

PENGEMBANGAN INSTRUMEN ASESMEN KOGNITIF SISTEM AIR CONDITIONING KENDARAAN RINGAN

Albiyan Niezar Rozi ^{1*}; Yusep Sukrawan ²; Ridwan Adam M. Noor ³

¹Program Studi Pendidikan Teknik Otomotif, Universitas Indonesia/Bandung, Indonesia

¹Program Studi Pendidikan Teknik Otomotif, Universitas Indonesia/Bandung, Indonesia

¹Program Studi Pendidikan Teknik Otomotif, Universitas Indonesia/Bandung, Indonesia

*Corresponding Author: albi16@upi.edu

Abstract

The dynamics of the automotive industry demand vocational graduates who are competent in diagnosing light vehicle Air Conditioning (A/C) system failures. However, existing assessment instruments are still dominated by memorization tests (Lower-Order Thinking Skills/LOTS) and have not optimally measured problem-solving abilities (Higher-Order Thinking Skills/HOTS). Therefore, this study aims to develop a valid and reliable HOTS-based cognitive assessment instrument for light vehicle A/C systems. This study uses the educational assessment development model by Mardapi, focusing on four main stages: (1) compiling test specifications, (2) writing test items, (3) qualitative item review (expert validation), and (4) preliminary empirical testing and item analysis. The resulting product is a set of 50 objective multiple-choice test items. The content validation results by three experts showed a Content Validity Ratio (CVR) value of 1.00, meaning the instrument is theoretically valid. Furthermore, the preliminary empirical field test results on 35 students showed that 44 items (88%) were declared valid with a Kuder-Richardson 20 (KR-20) reliability coefficient of 0.872 (very high). This instrument shows excellent initial feasibility as an objective diagnostic tool, although it still requires large-scale testing for advanced psychometric parameter analysis.

Keywords: Cognitive Assessment, Instrument, HOTS, A/C System, Vocational Education

Abstrak

Dinamika industri otomotif menuntut lulusan vokasi yang kompeten dalam mendiagnosis kerusakan berbasis studi sistem *Air Conditioning* (A/C) kendaraan ringan. Namun, instrumen asesmen yang ada saat ini masih didominasi oleh pengujian daya ingat (*Lower-Order Thinking Skills/LOTS*) dan belum mengukur kemampuan pemecahan masalah (*Higher-Order Thinking Skills/HOTS*) secara optimal. Oleh karena itu, penelitian ini bertujuan mengembangkan instrumen asesmen kognitif berbasis HOTS untuk sistem A/C kendaraan ringan yang valid dan reliabel. Penelitian ini menggunakan pendekatan pengembangan instrumen tes dari Mardapi yang difokuskan pada empat tahapan utama: (1) penyusunan spesifikasi tes, (2) penulisan butir soal, (3) telaah kualitatif soal (validasi ahli), dan (4) uji coba empiris serta analisis butir soal. Produk yang dihasilkan berupa 50 butir soal pilihan ganda objektif. Hasil validasi isi oleh tiga pakar menunjukkan nilai *Content Validity Ratio* (CVR) sebesar 1,00, yang berarti instrumen sangat valid secara teoretis. Selanjutnya, hasil uji coba empiris terbatas pada 35 mahasiswa menunjukkan bahwa 44 butir soal (88%) dinyatakan valid dengan koefisien reliabilitas *Kuder-Richardson 20* (KR-20) sebesar 0,872 (sangat tinggi). Instrumen ini menunjukkan mencapai kualitas psikometri yang baik sebagai alat ukur diagnostik yang objektif, sehingga berpotensi digunakan sebagai instrumen asesmen kognitif berbasis HOTS pada pembelajaran sistem A/C kendaraan ringan setelah melalui uji coba skala luas.

Kata Kunci: Asesmen Kognitif, Instrumen, HOTS, Sistem A/C, Pendidikan Vokasi

PENDAHULUAN

Perubahan teknologi pada industri otomotif telah meningkatkan tuntutan kompetensi lulusan pendidikan vokasi agar tidak hanya memenuhi prinsip *link and match*, tetapi juga mampu menguasai sistem kendaraan modern secara komprehensif. Kondisi tersebut menunjukkan bahwa kompetensi teknisi otomotif saat ini semakin kompleks sehingga memerlukan penguasaan pengetahuan konseptual dan kemampuan pemecahan masalah yang lebih tinggi (Husein, 2019; Erjavec, 2020; Hadiyanto, 2023). Salah satu kompetensi krusial yang wajib dikuasai mahasiswa di bidang ini adalah perawatan dan perbaikan sistem *Air Conditioning (A/C)* kendaraan ringan (Rohman & Setiadi, 2022). Pemahaman terhadap sistem ini tidak sekadar menuntut keterampilan bongkar pasang komponen dasar (Daryanto, 2013), tetapi juga memerlukan fondasi kognitif yang kuat mengenai termodinamika dan siklus refrigerasi (Tomczyk, 2017) guna menunjang kemampuan diagnosis kerusakan. Pembelajaran sistem A/C ini seringkali membutuhkan integrasi antara pemahaman teoretis yang kuat, instrumen yang tepat, interpretasi data manifold gauge, evaluasi alternatif kerusakan AC, dan pengambilan keputusan teknis agar hasil belajar mahasiswa dapat diukur dengan maksimal (Isnanto & Kustono, 2019).

