



ANALISIS SOAL HOTS PILIHAN GANDA: ALAT PENILAIAN KOMPETENSI PADA TEKNIK MESIN

Rachmad Syarifudin Hidayatullah¹, Sudirman Rizki Ariyanto¹, Akhmad Hafizh Ainur Rasyid¹, Maisarah Azizah¹

¹Jurusan Teknik Mesin, Univeritas Negeri Surabaya

Email: rachmadhidayatullah@unesa.ac.id

INFO ARTIKEL

Sejarah Artikel:

Diterima 23 April 2023

Disetujui 28 April 2023

Dipublikasikan 30 Mei 2023

Kata Kunci: hots, butir soal, siswa, analisis

Abstrak

Penelitian ini bertujuan untuk menganalisa dari butir soal pilihan ganda yang disusun dengan konsep *high order thinking skills* (HOTS). Penelitian adalah penelitian evaluatif dimana dalam prosesnya melewati beberapa tahapan kegiatan dalam pengumpulan datanya. Dalam penelitian ini peneliti menggunakan sejumlah 35 soal pilihan ganda, sedangkan sampel yang digunakan adalah siswa kelas XII TKRO 1 SMK Dharma Bahari Surabaya, pada mata pelajaran kompetensi sistem bahan bakar injeksi pada tahun ajaran 2020/2021, soal tersebut di berikan kepada siswa sejumlah 34 siswa. Instrumen berupa soal yang diberikan kepada siswa telah melewati tahapan analisis validasi, reliabilitas tingkat kesukaran, daya pembeda dan efektifitas pengecoh, dalam tahapan pengujian instrumen tersebut peneliti menggunakan bantuan *software Anates V4* dengan hasil analisis butir soal taraf signifikansi 5%.

Abstract

This study aims to analyze the multiple-choice items which are arranged using the concept of high-order thinking skills (HOTS). Research is evaluative research that goes through several stages of activity in collecting data. In this study, the researchers used 35 multiple choice questions, while the sample used was students of class XII TKRO 1 SMK Dharma Bahari Surabaya, in the fuel injection system competency subject in the 2020/2021 academic year, the questions were given to 34 students. The instrument in the form of questions given to students has gone through the stages of the validation analysis, difficulty level reliability, discriminating power, and effectiveness of the detractor. In the instrument testing stage, the researcher used the help of Anates V4 software with the results of item analysis at a significance level of 5%.

1. PENDAHULUAN

Penilaian hasil kinerja menjadi salah satu hal menarik dan selalu menjadi sesuatu penting dalam pengembangan sumber daya manusia (Ma'arif & Kartika, 2012)(Quan et al., 2018). Peningkatan kualitas pendidikan dapat dilakukan melalui evaluasi yang dilakukan guru dan pihak berwenang (Khonbi & Sadeghi, 2013). Kemampuan berfikir merupakan komponen yang berpengaruh terhadap prestasi belajar (Kamei & Pavlovic, 2021).

Secara umum, penilaian kompetensi

ditunjukkan dari hasil ujian akhir siswa(Astuti et al., 2012). Penilaian tersebut menggambarkan kemajuan belajar dan grafik peningkatan kemampuan siswa. Kualitas penilaian yang diberikan oleh guru dan pihak berwenang mempunyai hubungan positif terhadap pendekatan yang digunakan dalam pembelajaran (Gerritsen-van Leeuwenkamp et al., 2019).

Dalam menyusun soal membangun konstruksi soal yang tepat adalah sebuah tantangan, melakukan analisis item dapat

memberikan umpan balik berupa nilai validitas soal yang disusun (Shigli et al., 2018)(Hoang Thuan et al., 2019). Hal tersebut tentunya tidak lepas dari kualitas dan akurasi instrumen yang digunakan oleh guru. Karakteristik instrumen yang baik yakni harus valid, reliabel, dan objektif baik dari konten maupun struktur (Haryanto, 2020). Semakin tinggi skor yang didapatkan, maka kualitas instrumen semakin baik. Sebelum soal diberikan kepada siswa pembuat soal perlu melakukan analisa butir soal agar diperoleh soal yang berkualitas yaitu soal yang mempunyai nilai validitas dan reliabilitas (Kumar et al., 2021).

