaluasi soal dapat memberikan gambaran yang lebih jelas tentang kualitas soal dan memberikan rekomendasi untuk perbaikan yang dapat meningkatkan efektivitas evaluasi pembelajaran di sekolah.

KAJIAN TEORITIS

Higher Order Thinking Skills (HOTS) mengacu pada kemampuan berpikir yang melampaui pemahaman dasar atau pengetahuan fakta, yang melibatkan keterampilan analisis, evaluasi, dan sintesis. Dalam konteks pendidikan, HOTS adalah aspek penting yang bertujuan untuk mempersiapkan siswa menghadapi tantangan kehidupan nyata yang memerlukan keterampilan berpikir kritis dan kreatif. Anderson dan Krathwohl (2001) dalam *Taxonomy for Learning, Teaching, and Assessing* menyebutkan bahwa HOTS berkaitan dengan pemrosesan kognitif tingkat tinggi yang meliputi kemampuan untuk menganalisis informasi, mengevaluasi berbagai alternatif, dan menghasilkan ide atau solusi baru. Penerapan HOTS dalam soal ujian bertujuan untuk mengembangkan pemikiran siswa lebih mendalam daripada sekadar mengingat informasi. Oleh karena itu, soal-soal yang dirancang harus mampu merangsang siswa untuk berpikir secara kritis dan kompleks.

Kualitas soal merupakan faktor penting dalam menentukan efektivitas instrumen evaluasi. Soal HOTS yang baik harus mampu mengukur kemampuan analisis, evaluasi, dan sintesis siswa secara tepat. Menurut Wiggins (1998), soal yang efektif harus dapat menilai pemahaman mendalam dan tidak hanya mengukur ingatan faktual. Oleh karena itu, soal HOTS biasanya lebih kompleks dan membutuhkan pemikiran tingkat tinggi untuk menjawabnya dengan tepat. Pengukuran kualitas soal dapat dilakukan melalui beberapa indikator, seperti reliabilitas, validitas, daya pembeda, tingkat kesukaran, dan efektivitas pengecoh. Penggunaan aplikasi seperti Anates dapat membantu dalam menganalisis berbagai aspek ini dengan lebih objektif dan sistematis.

Reliabilitas mengukur konsistensi suatu tes dalam menghasilkan hasil yang stabil dan dapat diandalkan. Menurut Crocker dan Algina (2006), reliabilitas yang tinggi menunjukkan bahwa soal tersebut memberikan hasil yang konsisten jika digunakan dalam pengujian yang berbeda. Dalam penelitian ini, reliabilitas tes yang menunjukkan nilai 0,55 masih dianggap cukup baik, meskipun masih bisa ditingkatkan.

Tingkat kesukaran soal menunjukkan seberapa mudah atau sulit soal tersebut untuk dijawab oleh siswa. Dalam konteks HOTS, soal yang berada dalam kategori sedang, seperti yang ditemukan dalam penelitian ini, cenderung ideal untuk mendorong siswa berpikir kritis tanpa membuat mereka merasa terlalu terbebani atau terlalu mudah.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode deskriptif untuk menganalisis kualitas soal berbasis Higher Order Thinking Skills (HOTS) yang digunakan pada siswa kelas X Akuntansi di SMK Taman Sidoarjo. dengan sampel sebanyak 21 siswa. Instrumen yang digunakan adalah seperangkat soal pilihan ganda yang terdiri dari 15 butir soal dengan 5 alternatif jawaban (a, b, c, d, e) yang dirancang untuk mengukur kemampuan analisis, evaluasi, dan sintesis siswa. Penelitian ini berfokus pada evaluasi beberapa aspek kualitas soal, yaitu reliabilitas, validitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh, dengan menggunakan aplikasi Anates sebagai alat bantu analisis. Pengumpulan data dilakukan melalui Google Form (g-form), yang memungkinkan pengumpulan jawaban siswa secara otomatis dan efisien.

Instrumen utama dalam penelitian ini adalah soal ujian berbasis HOTS yang terdiri dari 15 soal pilihan ganda. Soal-soal tersebut dirancang untuk mengukur kemampuan siswa dalam analisis, evaluasi, dan sintesis, yang merupakan aspek-aspek utama dalam HOTS. Soal-soal tersebut kemudian dianalisis menggunakan aplikasi Anates (Analisis Test) untuk mengukur berbagai parameter kualitas soal, termasuk Reliabilitas: Koefisien Alpha Cronbach digunakan untuk mengukur konsistensi internal soal. Nilai reliabilitas diharapkan lebih besar dari 0,5 untuk menunjukkan tingkat konsistensi yang cukup baik (Arikunto, 2010). Validitas: Diuji dengan melihat korelasi antara skor setiap soal dengan skor total untuk mengukur sejauh mana soal-soal tersebut benar-benar mengukur kompetensi yang diinginkan (Sudijono, 2005). Tingkat Kesukaran: Menggunakan rumus proporsi siswa yang menjawab benar, soal dikategorikan dalam kategori mudah, sedang,