Beberapa penelitian sebelumnya telah mengembangkan instrumen penilaian pada kompetensi otomotif, seperti sistem starter, sistem pengapian, maupun penilaian kinerja praktik. Namun, belum ditemukan instrumen asesmen kognitif berbasis HOTS yang secara khusus dirancang untuk mengukur kemampuan diagnosis sistem Air Conditioning kendaraan ringan melalui analisis kasus. Praktik asesmen kognitif pada materi otomotif saat ini masih memiliki kelemahan mendasar karena cenderung berfokus pada *Lower-Order Thinking Skills (LOTS)* (Maulidia & Pahlevi, 2020). Evaluasi pembelajaran seringkali terhenti pada tahap mengingat dan memahami (C1-C2) dalam revisi Taksonomi Bloom (Anderson & Krathwohl, 2001). Di sisi lain, industri sangat membutuhkan lulusan dengan kemampuan *Higher-Order Thinking Skills (HOTS)* yang sanggup memecahkan masalah (problem-solving) secara mandiri dalam menghadapi kerusakan kendaraan (Nurtanto, Fawaid, & Sofyan, 2020). Asesmen vokasi yang tepat harus berfungsi sebagai alat ukur diagnostik yang andal dan tidak sekadar menguji daya ingat (Fitriyani et al., 2021; Mardapi, 2017). Penilaian diagnostik kognitif yang spesifik pada sistem kendaraan terbukti mampu memetakan letak kesulitan belajar dan kemampuan pemecahan masalah mahasiswa secara lebih akurat (Firmansyah & Mulyani, 2022). Sayangnya, ketiadaan instrumen standar yang mampu mengukur analisis diagnostik secara objektif kerap menimbulkan kesenjangan antara pencapaian akademis dan kompetensi riil di lapangan (Fauzi, Pardjono, & Soenarto, 2020; Suratno, 2016).

Berdasarkan celah penelitian (*research gap*) tersebut, penelitian ini bertujuan

mengembangkan instrumen asesmen kognitif sistem A/C kendaraan ringan yang valid dan terstruktur. Kebaruan penelitian ini terletak pada penciptaan seperangkat tes objektif berbobot yang secara spesifik menggeser fokus penilaian dari hafalan teoretis menjadi analisis diagnostik berbasis studi kasus dengan mengintegrasikan level C1–C6 dengan dominasi HOTS. Instrumen ini diharapkan mampu menjadi alat ukur terstandarisasi yang objektif dalam mengevaluasi capaian pembelajaran mahasiswa secara komprehensif.

METODE

Penelitian ini merupakan penelitian pengembangan (*developmental research*) yang secara spesifik difokuskan pada pengembangan instrumen asesmen. Prosedur pengembangan mengacu pada langkah-langkah penyusunan tes baku menurut Mardapi (2017). Mengingat fokus penelitian ini adalah untuk menghasilkan draf awal instrumen kognitif sistem *Air Conditioning* (A/C) kendaraan ringan yang teruji secara teoretis dan empiris awal, tahapan pengembangan dibatasi pada empat langkah krusial, yaitu: (1) penyusunan spesifikasi tes (kisi-kisi), (2) penulisan butir soal, (3) telaah soal secara kualitatif (*expert judgment*), dan (4) uji coba empiris skala terbatas beserta analisis psikometrik butir soal.



Gambar 1. Alur Proses Pengembangan Tes

Penelitian dilaksanakan di Program Studi Pendidikan Teknik Otomotif Universitas Pendidikan Indonesia. Subjek penelitian pada tahap validasi ahli terdiri atas tiga orang dosen pakar di bidang Evaluasi pendidikan, ahli sistem AC, dan ahli pendidikan otomotif yang dipilih secara *purposive*. Sementara itu, subjek untuk tahap uji coba lapangan terbatas adalah 35 mahasiswa semester 6 sampai semester 8 yang telah menempuh dan lulus mata kuliah Praktik Sistem A/C. Pemilihan sampel ini menggunakan teknik total sampling, di mana seluruh anggota populasi yang memenuhi kriteria inklusi dijadikan responden guna memperoleh data awal kelayakan instrumen sebelum digunakan secara luas.

