

## Penerapan Aplikasi Anates Untuk Analisis Soal HOTS Dalam Evaluasi Pembelajaran Spreadsheet Di SMKN 1 Kendal

Dwi Andrianik Susilowati<sup>1</sup>, Nidaria Saputri<sup>2</sup>, Khafidhotun Ni'mah<sup>3</sup>, Luqman Hakim<sup>4</sup>, Vivi Pratiwi<sup>5</sup>

<sup>1,2,3,4,5</sup>Universitas Negeri Surabaya

E-mail: dwi.23141@mhs.unesa.ac.id<sup>1</sup>, nidaria.23151@mhs.unesa.ac.id<sup>2</sup>,  
khafidhotun.23148@mhs.unesa.ac.id<sup>3</sup>, luqmanhakim@unesa.ac.id<sup>4</sup>, vivipratiwi@unesa.ac.id<sup>5</sup>

### Article History:

Received: 09 Januari 2025

Revised: 17 Januari 2025

Accepted: 25 Januari 2025

**Keywords:** *HOTS, Anates, validity, reliability, difficulty, discrimination.*

**Abstract:** *This study analyzes the quality of Higher Order Thinking Skills (HOTS)-based questions using the Anates application for class X Accounting students at SMKN 1 Kendal. The research involved 15 multiple-choice questions designed to assess students' analysis, evaluation, and synthesis skills, with data collected from 30 students. The analysis focused on reliability, difficulty level, discriminating power, distractor effectiveness, and validity. Results showed a reliability score of 0.76, indicating good consistency. However, only 3 out of 15 questions met validity criteria, while most had low correlations with total scores. The majority of questions were classified as very easy (>90%), failing to challenge critical thinking. Additionally, 80% of the questions had low discriminating power, and 60% of distractors were ineffective. These findings highlight the need for significant question revisions to align with HOTS standards. The Anates application proved useful in systematically identifying weaknesses and guiding improvements. This research aims to enhance the quality of evaluation instruments, particularly HOTS-based questions, in vocational high schools.*

## PENDAHULUAN

Pendidikan di era globalisasi menghadapi tantangan besar untuk menghasilkan lulusan yang tidak hanya memiliki pengetahuan dasar, tetapi juga mampu berpikir kritis, kreatif, dan analitis. Higher Order Thinking Skills (HOTS) menjadi salah satu keterampilan utama yang harus dimiliki siswa dalam menghadapi tantangan abad ke-21. HOTS mencakup kemampuan analisis, evaluasi, dan sintesis yang penting untuk menyelesaikan permasalahan dunia nyata secara inovatif dan efektif (Purwati et al., 2021). Kebutuhan akan soal berbasis HOTS dalam proses evaluasi pembelajaran semakin mendesak seiring diterapkannya Kurikulum 2013 di Indonesia, yang menekankan pembelajaran berbasis kompetensi dengan mengintegrasikan pengetahuan, keterampilan, dan sikap secara holistik (Fetri et al., 2021). Namun, penyusunan soal HOTS masih menjadi tantangan bagi pendidik, karena sebagian besar soal yang disusun masih berfokus pada Lower Order Thinking Skills (LOTS), seperti menghafal fakta atau informasi (Purwati et al., 2021). Padahal, soal HOTS memiliki peran penting dalam melatih siswa untuk berpikir mendalam,

memahami konsep kompleks, serta menerapkan pengetahuan dalam situasi baru (Fietri et al., 2021). Penilaian kualitas soal menjadi aspek krusial dalam evaluasi pembelajaran. Soal yang baik harus memenuhi kriteria validitas, reliabilitas, daya pembeda, dan tingkat kesukaran. Validitas menunjukkan sejauh mana soal dapat mengukur kemampuan yang seharusnya diukur, sedangkan reliabilitas mengacu pada konsistensi hasil pengukuran. Daya pembeda menentukan kemampuan soal dalam membedakan siswa berkemampuan tinggi dan rendah, sementara tingkat kesukaran menunjukkan sejauh mana soal tersebut dapat dijawab dengan benar oleh siswa (Purwati et al., 2021). Untuk membantu menganalisis kualitas soal, aplikasi Anates menjadi alat bantu yang efektif. Aplikasi ini dirancang untuk mempermudah pendidik dalam mengevaluasi kualitas soal secara kuantitatif. Melalui Anates, pendidik dapat mengidentifikasi kekuatan dan kelemahan setiap butir soal berdasarkan data statistik, seperti indeks validitas, reliabilitas, daya pembeda, dan tingkat kesukaran (Fietri et al., 2021). Dengan demikian, pendidik dapat melakukan revisi soal secara sistematis untuk memastikan instrumen evaluasi yang digunakan benar-benar sesuai dengan tujuan pembelajaran (Purwati et al., 2021). Penelitian ini dilakukan untuk menganalisis kualitas soal HOTS yang digunakan di SMKN 1 Kendal, khususnya pada siswa kelas X Akuntansi. Dengan menggunakan aplikasi Anates, penelitian ini bertujuan untuk mengidentifikasi kelemahan soal dan memberikan rekomendasi perbaikan agar soal dapat digunakan sebagai instrumen evaluasi yang lebih efektif. Analisis ini diharapkan dapat memberikan kontribusi positif bagi peningkatan kualitas pembelajaran, terutama dalam menyiapkan siswa menghadapi tantangan dunia kerja dan pendidikan lanjut.