Dari beberapa instrumen uji evaluasi kompetensi, soal pilihan ganda yang banyak digunakan oleh guru sekolah menengah kejuruan (SMK). Alasan pemilihan instrumen ini adalah sangat efisien, andal, dan mudah dalam proses standarisasi. Oleh karena itu, kualitas dari setiap butir soal perlu diperhatikan karena dapat berpengaruh terhadap pengetahuan dan pemahaman siswa saat proses penilaian. Dalam soal pilihan ganda pembuat soal dapat mengurangi kecurangan ketika proses ujian dengan cara mempersonalisasi pertanyaan dan pilihan jawaban (Manoharan, 2019). Selain itu pemilihan tes pilhan ganda karena jenis tes ini merupaj jenis tes ang paling populer di pendidikan.(Butler, 2018). Pertanyaan pilihan ganda yang dirancang dengan baik memungkinkan siswa untuk meningkatkan kemampuan Higher Order Thinking Skills

(HOTS), seperti interpretasi, analisis dan pemikiran kritis, aplikasi atau sintesis dalam kerangka Taksonomi Bloom dan Piramida Miller.

Menyusun setiap butir soal pilihan ganda merupakan tugas yang menantang. Bank soal pilihan ganda harus disusun dengan cermat dan dilakukan analisis pada setiap butir soalnya secara komprehensif. Hal tersebut bertujuan agar didapatkan instrumen dengan kualitas tingg dan dapat dimanfaatkan secara maksimal oleh institusi dalam melakukan penilaian.

Pada penelitian ini dilakukan analisis butir soal pilihan ganda untuk siswa SMK bidang keahlian Teknik Otomotif. Analisis butir soal ini digunakan untuk mengukur kualitas butir soal yang telah buat. Butir soal dengan kualitas tinggi akan dipertahankan sedangkan yang berkuliatas rendah akan dibuang atau diganti dengan butir soal yang lebih baik. Tujuan utama penelitian adalah untuk menghasilkan soal pilihan ganda HOTS yang layak digunakan sebagai instrumen evaluasi domain kognitif.

2. METODE PENELITIAN

Penelitian adalah penelitian evaluatif dimana dlaam prosesnya melewati beberapa tahapan kegiatan dalam pengumpulan datanya. Data yang diperoleh dari kegiatan pengambilan data selanjutnya dibandingkan dengan kriteria yang telah ditentukan. Dalam penelitian ini peneliti menggunakan sejumlah 35 soal pilihan ganda, sedangkan sampel yang

digunakan adalah siswa kelas XII TKRO 1 SMK Dharma Bahari Surabaya, pada mata pelajaran kompetensi sistem bahan bakar injeksi pada tahun ajaran 2020/2021, soal tersebut di berikan kepada siswa sejumlah 34 siswa. dalam pengambilan data pada penelitian ini peneliti menggunakan teknik dokumentasi dan instrumen penelitian yang digunakan yaitu: 1 set naskah soal jenis pilihan ganda dengan tipe soal HOTS lengkap dengan kunci jawaban. Instrumen berupa soal yang diberikan kepada siswa telah melewati tahapan analisis validasi, reliabilitas tingkat kesukaran, daya pembeda dan efektifitas pengecoh, dalam tahapan pengujian instrumen tersebut peneliti menggunakan bantuan software Anates V4 dengan hasil analisis butir soal taraf signfikansi 5%.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Hasil Validasi Instrumen Tes

Penilaian pada 3 aspek Instrumen Tes adalah aspek materi, aspek konstruksi dan aspek Bahasa atau budaya. Hasil penilaian validator ditunjukkan pada Tabel 1 di bawah ini.

Tabel. 1 Hasil Validasi Instrumen Hasil Belajar Ranah Kognitif

No	Aspek	Average	Kategori
1.	Materi	4	Valid
2.	Konstruksi	3,77	Valid
3.	Bahasa atau Budaya	4,17	Valid
Rata-rata Total		3,98	Valid

Berdasarkan data **Error! Reference source not found.** pada aspek materi memiliki nilai 4 dikatakan valid. Aspek konstruksi memiliki nilai 3,77 dikatakan valid dan aspek

Bahasa atau budaya mendapat nilai 4,17 dikatakan valid. Jadi, rata-rata total ketiga kategori bernilai 3,98 berkategori valid. Dilihat dari hasil tersebut, disimpulkan bahwa instrumen tersebut layak diaplikasikan untuk mendapatkan hasil belajar ranah kognitif.

Hasil Uji Coba Lapangan

Setelah seluruh instrumen penilaian dan perangkat pembelajaran divalidasi oleh validator ahli, maka selanjutnya dilakukan uji coba lapangan di SMK Dharma Bahari Surabaya dengan hasil sebagai berikut.