atau sukar. Soal dengan tingkat kesukaran sedang dianggap lebih efektif untuk menguji kemampuan berpikir tingkat tinggi siswa (Wiggins, 1993). Daya Pembeda: Mengukur kemampuan soal dalam membedakan siswa dengan kemampuan tinggi dan rendah. Soal dengan daya pembeda lebih dari 0,3 dianggap baik untuk membedakan tingkat kemampuan siswa (Prastowo, 2013). Efektivitas Pengecoh: Mengukur seberapa baik pilihan jawaban yang salah (pengecoh) berfungsi untuk mengeliminasi siswa yang tidak benar-benar memahami materi. Pengecoh yang tidak dipilih oleh lebih dari 50% siswa dianggap tidak efektif.

HASIL DAN PEMBAHASAN

1. Realibitas Tes

Karena reliabilitas digunakan untuk mengetahui konsistensi alat ukur, perhitungan estimasi reliabilitas dilakukan selama analisis soal. Untuk mendapatkan (Ayunita, 2018) menunjukkan bahwa keputusan dapat digunakan kembali atau harus diulang. Menurut Malik dan Chusni (2018), reliabilitas dapat didefinisikan sebagai ketepatan atau konsistensi suatu instrumen. Dari lima belas soal evaluasi mata pelajaran aplikasi pengolah angka pada siswa kelas 10 akuntansi yang di Analisa menggunakan aplikasi anates menunjukan indeks tingkat realibitas nya sebesar 0,55 yang dimasukkan ke dalam kategori baik dengan rentang 0,40-0,70 yang menunjukan tingkat korelasi cukup. Namun pada tingkat tersebut realibitas bisa di tingkatkan dengan revisi soal yang kurang untuk bisa meningkatkan indeks realibitas, berikut adalah hasil dari koefisien realibitas..

```
RELIABILITAS TES
-----
Rata2= 6,52
Simpang Baku= 1,78
KorelasiXY= -0,21
Reliabilitas Tes= 0,55
Nama berkas: C:\USERS\HP\ONEDRIVE\DOKUMEN\REVISI ANATES.ANA
```

No.Ucut	No.	Subyek	Kode>Nama Subyek	Skor Ganjil	Skor Genap	Skor Total
1	1	Ananda Anggre...		5	3	8
2	2	Andini Marcha...		4	0	4
3	3	Ananda Intan ...		2	4	6
4	4	Ayu Widya Sari		4	3	7
5	5	Musa		4	4	8
6	6	M.Dimas Jidan...		6	1	7
7	7	Wawang Agus S...		6	4	10
8	8	Achmad Muzakqi		5	2	7
9	9	Decca Wahyu M.Z		2	4	6
10	10	M.Fariz Erakoso		4	4	8
11	11	Ravya Rambo R...		2	4	6
12	12	M.Eatctur Rosi		4	5	9
13	13	Siti Yuliana ...		2	2	4
14	14	Esti Damara ...		2	4	6
15	15	Novita Erlina...		1	3	4
16	16	Yunita Anggraini		3	4	7
17	17	Halifatul Anlah		1	4	5
18	18	Milda Ulan Sari		1	3	4
19	19	Maulana		4	5	9
20	20	Nia Putri Yun...		3	2	5
21	21	Nadiva Nur Ni...		3	4	7

Gambar 1. Realibitas Tes

2. Tingkat Kesukaran

Tingkat kesukaran soal didefinisikan sebagai peluang siswa untuk menjawab soal dengan benar pada tingkat tertentu (Arikunto, 2010). Ini sejalan dengan pendapat Fatimah (2019) bahwa tingkat kesukaran soal adalah persentase jumlah siswa yang menjawab soal dengan benar atau salah. Kemampuan untuk mengetahui apakah sebuah soal mudah atau sulit diklasifikasikan sebagai kemampuan atau kemampuan. Indeks kesulitan biasanya digunakan untuk menunjukkan tingkat kesulitan soal. Indeks ini biasanya diwakili sebagai proporsi dengan besarnya berkisar antara 0,00 dan 1,00. Hasil analisis 15 butir soal penilaian pada kelas 10 akuntansi mata Pelajaran aplikasi pengolah angka menunjukkan bahwa 8 soal sedang, 5 soal sukar, 2 soal mudah. Jika di lihat dari hasil analisis menggunakan aplikasi anates soal lebih besar masuk dalam kategori sedang dengan tingkat persentase yaitu 53,33% pada soal kategori sedang, 33,33% sukar, dan 13,33% mudah. Dengan adanya hasil dari Analisa tersebut menunjukkan soal dapat merangsang kemampuan siswa untuk berfikir kritis. Berikut hasil dari tingkat kesukaran soal