## METODE PENELITIAN

Penelitian ini menggunakan pendekatan deskriptif kuantitatif yang bertujuan untuk menganalisis kualitas soal berbasis Higher Order Thinking Skills (HOTS). Subjek penelitian adalah siswa kelas X Akuntansi di SMKN 1 Kendal. Proses analisis dilakukan menggunakan program Anates versi 4.0 for Windows, mencakup validitas, reliabilitas, daya pembeda, tingkat kesukaran, serta analisis pengecoh. Metode serupa telah digunakan dalam penelitian lain untuk mengukur kualitas soal HOTS di berbagai mata pelajaran (Purwati et al., 2021; Fietri et al., 2021). Penyusunan soal berdasarkan pedoman Taxonomi Bloom dilakukan untuk memastikan soal dapat mengukur kemampuan berpikir tingkat tinggi, seperti analisis, 274 evaluasi, dan sintesis. Proses ini sejalan dengan penelitian terdahulu yang menekankan pentingnya desain instrumen soal berbasis HOTS dalam mendukung pengembangan kemampuan berpikir kritis siswa (Nufus, 2023; Gamar & Shamdas, 2022). Prosedur penelitian melibatkan distribusi soal melalui Google Form, di mana metode ini juga digunakan dalam pengembangan dan evaluasi soal berbasis HOTS di beberapa penelitian lain untuk efektivitas dan efisiensi (Pahlevi, 2021; Alifah & Sugilar, 2024). Program Anates menjadi alat yang sangat membantu dalam menganalisis kualitas butir soal, yang juga telah dibuktikan dalam berbagai studi analisis butir soal menggunakan perangkat lunak ini (Mardiyah & Rufi'i, 2024; Amelia et al., 2021). Metode analisis yang digunakan memberikan gambaran menyeluruh mengenai kualitas soal, sehingga hasilnya dapat digunakan sebagai acuan dalam meningkatkan kualitas instrumen evaluasi pembelajaran di masa mendatang.

## HASIL DAN PEMBAHASAN

Analisis soal HOTS menggunakan aplikasi Anates bertujuan untuk mengevaluasi kualitas instrumen penilaian yang digunakan dalam penelitian ini. Evaluasi kualitas mencakup beberapa aspek utama, yaitu validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Pendekatan ini telah digunakan dalam berbagai penelitian untuk mengukur sejauh mana

---

soal memenuhi kriteria kualitas, terutama pada soal berbasis HOTS (Purwati et al., 2021; Fetri et al., 2021; Gamar & Shamdas, 2022). Hasil analisis ini memberikan wawasan tentang sejauh mana soal-soal yang diuji memenuhi kriteria soal HOTS yang berkualitas.

### 1. Reliabilitas Tes

Perhitungan estimasi reliabilitas dilakukan dalam analisis soal karena reliabilitas digunakan untuk mengetahui konsistensi alat ukur, guna menentukan apakah instrumen tersebut layak digunakan atau perlu diperbaiki. Reliabilitas didefinisikan sebagai ketepatan atau konsistensi suatu instrumen. Dari 15 butir soal yang dianalisis menggunakan program aplikasi Anates untuk mata pelajaran Penggunaan Aplikasi Pengolah Angka kelas X Akuntansi, diperoleh indeks sebesar 0,76. Nilai ini menunjukkan bahwa instrumen memiliki konsistensi internal yang cukup baik. Reliabilitas dalam rentang 0,70 hingga 0,80 termasuk dalam kategori baik, yang mengindikasikan bahwa instrumen evaluasi ini mampu menghasilkan hasil yang konsisten jika digunakan kembali pada populasi yang sama. Pendekatan ini sejalan dengan temuan dalam penelitian sebelumnya yang menyatakan bahwa reliabilitas merupakan indikator penting untuk menentukan sejauh mana suatu instrumen mampu memberikan hasil yang konsisten (Akhmadi, 2021; Fetri et al., 2021; Purwati et al., 2021).

Namun, meskipun reliabilitas sudah cukup baik, terdapat ruang untuk peningkatan. Salah satu upaya yang dapat dilakukan adalah melakukan revisi pada soal-soal yang memiliki validitas rendah. Dengan meningkatkan validitas butir soal, reliabilitas keseluruhan pengujian juga dapat diperbaiki, sehingga hasil evaluasi akan lebih akurat dan dapat diandalkan. Penelitian lain juga menunjukkan bahwa peningkatan validitas butir soal memiliki korelasi positif terhadap reliabilitas instrumen secara keseluruhan (Lestari et al., 2024; Amelia et al., 2021). Berikut hasil koefisien reliabilitas.

No.Urut	No.	Subyek	Kode/Nama Subyek	Skor Ganjil	Skor Genap	Skor Total
1	1			7	7	14
2	2			7	5	12
3	3			7	5	12
4	4			8	6	14
5	5			7	6	13
6	6			8	6	14
7	7			8	6	14
8	8			8	6	14
9	9			8	6	14
10	10			7	5	12
11	11			7	6	13
12	12			7	6	13
13	13			8	6	14
14	14			6	6	12
15	15			8	6	14
16	16			8	6	14
17	17			8	6	14
18	18			8	6	14
19	19			7	6	13
20	20			4	4	8
21	21			6	6	12
22	22			5	5	10
23	23			7	6	13
24	24			8	6	14
25	25			7	6	13
26	26			8	6	14
27	27			8	6	14
28	28			8	6	14
29	29			8	6	14
30	30			7	6	13

**Gambar 1. Hasil Koefisien Reliabilitas Tes**

### 2. Tingkat Kesukaran

Tingkat kesukaran soal merupakan salah satu parameter penting dalam analisis kualitas soal.