Instrumen pengumpulan data terdiri atas dua jenis. Pertama, lembar telaah validasi ahli berskala Guttman (skor 1 untuk "Ya/Relevan" dan 0 untuk "Tidak/Tidak Relevan") disertai kolom catatan perbaikan kualitatif. Kedua, perangkat tes kognitif berupa 50 butir soal pilihan ganda yang diberikan kepada responden uji coba untuk menghasilkan data dikotomi.

Tabel 1 Skala Guttman

Skor	Pilihan	Keterangan
1	Ya	Relevan
0	Tidak	Tidak Relevan

Analisis data dilakukan secara deskriptif kuantitatif. Hasil telaah pakar dianalisis menggunakan *Content Validity Ratio* (CVR) dari Lawshe (1975). Dengan jumlah 3 validator panelis (N=3), nilai kritis minimum CVR agar sebuah butir dinyatakan valid secara teoretis adalah 0,99. Selanjutnya, data dari uji coba lapangan dianalisis menggunakan *Software Excel* untuk menentukan kualitas psikometrik butir soal yang meliputi: validitas empiris butir soal menggunakan korelasi *Point Biserial*, reliabilitas tes menggunakan estimasi *Kuder-Richardson 20* (KR-20), tingkat kesukaran yang dihitung dari proporsi jawaban benar, serta daya pembeda butir soal (Arikunto, 2019).

HASIL DAN PEMBAHASAN

Hasil

Pengembangan instrumen asesmen kognitif pada materi Sistem *Air Conditioning* (A/C) Kendaraan Ringan menghasilkan produk awal berupa seperangkat tes objektif pilihan ganda sebanyak 50 butir soal. Instrumen ini dikembangkan berdasarkan kisi-kisi yang mengintegrasikan level kognitif dari *Lower-Order Thinking Skills* (LOTS, C1-C2) hingga *Higher-Order Thinking Skills* (HOTS, C4-C6) dengan menelaah Capaian Pembelajaran Mata Kuliah (CPMK), Indikator Tes, Teori *Air Conditioning*.

Sebelum diujicobakan secara empiris, draf instrumen divalidasi oleh tiga orang ahli. Berdasarkan penilaian kuantitatif menggunakan angket berskala Guttman, seluruh 50 butir soal mendapatkan persetujuan penuh dari ketiga validator ($N_e = 3$). Hasil perhitungan *Content Validity Ratio* (CVR) menunjukkan nilai 1,00 untuk seluruh butir instrumen. Karena nilai tersebut melampaui batas kritis minimum Lawshe (0,99 untuk 3 panelis), seluruh butir soal dinyatakan valid secara isi (*content validity*).

Validasi isi tidak hanya berhenti pada perhitungan matematis, melainkan juga melibatkan perbaikan instrumen berdasarkan evaluasi kualitatif para ahli. Validator menelaah kesesuaian redaksi, kejelasan gambar, serta ketepatan konsep otomotif. Rangkuman masukan kualitatif dan tindak lanjut perbaikan disajikan pada Tabel 2.

Tabel 2 Masukan Kualitatif Validator dan Tindak Lanjut Revisi

Validator	Saran dan Masukan Perbaikan	Tindak Lanjut (Revisi Instrumen)
Validator 1	Redaksi pada soal studi kasus (HOTS) perlu diperjelas agar tidak menimbulkan penafsiran ganda.	Peneliti merevisi struktur kalimat pada soal studi kasus (C4-C6) menjadi lebih lugas dan spesifik pada satu gejala kerusakan.
Validator 2	Kualitas gambar komponen/skema kelistrikan perlu ditingkatkan resolusinya.	Peneliti mengganti gambar ilustrasi sistem A/C dengan resolusi tinggi dan menambahkan label komponen yang lebih jelas.
Validator 3	Pilihan jawaban jangan terlalu berlawanan dengan prinsip dasar sistem A/C; kalimat disesuaikan dengan ejaan yang benar.	Peneliti memperbaiki <i>distractor</i> (pengecoh) agar lebih logis secara teknis dan memperbaiki eror pengetikan pada instrumen.

Instrumen yang telah direvisi kemudian diujicobakan secara terbatas kepada 35 mahasiswa. Uji coba skala kecil ini bertujuan untuk mendapatkan gambaran awal mengenai kualitas psikometrik butir soal. Pengujian validitas empiris dilakukan menggunakan korelasi Point Biserial r_{pbi} dengan membandingkan r_{hitung} terhadap r_{tabel} (0,334 pada taraf signifikansi 5% dengan $df = 33$). Hasil analisis validitas butir dirangkum pada Tabel 3.

