

## Analisis Butir Soal K3LH Pada Tingkat SMK Menggunakan Aplikasi ANATES

Anindya Paramesti Widhi Pangastuti<sup>1\*</sup>, Luqman Hakim<sup>2</sup>, Vivi Pratiwi<sup>3</sup>, Putri Nazhafah An’Nur<sup>4</sup>, Ananda Dian Febrianti Refany Putri<sup>5</sup>, Early Desti Mufida<sup>6</sup>, Naila Asalin Mushoffa<sup>7</sup>, Deandra Vernnata Emka<sup>8</sup>, dan Nala Arina Qifar<sup>9</sup>

<sup>1,2,3,4,5,6,7,8,9</sup> Universitas Negeri Surabaya

E-mail Korespondensi: 24080304129@mhs.unesa.ac.id

---

### Information Article

*History Article*

*Submission: 25-11-2025*

*Revision: 29-11-2025*

*Published: 30-11-2025*

---

**DOI Article:**

**10.24905/permana.v17i3.1251**

---

### ABSTRAK

Efektivitas sistem manajemen K3LH sangat bergantung pada evaluasi pendidikan K3LH (Keselamatan, Kesehatan Kerja, dan Lingkungan Hidup). Artikel ini membahas penggunaan aplikasi ANATES (Analysis of Effects and Action for System Evaluation) untuk menganalisis validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan kualitas materi pelajaran. Hasil penelitian terhadap 30 mata pelajaran menunjukkan bahwa, meskipun sebagian besar topik memiliki validitas dan reliabilitas yang baik, beberapa topik harus dibandingkan dengan tingkat kesulitan dan kualitas pengecoh. Penggunaan ANATES dapat meningkatkan efektivitas K3LH dan menyediakan lingkungan kerja yang lebih aman dan sehat..

**Kata Kunci:** Efektivitas, K3LH, Evaluasi, Pendidikan, ANATES, Validitas, Reliabilitas, Kualitas

### ABSTRACT

*The effectiveness of the K3LH management system greatly depends on the evaluation of K3LH education (Keselamatan, Kesehatan Kerja, dan Lingkungan Hidup). The article discusses the use of the ANATES (Analysis of Effects and Action for System Evaluation) application to analyze validity, reliability, daya pembeda, tingkat kesukaran, and quality of the subject matter. The results of the study on 30 subjects indicate that, even though the majority of the topics have good validity and reliability, some of the topics should be compared to the tingkat kesulitan and pengecoh quality. Utilizing ANATES can increase K3LH effectiveness and provide a more secure and healthy work environment*

**Keywords:** Effectiveness, K3LH, Evaluation, Education, ANATES, Validity, Reliability, Quality

---

### Acknowledgment

## PENDAHULUAN

Keselamatan, Kesehatan Kerja, dan Lingkungan Hidup (K3LH) merupakan aspek yang sangat penting untuk menunjang peningkatan produktivitas serta menciptakan lingkungan kerja yang aman. Menurut sisi akademik, K3 merupakan bidang keilmuan dan praktik yang berfokus pada upaya pencegahan berbagai potensi bahaya di tempat kerja seperti kecelakaan, kebakaran, ledakan, pencemaran lingkungan, serta gangguan kesehatan. Sedangkan pada sisi filosofisnya, K3 dapat dipahami sebagai seperangkat nilai dan tindakan yang bertujuan melindungi kesejahteraan jasmani maupun rohani para pekerja serta masyarakat luas, termasuk menjaga hasil kerja dan warisan budaya demi terciptanya kehidupan yang sejahtera dan berkeadilan. K3 adalah disiplin ilmu beserta implementasinya yang bertujuan menghindari berbagai risiko di tempat kerja seperti kecelakaan, kebakaran, ledakan, kontaminasi lingkungan, dan gangguan kesehatan (Lumbangaol dkk., 2022).

K3 menjadi bagian integral dari sistem ketenagakerjaan dan pengelolaan sumber daya manusia (Putri dkk., 2021). Di tingkat Sekolah Menengah Kejuruan (SMK), pendidikan K3LH disusun untuk membekali peserta didik dengan kemampuan dan wawasan yang sesuai dengan tuntutan dunia industri. Kesadaran tenaga kerja terhadap pentingnya K3 dapat tercermin melalui perilaku mereka dalam menaati aturan serta menjaga keamanan selama bekerja. Berdasarkan Undang-Undang Nomor 1 Tahun 1970 tentang Keselamatan Kerja, kecelakaan kerja diartikan sebagai kejadian yang tidak diinginkan serta berpotensi mengganggu kelancaran aktivitas dan menimbulkan kerugian baik berupa korban jiwa maupun materi (Herawati dkk., 2023).