Validitas Butir Soal

Analisis butir soal dilakukan di SMK Dharma Bahari Surabaya pada kelas XII TKR 1 dengan responden sebanyak 34 siswa. Tujuan dilakukan analisis butir soal yaitu untuk mengetahui bagaimana kualitas butir soal yang kedepannya akan digunakan dalam pengambilan data untuk hasil penelitian ini. Jenis soal yang digunakan pada analisis butir soal ini yaitu soal pilihan ganda dengan 5 opsi pilihan, dengan jumlah butir soal yang sejumlah 35.

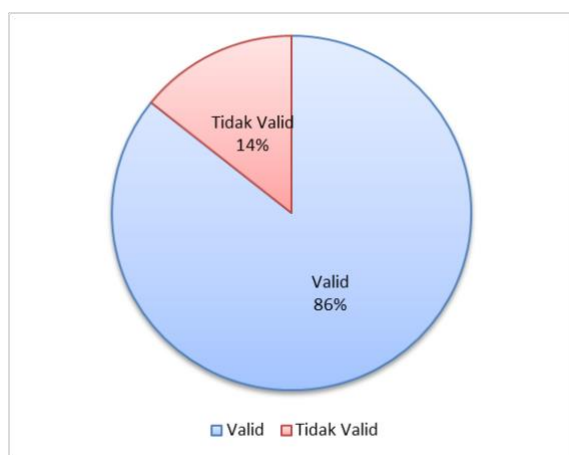
Hasil analisis butir soal diuji dengan taraf signifikansi 5%, sehingga karena jumlah (N) sebanyak 34 siswa maka nilai rtabel product moment sebesar 0,349. Berdasarkan hasil analisis, butir soal dapat dikategorikan valid apabila nilai $Y_{pbi} > 0,349$ dan butir soal dikategorikan tidak valid apabila nilai $Y_{pbi} < 0,349$. Distribusi soal yang dibagi berdasarkan indeks validitas dapat dilihat pada.2 berikut.

Tabel. 2

Distribusi Soal Berdasarkan Indeks Validitas		
Indeks	Soal Valid	Soal Tidak

Validitas	($Y_{pbi} \geq 0,349$)	Valid ($Y_{pbi} < 0,349$)
No. Butir Soal	1, 2, 3, 5, 6, 7, 8, 9, 10, 12, 13, 14, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 31, 32, 33, 34	4, 11, 15, 30, 35
Jumlah	30	5
Persentase	86%	14%

Tabel.2 di atas dapat ditampilkan dalam bentuk diagram lingkaran atau diagram pie seperti pada Gambar dibawah ini



Gambar. 1 Diagram *Pie* Distribusi Soal Berdasarkan Indeks Validitas

Berdasarkan diagram pie pada Gambar, dapat diketahui bahwa dari total 35 soal pilihan ganda sebanyak 86% dikategorikan dalam kategori “Valid”, yaitu terdapat 30 soal yang masuk dalam kategori “Valid”, sedangkan persentase sebesar 14% masuk pada kategori “Tidak Valid” yaitu sebanyak 5 soal. Maka dapat dikatakan bahwa 30 soal dari total 35 soal keseluruhan soal pilihan ganda pada Kompetensi sistem bahan bakar injeksi layak digunakan karena termasuk dalam kategori “Valid”.

Tingkat Kesukaran Butir Soal

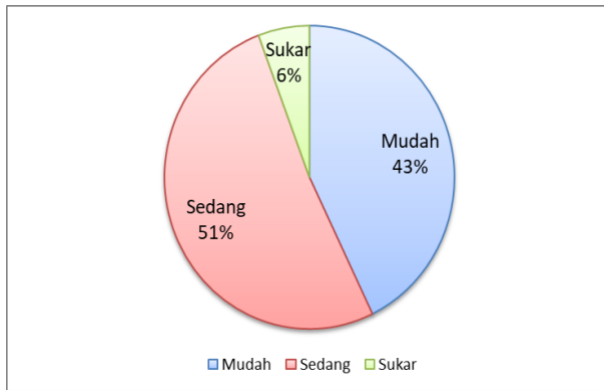
Soal tes dikategorikan baik apabila soal yang diujikan tidak terlalu mudah atau terlalu sukar (Sahoo & Singh, 2017). Agar dapat mengetahui bagaimana tingkat kesukaran butir soal, seorang guru perlu melakukan pengujian. Adapun tingkatan kesukaran butir soal dibagi menjadi tiga, yaitu mudah, sedang, dan sukar.