Parameter ini digunakan untuk menentukan sejauh mana suatu soal dapat dijawab dengan benar oleh siswa, yang mencerminkan tingkat pemahaman mereka terhadap materi yang diujikan. Tingkat kesukaran soal menggambarkan persentase siswa yang menjawab soal dengan benar. Semakin tinggi persentase jawaban benar, maka semakin mudah soal tersebut, dan sebaliknya. Tingkat kesukaran juga dapat diartikan sebagai peluang bagi siswa untuk menjawab benar suatu soal dalam suatu tingkat kemampuan tertentu. Indeks tingkat kesukaran biasanya dinyatakan dalam rentang 0,00 hingga 1,00. Analisis tingkat kesukaran telah digunakan dalam berbagai penelitian untuk mengevaluasi kualitas soal, terutama dalam memastikan bahwa instrumen penilaian mencakup soal yang sesuai dengan kemampuan siswa (Purwati et al., 2021; Alifah & Sugilar, 2024). Kriteria tingkat kesukaran dapat dijelaskan pada table 1.

**Tabel.1 Kriteria Tingkat Kesukaran**

Interval	Kriteria
0,00 ≤ TK ≤ 0,30	Sukar
0,31 ≤ TK ≤ 0,70	Sedang
0,71 ≤ TK ≤ 1,00	Mudah

TINGKAT KESUKARAN						
=====						
Jumlah Subyek= 30						
Butir Soal= 15						
Nama berkas: C:\USERS\PROBOOK\DOCUME~1\KELOMPOK~1.ANA						
No	Butir Baru	No Butir Asli	Jml Betul	Tkt. Kesukaran (%)	Tafsiran	
1	1	1	30	100,00	Sangat Mudah	
2	2	2	29	96,67	Sangat Mudah	
3	3	3	28	93,33	Sangat Mudah	
4	4	4	29	96,67	Sangat Mudah	
5	5	5	28	93,33	Sangat Mudah	
6	6	6	29	96,67	Sangat Mudah	
7	7	7	28	93,33	Sangat Mudah	
8	8	8	27	90,00	Sangat Mudah	
9	9	9	21	70,00	Sedang	
10	10	10	30	100,00	Sangat Mudah	
11	11	11	28	93,33	Sangat Mudah	
12	12	12	1	3,33	Sangat Sukar	
13	13	13	29	96,67	Sangat Mudah	
14	14	14	30	100,00	Sangat Mudah	
15	15	15	26	86,67	Sangat Mudah	

**Gambar.2 Hasil Analisis Tingkat Kesukaran Butir Soal**

Tingkat kesukaran soal menunjukkan bahwa sebagian besar soal tergolong sangat mudah. Sebanyak 80% soal memiliki tingkat kesukaran di atas 90%, yang berarti sebagian besar siswa dapat menjawab soal tersebut dengan benar. Hanya satu soal yang tergolong pada kategori sedang dengan tingkat kesukaran antara 70% hingga 90%. Tidak ada soal yang masuk kategori sukar. Tingkat kesukaran yang terlalu mudah ini menunjukkan bahwa soal-soal tersebut kurang mampu menantang siswa untuk berpikir kritis, yang merupakan salah satu tujuan dari soal berbasis HOTS. Soal-soal yang sangat mudah dapat mengurangi kemampuan tes untuk mengevaluasi keterampilan berpikir kritis dan analitis siswa. Oleh karena itu, diperlukan revisi soal untuk meningkatkan tingkat kesulitan, seperti menambahkan konteks masalah yang lebih kompleks atau menggunakan bentuk soal yang menuntut analisis lebih mendalam. Pendekatan ini sesuai dengan rekomendasi

dalam penelitian lain yang menyatakan bahwa soal berbasis HOTS harus dirancang dengan mempertimbangkan keseimbangan tingkat kesukaran untuk mendukung pengembangan kemampuan berpikir tingkat tinggi (Purwati et al., 2021; Nufus, 2023; Sofia, 2023; Lestari et al., 2024).

### 3. Daya Pembeda

Daya pembeda merupakan salah satu parameter penting dalam analisis kualitas soal, terutama pada soal pilihan ganda. Parameter ini menunjukkan kemampuan suatu butir soal untuk membedakan siswa dengan kemampuan tinggi dan rendah. Daya pembeda menggambarkan seberapa baik sebuah soal dapat mengelompokkan siswa berdasarkan tingkat penguasaan materi. Daya pembeda merupakan kemampuan soal untuk memisahkan kelompok siswa yang berada di kategori atas dan bawah berdasarkan hasil tes. Daya pembeda yang baik mencerminkan butir soal yang mampu mengevaluasi perbedaan kemampuan siswa secara akurat, sementara daya pembeda yang rendah menunjukkan bahwa soal tidak mampu memberikan informasi yang berarti untuk tujuan tersebut.