Tabel 3 Rekapitulasi Validitas Empiris Butir Soal

Kriteria	Jumlah Soal	Persentase	Nomor Butir Soal
Valid $r_{hitung} > r_{tabel}$	44	88%	1-8, 10-12, 14-29, 31-32, 34-45, 47-48, 50
Tidak Valid (Gugur)	6	12%	9, 13, 30, 33, 46, 49
Total	50	100%	

Sebanyak 44 butir soal dinyatakan valid secara statistik, sedangkan 6 butir soal dinyatakan gugur karena tidak mampu menunjukkan korelasi yang signifikan dengan skor total. Selanjutnya, perhitungan reliabilitas instrumen secara keseluruhan menggunakan rumus Kuder-Richardson 20 (KR-20) menghasilkan koefisien r_{11} sebesar 0,872. Mengacu pada kriteria penafsiran reliabilitas, nilai tersebut berada pada rentang 0,80 – 1,00 yang mengindikasikan bahwa instrumen memiliki tingkat konsistensi yang Sangat Tinggi.

Analisis tingkat kesukaran dilakukan untuk mengetahui proporsi responden yang dapat menjawab benar pada tiap butir soal. Distribusi tingkat kesukaran dari 50 butir soal uji coba disajikan pada Tabel 4.

Tabel 4 Rekapitulasi Tingkat Kesukaran Butir Soal

Kategori	Rentang Indeks (p)	Jumlah Soal	Persentase
Mudah	$p > 0,70$	28	56%

Sedang	$0,30 < p \leq 0,70$	20	40%
Sukar	$p \leq 0,30$	2	4%
Total		50	100%

Secara empiris, instrumen ini masih didominasi oleh butir soal dengan kategori mudah (56%). Temuan statistik pada uji coba terbatas ini menjadi dasar bagi proses analisis lanjutan pada tahap pembahasan.

Analisis daya pembeda dilakukan untuk mengukur sejauh mana butir soal mampu membedakan antara mahasiswa kelompok atas (kemampuan tinggi) dan kelompok bawah (kemampuan rendah). Rekapitulasi hasil analisis daya pembeda disajikan pada Tabel 5.

Tabel 5 Rekapitulasi Daya Pembeda Butir Soal

Kategori	Rentang Indeks (D)	Jumlah Soal	Persentase
Sangat Baik	$D \geq 0,40$	10	20%
Baik	$0,30 \leq D < 0,40$	20	40%
Cukup	$0,20 \leq D < 0,30$	14	28%
Jelek	$D < 0,20$	6	12%
Total		50	100%

Berdasarkan Tabel 5, sebanyak 30 butir soal (60%) memiliki daya pembeda pada rentang baik hingga sangat baik, yang berarti butir-butir tersebut sangat efektif dalam membedakan tingkat pemahaman mahasiswa. Sebanyak 14 butir (28%) berada pada kategori cukup dan masih dapat dipertahankan. Sementara itu, 6 butir soal (12%) yang berada pada kategori jelek merupakan butir soal yang sama dan linier dengan hasil uji validitas empiris yang dinyatakan gugur (butir nomor 9, 13, 30, 33, 46, 49).

Selain validitas, reliabilitas, tingkat kesukaran, dan daya pembeda, analisis efektivitas pengecoh (*distractor*) juga dilakukan untuk mengetahui apakah pilihan jawaban salah telah berfungsi dengan semestinya. Sebuah pengecoh dianggap berfungsi secara efektif apabila dipilih oleh minimal 5% dari total peserta tes, terutama oleh responden dari kelompok bawah. Mengingat instrumen ini memiliki 5 pilihan jawaban (1 kunci jawaban dan 4 pengecoh) pada setiap butir soal, maka secara keseluruhan terdapat 200 pilihan jawaban pengecoh yang dianalisis dari 50 butir instrumen. Rekapitulasi efektivitas pengecoh disajikan pada Tabel 6.

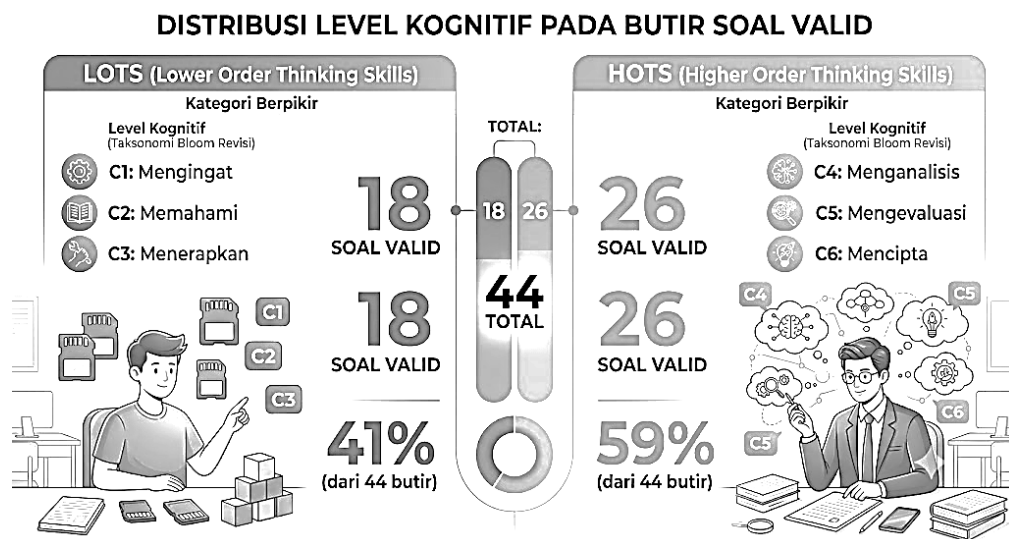
Tabel 6 Rekapitulasi Efektivitas Pengecoh (*Distractor*)