No Butir	Bacu	No Butir Asli	Jml Betul	Tkt. Kesukaran (%)	Tafsiran
1	1	1	6	28,57	Sukar
2	2	2	15	71,43	Mudah
3	3	3	16	76,19	Mudah
4	4	4	13	61,90	Sedang
5	5	5	7	33,33	Sedang
6	6	6	13	61,90	Sedang
7	7	7	5	23,81	Sukar
8	8	8	9	42,86	Sedang
9	9	9	10	47,62	Sedang
10	10	10	8	38,10	Sedang
11	11	10	47,62	Sedang	
12	12	5	23,81	Sukar	
13	13	6	28,57	Sukar	
14	14	6	28,57	Sukar	
15	15	8	38,10	Sedang	

Gambar 2. Tingkat Kesukaraan

3. Daya Pembeda

Daya Pembeda merujuk pada kemampuan soal untuk membedakan antara siswa dengan kemampuan tinggi dan rendah. Dalam hal ini, daya pembeda yang baik akan menunjukkan seberapa efektif soal tersebut dalam mengidentifikasi siswa yang menguasai materi dengan baik dan mereka yang kurang menguasai materi tersebut. Hasil analisis daya pembeda yang diperoleh dari 15 soal dalam tes ini menunjukkan variasi yang cukup beragam, yang menggambarkan bahwa sebagian soal tidak mampu membedakan kemampuan siswa secara optimal.

```
DAYA PEMBEDA
=====
Jumlah Subyek= 21
Klp atas/bawah(n)= 6
Butir Soal= 15
Nama berkas: C:\USERS\HP\ONEDRIVE\DOKUMEN\REVISI ANATES.ANA
```

No Butir	Bany	No Butir	Asli	Kel. Atas	Kel. Bawah	Beda	Indeks	DP (%)
1	1	1	1	1	0	1	16,67	16,67
2	2	2	2	5	5	0	0,00	0,00
3	3	3	4	5	3	2	33,33	33,33
4	4	4	5	6	0	6	100,00	100,00
5	5	5	5	5	1	4	66,67	66,67
6	6	6	4	4	3	1	16,67	16,67
7	7	7	4	4	1	3	50,00	50,00
8	8	8	4	4	1	3	50,00	50,00
9	9	9	2	2	3	-1	-16,67	-16,67
10	10	10	1	3	3	-2	-33,33	-33,33
11	11	11	5	0	5	5	83,33	83,33
12	12	12	1	1	0	0	0,00	0,00
13	13	13	2	2	0	0	0,00	0,00
14	14	14	4	1	3	3	50,00	50,00
15	15	15	3	2	1	1	16,67	16,67

Gambar 3. Daya Pembeda

Dari hasil analisis, terlihat bahwa beberapa soal memiliki Indeks Daya Pembeda yang cukup baik, sementara lainnya menunjukkan nilai negatif atau sangat rendah, yang mengindikasikan bahwa soal tersebut justru gagal membedakan kemampuan siswa dengan benar.

Berikut adalah analisis dari hasil daya pembeda masing-masing soal berdasarkan data yang diberikan:

Soal 4 (Indeks Daya Pembeda 100%) menunjukkan kemampuan yang sangat baik dalam membedakan siswa dengan kemampuan tinggi dan rendah. Ini berarti bahwa soal ini hanya dapat dijawab dengan benar oleh siswa yang memiliki pemahaman yang baik tentang materi yang diuji.

Soal 5, 11, 7, 8, 14 (Indeks Daya Pembeda 66,67%, 83,33%, 50,00%, 50,00%, 50,00%) memiliki daya pembeda yang cukup baik. Soal-soal ini mampu membedakan siswa yang lebih mampu dengan yang kurang mampu, namun ada kemungkinan bahwa beberapa siswa dengan pemahaman sedang juga dapat menjawab soal-soal ini dengan benar.

Soal 3 (Indeks Daya Pembeda 33,33%) juga menunjukkan daya pembeda yang cukup, meskipun hasilnya tidak sebaik soal dengan DP lebih tinggi. Soal ini bisa jadi lebih sulit untuk membedakan siswa dengan kemampuan tinggi dan rendah.