Analisis daya pembeda telah banyak digunakan dalam berbagai penelitian untuk menilai efektivitas soal dalam membedakan kemampuan siswa. Misalnya, penelitian oleh Akhmad (2021) dan Fietri et al. (2021) menunjukkan bahwa daya pembeda yang baik membantu dalam mengidentifikasi kualitas instrumen evaluasi. Selain itu, penelitian lain menegaskan pentingnya revisi soal dengan daya pembeda rendah untuk meningkatkan akurasi evaluasi siswa (Bakri, 2024; Amelia et al., 2021; Kaka et al., 2024).. Penelitian ini kriteria yang digunakan dapat dilihat pada table 2.

**Tabel.2 Kriteria Daya Pembeda**

Interval	Kriteria
0,00 - 0,20	Jelek (soal diperbaiki)
0,20 - 0,40	Cukup (soal diterima dan diperbaiki)
0,40 – 0,70	Sangat baik (soal diterima)
0,70 – 1,00	Semuanya tidak baik (soal dibuang)
Negatif	Semuanya tidak baik (soal dibuang)

DAYA PEMBEDA							
=====							
<u>Jumlah Subyek= 30</u>							
<u>Klp atas/bawah(n)= 8</u>							
<u>Butir Soal= 15</u>							
<u>Nama berkas: C:\USERS\PROBOOK\DOCUMENTS\KELOMPOK 6-ANALISIS BUTIR SOAL.ANA</u>							
No	Butir Baru	No	Butir Asli	Kel. Atas	Kel. Bawah	Beda	Indeks DP (%)
1		1		8	8	0	0,00
2		2		8	7	1	12,50
3		3		8	6	2	25,00
4		4		8	7	1	12,50
5		5		7	7	0	0,00
6		6		8	7	1	12,50
7		7		8	7	1	12,50
8		8		8	5	3	37,50
9		9		8	2	6	75,00
10		10		8	8	0	0,00
11		11		8	6	2	25,00
12		12		1	0	1	12,50
13		13		8	8	0	0,00
14		14		8	8	0	0,00
15		15		8	5	3	37,50

**Gambar.3 Hasil Analisis Daya Pembeda Butir Soal**

Berdasarkan hasil analisis daya pembeda, mayoritas soal memiliki daya pembeda yang rendah, yang berarti soal-soal tersebut tidak mampu membedakan siswa dengan kemampuan tinggi dan rendah secara efektif. Hanya sebagian kecil soal yang memiliki daya pembeda yang baik, sedangkan sisanya perlu diperbaiki. Soal dengan daya pembeda rendah cenderung tidak memberikan informasi yang cukup untuk mengidentifikasi perbedaan kemampuan siswa, sehingga mengurangi efektivitas evaluasi. Perbaikan dapat dilakukan dengan meningkatkan tingkat kesulitan pada beberapa butir soal untuk menantang siswa dengan kemampuan tinggi tanpa mengabaikan siswa dengan kemampuan sedang atau rendah. Dengan demikian, daya pembeda soal dapat ditingkatkan untuk mendukung kualitas evaluasi pembelajaran secara keseluruhan.

#### **4. Kualitas Pengecoh**

Kualitas pengecoh, yaitu kemampuan pilihan jawaban salah untuk menarik perhatian siswa dengan pemahaman yang kurang, juga memerlukan perhatian khusus. Sebagian besar pengecoh dalam soal yang diuji tidak berfungsi dengan baik. seluas 60% pengecoh tidak dipilih oleh siswa sama sekali, menunjukkan bahwa pilihan jawaban tersebut tidak dianggap relevan atau terlalu mudah dikenal sebagai jawaban yang salah. Pengecoh yang efektif harus mampu mengecoh siswa dengan pemahaman yang kurang dan memaksa mereka untuk berpikir lebih kritis sebelum memilih jawaban. Perbaikan pengecoh dapat dilakukan dengan membuat alternatif jawaban yang lebih masuk akal dan relevan dengan konteks soal. Hal ini sejalan dengan penelitian yang menyarankan agar pengecoh dibuat dengan cara yang dapat memotivasi siswa untuk berpikir lebih mendalam dan menganalisis setiap pilihan jawaban secara lebih kritis (Fiska et al., 2021; Gamar & Shamdas, 2022; Hidayah et al., 2022). Pengecoh yang dirancang dengan baik dapat meningkatkan efektivitas soal dalam mengukur kemampuan berpikir tingkat tinggi siswa.

KUALITAS PENGECHOH									
<hr/>									
<u>Jumlah Subyek= 30</u>									
<u>Butir Soal= 15</u>									
<u>Nama berkas:</u>	<u>C:\USERS\PROBOOK\DOCUMENTS\KELOMPOK 6-ANALISIS BUTIR SOAL.ANA</u>								
No	Butir Baru	No	Butir Asli	a	b	c	d	e	*
1		1	30**	0	0	0	0	0	0
2		2	29**	1---	0---	0---	0---	0	0
3		3	0--	0--	0--	28**	2---	0	0
4		4	0--	29**	1---	0--	0--	0	0
5		5	0--	28**	1--	1--	0--	0	0
6		6	0--	29**	0--	0--	1---	0	0
7		7	1--	28**	1--	0--	0--	0	0
8		8	1+	0--	27**	1+	1+	0	0
9		9	5---	2++	21**	2++	0--	0	0
10		10	0	30**	0	0	0	0	0
11		11	0--	2---	28**	0--	0--	0	0
12		12	0--	29---	1**	0--	0--	0	0
13		13	1---	29**	0--	0--	0--	0	0
14		14	30**	0	0	0	0	0	0
15		15	26**	1++	2--	0--	1++	0	0