Kategori Fungsi	Kriteria Pemilih	Jumlah Pengecoh	Persentase
Berfungsi Sangat Baik	$> 5\%$	162	81%
Tidak Berfungsi (Revisi/Ganti)	$\leq 5\%$ atau 0%	38	19%
Total Pengecoh		200	100%

Berdasarkan Tabel 6, mayoritas pengecoh (81%) telah berfungsi dengan baik dan mampu mengecoh responden yang tidak memahami materi. Namun, terdapat 38 pengecoh

(19%) yang dinyatakan tidak berfungsi karena dipilih oleh kurang dari 5% responden, atau bahkan tidak dipilih sama sekali. Sebaran pengecoh yang tidak berfungsi ini ditemukan mendominasi pada 6 butir soal yang dinyatakan gugur secara empiris (butir nomor 9, 13, 30, 33, 46, dan 49), sehingga mempermudah mahasiswa dalam menebak kunci jawaban secara acak.

Untuk mempertegas karakteristik instrumen sebagai alat ukur berbasis *Higher-Order Thinking Skills* (HOTS), pemetaan distribusi level kognitif dilakukan secara spesifik pada 44 butir soal yang telah dinyatakan valid secara empiris. Berdasarkan rancangan awal (kisi-kisi), instrumen menetapkan proporsi 40% (20 butir) untuk LOTS dan 60% (30 butir) untuk HOTS. Setelah proses uji coba empiris terbatas dan penyaringan 6 butir soal yang gugur, sebaran level kognitif pada butir soal yang layak dipertahankan disajikan pada Gambar 2.



Gambar 2 Distribusi Level Kognitif pada Butir Soal Valid

Berdasarkan Gambar 2, instrumen yang telah teruji validitas empirisnya secara jelas didominasi oleh butir soal berkategori HOTS sebanyak 26 butir (59%). Data ini memberikan bukti empiris bahwa instrumen asesmen yang dikembangkan telah memenuhi tujuannya untuk bergeser dari sekadar evaluasi hafalan teoretis menuju pengukuran kemampuan pemecahan masalah (diagnosis kerusakan) tingkat tinggi, sesuai dengan tuntutan kompetensi industri.

Tabel 7 kisi – kisi

No.	Capaian Pembelajaran	Indikator Capaian Pembelajaran	Level Kognitif	Bentuk Soal	No. Butir
1.	Memahami prinsip dasar dan komponen sistem AC	Identifikasi Komponen Utama	C1	PG	1,2,4,5,10
2.		Identifikasi Komponen Pendukung	C1	PG	3,14,15
3.		Fungsi Sistem AC	C2	PG	6-7
4.		Siklus Refrigerasi	C2	PG	8.9.13.18.19.20
5.		Prinsip Kerja	C2	PG	11,12,16,17

No.	Capaian Pembelajaran	Indikator Capaian Pembelajaran	Level Kognitif	Bentuk Soal	No. Butir
6.	Menerapkan prosedur perawatan dan diagnosis	Pemeriksaan Visual (SOP) Sistem AC	C3	PG	21,22,23,24,25
7.	Menerapkan prosedur perawatan dan diagnosis sistem AC	Penggunaan Alat Ukur	C3	PG	26,27,28,30
8.		Pengisian Refrigeran	C3	PG	29,31,32
9.	Menganalisis gangguan sistem AC	Analisis Gejala (Troubleshooting)	C4	PG	33,34,35,36
10.		Analisis Tekanan Manifold	C4	PG	45,47,48
11.		Analisis Kelistrikan	C4	PG	37,38,39,40
12.	Mengevaluasi dan Merancang prosedur	Evaluasi Kerusakan	C5	PG	41,43,44,46
13.		Merancang Solusi	C6	PG	42,49,50