Jika dibagi berdasarkan kelompok indeks tingkat kesukaran, maka dapat diketahui berapa besar persentase dari tingkat kesukaran masing-masing kelompok. Tabel di bawah ini menunjukkan distribusi butir soal berdasarkan indeks tingkat kesukaran.

Tabel. 3
Distribusi Soal Berdasarkan Indeks Tingkat Kesukaran

Indeks Validitas	Mudah	Sedang	Sukar
	(0,70 – 1)	(0,3 – 0,69)	(0,00 – 0,29)
No. Butir Soal	1, 2, 3, 5, 6, 9, 13, 14, 19, 21, 28, 29, 30, 33, 34	4, 7, 8,10, 12, 16, 17, 18, 20, 22, 23, 24, 25, 26, 27, 31, 32, 35	11, 15
Jumlah	15	18	2
Persentase	44%	50%	6%

Jika data pada Tabel2 ditampilkan dalam bentuk diagram lingkaran atau diagram pie, maka terlihat seperti Gambar. 1. 2 dibawah ini.



Gambar. 1 Diagram Pie Distribusi Soal Berdasarkan Indeks Tingkat Kesukaran

Gambar. 1 diatas menunjukkan bahwa dari 35 soal pilihan ganda Kompetensi sistem bahan bakar injeksi, soal dengan kategori “Mudah” memiliki persentase 44% yaitu sebanyak 15 soal, soal dengan kategori “Sedang” memiliki persentase 50% yaitu sebanyak 17 soal. Dan untuk soal dengan kategori “Sukar” memiliki persentase 6% yaitu sebanyak 2 soal.

Daya Bada Butir Soal

Daya pembeda butir soal adalah kemampuan dari suatu soal dalam membedakan siswa yang masuk pada kategori pintar hingga kurang pintar (Andraini et al., 2021). Daya beda yang digunakan dalam penelitian ini terdapat lima kategori, yaitu: (1) baik sekali; (2) baik; (3) cukup; (4) buruk; dan (5) soal tidak baik.

Berdasarkan hasil pengujian daya beda butir soal, apabila hasil analisis dibagi dalam lima kategori, maka diketahui seberapa besar persentase daya beda pada setiap butir soal. tabel 4 di bawah ini menunjukkan distribusi soal berdasarkan indeks daya beda.

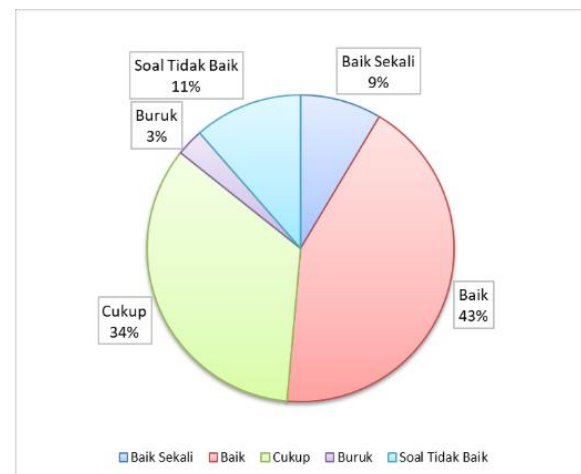
Tabel 4

Distribusi Soal Berdasarkan Indeks Daya Bada

Indeks Validita	No. Butir Soal	Jumla h	Persentas e
-----------------	----------------	---------	-------------

s			
Baik Sekali	8, 22, 32	3	9%
Baik	1, 6, 10, 12, 13, 17, 18, 20, 23, 24, 25, 26, 28, 31, 34	15	43%
Cukup	2, 3, 5, 9, 14, 16, 19, 21, 27, 29, 33	12	34%
Buruk	4	1	3%
Soal Tidak Baik	11, 15, 30, 35	4	11%

Tabel 4 jika ditampilkan dalam bentuk diagram lingkaran atau diagram pie, maka akan terlihat seperti **Error! Reference source not found.** dibawah ini



Gambar. 3 Diagram Pie Distribusi Soal Berdasarkan Indeks Daya Bada

Berdasarkan **Error! Reference source not found.** diagram pie diatas dapat diketahui bahwa dari 35 soal pilihan ganda Kompetensi sistem bahan bakar injeksi, soal dengan kategori “Baik” memiliki kontribusi yang cukup signifikan dimana memiliki persentase sebesar 43%. Sedangkan soal dengan kategori

“Baik Sekali” persentase sebesar 9%, soal dengan kategori “Cukup” persentase sebesar 34%, dan soal dengan kategori “Soal Tidak Baik” memiliki persentase sebesar 11%. Kemudian soal dengan kategori “Buruk” memiliki persentase yang paling rendah yakni sebesar 3%.