**Keterangan:**

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

**Gambar.4 Hasil Analisis Kualitas Pengecoh Butir Soal**

Kesimpulan mengenai kualitas pengecoh pada setiap butir soal dalam penelitian ini didasarkan pada kriteria yang dimodifikasi dari Skala Likert. Soal algoritma dikategorikan memiliki efektivitas pengecoh yang sangat baik jika terdapat tiga jawaban pengecoh yang berfungsi, baik jika terdapat dua jawaban pengecoh yang berfungsi, cukup baik jika hanya satu jawaban pengecoh yang berfungsi, dan tidak baik jika tidak ada jawaban pengecoh yang berfungsi sama sekali. Berdasarkan hasil analisis, mayoritas pengecoh pada soal-soal yang diuji tidak berfungsi secara optimal. Berdasarkan penyebaran jawaban siswa, sebagian besar soal menunjukkan efektivitas pengecoh yang kurang baik karena tidak semua pilihan jawaban alternatif menarik perhatian siswa. Hal ini terlihat pada beberapa butir soal yang hanya memiliki satu atau bahkan tidak ada pengecoh yang berfungsi. Rekapitulasi hasil efektivitas pengecoh menunjukkan bahwa perbaikan pada jawaban alternatif sangat diperlukan, terutama untuk meningkatkan daya tarik dan relevansi pilihan pengecoh terhadap konteks soal. Dengan perbaikan tersebut, pengecoh diharapkan mampu berfungsi secara efektif untuk mendukung tujuan evaluasi yang lebih komprehensif.

## 5. Validitas Korelasi Skor Butir dengan Skor Total

Validitas butir soal merupakan aspek penting yang menentukan sejauh mana sebuah soal mampu mengukur kemampuan yang dimaksud secara akurat. Validitas mencerminkan kecermatan suatu instrumen dalam mengukur suatu variabel. Validitas bertujuan untuk menentukan kelayakan sebuah instrumen apakah dapat digunakan atau tidak dalam evaluasi pembelajaran. Pada penelitian ini, validitas soal dihitung menggunakan korelasi antara skor butir dengan skor total melalui

aplikasi Anates versi 4.0 for Windows. Korelasi ini mengindikasikan seberapa besar hubungan antara jawaban siswa pada butir soal tertentu dengan hasil keseluruhan tes. Nilai korelasi yang tinggi menunjukkan soal yang valid dan dapat digunakan sebagai instrumen evaluasi yang representatif, sedangkan nilai korelasi yang rendah menunjukkan perlunya revisi pada soal tersebut. Berikut adalah hasil analisis validitas berdasarkan korelasi skor butir dengan skor total.

Validitas soal sangat penting untuk memastikan bahwa soal yang digunakan dalam evaluasi dapat mengukur kemampuan siswa secara tepat. Penelitian sebelumnya menunjukkan bahwa validitas yang baik memungkinkan soal untuk memberikan hasil yang konsisten dan relevan dengan tujuan pembelajaran (Akhmadi, 2021; Fiska et al., 2021; Hidayatullah et al., 2023). Korelasi yang tinggi antara butir soal dan skor total memberikan indikasi bahwa instrumen tes tersebut dapat diandalkan untuk mengevaluasi kemampuan siswa secara objektif (Purwati et al., 2021; Gamar & Shamdas, 2022; Mulyani et al. 2022).

KORELASI SKOR BUTIR DG SKOR TOTAL  
 =====

Jumlah Subyek= 30  
 Butir Soal= 15  
 Nama berkas: C:\USERS\PROBOOK\DOCUME~1\KELOMPOK~1.ANA

No Butir Baru	No Butir Asli	Korelasi	Signifikansi
1	1	NAN	NAN
2	2	0,151	-
3	3	0,614	Sangat Signifikan
4	4	0,151	-
5	5	0,020	-
6	6	0,701	Sangat Signifikan
7	7	0,119	-
8	8	0,765	Sangat Signifikan
9	9	0,641	Sangat Signifikan
10	10	NAN	NAN
11	11	0,811	Sangat Signifikan
12	12	0,124	-
13	13	0,014	-
14	14	NAN	NAN
15	15	0,683	Sangat Signifikan

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	0,708	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

Chapter 5: Health Analytics Validity, Reliability, and Scale

Berdasarkan tabel korelasi skor butir dengan skor total pada laporan, analisis menunjukkan bahwa dari 15 soal yang diuji, hanya 3 soal (nomor 3, 6, dan 11) yang memiliki validitas signifikan dengan korelasi  $r$  hitung lebih besar dari  $r$  tabel pada taraf signifikansi 5%. Hal ini menunjukkan bahwa hanya sebagian kecil soal yang mampu mengukur kemampuan siswa secara akurat sesuai dengan tujuan evaluasi.