Pembahasan

Tahap telaah soal secara kualitatif (*expert judgment*) menghasilkan nilai CVR 1,00 dari ketiga pakar. Secara teoretis, pencapaian ini mengonfirmasi bahwa spesifikasi tes dan butir soal yang ditulis telah selaras dengan tuntutan kompetensi diagnosis industri (Husein, 2019). Hasil ini sejalan dengan penelitian pengembangan instrumen otomotif sebelumnya oleh Suratno (2016) dan Rohman & Setiadi (2022), yang menegaskan bahwa validitas pakar sangat bergantung pada kejelasan rumusan indikator kinerja teknis. yang menegaskan bahwa validitas pakar sangat bergantung pada kejelasan rumusan indikator kinerja teknis. Temuan ini juga didukung oleh Nugroho, Wagiran, dan Mustaghfirin (2021) yang menyimpulkan bahwa instrumen asesmen berbasis HOTS di bidang otomotif membutuhkan telaah konstruk yang mendalam agar tidak bias makna saat diujikan kepada mahasiswa. Lebih dari sekadar angka matematis, proses revisi kualitatif pada instrumen ini terutama perbaikan redaksional pada butir *Higher-Order Thinking Skills* (HOTS) terbukti krusial untuk mencegah bias penafsiran ganda saat mahasiswa melakukan diagnosis kerusakan sistem *Air Conditioning* (A/C).

Pada tahap uji coba empiris terbatas, ditemukan bahwa 44 butir (88%) dinyatakan valid, sementara 6 butir (12%) dinyatakan gugur. Berdasarkan pemetaan level kognitif (Anderson & Krathwohl, 2001), keenam butir yang tidak valid tersebut (nomor 9, 13, 30, 33, 46, 49) ternyata tersebar lintas tingkatan, mulai dari level pemahaman (C2), penerapan (C3), hingga tingkat analisis dan perancangan diagnosis (C4-C6). Gugurnya keenam butir soal ini berbanding lurus dengan hasil analisis daya pembeda yang jelek ($D < 0,20$) serta didukung oleh temuan 38 pengecoh (distractor) yang tidak berfungsi (Tabel 6). Lemahnya pengecoh ini, khususnya pada soal level kognitif tinggi, menyebabkan butir soal tidak mampu menjebak mahasiswa yang kurang kompeten (Retnawati, 2016). Pada evaluasi teknis kejuruan, pengecoh seharusnya merepresentasikan kesalahan diagnosis yang lazim terjadi di bengkel, bukan sekadar jawaban yang salah secara teoretis, sehingga meminimalisasi peluang mahasiswa menebak secara acak.

Analisis tingkat kesukaran menunjukkan dominasi kategori mudah sebesar 56% (28 butir).

Hasil ini perlu disikapi secara kritis. Dominasi tersebut sebagian memang dapat dijelaskan oleh proporsi soal *Lower-Order Thinking Skills* (LOTS) pada rancangan kisi-kisi yang mencapai 40%. Namun, secara empiris, jumlah 28 butir mudah mengisyaratkan bahwa sebagian soal yang dirancang sebagai HOTS (C4-C5) pada praktiknya belum cukup menantang. Kondisi ini relevan dengan temuan Nurtanto, Fawaid, dan Sofyan (2020) serta Maulidia & Pahlevi (2020), di mana perancangan evaluasi vokasi seringkali kesulitan menghadirkan masalah (*problem-solving*) yang benar-benar autentik dan kompleks. Tingginya persentase butir mudah ini juga sangat mungkin dipengaruhi oleh faktor retensi ingatan subjek uji coba yang baru saja lulus dan menyelesaikan mata kuliah Praktik Sistem A/C.

Koefisien reliabilitas instrumen mencapai angka yang sangat tinggi, yaitu 0,872 (Arikunto, 2019). Keandalan suatu tes instrumen sangat dipengaruhi oleh konsistensi butir soal dalam mengukur satu dimensi kompetensi yang spesifik (Azwar, 2015). Tingkat keajegan ini mengindikasikan potensi yang kuat bagi instrumen ini untuk digunakan sebagai alat ukur yang konsisten (Mardapi, 2017). Walaupun demikian, pemaknaan terhadap temuan ini harus mengakui adanya keterbatasan metodologis. Uji coba masih berskala terbatas (*preliminary field testing*) dengan melibatkan 35 responden yang relatif homogen. Parameter psikometrik yang dihasilkan belum cukup stabil untuk merepresentasikan populasi yang lebih luas. Oleh karena itu, pengujian tahap lanjut dengan skala luas sangat diperlukan untuk memantapkan efektivitas daya pembeda butir soal dan fungsi pengecoh secara lebih komprehensif (Fitriyani et al., 2021).

Instrumen yang telah tervalidasi secara isi dan memiliki indikasi empiris yang baik ini tidak dapat langsung diklaim sempurna. Meskipun analisis awal menunjukkan mayoritas butir soal memiliki daya pembeda yang memadai, diperlukan uji coba skala luas (*main field testing*) dengan melibatkan karakteristik peserta didik vokasi yang lebih heterogen untuk menghasilkan analisis psikometrik khususnya pengujian stabilitas butir dan efektivitas pengecoh yang lebih komprehensif.