Soal 9, 10, 12, 13 (Indeks Daya Pembeda -16,67%, -33,33%, 0%, 0%) menunjukkan nilai yang sangat rendah atau negatif, yang mengindikasikan bahwa soal-soal ini gagal dalam membedakan siswa dengan kemampuan tinggi dan rendah. Bahkan, soal 9 dan 10 memiliki nilai negatif, yang bisa jadi disebabkan oleh faktor-faktor lain seperti tingkat kesukaran yang tidak tepat atau kesalahan dalam penyusunan soal.

Secara keseluruhan, sebagian besar soal dalam tes ini memiliki daya pembeda yang rendah, dengan beberapa soal tidak mampu membedakan siswa yang memiliki kemampuan berbeda secara jelas. Hal ini menandakan perlunya perbaikan dan revisi soal, khususnya pada soal dengan nilai Daya Pembeda rendah atau negatif, untuk memastikan bahwa soal dapat mengukur kemampuan siswa secara lebih efektif.

4. Kualitas Pengecoh

Pengecoh adalah pilihan jawaban yang salah dalam soal pilihan ganda yang dirancang untuk menggoda siswa agar memilihnya, meskipun jawaban yang benar berbeda. Efektivitas pengecoh dapat diukur dengan melihat seberapa baik pengecoh ini

menarik perhatian siswa yang kurang memahami materi, tanpa mempengaruhi siswa yang memahami materi dengan baik. Berikut adalah penjelasan berdasarkan hasil analisis kualitas pengecoh dari soal-soal yang diuji.

```
KUALITAS PENGECOH
=====
Jumlah Subyek- 21
Butir Soal= 15
Name Backlog: C:\USERS\HP\ONEDRIVE\DOKUMEN\REVISI ANATES.ANA

No Butir Baku No Butir Asli a b c d e *
```

No	Butir Baku	No Butir Asli	a	b	c	d	e	*
1	1	1	1-	1-	7--	6**	6-	0
2	2	4---	2+	15**	0--	0--	0	
3	3	16**	1++	1++	3---	0--	0	
4	4	13**	3+	3+	0--	2++	0	
5	5	6-	7**	3++	0--	5+	0	
6	6	5---	2++	13**	0--	1-	0	
7	7	5**	2-	3+	7-	4++	0	
8	8	6--	9**	2+	3++	1-	0	
9	9	10**	6---	3++	2+	0--	0	
10	10	7---	8**	1-	1-	4++	0	
11	11	7---	10**	3++	1-	0--	0	
12	12	5**	7-	7-	2-	0--	0	
13	13	0**	5+	9---	1-	0--	0	
14	14	6**	10---	5+	0--	0--	0	
15	15	6--	8**	6--	1-	0--	0	

Gambar 4. Kualitas Pengecoh

Pada soal 1 Pilihan jawaban a (1-), b (1-), c (7--), d (6**) dan e (6-) menunjukkan bahwa sebagian besar siswa memilih jawaban d (6**) yang benar, namun pilihan lainnya memiliki jumlah pemilih yang sangat rendah. Pengecoh ini sangat tidak efektif karena tidak menarik perhatian siswa. Jawaban yang salah hampir tidak dipilih, yang menunjukkan bahwa soal ini tidak cukup menantang.

Pada soal 2 Pilihan a (4---), b (2+), c (15**) menunjukkan bahwa hampir semua siswa memilih jawaban c (15**) sebagai jawaban yang benar. Pengecoh pada soal ini (a dan b) gagal menarik perhatian siswa karena jumlah pemilihnya sangat rendah atau tidak ada sama sekali. Soal ini menunjukkan kualitas pengecoh yang buruk.

Pada soal 3 Pilihan jawaban a (16**),*b (1++), c (1++), d (3---) menunjukkan bahwa jawaban a (16**) dipilih oleh sebagian besar siswa, sementara pilihan b dan c sedikit lebih menarik, namun masih jauh dari jumlah yang cukup untuk dianggap efektif. Pengecoh b dan c memberikan tantangan minimal, dan kemungkinan besar terlalu mudah untuk dideteksi sebagai salah.

Pada soal 6 Pilihan a (5---), b (2++), c (13**) menunjukkan bahwa jawaban yang benar adalah c (13**) dan pilihan b (2++) sedikit efektif, dengan dua siswa memilihnya. Pengecoh b mungkin dapat menarik perhatian sebagian siswa, namun sebagian besar memilih jawaban yang benar.