Peralatan keselamatan yang digunakan oleh pekerja untuk melindungi diri dari potensi bahaya di tempat kerja disebut Alat Pelindung Diri (APD). APD berfungsi sebagai perlindungan individu agar terhindar dari cedera maupun gangguan kesehatan yang mungkin terjadi saat bekerja. Penggunaan APD yang memenuhi standar keselamatan memberikan manfaat utama berupa peningkatan perlindungan bagi pekerja. Contoh APD yang umum digunakan antara lain pakaian kerja khusus, helm keselamatan, dan sepatu pelindung yang semuanya dirancang untuk meminimalkan risiko cedera akibat terjatuh, tertimpa benda, atau kontak dengan mesin berbahaya (Tanjung & Susilawati, 2024).

Efektivitas pembelajaran K3LH sangat dipengaruhi oleh instrumen evaluasi yang digunakan. Instrumen yang tidak valid dan reliabel akan menghasilkan penilaian yang kurang akurat dan berdampak pada mutu pembelajaran secara keseluruhan. Dalam konteks ini, penggunaan perangkat lunak ANATES menjadi alternatif yang menjanjikan untuk membantu menganalisis butir soal, khususnya dalam menilai aspek validitas, reliabilitas, tingkat kesulitan, daya pembeda, serta efektivitas pengecoh.

ANATES merupakan salah satu software komputer yang dirancang untuk menganalisis butir soal secara mendalam, aplikasi ini dilengkapi dengan berbagai fitur komprehensif seperti penghitungan reliabilitas tes, penentuan tingkat kesulitan, pengukuran daya pembeda, dan evaluasi efektivitas pengecoh pada soal pilihan ganda (Hartono dkk., 2024).

Berbagai penelitian terdahulu telah menunjukkan bahwa ANATES mampu meningkatkan mutu analisis butir soal. Program ini mempermudah proses evaluasi karena dapat memeriksa jawaban secara otomatis dengan hasil yang cepat dan efisien. Kelebihan lain dari ANATES adalah antarmuka berbahasa Indonesia yang mudah dipahami serta kemampuan untuk mengekspor hasil analisis ke Microsoft Excel guna perhitungan lanjutan (Elviana, 2020). Namun, sebagian besar penelitian tersebut masih membahas penerapan ANATES secara umum tanpa menyoroti penggunaannya pada pembelajaran K3LH di SMK. Selain itu, kajian tentang efektivitas pengecoh dalam soal-soal K3LH masih jarang dilakukan.

Meskipun efektivitas ANATES telah banyak dibuktikan, penelitian ini mengidentifikasi adanya kesenjangan penelitian dalam penerapannya pada konteks pendidikan K3LH di SMK. Hingga saat ini, belum ada penelitian yang secara menyeluruh menggabungkan analisis validitas, reliabilitas, daya pembeda, tingkat kesulitan, dan kualitas pengecoh dalam satu studi terpadu. Oleh karena itu, penelitian ini berupaya menghadirkan pendekatan baru dengan fokus pada penerapan ANATES untuk menganalisis butir soal K3LH di SMK. Hasilnya diharapkan dapat menyediakan instrumen evaluasi yang lebih akurat dan sesuai dengan standar kebutuhan dunia kerja.

Tujuan utama dilakukannya penelitian ini adalah untuk menjawab kebutuhan terhadap alat evaluasi yang lebih berkualitas untuk pembelajaran K3LH dengan memanfaatkan aplikasi ANATES. Penelitian ini dibuat guna mengevaluasi kualitas soal-

soal K3LH di tingkat SMK.

## METODE

Penelitian butir soal K3LH (Keselamatan, Kesehatan Kerja, dan Lingkungan Hidup) menggunakan aplikasi ANATES ini menggunakan pendekatan deskriptif kuantitatif, metode ini cocok untuk mengevaluasi kualitas dari butir soal secara objektif berdasarkan data numerik. Penelitian ini akan menganalisis data hasil tes siswa untuk mengetahui karakteristik soal, melalui pendekatan ini peneliti berupaya menggambarkan karakteristik butir soal berdasarkan beberapa indikator, yaitu validitas, reliabilitas, daya pembeda, tingkat kesukaran, serta efektivitas pengecoh. Proses analisis dilakukan menggunakan perangkat lunak ANATES (*Analysis of Test Items*) untuk memperoleh hasil evaluasi yang lebih sistematis dan menyeluruh mengenai kualitas soal

Kegiatan penelitian ini dilakukan dengan beberapa tahapan. Tahap pertama adalah pengumpulan data dengan cara menyebarkan Google Form yang berisi soal-soal K3LH ke sejumlah responden dengan kriteria: 1) Berasal dari SMK/MAK, dan 2) Telah mengikuti atau sedang mengikuti pelajaran K3LH (Keselamatan, Kesehatan Kerja, dan Lingkungan Hidup). Kemudian dilakukan pengolahan data menggunakan aplikasi ANATES, yang juga dilakukan melalui beberapa tahap berikut tahap Analisis Data, Mengukur Validitas, Reliabilitas, Daya Pembeda, dan Tingkat Kesukaran, dan Tahap Interpretasi Data hingga terakhir adalah pembuatan laporan penelitian. Jumlah butir soal yang diberikan sejumlah 63 butir dengan responden sejumlah 20 orang. Pemilihan soal diambil dari materi yang mewakili dari setiap indikator pembelajaran