Efektivitas Pengecoh (Distraktor)

Pada umumnya pengecoh (distraktor) digunakan dengan tujuan untuk mengecoh melalui pilihan jawaban yang tersedia pada soal. Menurut Arikunto (2013) suatu pengecoh dikatakan berfungsi apabila suatu jawaban paling sedikit dipilih oleh 5% responden.

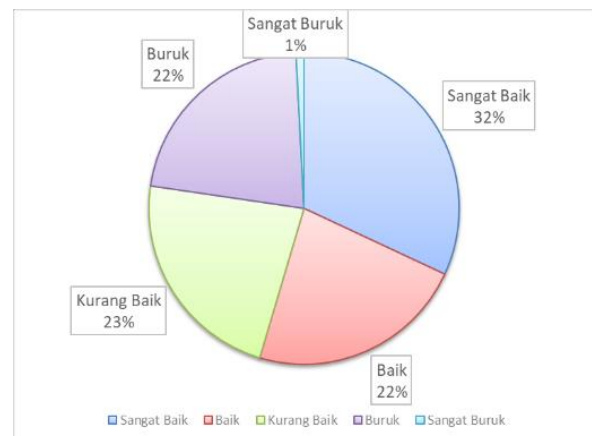
Berdasarkan hasil dari analisis pengecoh, jika dibagi ke dalam empat kelompok indeks pengecoh, maka akan dapat diketahui berapa persentase efektivitas pengecoh pada tiap kategori (Riani & Almujab, 2020). Perhitungan efektivitas pengecoh dilakukan berdasarkan jumlah pilihan jawaban yang kemudian akan dikalikan dengan jumlah butir soal. Namun, kunci jawaban tidak termasuk dalam perhitungan sehingga total jumlahnya menjadi 110 pengecoh. Distribusi soal berdasarkan indeks pengecoh ditunjukkan pada di bawah ini.

Tabel. 5.
Distribusi Soal Berdasarkan Indeks Pengecoh

Indeks Validitas	Jumlah	Persentase
Baik Sekali	38	0.32
Baik	27	0.23
Kurang Baik	27	0.23
Buruk	26	0.22
Soal Buruk	1	0.01

Berdasarkan Tabel. 5 diatas, jika

digambarkan dalam bentuk diagram lingkaran atau diagram pie, maka akan terlihat seperti Gambar 4 Gambar 4 berikut ini.



Gambar 4 Diagram Pie Distribusi Soal Berdasarkan Indeks Pengecoh

Berdasarkan Gambar 44 diagram pie diatas, dapat diketahui berapa persentase masing-masing efektivitas pengecoh pada setiap kategori. Hasil diatas menunjukkan bahwa dari 110 pengecoh yang ada pada 35 butir soal, pengecoh dengan kategori “Sangat Baik” yaitu sebanyak 38 soal atau 32%. Kategori “Baik” yaitu sebanyak 27 soal atau 23% pengecoh. Kemudian pengecoh dengan kategori “Kurang baik” yaitu sebanyak 27 soal atau 23%. Sedangkan untuk pengecoh dengan kategori “Buruk” yaitu sebanyak 26 soal atau 22%, dan pengecoh dengan kategori “Sangat Buruk” yaitu hanya berjumlah 1 soal atau 1%.

Reliabilitas Instrumen Hasil Belajar Ranah Kognitif

Reliabilitas Instrumen Hasil Belajar Ranah Kognitif berdasarkan dari hasil validasi 3 validator ahli terhadap kualitas 35 butir soal. Jumlah pengamat lebih dari 3 orang, sehingga reliabilitas diuji menggunakan dua teknik yaitu

teknik *Cronbach's Alpha* dan teknik *intraclass correlation* (ICC) melalui *software* SPSS 24. Nilai hasil pengujian reliabilitas terhadap hasil belajar ranah kognitif bisa diperhatikan pada Tabel 6

Tabel 6
Hasil Uji Reliabilitas Butir Soal Hasil Belajar Ranah Kognitif

Reliability Statistics	
Cronbach's Alpha	N of Items
0,831	3

Hasil pengujian pada Tabel 5, butir soal yang ada pada hasil belajar ranah kognitif mendapatkan nilai koefisien reliabilitas sebesar 0,831. Nilai tersebut apabila disesuaikan dengan klasifikasi kategori teknik *Cronbach's Alpha* pada Tabel 5 disimpulkan bahwa instrumen termasuk kategori istimewa sehingga, butir soal hasil belajar ranah kognitif layak digunakan dan dapat dipercaya.