Keberhasilan instrumen ini dalam mempertahankan dominasi butir soal HOTS pasca uji coba empiris memberikan implikasi penting terhadap praktik evaluasi di pendidikan vokasi. Karakteristik utama yang membedakan instrumen ini dari tes konvensional terletak pada pergeseran desain stimulus soal, dari sekadar *recall knowledge* menuju analisis diagnostik berbasis kasus riil (*case-based testing*). Sebagai ilustrasi, instrumen konvensional yang didominasi level LOTS umumnya sebatas menguji ingatan teoretis, misalnya dengan menanyakan: "Apa fungsi kompresor pada sistem A/C?" atau "Apa jenis refrigeran yang standar digunakan?". Pertanyaan semacam ini terbukti tidak cukup untuk mengukur kesiapan kerja teknisi (Nurtanto dkk., 2020).

Sebaliknya, butir soal HOTS (C4-C6) pada instrumen yang dikembangkan ini mensimulasikan pemecahan masalah autentik. Mahasiswa disajikan stimulus berupa data gejala malafungsi, seperti pembacaan spesifik pada *manifold gauge* (misalnya: tekanan sisi *Low* terlalu tinggi dan sisi *High* terlalu rendah) yang disertai deskripsi keluhan bahwa udara kabin tidak dingin. Untuk menjawab dengan benar, mahasiswa tidak bisa sekadar mengandalkan hafalan, melainkan harus menganalisis korelasi termodinamika antar-komponen, mengevaluasi probabilitas kerusakan, dan memutuskan diagnosis teknis secara presisi (contoh: menyimpulkan adanya kebocoran internal pada katup kompresor). Konstruksi instrumen diagnostik semacam ini secara langsung menjembatani kesenjangan antara tes akademis di kampus dengan realitas pemecahan masalah (*problem-solving*) yang kompleks dan dinamis di industri otomotif yang sesungguhnya.

KESIMPULAN

Berdasarkan tahapan pengembangan instrumen tes dan hasil analisis data, dapat disimpulkan bahwa draf instrumen asesmen kognitif pada materi Sistem *Air Conditioning* (A/C) Kendaraan Ringan telah berhasil disusun dan menunjukkan potensi kelayakan awal yang baik. Melalui tahapan telaah kualitatif, draf instrumen berupa 50 butir soal pilihan ganda ini tervalidasi secara mutlak oleh pakar (CVR = 1,00), yang mengonfirmasi relevansi konstruksi tes dalam mengukur kemampuan pemecahan masalah diagnostik (*Higher-Order Thinking Skills*). Secara empiris, hasil uji coba lapangan terbatas memberikan indikasi awal yang positif dengan tingkat konsistensi sangat tinggi (KR-20 = 0,872), di mana 44 butir soal dinyatakan valid secara statistik dan 60% butir soal memiliki daya pembeda yang memadai.

Meskipun instrumen ini menunjukkan indikasi parameter psikometrik yang positif, simpulan efektivitasnya secara ketat dibatasi pada ruang lingkup uji coba awal (*preliminary testing*) terhadap 35 responden. Mengingat ukuran dan homogenitas sampel tersebut, instrumen ini belum dapat dinyatakan siap untuk diimplementasikan secara luas. Karakteristik psikometrik tingkat lanjut, khususnya stabilitas validitas butir pada populasi yang beragam serta efisiensi fungsi pengecoh (*distractor*), masih memerlukan pembuktian empiris. Sebagai rekomendasi, penelitian ini mewajibkan dilakukannya pengujian skala luas (*main field testing*) dengan sampel mahasiswa yang lebih representatif dan heterogen sebagai prasyarat sebelum instrumen ini distandardisasi secara utuh dalam evaluasi pendidikan vokasi otomotif. Sebagai prospek pengembangan ke depan, instrumen ini merepresentasikan langkah maju dalam menggeser paradigma evaluasi vokasi dari sekadar hafalan (LOTS) menuju evaluasi pemecahan masalah (HOTS). Namun, untuk menghasilkan instrumen yang benar-benar mapan dan siap diimplementasikan secara luas, direkomendasikan agar penelitian selanjutnya melakukan uji coba skala luas (*main field testing*) dengan sampel mahasiswa yang lebih besar

dan heterogen.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Program Studi Pendidikan Teknik Otomotif Universitas Pendidikan Indonesia atas dukungan fasilitas selama pelaksanaan penelitian. Apresiasi juga disampaikan kepada para validator ahli atas masukan dalam proses pengembangan instrumen, serta kepada seluruh mahasiswa yang telah berpartisipasi sebagai responden pada uji coba instrumen.