## HASIL

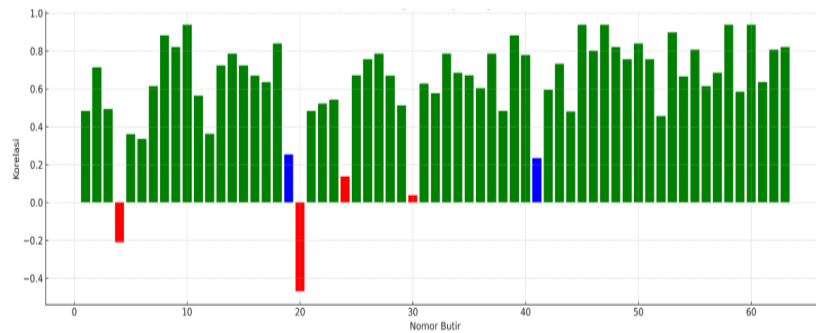
Dari penelitian yang kelompok kami laksanakan, diperoleh hasil penelitian hasil analisis dan interpretasi dari Validitas, Reliabilitas, Daya Pembeda, Tingkat kesukaran, Kualitas Pengecoh, dan Relevansi hasil penelitian untuk pendidikan K3LH.

### Validitas

Validitas instrumen adalah bagian penting dalam evaluasi pendidikan. Suatu instrumen dikatakan memiliki validitas ketika mampu melakukan pengukuran dengan akurat terhadap hal yang ingin diukur (Alista & Syahzanani, 2023). Berdasarkan kriteria validitas, soal dianggap baik apabila mampu mengukur hal yang diinginkan secara tepat. Kemudian validitas butir soal dianggap tinggi jika skor setiap butir soal memiliki arah yang sejajar dengan skor total. Hasil penelitian ini menunjukkan bahwa sebagian besar butir soal (59 dari 63) memiliki korelasi yang tinggi dengan skor total dan dikategorikan sebagai “sangat signifikan”. Artinya sebesar 93% butir soal dinyatakan valid berdasarkan korelasi dengan skor total dalam analisis ini. Hal ini menunjukkan bahwa hampir semua butir soal layak dipertahankan dan dapat digunakan untuk mengukur prestasi siswa. Namun, empat butir soal lainnya (4, 20, 24, dan 30) tidak mencapai tingkat validitas ini. Validitas yang tidak mencukupi biasanya disebabkan :

- 1.) Soal yang dihasilkan tidak tepat dalam mengukur aspek yang diinginkan, sehingga hubungan antara skor soal individu dengan skor total menjadi rendah atau tidak signifikan
- 2.) Soal memiliki kalimat yang tidak jelas atau pengecoh yang tidak efektif, sehingga menurunkan kualitas validitas butir soal
- 3.) Kualitas soal yang masih sangat rendah secara keseluruhan, termasuk soal-soal yang tidak sesuai dengan indikator pembelajaran atau kompetensi dasar

Untuk mengatasi masalah ini, revisi harus dilakukan dengan memvalidasi kembali soal menggunakan pendapat para ahli atau praktisi K3LH, memperbaiki cara penyusunan soal, serta meningkatkan kualitas pengecoh. Dengan demikian, validitas instrumen dapat ditingkatkan, sehingga menghasilkan evaluasi yang lebih tepat dalam mengukur pencapaian siswa.



**Gambar 1. Diagram Hasil Analisis Korelasi Validitas**

Sumber: Hasil Analisis Korelasi Validitas (2025)

## Reliabilitas

Reliabilitas berasal dari kata "reliabilitas", yang berarti sejauh mana hasil pengukuran dapat dipercaya. Hasil pengukuran dapat dipercaya apabila diperoleh hasil yang relatif sama dalam beberapa kali pengukuran terhadap kelompok subjek yang sama, selama aspek yang diukur dalam diri subjek tidak berubah (Ida & Musyarofah, 2021).

Berdasarkan hasil analisis, diperoleh koefisien korelasi split-half ( $r_{XY}$ ) sebesar 0,97 dan koefisien reliabilitas tes sebesar 0,98. Nilai ini menunjukkan bahwa instrumen memiliki tingkat reliabilitas yang sangat tinggi, karena mendekati angka 1, sehingga dapat disimpulkan bahwa alat ukur konsisten dalam menghasilkan data. Rata-rata skor peserta adalah 43,70 dengan simpangan baku 17,10, yang menunjukkan adanya variasi skor antar peserta. Dengan reliabilitas yang sudah sangat tinggi, instrumen ini dapat dipercaya untuk digunakan dalam pengukuran kemampuan siswa. Meskipun demikian, penguatan reliabilitas tetap dapat dilakukan melalui beberapa langkah, seperti:

- 1) **Penyeimbangan Tingkat Kesukaran Soal:** Mengatur proporsi butir mudah, sedang, dan sulit agar distribusi skor lebih stabil.
- 2) **Menambah Jumlah Subjek Uji:** Sampel yang lebih besar dapat memberikan estimasi reliabilitas yang lebih akurat.