Sebaliknya, 12 soal lainnya memiliki validitas rendah atau tidak signifikan, seperti terlihat pada soal nomor 2, 4, 5, 7, 12, dan 13, dengan nilai korelasi di bawah ambang batas signifikan. Ketidakvalidan ini disebabkan oleh beberapa faktor, termasuk formulasi soal yang kurang jelas, bahasa yang ambigu, dan ketidaksesuaian antara alternatif jawaban dengan materi yang diukur. Simpulan ini menunjukkan perlunya revisi pada sebagian besar soal untuk meningkatkan

validitasnya. Dengan perbaikan yang menyeluruh, soal dapat digunakan sebagai instrumen evaluasi yang lebih representatif dan akurat dalam mengukur kemampuan siswa. Validitas yang baik akan meningkatkan kualitas hasil evaluasi sebagai dasar pengambilan keputusan dalam proses pembelajaran. Penelitian sebelumnya juga mengungkapkan bahwa revisi soal yang dilakukan untuk meningkatkan validitas dapat meningkatkan efektivitas instrumen tes dalam mengevaluasi hasil belajar siswa (Aden Arif Gaffar, 2024; Hidayah et al., 2022; Triesninda Pahlevi, 2021). Dengan meningkatkan validitas soal, hasil evaluasi dapat lebih menggambarkan tingkat kemampuan siswa secara objektif dan akurat (Alifah & Sugilar, 2024; Meita Fitrianawati, 2023).

## 6. Rekap Analisis Butir Soal

Kegiatan analisis yang telah dilakukan sebelumnya mencakup reliabilitas, daya pembeda, tingkat kesukaran, efektivitas pengecoh secara parsial, dan validitas untuk setiap butir soal. Analisis tersebut memberikan gambaran terperinci tentang kualitas masing-masing butir soal berdasarkan aspek-aspek tersebut. Langkah terakhir adalah melakukan rekapitulasi analisis secara keseluruhan, yang bertujuan untuk menentukan kelayakan butir soal yang dapat digunakan, direvisi, atau bahkan dibuang. Rekapitulasi ini mempermudah identifikasi butir soal mana yang memenuhi kriteria sebagai instrumen evaluasi yang baik. Berikut adalah hasil rekapitulasi analisis butir soal secara menyeluruh berdasarkan data yang telah dianalisis.

Rekapitulasi ini sejalan dengan prinsip-prinsip analisis soal yang telah dibahas dalam berbagai studi sebelumnya, yang menunjukkan pentingnya evaluasi menyeluruh terhadap soal untuk memastikan kelayakan dan efektivitasnya dalam mengukur kemampuan siswa (Sofia, 2023; Lestari et al., 2024; Hidayah et al., 2022). Hal ini penting untuk meningkatkan kualitas soal agar dapat memberikan evaluasi yang akurat dan dapat diandalkan (Mardiyah & Rufi'i, 2024; Nurjanah et al., 2024). Berikut rekapitulasi hasil analisis butir soal.

REKAP ANALISIS BUTIR							
<hr/>							
Rata2= 13,10							
Simpang Baku= 1,37							
KorelasiXY= 0,62							
Reliabilitas Tes= 0,76							
Butir Soal= 15							
Jumlah Subjek= 30							
Nama berkas: C:\USERS\PROBOOK\DOCUMENTS\KELOMPOK 6-ANALISIS BUTIR SOAL.ANA							
Btr Baru	Btr Asli	D. Pembeda (%)	T. Kesukaran	Korelasi	Sign.	Korelasi	
1	1	0,00	Sangat Mudah	NAN	NAN		
2	2	12,50	Sangat Mudah	0,151	-		
3	3	25,00	Sangat Mudah	0,614	Sangat Signifikan		
4	4	12,50	Sangat Mudah	0,151	-		
5	5	0,00	Sangat Mudah	0,020	-		
6	6	12,50	Sangat Mudah	0,701	Sangat Signifikan		
7	7	12,50	Sangat Mudah	0,119	-		
8	8	37,50	Sangat Mudah	0,765	Sangat Signifikan		
9	9	75,00	Sedang	0,641	Sangat Signifikan		
10	10	0,00	Sangat Mudah	NAN	NAN		
11	11	25,00	Sangat Mudah	0,811	Sangat Signifikan		
12	12	12,50	Sangat Sukar	0,124	-		
13	13	0,00	Sangat Mudah	0,014	-		
14	14	0,00	Sangat Mudah	NAN	NAN		
15	15	37,50	Sangat Mudah	0,683	Sangat Signifikan		

**Gambar.7 Hasil Rekap Analisis Butir Soal**

Rata-rata skor yang diperoleh dalam penelitian ini adalah 13,10, dengan simpangan baku

sebesar 1,37 dan nilai korelasi XY mencapai 0,62. Dari 15 butir soal yang dianalisis menggunakan program Anates, diperoleh indeks reliabilitas sebesar 0,76. Berdasarkan hasil analisis, 6 butir soal dinyatakan valid, sehingga dapat digunakan kembali untuk tes hasil belajar di masa mendatang. Di sisi lain, 9 butir soal tergolong tidak valid dan memerlukan perbaikan. Langkah perbaikan dapat berupa penghapusan, peninjauan ulang, atau penyempurnaan, terutama pada aspek bahasa, seperti kalimat yang tidak jelas, soal yang sulit dipahami, atau penggunaan istilah yang kurang tepat. Setelah melalui proses perbaikan, soal-soal tersebut dapat diintegrasikan kembali ke dalam tes di masa yang akan datang. Hal ini sejalan dengan penelitian yang menekankan pentingnya revisi soal untuk meningkatkan kualitasnya, baik dari segi kejelasan bahasa maupun kesesuaian dengan tujuan evaluasi (Fietri et al., 2021; Alifah & Sugilar, 2024; Hidayatullah et al., 2023). Revisi yang menyeluruh pada aspek bahasa dapat meningkatkan pemahaman siswa terhadap soal dan memastikan bahwa soal tersebut dapat mengukur kemampuan siswa secara akurat (Fiska et al., 2021; Markhamah, 2021).