Berdasarkan hasil pengolahan data reliabilitas antar pengamat dilakukan menggunakan teknik ICC, dengan hasil uji reliabilitasnya di dalam Tabel 7 berikut.

Tabel 7. ICC Reliabilitas

	Intraclass Correlation Coefficient		
	Intraclass Correlation	95% Confidence Interval	
		Lower Bound	Upper Bound
Single Measures Average	0,622	0,353	0,825
Measures	0,831	0,621	0,934

Berdasarkan hasil uji Tabel 6 di atas, diperoleh nilai reliabilitas hasil belajar ranah kognitif antar pengamat sebesar 0,622. Nilai

tersebut disesuaikan dengan klasifikasi kategori teknik ICC pada Tabel 5 maka dapat dinyatakan reliabilitas antar pengamat masuk dalam kategori sedang.

3.2 Pembahasan

Validitas Instumen Penelitian

Hasil validasi instrumen tes menunjukkan bahwa instrumen yang disusun layak digunakan dalam proses pembelajaran. Instrumen hasil belajar ranah kognitif mendapatkan nilai sebesar 3,98 dan masuk kategori valid. Berdasarkan hasil validasi yang telah dilakukan, maka dinyatakan bahwa instrumen tes yang digunakan dalam proses pembelajaran layak digunakan.

Hasil penelitian ini sejalan dengan penelitian Connell et al. (2018) yang menyatakan bahwa suatu instrumen penilaian dapat dikategorikan layak digunakan apabila masuk dalam kategori minimal valid. Memperkuat pernyataan tersebut, Ariyanto et al. (2019a) menjelaskan bahwa kunci dari validitas suatu instrumen terletak dari sejauhmana instrumen tersebut mampu menangkap apa yang dimaksudkan untuk diukur. Hasil uji validasi, dalam hal ini memberikan definisi operasional dari tingkat kelayakan instrumen, yang memenuhi berbagai persyaratan yang dikriteriakan untuk mendapatkan kategori valid (Buntins et al., 2017).

Hasil Uji Coba Lapangan

Hasil uji coba lapangan yang telah dilakukan pada kelas eksperimen, yang perama untuk validitas butir soal menunjukkan bahwa

dari 35 soal yang diujikan kepada siswa menunjukkan bahwa sejumlah 30 soal valid atau jika dipersentasekan yaitu 86% dan 5 soal tidak valid atau jika dipersentasekan yaitu 14%. Validitas butir soal merupakan salah satu indikator yang perlu diperhatikan dalam pembelajaran, apabila butir soal memiliki tingkat validitas yang tinggi maka dapat dikatakan butir soal tersebut dapat mengukur seberapa jauh kemampuan siswa dalam memahami pembelajaran. Arikunto (2009) mengatakan bahwa suatu tes akan mempunyai tingkat validitas tinggi jika mampu mencapai dan sesuai dengan indikator pembelajaran dalam kompetensi dasar.

Untuk tingkat kesukaran butir soal didapatkan hasil bahwa 3 atau 9% dengan kategori baik sekali, 15 soal atau 44% dengan kategori mudah, 18 soal atau 50% dengan kategori sedang, dan 2 soal atau 6% dengan kategori sukar. Menurut Arikunto (2009) soal yang termasuk baik yaitu soal yang tidak terlalu sulit atau juga terlalu mudah. Permasalahan yang penting saat menganalisis tingkat kesukaran soal yaitu menentukan kriteria soal dan juga proporsi soal yang termasuk sukar, sedang, dan mudah.

Selanjutnya untuk daya beda butir soal menghasilkan soal dengan kategori baik sekali 9% yaitu sebanyak 3 soal, kategori baik 43% yaitu sebanyak 15 soal, kategori cukup 34% yaitu sebanyak 12 soal, kategori buruk 11% yaitu sebanyak 1 soal, dan kategori soal tidak baik 11% yaitu sebanyak 4 soal. Hasil yang cukup signifikan diperoleh kategori baik,

artinya pada butir soal ini dapat menentukan antara siswa yang pandai hingga kurang pandai.

Sementara itu, distraktor dengan hasil kategori baik sekali 32% yaitu sebanyak 38, kategori baik 23% yaitu sebanyak 27, kategori kurang baik 23% yaitu sebanyak 27, kategori buruk 22% yaitu sebanyak 26, kategori soal buruk 1% yaitu sebanyak 1. Hasil yang cukup signifikan yaitu ditunjukkan pada kategori baik sekali, artinya distraktor pada butir soal ini dinyatakan efektif.