- 3) Meningkatkan Keterpaduan Materi: Menyusun butir yang benar-benar sesuai dengan kompetensi yang diukur.
- 4) Menggunakan Teknik Pengujian Tambahan: Misalnya tes paralel atau metode split-half lainnya untuk memperkuat konsistensi hasil.

Dengan demikian, instrumen pada penelitian ini sudah berada pada kategori sangat reliabel dan dapat digunakan sebagai alat ukur yang andal.

RELIABILITAS TES  
=====

Rata2= 43,70  
Simpang Baku= 17,10  
KorelasiXY= 0,97  
Reliabilitas Tes= 0,98  
Nama berkas: D:\NAILA\BUTIR SOAL.ANA

No.Urut	No. Subyek	Kode>Nama Subyek	Skor Ganjil	Skor Genap	Skor Total
1	1	intan	5	8	13
2	2	fariha	10	6	16
3	3	hanimatul	19	21	40
4	4	dwi	19	21	40
5	5	aufa	11	15	26
6	6	bella	28	23	51
7	7	farulya	30	28	58
8	8	nadya	30	26	56
9	9	luthfiyah	28	26	54
10	10	sofy	30	27	57
11	11	safira	27	24	51
12	12	taffana	30	29	59
13	13	intan nur	31	29	60
14	14	cinta	31	28	59
15	15	wardana	7	6	13
16	16	lutviyatul	26	23	49
17	17	virna	27	23	50
18	18	levina	28	26	54
19	19	ara	8	7	15
20	20	fannia	28	25	53

**Gambar 2. Hasil Analisis Reliabilitas Tes**

Sumber: Hasil Analisis Reliabilitas (2025)

## Daya Pembeda

Daya pembeda adalah kapasitas sebuah soal dalam tes hasil belajar untuk mengidentifikasi perbedaan antara peserta tes yang pandai dan kurang pandai. Peserta dengan kemampuan tinggi umumnya akan berhasil menjawab soal dengan benar, sedangkan peserta dengan kemampuan rendah cenderung tidak mampu memberikan jawaban yang tepat. Dalam penelitian ini, daya pembeda butir soal dapat dilihat pada kolom nilai koefisiennya, di mana semua butir soal yang terlihat di dalam memiliki kategori daya pembeda Sangat Signifikan. Dari no 46,47,48,49,50,52,53, dan 55 memiliki daya pembeda (> 60%). Jadi

1705

sebagian besar butir soal (59 dari 63 butir) dapat diterapkan kepada siswa karena menunjukkan kualitas yang baik hingga sangat baik, khususnya pada Daya Pembeda (DP) dan Korelasi (Validitas). Maka dari itu, butir-butir soal yang terlihat pada cuplikan memiliki daya pembeda yang Sangat Signifikan, yang berarti butir-butir soal tersebut sangat efektif dalam membedakan antara kelompok siswa yang pandai dan kelompok siswa yang kurang pandai.



DAYA PEMBEDA  
\*\*\*\*\*

Jumlah Subyek= 20  
Klp atas/bawah(n)= 5  
Butir Soal= 63  
Nama berkas: D:\NAILA\BUTIR SOAL.ANA

No	Butir Baru	No Butir Asli	Kel. Atas	Kel. Bawah	Beda	Indeks	DP (%)
1	1	1	5	3	2		40,00
2	2	2	5	0	5		100,00
3	3	3	5	2	3		60,00
4	4	4	0	1	-1		-20,00
5	5	5	2	0	2		40,00
6	6	6	5	3	2		40,00
7	7	7	5	3	2		40,00
8	8	8	5	1	4		80,00
9	9	9	5	0	5		100,00
10	10	10	5	0	5		100,00
11	11	11	5	3	2		40,00
12	12	12	4	3	1		20,00
13	13	13	5	1	4		80,00
14	14	14	5	1	4		80,00
15	15	15	5	1	4		80,00
16	16	16	5	1	4		80,00
17	17	17	4	1	3		60,00
18	18	18	5	0	5		100,00
19	19	19	3	1	2		40,00
20	20	20	4	2	-2		-40,00
21	21	21	5	3	2		40,00
22	22	22	5	1	4		80,00
23	23	23	5	1	4		80,00
24	24	24	4	2	2		40,00
25	25	25	5	1	4		80,00

Gambar 3. Hasil Analisis Daya Pembeda

Sumber: Hasil Analisis Daya Pembeda (2025)

### Tingkat Kesukaran

Tingkat kesukaran merupakan keberadaan suatu item soal apakah dipandang sukar, sedang, atau mudah dalam mengerjakannya (Sarkawi, 2018). Berdasarkan hasil analisis tingkat kesukaran menggunakan aplikasi ANATES terhadap 63 butir soal dengan jumlah responden sebanyak 20 siswa, diketahui bahwa mayoritas soal berada pada kategori sedang (49,2%), yang berarti bahwa sebagian besar butir soal memiliki tingkat kesulitan yang cukup seimbang dan mampu mengukur kemampuan peserta didik secara proporsional. Selain itu, terdapat 22 butir soal (34,9%) yang tergolong mudah, 6 butir soal (9,5%) sangat mudah, 2 butir soal (3,2%) sukar, dan 1 butir soal (1,6%) sangat sukar.