## KESIMPULAN

Berdasarkan hasil penelitian, analisis terhadap butir soal HOTS menggunakan program komputer Anates versi 4.0 for Windows menunjukkan bahwa reliabilitas tes mencapai 0,76, yang termasuk dalam kategori cukup baik. Validasi terhadap butir soal dilakukan pada taraf signifikansi 5%, dan hasilnya menunjukkan bahwa hanya 3 dari 15 soal yang memenuhi kriteria valid. Sebaliknya, sebagian besar soal memiliki karakteristik dengan tingkat kesukaran yang dikategorikan "sangat mudah," daya pembeda yang rendah, serta efektivitas pengecoh yang kurang optimal. Dari hasil analisis, ditemukan bahwa sekitar 60% pengecoh dalam soal-soal tersebut tidak berfungsi sebagaimana yang diharapkan. Hasil rekapitulasi analisis menunjukkan bahwa 6 butir soal dapat digunakan kembali dalam tes hasil belajar pada waktu yang akan datang, sedangkan 9 butir soal memerlukan perbaikan atau dibuang. Perbaikan meliputi revisi formulasi soal, peningkatan tingkat kesulitan, dan optimasi pengecoh agar lebih efektif. Setelah dilakukan revisi, soal-soal tersebut diharapkan mampu memenuhi kriteria HOTS yang baik, meningkatkan validitas dan daya pembeda, serta mendukung pengukuran kemampuan siswa secara lebih akurat dan komprehensif. Dengan perbaikan menyeluruh, instrumen evaluasi ini dapat menjadi alat ukur yang lebih representatif dan reliabel dalam mendukung proses pembelajaran berbasis HOTS. Hal ini juga didukung oleh penelitian sebelumnya yang menunjukkan bahwa revisi dan penyempurnaan soal dapat meningkatkan kualitas tes berbasis HOTS, terutama dalam aspek validitas dan reliabilitas (Mulyani et al., 2022; Gamar et al., 2022). Revisi soal yang terfokus pada peningkatan aspek daya pembeda dan efektivitas pengecoh sangat penting dalam menghasilkan instrumen evaluasi yang dapat membedakan siswa dengan kemampuan tinggi dan rendah secara lebih akurat, sehingga meningkatkan keandalan evaluasi pembelajaran berbasis HOTS (Purwati, 2021; Supriadi & Wulandari, 2024).

## DAFTAR REFERENSI

Aden Arif Gaffar, F. A. (2024). Pengembangan Soal-Soal Pilihan Ganda Beralasan Menggunakan Google Form Untuk Mengukur Kemampuan Berpikir Kritis Siswa SMA Pada Konsep Sistem Pencernaan Pada Manusia. *Jurnal Pendidikan dan Pembelajaran Biologi*, 9.

Akhmadi, M. N. (2021). ANALISIS BUTIR SOAL EVALUASI TEMA 1 KELAS 4 SDN PLUMBUNGANMENGGUNAKAN PROGRAM ANATES.

Alifah, S. N., & Sugilar, H. (2024). Analisis Tingkat Kesukaran Soal Matematika Pada Kemampuan Penyelesaian Masalah Matematis Siswa Kelas VII. *Gunung Djati Conference*

Series 41, 96-106.

Amelia, N., Paridjo, & Sina, I. (2021). Analisis butir soal penilaian akhir tahun mata pelajaran matematika menggunakan software anates. *AKSIOMA: Jurnal Matematika dan Pendidikan Matematika*.

Azzahroh, S., Iman, F. L., Anwar, B., & Azis, R. (2022). Analisis butir soal ujian akhir semester mata kuliah psikologi belajar menggunakan software Anates. *Indonesian Psychological Science*.

Bunga Setyaning Fahma, I. (2024). Analisis Butir Soal Asesmen Sumatif Pendidikan Agama Islam sekolah Dasar. *Jurnal Studi Edukasi Integratif*, 9.

Cahyanignrum, I. Y., Fuady, A., & Sunismi. (2023). Analisis Butir Soal Sumatif Akhir Semester Ganjil MataPelajaran Matematika Kelas VII dengan Berbantuan Aplikasi Software Anates. *MATHEMA JURNAL*.

Desi Sukenti, L. R. (2024). Kualitas Soal Bahasa Indonesia Kelas XI SMAN 2 Bangko Pusako: Analisis Butir Soal. *Jurnal Onoma*, 8.

Faldi Bakri, I. K. (2024). Analisis Karakteristik Butir Soal Dengan Classical Test Teory Pada Pelajaran PKn di Sekolah Dasar. *Jurnal Penelitian Pendidikan Dasar*, 11.