Reliabilitas Instrumen Penelitian

Uji reliabilitas dilakukan dengan dua teknik, yakni *Cronbach's Alpha* dan *intraclass correlation* (ICC). Teknik *Cronbach's Alpha* digunakan mengukur keandalan indikator-indikator yang digunakan dalam instrumen penelitian, sedangkan teknik ICC digunakan untuk menilai reliabilitas antar dua atau lebih pengamat. Hasil uji reliabilitas instrument hasil belajar ranah kognitif menggunakan teknik *Cronbach's Alpha* adalah sebesar 0,831 dan masuk dalam kategori istimewa.

Hasil uji reliabilitas instrument hasil belajar ranah kognitif menggunakan teknik ICC adalah sebesar 0,622. Instrumen hasil belajar ranah kognitif masuk dalam kategori sedang Berdasarkan hasil tersebut maka dinyatakan bahwa ketiga validator mencapai kesepakatan terkait instrument yang telah disusun.

Taber (2018) mengemukakan bahwa uji reliabilitas banyak digunakan oleh peneliti sebagai salah satu indikator untuk menilai kualitas dari suatu instrumen. Menguatkan pernyataan tersebut, Khan et al. (2019)

menjelaskan bahwa dengan melakukan uji reliabilitas kita dapat mengukur keajegan dari setiap butir-butir pertanyaan. Sejalan dengan dua pernyataan tersebut, Chen et al. (2020) juga mengungkapkan bahwa uji reliabilitas banyak dilakukan untuk menguji kestabilan dan konsistensi dari instrumen yang sedang dikembangkan.

4. KESIMPULAN

Berdasarkan penjelasan dari hasil dan pembahasan di atas, kemudian di korelasikan dengan latar belakang dan tujuan penelitian, dapat ditarik kesimpulan yaitu soal pilihan ganda HOTS yang disusun dinyatakan valid baik secara validitas instrumen maupun diuji di lapangan atau pada peserta didik. Hal tersebut ditinjau dari beberapa aspek yaitu validitas yang mendapat sebanyak 86% soal masuk dalam kategori Valid. Kemudian untuk tingkat kesukaran butir soal dengan kategori sukar sebanyak 6%, kategori sedang sebanyak 51% dan kategori mudah sebanyak 43%. Untuk daya beda untuk katgeori buruk hanya terdapat 3%. Pada soal pengecoh didapatkan pengecoh dengan kategori “Sangat Baik” yaitu sebanyak 38 soal atau 32%. Kategori “Baik” yaitu sebanyak 27 soal atau 23% pengecoh. Kemudian pengecoh dengan kategori “Kurang baik” yaitu sebanyak 27 soal atau 23%. Sedangkan untuk pengecoh dengan kategori “Buruk” yaitu sebanyak 26 soal atau 22%, dan pengecoh dengan kategori “Sangat Buruk” yaitu hanya berjumlah 1 soal atau 1%. Kemudian untuk reliabilitas mendapatkan nilai

koefisien reliabilitas sebesar 0,831 dengan klasifikasi kategori teknik *Cronbach's Alpha* masuk kategori istimewa atau layak, dan jika dengan klasifikasi teknik ICC didapat sebesar 0,622 atau masuk kategori sedang.

5. DAFTAR PUSTAKA

- Andraini, M. R., Rohiat, S., & Elvia, R. (2021). ANALISIS KEMAMPUAN BERPIKIR KRITIS SISWA PADA MATERI REAKSI REDUKSI OKSIDASI (REDOKS) DI MAN 1 KOTA BENGKULU. *Alotrop*, 5(1), 35–41. <https://doi.org/10.33369/atp.v5i1.16484>
- Ariyanto, S. R., Munoto, M., & Muhaji, M. (2019). Development of affective authentic assessment instruments for automotive engineering expertise in vocational school. *Jurnal Taman Vokasi*, 7(1), 42. <https://doi.org/10.30738/jtv.v7i1.4777>
- Astuti, W. P., Prasetyo, A. P. B., & Rahayu, E. S. (2012). PENGEMBANGAN INSTRUMEN ASESMEN AUTENTIK BERBASIS LITERASI SAINS PADA MATERI SISTEM EKSKRESI. *Journal of Educational Research*, 41(01). <https://doi.org/10.15294/lik.v41i1.2228>
- Buntins, M., Buntins, K., & Eggert, F. (2017). Clarifying the concept of validity: From measurement to everyday language. *Theory & Psychology*, 27(5), 703–710. <https://doi.org/10.1177/0959354317702256>
- Butler, A. C. (2018). Multiple-Choice Testing in Education: Are the Best Practices for Assessment Also Good for Learning? *Journal of Applied Research in Memory and Cognition*, 7(3), 323–331. <https://doi.org/10.1016/j.jarmac.2018.07.002>
- Chen, T., Peng, L., Yin, X., Rong, J., Yang, J., & Cong, G. (2020). Analysis of User Satisfaction with Online Education Platforms in China during the COVID-19 Pandemic. *Healthcare*, 8(3), 200. <https://doi.org/10.3390/healthcare8030200>