Distribusi tingkat kesukaran tersebut menunjukkan bahwa tes ini cenderung memiliki proporsi butir yang relatif baik, meskipun sedikit didominasi oleh soal dengan tingkat kesukaran sedang dan mudah. Hal ini menunjukkan bahwa instrumen tes sudah cukup efektif dalam mengukur kemampuan rata-rata siswa, namun masih perlu penyempurnaan untuk meningkatkan variasi tingkat kesulitan agar hasil penilaian lebih representatif terhadap seluruh kemampuan peserta didik.

Soal yang tergolong terlalu mudah biasanya disebabkan oleh materi yang terlalu sederhana atau konteks soal yang sudah sering muncul dalam latihan siswa, sehingga hampir semua peserta dapat menjawabnya dengan benar. Sebaliknya, soal yang termasuk terlalu sukar sering kali dipengaruhi oleh redaksi yang kurang jelas, pilihan jawaban yang terlalu mirip, atau konteks soal yang terlalu kompleks. Kondisi ini berpotensi menurunkan daya

pembeda dan reliabilitas tes secara keseluruhan. Untuk memperbaiki distribusi tingkat kesukaran agar lebih seimbang, beberapa langkah dapat dilakukan, antara lain:

1. Meninjau ulang redaksi soal agar tidak terlalu sederhana maupun terlalu kompleks.
2. Menyesuaikan konteks dan indikator soal agar tetap relevan dengan tujuan pembelajaran.
3. Melakukan uji coba ulang setelah revisi untuk memastikan tingkat kesukaran soal telah proporsional.

Dengan demikian, tingkat kesukaran pada instrumen ini secara umum sudah baik, namun masih diperlukan penyempurnaan untuk memperkaya variasi tingkat kesulitan soal sehingga dapat mengukur kemampuan siswa secara lebih menyeluruh.

TINGKAT KESUKARAN  
-----

Jumlah Subyek= 20  
Butir Soal= 63  
Nama berkas: D:\NAILA\BUTIR SOAL.ANA

No Butir Baru	No Butir Asli	Jml Betul	Tkt. Kesukaran(%)	Tafsiran
1	1	18	90,00	Sangat Mudah
2	2	12	60,00	Sedang
3	3	16	80,00	Mudah
4	4	3	15,00	Sangat Sukar
5	5	4	20,00	Sukar
6	6	14	70,00	Sedang
7	7	18	90,00	Sangat Mudah
8	8	16	80,00	Mudah
9	9	14	70,00	Sedang
10	10	15	75,00	Mudah
11	11	18	90,00	Sangat Mudah
12	12	17	85,00	Mudah
13	13	13	65,00	Sedang
14	14	16	80,00	Mudah
15	15	13	65,00	Sedang
16	16	15	75,00	Mudah
17	17	13	65,00	Sedang
18	18	14	70,00	Sedang
19	19	6	30,00	Sukar
20	20	8	40,00	Sedang
21	21	18	90,00	Sangat Mudah
22	22	10	50,00	Sedang
23	23	12	60,00	Sedang
24	24	10	50,00	Sedang
25	25	13	65,00	Sedang
--	--	--	--	--

**Gambar 4. Hasil Analisis Tingkat Kesukaran**

Sumber: Hasil Analisis Tingkat Kesukaran (2025)

### Kualitas Pengecoh

Pengecoh memiliki peran penting dalam mengevaluasi sejauh mana suatu pertanyaan dapat membedakan antara siswa yang sudah memahami materi atau belum. Efektivitas pengecoh merupakan pengukuran terhadap alternatif jawaban pada soal bentuk pilihan ganda (Ashari, 2021). Pengecoh yang baik seharusnya mampu menarik perhatian siswa dengan kemampuan rendah tanpa mengganggu siswa yang lebih unggul. Namun, penelitian ini menunjukkan bahwa beberapa pengecoh tidak berfungsi dengan baik, terlihat dari sedikitnya siswa yang memilih opsi tersebut. Kondisi ini menandakan bahwa pengecoh

kurang logis, tidak menarik, atau tidak sesuai dengan pola pikir siswa.

Hasil penelitian lain yang dilakukan oleh (Nafs, Sridana, Hikmah, dan Soeprianto, 2023) juga menunjukkan bahwa banyak pengecoh dalam soal pilihan ganda belum efektif karena tidak mampu menyesatkan peserta yang kurang memahami materi. Selain itu, penelitian oleh (Novriyanti, Sarifah, dan Riyadi, 2024).