DAFTAR PUSTAKA

- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of educational objectives*. New York, NY: Longman.
- Arikunto, S. (2019). *Prosedur Penelitian: Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta.
- Azwar, S. (2015). *Reliabilitas dan Validitas (Edisi 4)*. Yogyakarta: Pustaka Pelajar.
- Daryanto. (2013). *Teknik Servis Mobil*. Bandung: Alfabeta.
- Erjavec, J. (2020). *Automotive Technology: A Systems Approach (7th ed.)*. Boston, MA: Cengage Learning.
- Fauzi, A., Pardjono, P., & Soenarto, S. (2020). The Development of Performance Assessment Instruments for Engine Overhaul Practice in Vocational High Schools. *Journal of Education and Learning (EduLearn)*, 14(1), 123-130, from <https://edulearn.intelektual.org/index.php/edulearn/article/view/14488>.
- Firmansyah, H., & Mulyani, F. (2022). Pengembangan Instrumen Penilaian Diagnostik Kognitif pada Materi Sistem Starter dan Pengisian. *Jurnal Inovasi Pendidikan Teknik dan Kejuruan*, 2(1), 45-53.
- Fitriyani, Y., Nurtanto, M., & Ramdani, S. D. (2021). Pendekatan Evaluasi Pendidikan Vokasi: Sebuah Tinjauan Literatur Sistematis. *Jurnal Pendidikan Vokasi Otomotif*, 4(1), 21–30.
- Hadiyanto, R. (2023). Pengembangan Instrumen Asesmen Uji Kompetensi Mahasiswa SMK untuk Jabatan Teknisi di Dunia Kerja Bidang Otomotif. *Jurnal Inovasi Pembelajaran Karakter*, 8(3), 38-47.
- Husein, M. T. (2019). Link and Match Pendidikan Sekolah Kejuruan. *Rausyan Fikr: Jurnal Pemikiran dan Pencerahan*, 15(2), 39–47, from <https://jurnal.umt.ac.id/index.php/RausyanFikr/article/view/1801>[cite: 1, 2].

- Isnanto, R. R., & Kustono, D. (2019). Pengembangan Media Pembelajaran Trainer Sistem A/C Mobil untuk Meningkatkan Hasil Belajar Mahasiswa. *Jurnal Pendidikan: Teori, Penelitian, dan Pengembangan*, 4(8), 1059-1065.
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563–575, from <https://doi.org/10.1111/j.1744-6570.1975.tb01393.x>[cite: 1, 2, 3].
- Mardapi, D. (2017). *Pengukuran, Penilaian, dan Evaluasi Pendidikan: Edisi Kedua*. Yogyakarta: Parama Publishing.
- Maulidia, F., & Pahlevi, T. (2020). Pengembangan Instrumen Penilaian Tes Soal Pilihan Ganda Berbasis HOTS Pada Mata Pelajaran Administrasi Umum Jurusan OTKP SMK Negeri 1 Lamongan. *Jurnal Pendidikan Administrasi Perkantoran (JPAP)*, 8(1), 136-145, from <https://ejournal.unesa.ac.id/index.php/jpap/article/view/33425>[cite: 1, 2].
- Nugroho, A., Wagiran, W., & Mustaghfirin, M. (2021). Pengembangan Instrumen Penilaian Berbasis HOTS pada Kompetensi Sistem Pengapian Elektronik. *Jurnal Pendidikan Vokasional Teknik Otomotif*, 4(2), 78-86.
- Nurtanto, M., Fawaid, M., & Sofyan, H. (2020). Problem-Based Learning (PBL) in Automotive Technology Vocational Education: The Impact on Higher-Order Thinking Skills (HOTS). *Journal of Education and Learning (EduLearn)*, 14(3), 428-436, from <https://edulearn.intelektual.org/index.php/edulearn/article/view/15787>[cite: 1, 2, 3].
- Retnawati, H. (2016). *Validitas, Reliabilitas, dan Karakteristik Butir*. Yogyakarta: Parama Publishing.
- Rohman, A. F., & Setiadi, A. (2022). Pengembangan Instrumen Penilaian Kinerja pada Praktik Sistem Air Conditioner (A/C). *Jurnal Pendidikan Vokasional Teknik Otomotif*, 5(1), 12-21, from <https://journal.uny.ac.id/index.php/jpvto/article/view/48771>.
- Sugiyono. (2019). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Suratno. (2016). Pengembangan Instrumen Penilaian Kompetensi Praktikum Engine Mahasiswa SMK Program Keahlian Teknik Otomotif. *VANOS Journal of Mechanical Engineering Education*, 1(1), 69-78, from <https://jurnal.untirta.ac.id/index.php/vanos/article/view/826>.
- Tomczyk, J. (2017). *Refrigeration and Air Conditioning Technology (8th ed.)*. Boston, MA: Cengage Learning.