Pada soal 10 Pilihan a (7---), b (8**) menunjukkan bahwa jawaban yang benar adalah b (8**) dengan 8 siswa memilihnya, sementara pilihan lainnya (a, c, d, e) hampir tidak dipilih. Ini menunjukkan bahwa pengecoh tidak cukup efektif dan cenderung lebih mudah untuk diidentifikasi oleh siswa

Berdasarkan data yang diberikan, sebagian besar soal menunjukkan pengecoh yang tidak efektif. Pengecoh yang baik seharusnya memiliki kemampuan untuk menarik perhatian siswa yang tidak sepenuhnya menguasai materi, namun data di atas menunjukkan bahwa banyak soal yang memiliki pengecoh yang sangat sedikit dipilih. Beberapa soal bahkan menunjukkan pilihan yang salah hampir tidak dipilih sama sekali oleh siswa, yang berarti soal tersebut tidak cukup menantang.

Beberapa soal seperti Soal 4 dan Soal 5 menunjukkan pengecoh yang lebih efektif, karena pilihan salahnya mendapat beberapa pemilih, meskipun jumlahnya tidak terlalu banyak. Meskipun demikian, banyak soal yang menunjukkan bahwa pengecohannya tidak cukup bervariasi atau menantang untuk mengeliminasi siswa yang hanya memiliki pemahaman minimal.

5. Validitas Korelasi Skor Butir Dengan Skor Total

Korelasi antara skor butir dan skor total merupakan salah satu indikator penting dalam mengevaluasi kualitas soal dalam suatu instrumen tes. Korelasi ini mengukur sejauh mana hubungan antara kinerja peserta didik pada satu butir soal dengan kinerja mereka secara keseluruhan pada tes. Korelasi yang baik menunjukkan bahwa soal tersebut berkontribusi secara signifikan terhadap evaluasi keseluruhan kemampuan peserta didik. Sebaliknya, korelasi yang rendah menunjukkan bahwa soal tersebut tidak terlalu berkaitan dengan hasil tes secara keseluruhan, yang bisa menjadi indikasi bahwa soal tersebut kurang relevan atau tidak efektif dalam mengukur kemampuan yang dimaksud.

```
KORELASI SKOR BUTIR DG SKOR TOTAL
=====
Jumlah Subjek= 21
Butir Soal= 15
Nama berkas: C:\USERS\HP\ONEDRIVE\DOKUMEN\REVISI ANATES.AMA
```

No Butir Baru	No Butir Asli	Korelasi	Signifikansi
1	1	0,174	-
2	2	0,009	-
3	3	0,298	-
4	4	0,689	Sangat Signifikan
5	5	0,002	Signifikan
6	6	0,100	-
7	7	0,476	-
8	8	0,404	-
9	9	-0,068	-
10	10	-0,293	-
11	11	0,536	Signifikan
12	12	-0,104	-
13	13	0,113	-
14	14	0,417	-
15	15	0,215	-

Gambar 5. Korelasi Skor Butir Dengan Skor Total

Catatan: Batas signifikansi koefisien korelasi sebagai berikut:

df	(N-2)	P=0,05	P=0,01	df	(N-2)	P=0,05	P=0,01
10	0,576	0,706	60	0,250	0,325		
15	0,482	0,606	70	0,233	0,302		
20	0,423	0,549	80	0,217	0,283		
25	0,381	0,496	90	0,205	0,267		
30	0,349	0,449	100	0,195	0,254		
40	0,304	0,393	125	0,174	0,228		
50	0,273	0,354	>150	0,159	0,208		

Bila koefisien = 0,000 berarti tidak dapat dihitung.

Gambar 6. Catatan Signifikansi Koefisien Korelasi

Berdasarkan data diatas, berikut adalah analisis korelasi antara skor butir dengan skor total pada 15 butir soal yang diuji:

Soal 1 (Korelasi 0,174): Korelasi antara skor butir soal pertama dengan skor total sangat rendah. Nilai ini lebih kecil dari batas signifikansi untuk $p=0,05$ (0,423 untuk $df=20$), yang menunjukkan bahwa soal ini tidak cukup baik dalam mencerminkan kinerja keseluruhan peserta didik. Soal ini perlu direvisi agar dapat mengukur kompetensi siswa secara lebih efektif.

Soal 2 (Korelasi 0,069): Korelasi yang sangat rendah ini mengindikasikan bahwa soal ini hampir tidak berhubungan dengan skor total. Hasil ini menunjukkan bahwa soal ini tidak efektif dalam mengukur pemahaman siswa secara keseluruhan dan kemungkinan besar perlu diperbaiki.