Mengungkapkan bahwa ketidakefektifan pengecoh sering terjadi ketika alternatif jawaban tidak disusun berdasarkan kesalahan umum yang sering dilakukan siswa, sehingga pengecoh menjadi mudah dikenali sebagai jawaban salah. Penelitian terbaru oleh (Rahmawati dan Rahman, 2025) turut menegaskan bahwa pengecoh harus dirancang dengan mempertimbangkan homogenitas pilihan jawaban dan daya tarik yang seimbang untuk memastikan kualitas soal tetap optimal.

Dengan demikian, perbaikan dapat dilakukan dengan menciptakan pengecoh yang relevan dengan materi, sulit ditebak sebagai jawaban salah, dan disusun dengan variasi yang tepat. Pengecoh yang dirancang secara cermat akan membantu meningkatkan efektivitas instrumen penilaian dalam mengukur pemahaman siswa secara lebih akurat dan komprehensif.

**Tabel 1. Kualitas Pengecoh**

Kategori	Jumlah	Presentase
Pengecoh Efektif (++,+)	90	32,6 %
Pengecoh Tidak Efektif (-, -, —)	162	58,8%

Sumber: Hasil Analisis Kualitas Pengecoh (2025)

### **Relevansi Hasil Penelitian Untuk Pendidikan K3LH**

Hasil penelitian ini memiliki relevansi yang signifikan terhadap peningkatan mutu pendidikan K3LH di tingkat Sekolah Menengah Kejuruan (SMK). Melalui analisis butir soal menggunakan aplikasi ANATES, pendidik dapat meningkatkan validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan kualitas pengecoh.

Secara praktis, hasil penelitian ini membantu guru dan pengembang kurikulum dalam menyusun instrumen evaluasi pembelajaran yang lebih berkualitas dan sesuai dengan standar kompetensi industri, sehingga proses pembelajaran K3LH tidak hanya berorientasi pada teori, tetapi juga relevan dengan kebutuhan keselamatan dan kesehatan

kerja di dunia nyata. Analisis butir soal yang dilakukan dengan ANATES memungkinkan guru mengevaluasi sejauh mana peserta didik memahami prinsip-prinsip keselamatan kerja, penggunaan alat pelindung diri (APD), serta penerapan prosedur kerja aman di lingkungan industri.

Secara teoritis, penelitian ini memperkuat kajian mengenai pentingnya evaluasi berbasis data dan analisis kuantitatif merupakan bagian penting dalam pendidikan vokasi, khususnya pada bidang K3LH yang menuntut ketelitian, akurasi, dan konsentrasi. Dengan memperbaiki kualitas soal melalui analisis validitas dan reliabilitas, pendidikan K3LH akan lebih efektif dalam membangun kesadaran, sikap, dan perilaku aman di tempat kerja

## **SIMPULAN**

Simpulan dari penelitian ini menunjukkan bahwa penerapan aplikasi ANATES dalam menganalisis butir soal K3LH pada tingkat Sekolah Menengah Kejuruan (SMK) memberikan kontribusi besar terhadap peningkatan kualitas instrumen evaluasi pembelajaran. Melalui analisis terhadap 63 butir soal, ditemukan bahwa sebagian besar soal memiliki tingkat validitas dan reliabilitas yang memadai, meskipun beberapa butir masih membutuhkan perbaikan agar selaras dengan indikator pembelajaran serta kompetensi dasar yang dituju.

Hasil uji validitas menunjukkan bahwa dari keseluruhan butir soal, hanya empat butir yaitu nomor 4, 20, 24, dan 30 yang dinyatakan tidak valid karena memiliki korelasi negatif atau berada di bawah batas signifikansi. Dengan demikian, sekitar 93% butir soal dinyatakan valid, menandakan bahwa instrumen evaluasi ini memiliki kemampuan yang baik dalam mengukur kompetensi yang diharapkan. Hal tersebut diperkuat oleh hasil reliabilitas yang mencapai nilai 0,98, yang dikategorikan sangat tinggi. Temuan ini menegaskan bahwa instrumen tes tersebut mampu memberikan hasil yang stabil dan konsisten apabila diujikan kembali pada kelompok peserta didik yang memiliki karakteristik serupa.

Selain itu, analisis daya pembeda menunjukkan bahwa mayoritas butir soal berada dalam kategori baik hingga sangat baik, dengan indeks berkisar antara 40% hingga 100%. Meskipun demikian, terdapat beberapa butir—seperti nomor 4 dan 20—yang memiliki daya pembeda negatif. Hal ini menunjukkan bahwa kedua butir tersebut tidak mampu membedakan peserta didik yang memiliki kemampuan tinggi dan rendah, sehingga memerlukan revisi agar lebih efektif. Secara umum, hanya sebagian kecil butir yang harus

diperbaiki, sementara selebihnya sudah cukup efektif untuk mengukur variasi kemampuan siswa.