Fietri, W. A., Zulyusri, & Violita. (2021). Analisis Butir Soal Biologi Kelas XI Madrasah Aliyah Sakinah Kerinci Menggunakan Program Komputer Anates 4.0 For Windows. *Natural Science: Jurnal Penelitian Bidang IPA dan Pendidikan IPA*.

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. *Jurnal Natural Science Educational Research*.

Gamar B.N Shamdas, N. D. (2022). Analisis Soal Ujian Biologi Berbasis HOTS (Higher Order Thinking Skill) di SMAN 9 Palu. *Journal of Biology Science and Education (JBSE)*, 7.

Hayatun Nufus, R. Z. (2023). Pengembangan Instrumen Kemampuan Literasi Numerasi Berbasis HOTS. *Jurnal Cendekia*, 12.

Hendrayadi, Kustati, M., & Amelia, R. (2024). ANALISIS ULANGAN HARIAN MATA PELAJARAN PAI DI SMA NEGERI 10 PADANG TAHUN PELAJARAN 2023/2024 ANALISIS ULANGAN HARIAN MATA PELAJARAN PAI DI SMA NEGERI 10 PADANG TAHUN PELAJARAN 2023/2024 (TELAAH TERHADAP RELIABILITAS, DAYA BEDA DAN TINGKAT KESUKARAN MENGGUNAKAN SOFTWARE ANATES. *Jurnal Review Pendidikan dan Pengajaran (JRPP)*.

Hidayah, N., Turen, D., Astija, & Shamdas, G. (2022). Analisis Soal Ujian Biologi Berbasis HOTS (Higher Order Thinking Skill) di SMAN 9 Palu. *Journal of Biology Science and Education (JBSE)*.

Hidayatullah, R. S., Ariyanto, S. R., Rasyid, A. H., & Azizah, M. (2023). ANALISIS SOAL HOTS PILIHAN GANDA: ALAT PENILAIAN KOMPETENSI PADA TEKNIK MESIN. *Journal of Mechanical Engineering Learning*.

Kaka, L., Bano, V. O., & Njoeroemana, Y. (2024). EFEKTIVITAS ANALISIS BUTIR SOAL PILIHAN GANDA MENGGUNAKAN APLIKASI ANATES DI SMPN 2 KANATANG. *Journal Inovasi Penelitian*.

Lestari, I., K, R., Dewi, S. A., M, E., Nurhalizah, P., & Rosalina, S. (2024). Analisis Kualitas Butir Soal HOTS pada Mata Pelajaran Biologi yang Diberikan pada Siswa Kelas 12 dan Dievaluasi Menggunakan Aplikasi Anates. *EduInovasi: Journal of Basic Educational Studies*.

Markhamah , N. (2021). Pengembangan Soal Berbasis HOTS (Higher OrderThinking Skills) pada Kurikulum 2013. *Nusantara: Jurnal Pendidikan Indonesia*.

Mardiyah, H., & Rufi'i. (2024). PENGGUNAAN APLIKASI ANATES PADA ANALISIS BUTIR SOAL PENILAIAN AKHIR JENJANG. *JURNAL WIDYALOKA*.

Meita Fitrianawati, I. (2023). Pengembangan Instrumen Soal berbasis Keterampilan Berpikir Tingkat Tinggi. 18.

Muhamad Abdul Gofur, N. (2024). PELATIHAN PENULISAN TES TERTULIS TERSTANDAR BERBASIS HIGHER ORDER THINKING SKILLS (HOTS) UNTUK GURU SDIT NURUL IKHLAS BEKASI. : *Jurnal Pengabdian kepada Masyarakat*, 7.

Mulyani, S., Krismonita, M., & Yamtinah, S. (2022). ANALISIS BUTIR SOAL DAN KECUKUPAN HOTS SOAL UJIAN AKHIR SEMESTER MATA PELAJARAN KIMIASMK KELAS X. *Jurnal Penelitian Pendidikan*.

Nurjanah, W., Sari, I., & Saepuzaman, D. (2024). Analisis Pemahaman Konsep Peserta Didik Pada Topik Perambatan Kalor Menggunakan Rasch Model. *Jurnal Kependidikan*.

Pamungkasih, R. S., & Nawawi, E. (2021). ANALISIS KUALITAS BUTIR SOAL UJIAN AKHIR SEMESTER GANJIL PADA MATA PELAJARAN KIMIA KELAS X DI SMA NEGERI 8 PALEMBANG TAHUN AJARAN 2020/2021.

Prastika, Y. D. (2021). Pengaruh Validitas Reliabilitas Dan Tingkat Kesukaran Terhadap Kualitas Butir Soal Ekonomi Menggunakan Software Anates Di SMKN 3 Bangkalan.

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 Universitas Garut*.

Rahmawati, D. A. (2023). Analisis Kualitas Butir Soal Sistem Reproduksi Manusia Menggunakan Aplikasi Anates.

Sofia, M. E. (2023). Pengembangan Soal HOTS (High Order Thinking Skill) Materi Sel Volta Kelas XII SMA/ MA. *Jurnal Inovasi Pendidikan Kimia*, 6.

Triesninda Pahlevi, T. (2021). Pengembangan Instrumen Penilaian Berbasis HOTS Berbantuan Google form pada Mata Pelajaran Kearsipan Kelas X Jurusan OTKP SMKN 2 Kediri. *Journal of Office Administration: Education and Practice*, 15.