Soal 3 (Korelasi 0,298): Meskipun korelasi ini sedikit lebih tinggi dibandingkan soal sebelumnya, nilai ini masih kurang signifikan karena berada di bawah batas signifikansi (0,423). Hal ini mengindikasikan bahwa meskipun soal ini mungkin sedikit lebih relevan dengan tes keseluruhan, soal ini tetap perlu perbaikan.

Soal 4 (Korelasi 0,689 – Sangat Signifikan): Korelasi ini sangat signifikan, jauh di atas batas signifikansi (0,423). Ini menunjukkan bahwa soal keempat sangat relevan dengan skor total dan memiliki kontribusi yang kuat dalam mengukur kompetensi siswa. Soal ini sangat baik dan tidak perlu perubahan.

Soal 5 (Korelasi 0,602 – Signifikan): Korelasi ini juga menunjukkan hubungan yang signifikan dengan skor total (lebih tinggi dari 0,423), yang mengindikasikan bahwa soal ini cukup baik dalam mengukur kemampuan siswa. Meskipun ada ruang untuk perbaikan, soal ini sudah berfungsi dengan baik dalam konteks tes.

Soal 6 (Korelasi 0,180): Korelasi ini sangat rendah dan tidak signifikan. Soal ini mungkin tidak relevan dengan tes keseluruhan, dan perlu direvisi untuk meningkatkan kualitasnya dalam mengukur kemampuan siswa.

Soal 7 (Korelasi 0,476): Korelasi ini masih berada di bawah batas signifikansi yang diperlukan untuk $p=0,05$, yang berarti soal ini belum cukup baik dalam mengukur kemampuan siswa. Revisi soal ini mungkin diperlukan agar dapat lebih berkontribusi pada evaluasi keseluruhan.

Soal 8 (Korelasi 0,404): Walaupun korelasi ini lebih baik dibandingkan dengan beberapa soal lainnya, nilai ini masih berada di bawah batas signifikansi (0,423), yang menunjukkan bahwa soal ini kurang efektif dalam mengukur kemampuan siswa.

Soal 9 (Korelasi -0,068): Korelasi negatif ini mengindikasikan bahwa soal ini bahkan dapat mengurangi hubungan antara kinerja siswa pada soal tersebut dengan skor

total mereka. Hal ini menunjukkan bahwa soal ini mungkin sangat tidak relevan dan perlu diperbaiki.

Soal 10 (Korelasi $-0,293$): Seperti soal 9, korelasi negatif ini menunjukkan ketidakefektifan soal ini dalam mengukur kemampuan siswa secara keseluruhan. Soal ini sangat perlu diperbaiki agar lebih relevan dengan tes secara keseluruhan.

Soal 11 (Korelasi $0,536$ – Signifikan): Korelasi yang signifikan ini menunjukkan bahwa soal ke-11 memiliki hubungan yang baik dengan skor total dan efektif dalam mengukur kemampuan siswa. Soal ini berfungsi dengan baik, meskipun bisa ada sedikit ruang untuk peningkatan.

Soal 12 (Korelasi $-0,104$): Korelasi yang sangat rendah dan negatif ini menunjukkan bahwa soal ini sangat tidak relevan dengan skor total dan tidak memberikan kontribusi yang baik pada tes. Soal ini perlu direvisi secara signifikan.

Soal 13 (Korelasi $0,113$): Korelasi yang rendah ini menunjukkan bahwa soal ini kurang efektif dalam mengukur kompetensi siswa. Soal ini perlu diperbaiki agar lebih relevan dan berfungsi lebih baik dalam evaluasi keseluruhan.

Soal 14 (Korelasi $0,417$): Korelasi ini cukup baik meskipun sedikit di bawah batas signifikansi ($0,423$), menunjukkan bahwa soal ini relevan tetapi masih ada ruang untuk perbaikan dalam hal kontribusinya terhadap evaluasi keseluruhan.

Soal 15 (Korelasi $0,215$): Korelasi yang rendah ini menunjukkan bahwa soal ini kurang relevan dalam mengukur kemampuan siswa dan perlu direvisi untuk meningkatkan kualitasnya

Dari hasil analisis korelasi skor butir dengan skor total, didapatkan bahwa sebagian besar soal tidak memiliki korelasi yang signifikan dengan skor total tes. Beberapa soal, seperti Soal 4 (Korelasi $0,689$) dan Soal 5 (Korelasi $0,602$), menunjukkan korelasi yang baik dan berkontribusi secara signifikan terhadap hasil tes secara keseluruhan. Sebaliknya, banyak soal lainnya menunjukkan korelasi yang rendah atau bahkan negatif, yang menunjukkan bahwa soal-soal tersebut kurang efektif dalam mengukur kompetensi siswa. Oleh karena itu, sebagian besar soal perlu direvisi agar lebih relevan dan efektif dalam mengukur kompetensi yang diinginkan.