Ditinjau dari tingkat kesukaran, sebagian besar butir soal berada pada kategori sedang hingga mudah, dengan rentang kesukaran antara 45%–85%. Meskipun demikian, beberapa soal, seperti nomor 4 dan 5, termasuk dalam kategori sukar. Komposisi tingkat kesukaran ini menunjukkan bahwa tes cukup representatif, tetapi akan lebih optimal jika ditambahkan beberapa soal dengan tingkat kesukaran tinggi guna meningkatkan kemampuan tes dalam membedakan peserta didik secara lebih spesifik.

Dari aspek kualitas pengecoh, sebagian besar pengecoh berfungsi dengan baik, sebagaimana ditunjukkan oleh variasi pemilihan jawaban oleh siswa. Namun, terdapat beberapa pengecoh yang kurang efektif, ditandai dengan simbol “--” atau “---”, yang menunjukkan bahwa pilihan jawaban tersebut kurang menarik bagi peserta. Oleh karena itu, revisi diperlukan untuk memastikan setiap alternatif jawaban memiliki daya tarik yang seimbang.

Secara keseluruhan, hasil penelitian menegaskan bahwa instrumen tes yang dianalisis memiliki validitas tinggi, reliabilitas sangat baik, daya pembeda yang memadai, serta tingkat kesukaran yang cukup seimbang. Hanya sedikit butir soal yang perlu diperbaiki agar kualitas instrumen semakin optimal. Dengan perbaikan tersebut, instrumen evaluasi ini dinilai layak digunakan dalam penilaian pembelajaran, khususnya pada mata pelajaran K3LH, karena mampu mengukur kemampuan siswa secara akurat, adil, dan konsisten.

## DAFTAR PUSTAKA

- Ainun, N. R., Mannyullei, S., & Amqam, H. (2021). Hasanuddin Journal of Public Health. *Hasanuddin Journal of Public Health*, 3(1), 99–114.
- Akhmadi, M. N. (2021). Analisis Butir Soal Evaluasi Tema 1 Kelas 4 SD NPlumbungan menggunakan Program Anates. *Ed- Humanistics : Jurnal Ilmu Pendidikan*, 6(1), 799–806.
- Alista, Y. F., & Syahzanani, R. A. (2023). Analisis Butir Soal Ulangan Harian Fisika dengan Pendekatan Teori Tes Klasik menggunakan Program Anates. *Seminar Nasional ...*, 1–11.  
<https://proceeding.unesa.ac.id/index.php/siptek/article/view/173%0Ahttps://proceeding.unesa.ac.id/index.php/siptek/article/download/173/180>
- Andryan Fitraynsyah, M., Hilmiyati, F., & Sultan Maulana Hasanuddin, U. (2024). Peran Tingkat Kesukaran dan Daya Pembeda dalam Analisis Butir Tes: Kajian Literatur untuk

- Pendidikan Menengah. *Jrep :Jurnal Riset Dan Evaluasi Pendidikan*, 1(4), 252–262.
- Augustia, A. D., Agustia, C. N., Azzahra, D., Hakim, L., & Pratiwi, V. (2025). Analisis Validitas dan Reliabilitas Soal Pilihan Ganda dengan Menggunakan Software Anates pada Mata Pelajaran Perpajakan. *AKADEMIK: Jurnal Mahasiswa Ekonomi & Bisnis*, 5(1), 250–265. <https://doi.org/10.37481/jmeb.v5i1.1165>
- Biologi, P., & Tanjungpura, U. (2024). *Jurnal kajian pembelajaran dan keilmuan*. 8(20), 162–171. <https://doi.org/10.26418/jurnalkpk.v8i2.78282>
- Collins, S. P., Storrow, A., Liu, D., Jenkins, C. A., Miller, K. F., Kampe, C., & Butler, J. (2021). *No Title 濟無No Title No Title No Title*. 4(1), 167–186.
- Elviana. (2020). Analisis Butir Soal Evaluasi Pembelajaran Pendidikan Agama Islam Menggunakan Program Anates.
- Eri Dia, E., Aldi Subangkit Syah, M., & PGRI Jombang, S. (2022). *Fourth Conference on Research and Community Services STKIP PGRI Jombang Transformasi Pendidikan Berbasis Hasil Penelitian dan Pengabdian Masyarakat di Era Merdeka Belajar Tingkat Kesukaran Soal Penilaian Tengah Semester Mata Pelajaran Bahasa Indonesia Di. September*, 14–23.
- Fiska, J. M., Hidayati, Y., Qomaria, N., & Hadi, W. P. (2021). Analisis Butir Soal Ulangan Harian Ipa Menggunakan Software Anates Pada Pendekatan Teori Tes Klasik. *Natural Science Education Research*, 4(1), 65–76. <https://doi.org/10.21107/nser.v4i1.8133>
- Hera, A. S., Zulhijrah, Nabila, J. L., & Shaleh. (2023). 2268-Article Text-9893-1-10-20231224. *Jurnal Ilmiah PGSD FKIP Universitas Mandiri*, 9, 2990–2991.
- Jurnal Visi Eksakta*, 3(1), 59–70.
- Karindi, D. R., & Rofi'i. (2024). Penggunaan ANATES 4.0.9 dalam menguji kelayakan soal pilihan ganda sumatif matematika. *LINEAR: Journal of Mathematics Education*, 5(2), 181–196. <https://doi.org/10.32332/2zg4f753>
- Lumban Gaol, P., Saragih, T., & Hasibuan, P. S. (2022). Keselamatan, Kesehatan Kerja (K3) Pada Proyek Galangan Kapal.
- Mardiyah, H., & Rofi'i. (2024). Penggunaan Aplikasi Anates Pada Analisis Butir Soal Penilaian Akhir Jenjang. *Jurnal Widyaloka*, 11(2), 26–33.
- Mudarrisuna*, 10(2), 58–74. <https://jurnal.ar-raniry.ac.id/index.php/mudarrisuna/article/view/7839>
- Nafs, H., Sridana, N., Hikmah, N., & Soeprianto, H. (2023). Analisis Kualitas Butir Soal Ulangan Akhir Semester Genap Mata Pelajaran Matematika Kelas VII SMPN 6 Mataram Tahun Ajaran 2022/2023. *Jurnal Ilmiah Profesi Pendidikan*, 8(4), 2324–2331. <https://doi.org/10.29303/jipp.v8i4.1785>
- Nidianti, E., & Lukiyono, Y. T. (2022). Edukasi Penggunaan Masker Pada Masyarakat Sebagai Upaya Preventif Terhadap Infeksi Pernapasan Sebagai Akibat Adanya Paparan Co Di Wilayah Gresik. In *SWARNA: Jurnal Pengabdian Kepada Masyarakat* (Vol. 1, Issue 4, pp. 480–486). <https://doi.org/10.55681/swarna.v1i4.186>