- Connell, J., Carlton, J., Grundy, A., Taylor Buck, E., Keetharuth, A. D., Ricketts, T., Barkham, M., Robotham, D., Rose, D., & Brazier, J. (2018). The importance of content and face validity in instrument development: lessons learnt from service users when developing the Recovering Quality of Life measure (ReQoL). *Quality of Life Research*, 27(7), 1893–1902. <https://doi.org/10.1007/s11136-018-1847-y>
- Gerritsen-van Leeuwenkamp, K. J., Joosten-ten Brinke, D., & Kester, L. (2019). Students' perceptions of assessment quality related to their learning approaches and learning outcomes. *Studies in Educational Evaluation*, 63, 72–82. <https://doi.org/10.1016/j.stueduc.2019.07.005>
- Haryanto. (2020). *Evaluasi Pembelajaran*. UNY Press.
- Hoang Thuan, N., Drechsler, A., & Antunes, P. (2019). Construction of Design Science Research Questions. *Communications of the Association for Information Systems*, 44(20), 332–363. <https://doi.org/10.17705/1CAIS.04420>
- Kamei, T., & Pavlovic, M. (2021). Investigating differential item functioning to validate a thinking skills learning progression for students with intellectual disability and autism spectrum disorder. *International Journal of Educational Research*, 106, 101726. <https://doi.org/10.1016/j.ijer.2020.101726>
- Khan, T., Johnston, K., & Ophoff, J. (2019). The Impact of an Augmented Reality Application on Learning Motivation of Students. *Advances in Human-Computer Interaction*, 2019, 1–14. <https://doi.org/10.1155/2019/7208494>
- Khonbi, Z. A., & Sadeghi, K. (2013). The Effect of Assessment Type (Self Vs. Peer) on Iranian University EFL Students' Course Achievement. *Procedia - Social and Behavioral Sciences*, 70, 1552–1564. <https://doi.org/10.1016/j.sbspro.2013.01.223>
- Kumar, D., Jaipurkar, R., Shekhar, A., Sikri, G., & Srinivas, V. (2021). Item analysis of multiple choice questions: A quality assurance test for an assessment tool. *Medical Journal Armed Forces India*, 77, S85–S89. <https://doi.org/10.1016/j.mjafi.2020.11.007>
- Ma'arif, M. S., & Kartika, L. (2012). *Manajemen Kinerja Sumber Daya Manusia*. Percetakan IPB.
- Manoharan, S. (2019). Cheat-resistant multiple-choice examinations using personalization. *Computers & Education*, 130, 139–151. <https://doi.org/10.1016/j.compedu.2018.11.007>
- Quan, P., Shi, Y., Niu, L., Liu, Y., & Zhang, T. (2018). Automatic Chinese Multiple-Choice Question Generation for Human Resource Performance Appraisal. *Procedia Computer Science*, 139, 165–172. <https://doi.org/10.1016/j.procs.2018.10.235>
- Riani, D., & Almujab, S. (2020). ANALISIS BUTIR SOAL DAN KEMAMPUAN SISWA DALAM MENJAWAB SOAL UJIAN NASIONAL PADA MATA PELAJARAN EKONOMI. *OIKOS Jurnal Kajian Pendidikan Ekonomi Dan Ilmu Ekonomi*. <https://doi.org/10.23969/oikos.v4i1.2425>
- Shigli, K., Nayak, S. S., Gali, S., Sankeshwari, B., Fulari, D., Shyam Kishore, K., Upadhyaya P., N., & Jirge, V. (2018). Are Multiple Choice Questions for Post Graduate Dental Entrance Examinations Spot On?-Item Analysis of MCQs in Prosthodontics in India. *Journal of the National Medical Association*, 110(5), 455–458. <https://doi.org/10.1016/j.jnma.2017.11.001>
- Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, 48(6), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>