KESIMPULAN DAN SARAN

Artikel ini membahas tentang adanya beberapa aspek yang perlu diperhatikan untuk meningkatkan kualitas soal dalam tes yang diujikan oleh aplikasi anates . Hasil analisis reliabilitas menunjukkan bahwa soal tes memiliki tingkat reliabilitas yang cukup baik, dengan indeks sebesar $0,55$, yang masuk dalam kategori "baik". Meskipun demikian, untuk meningkatkan reliabilitas, perlu dilakukan revisi pada beberapa soal agar konsistensi alat ukur lebih terjaga (Ayunita, 2018; Malik & Chusni, 2018). Berdasarkan analisis tingkat kesukaran, sebagian besar soal berada pada kategori "sedang" ($53,33\%$), diikuti dengan kategori "sukar" ($33,33\%$) dan "mudah" ($13,33\%$). Hasil ini menunjukkan

bahwa soal-soal tersebut dapat merangsang kemampuan berpikir kritis siswa, meskipun masih ada ruang untuk memperbaiki soal yang lebih sulit atau mudah. Daya pembeda soal bervariasi, dengan beberapa soal menunjukkan kemampuan yang baik dalam membedakan siswa dengan kemampuan tinggi dan rendah, namun banyak soal memiliki daya pembeda yang rendah atau negatif. Hal ini menunjukkan perlunya revisi pada soal dengan daya pembeda yang kurang efektif agar soal dapat lebih efektif membedakan siswa berdasarkan pemahaman materi. Analisis kualitas pengecoh mengungkapkan bahwa banyak pengecoh pada soal-soal yang tidak efektif, karena sebagian besar siswa cenderung memilih jawaban yang benar, sementara pilihan yang salah hampir tidak dipilih. Beberapa soal memiliki pengecoh yang sedikit lebih efektif, tetapi sebagian besar soal membutuhkan perbaikan dalam hal variasi dan daya tarik pengecoh. Hasil analisis korelasi menunjukkan bahwa sebagian besar soal tidak memiliki korelasi yang signifikan dengan skor total tes. Beberapa soal, seperti soal 4 dan soal 5, menunjukkan korelasi yang baik dan berkontribusi signifikan terhadap hasil tes keseluruhan, namun banyak soal lainnya memiliki korelasi yang rendah atau bahkan negatif. Hal ini menandakan bahwa soal-soal tersebut kurang relevan dan perlu revisi untuk meningkatkan efektivitasnya dalam mengukur kompetensi siswa. Secara keseluruhan, hasil analisis ini menunjukkan bahwa tes ini memiliki beberapa kelemahan yang perlu diperbaiki, terutama dalam hal daya pembeda, kualitas pengecoh, dan validitas korelasi. Revisi terhadap soal-soal yang kurang efektif sangat diperlukan agar tes dapat lebih tepat dalam mengukur kompetensi siswa.

Secara keseluruhan, aplikasi ANATES terbukti efektif dalam membantu para bapak ibu guru untuk mengidentifikasi dan memperbaiki alat penilaian yang digunakan untuk membantu menguji soal HOTS yang dibuat. Dengan memperbaiki butir soal yang tidak valid, meningkatkan reliabilitas, dan memenuhi tingkat kesulitan soal, sistem evaluasi soal dapat menjadi lebih akurat, efisien, dan efektif.

DAFTAR REFERENSI

- 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>
- Arif, M. (2014). Penerapan aplikasi Anates bentuk soal pilihan ganda. *Jurnal Ilmiah Edutic*, 1, 1–9.
- Arikunto, S. (2010). *Dasar-dasar evaluasi pendidikan*. Bumi Aksara.