- Novriyanti, E., Sarifah, I., & Riyadi, R. (2024). Analisis Kualitas Butir Soal Ujian Tengah Semester Biologi Umum Menggunakan Teori Tes Klasik. *Jurnal Riset Dan Inovasi Pembelajaran*, 4(3), 1844–1860. <https://doi.org/10.51574/jrip.v4i3.2245>
- Nurdiana Tanjung, & Susilawati Susilawati. (2024). Pentingnya Penggunaan Alat Pelindung Diri (APD) pada Pekerja Bangunan terhadap Keselamatan Kerja. *Corona: Jurnal Ilmu Kesehatan Umum, Psikolog, Keperawatan Dan Kebidanan*, 2(2), 86–96. <https://doi.org/10.61132/corona.v2i2.403>
- Puteri, N. D., Putricia, N. D., Arsyanti, K., Putricia, N. D., Hakim, L., & Pratiwi, V. (2025). Analisis Kelayakan Butir Soal Pada Handout Akuntansi Keuangan Dana Kas Kecil Berbasis E-Book Menggunakan Software Anates. *JIBEMA: Jurnal Ilmu Bisnis, Ekonomi, Manajemen, Dan Akuntansi*, 2(3), 205–214. <https://doi.org/10.62421/jibema.v2i3.108>
- Rahmawati, N. S., & Rahman, M. F. (2025). Analisis Butir Soal Objektif Penilaian Sumatif Akhir Semester Gasal Mata Pelajaran Ekonomi. *JiIP - Jurnal Ilmiah Ilmu Pendidikan*, 8(8), 8863–8873. <https://doi.org/10.54371/jiip.v8i8.8837>
- Rezha Nur Amalia, Ragil Setia Dianingati, & Eva Annisaa. (2022). Pengaruh Jumlah Responden Terhadap Hasil Uji Validitas Dan Reliabilitas Kuesioner Pengetahuan Dan Perilaku Swamedikasi. *Generics : Journal of Research in Pharmacy*, 2(1), 9–15.
- Sihombing, F., Yuanita, Susilowati, A., Intan, Silitonga, R., Hernawati, E., Yayi, Haeriyah, S., & Zulkarnaini. (2021). *Buku Ajar Keselamatan Pasien dan Keselamatan Perawat*.
- Siregar, T. M., Agustina, N., Siahaan, F. B., Sinaga, L. L., Prabawa, I., & Sagala, R. Z. (2024). Analisis Validitas dan Reliabilitas Butir Soal Pilihan Ganda Dalam Mata Pelajaran Matematika. *Journal of Education and Learning Evaluation*, 1(2), 725–730.
- Sugiarto, Y. A., Wibowo, S. R., & Pratiwi, V. (2024). Analisis Kelayakan Soal Pilihan Ganda Elemen Akuntansi Lembaga Menggunakan Anates untuk Meningkatkan Efektivitas Evaluasi Pembelajaran. 8(2016), 49499–49508.
- Zuhri, N. Z., Syihabuddin, S., & Tatang, T. (2024). Analisis Validitas, Reliabilitas, dan Tingkat Kesukaran Soal Bahasa Arab Tingkat SMP Berbasis Artificial Intelligence (AI) melalui Platform QuestionWell. *Jurnal Pendidikan Dan Pembelajaran Indonesia (JPPI)*, 4(2), 693–704. <https://doi.org/10.53299/jppi.v4i2.576>