- Ayunita, D. (2018). Modul uji validitas dan reliabilitas. *Jurnal Tarbiyah: Jurnal Ilmiah Kependidikan*, 7(1), 17–23.
- Depdiknas. (2007). Panduan penulisan soal pilihan ganda. Pusat Penilaian Pendidikan Balitbang-Depdiknas.
- Fanani, M. Z. (2018). Strategi pengembangan soal HOTS pada kurikulum 2013. *Edudeena*, 2(1), 57–76. <https://doi.org/10.30762/ed.v2i1.582>
- Fatimah, L., & A. K. (2019). Analisis kesukaran soal, daya pembeda dan fungsi distraktor. *Jurnal Komunikasi Dan Pendidikan Islam*, 8, 37–64.
- Fietri, W. A. (2021). Analisis butir soal biologi kelas VIII Madrasah Tsanawiyah Negeri 6 Kerinci. XX(X).
- Gunawan, A. W. (2003). *Genius learning strategy: Petunjuk praktis untuk menerapkan accelerated learning*. Gramedia Pustaka Utama.
- Hanifah, N. (2014). Perbandingan tingkat kesukaran, daya pembeda butir soal pilihan ganda asosiasi mata pelajaran ekonomi. *SOSIO E-KONS*, 6(1), 41–55.
- Hidinia, D. P., & Suprijono, A. (2020). Analisa tingkat kesulitan soal latihan USBN 2019 melalui aplikasi Anates. *Avatara, e-Journal Pendidikan Sejarah*, 8(1), 1–5.
- Jia, B., He, D., & Zhu, Z. (2020). Quality and feature of multiple-choice questions in education. *Problems of Education in the 21st Century*, 78(4), 576–594. <https://doi.org/10.33225/pec/20.78.576>
- Kemendikbud. (2017). Modul penyusunan soal higher order thinking skill (HOTS). Direktorat Pembinaan SMA Ditjen Pendidikan Dasar dan Menengah.
- Kuswana, W. S. (2011). Taksonomi berpikir. Rosdakarya.
- Lestari, D. (2016). Pengembangan soal tes berbasis HOTS pada model pembelajaran latihan penelitian di sekolah dasar. *Journal Universitas Pendidikan Indonesia*.
- Malik, A., & Chusni, M. M. (2018). Pengantar statistika pendidikan. *International Journal of Physiology*.
- Naga, D. S. (2002). Pengantar teori skor pada pengukuran pendidikan. Gunadarma.
- OECD. (2019). “PISA 2018 Mathematics Framework”, in PISA 2018 Assessment and Analytical Framework. OECD.
- Purwati, L. M., Arianty, R., Syakilah, D. M., Ridlo, S., & Susilaningsih, E. (2018). Analisis soal tes pilihan ganda berbasis higher order thinking skill menggunakan aplikasi Anates Windows versi 4.0.9 for Windows. 2015, 1–11.
- Putri, Resty A., & Rosliyah, Y. (2020). Analisis distraktor butir soal Bunpou Shokyuu Kouhan Semester 2 tahun 2018/2019 Prodi Pendidikan Bahasa Jepang Unnes. *Kiryoku*, 4, 20–28.
- Reniasih, G. N. (2020). Penerapan bimbingan berkelanjutan untuk meningkatkan kemampuan melaksanakan asesmen autentik pada guru di SD. *Jurnal Ilmiah Sekolah Dasar*, 4(1), 41. <https://doi.org/10.23887/jisd.v4i1.24278>
- Rintayati, P., Lukitasari, H., & Syawaludin, A. (2020). Development of two-tier multiple choice test to assess Indonesian elementary students’ higher-order thinking skills.

International Journal of Instruction, 14(1), 555–566.
<https://doi.org/10.29333/IJI.2021.14133A>

Sari, A. I. C., & Herawati, M. (2014). Aplikasi ANATES versi 4 dalam menganalisis butir soal. *Faktor Jurnal Ilmiah Kependidikan*, 1(2), 203–214.

Setiawaty. (2017). Indeks tingkat kesulitan (ITK) dan keberfungsian distraktor soal pilihan ganda UAS genap mata pelajaran Bahasa Indonesia. *SEMNASBAHTERA*, 68–78.

Sugiyono. (2007). *Metode penelitian pendidikan pendekatan kuantitatif, kualitatif, dan R&D*. Alfabeta.

Wiguna, S. (2021). *Aplikasi Anates dalam evaluasi pembelajaran*. Pena Persada.

Wilson, V. (2000). *Education forum on teaching thinking skills*. Edinburgh Report, the Scottish Council for Research in Education.

Yusup, F. (2018). Uji validitas dan reliabilitas instrumen penelitian kuantitatif. *Jurnal Terbiyah: Jurnal Ilmiah Kependidikan*, 17–23.

Yazid, A. J. (2021). Analisis kualitas butir soal ujian akhir semester pemindah tenaga kendaraan ringan kelas XI TKR di SMK Negeri 2 Praya Tengah. 3, 35–44.

Dewi, N. P., Rahmi, Y. L., Alberida, H., & Darussyamsu, R. (2020). Validitas dan reliabilitas instrumen penilaian kemampuan berpikir tingkat tinggi tentang materi hereditas untuk peserta didik SMA/MA. *Jurnal Eksakta Pendidikan (JEP)*, 4(2), 138.
<https://doi.org/10.24036/jep/vol4-iss2/512>