

Analisis Kualitas Butir Soal Esp (English For Specific Purposes) Berbasis Taksonomi Bloom Pada Program Keahlian Teknik Otomotif Smk Menggunakan Program Anates

Purwaty Mahareni*¹, Rufi'i Rufi'i²

^{1,2}Program Studi Magister Teknologi Pendidikan, Universitas PGRI Adibuan Surabaya, Indonesia
Email: ¹250020045@student.unipasby.ac.id, ²rufii@unipasby.ac.id

Abstrak

Penelitian ini bertujuan menganalisis kualitas butir soal ESP (English for Specific Purposes) berbasis Taksonomi Bloom pada Program Keahlian Teknik Otomotif SMK menggunakan program ANATES Versi 4.0. Pendekatan yang digunakan adalah kuantitatif deskriptif-evaluatif dengan subjek 30 peserta didik kelas XI. Instrumen berupa 25 butir soal pilihan ganda yang mencakup level kognitif C1–C4. Hasil analisis menunjukkan: (1) reliabilitas tes sebesar 0,75 (kategori tinggi); (2) distribusi tingkat kesukaran mendekati ideal (24% mudah, 52% sedang, 24% sukar); (3) 80% butir memiliki daya pembeda memadai; (4) 36% butir memiliki korelasi skor signifikan; dan (5) kualitas pengecoh bervariasi. Secara keseluruhan instrumen layak digunakan dengan perbaikan pada butir-butir yang teridentifikasi bermasalah.

Kata kunci: analisis butir soal, ANATES, ESP, SMK Teknik Otomotif, Taksonomi Bloom.

Analysis of ESP (English for Specific Purposes) Test Item Quality Based on Bloom's Taxonomy in the Automotive Engineering Vocational High School Program Using the ANATES Program

Abstract

This study aims to analyze the quality of ESP (English for Specific Purposes) test items based on Bloom's Taxonomy in the Automotive Engineering Program of Indonesian Vocational Schools (SMK) using the ANATES Version 4.0 program. A descriptive-evaluative quantitative approach was employed with 30 Grade XI students as subjects. The instrument consisted of 25 multiple-choice items covering cognitive levels C1–C4. The analysis results indicate: (1) test reliability of 0.75 (high category); (2) near-ideal difficulty distribution (24% easy, 52% medium, 24% difficult); (3) 80% of items have adequate discrimination; (4) 36% of items show significant item-total correlations; and (5) distractor quality varies. Overall, the instrument is suitable for use with revisions on the identified problematic items.

Keywords: item analysis, ANATES, ESP, Automotive Engineering SMK, Bloom's Taxonomy.

1. PENDAHULUAN

Evaluasi pembelajaran merupakan komponen esensial dalam sistem pendidikan yang berfungsi sebagai alat ukur pencapaian kompetensi peserta didik. Kualitas instrumen evaluasi yang digunakan secara langsung mempengaruhi validitas simpulan yang diambil dari proses penilaian tersebut [1], [2]. Di lingkungan Sekolah Menengah Kejuruan (SMK), evaluasi pembelajaran tidak hanya mengukur penguasaan konsep akademis, melainkan juga kompetensi vokasional yang relevan dengan dunia kerja.

English for Specific Purposes (ESP) merupakan pendekatan pembelajaran bahasa Inggris yang dirancang untuk memenuhi kebutuhan komunikasi spesifik dalam bidang profesi tertentu [3], [4]. Dalam konteks SMK Program Keahlian Teknik Otomotif, penguasaan kosakata teknis berbahasa Inggris menjadi kompetensi yang semakin krusial mengingat mayoritas manual servis, diagram teknik, dan dokumentasi kendaraan menggunakan bahasa Inggris. Kemendikbudristek [5] melalui Capaian Pembelajaran Fase F secara eksplisit mengintegrasikan ESP ke dalam pembelajaran Bahasa Inggris di SMK.

Meskipun demikian, pengembangan dan evaluasi instrumen soal ESP di SMK masih menghadapi berbagai tantangan. Penelitian Prasetyo & Sari [6] menemukan bahwa sebagian besar soal ujian Bahasa Inggris di SMK belum dianalisis secara sistematis dari sisi kualitas psikometrik. Hal ini menyebabkan instrumen yang digunakan tidak tervalidasi secara empiris dan berpotensi menghasilkan penilaian yang bias. Lebih lanjut, integrasi

Taksonomi Bloom sebagai kerangka pengembangan soal seringkali dilakukan secara normatif tanpa verifikasi empiris terhadap distribusi level kognitif yang sesungguhnya diukur oleh instrumen tersebut.

Analisis butir soal merupakan prosedur psikometrik yang memungkinkan evaluator untuk menilai kualitas setiap butir soal berdasarkan data respons nyata peserta didik. Program ANATES (Analisis Tes) yang dikembangkan oleh Karno To dan Yudi Wibisono merupakan perangkat lunak analisis butir soal yang telah banyak digunakan dalam konteks pendidikan Indonesia dan terbukti efektif untuk menganalisis reliabilitas tes, tingkat kesukaran, daya pembeda, dan kualitas pengecoh pada soal pilihan ganda [7], [8].

Penelitian ini bertujuan untuk: (1) mendeskripsikan reliabilitas instrumen soal ESP berbasis Taksonomi Bloom yang dikembangkan; (2) menganalisis distribusi tingkat kesukaran butir soal; (3) mengevaluasi daya pembeda setiap butir soal; (4) mengkaji kualitas pengecoh dan korelasi skor butir dengan skor total; serta (5) memberikan rekomendasi perbaikan butir soal yang belum memenuhi standar kualitas psikometrik.

2. KAJIAN TEORI

2.1. English for Specific Purposes (ESP) dalam Konteks SMK

ESP didefinisikan sebagai pembelajaran bahasa Inggris yang tujuan, isi, dan metodenya ditentukan oleh kebutuhan spesifik peserta didik dalam bidang akademis atau profesi tertentu [3]. Dalam konteks SMK Teknik Otomotif, ESP mencakup kosakata dan struktur bahasa yang relevan dengan sistem mesin, kelistrikan, sistem rem, transmisi, dan berbagai komponen teknis lainnya.

Rahardjo & Santoso[9] menekankan bahwa penilaian ESP harus mengukur kemampuan peserta didik untuk menggunakan bahasa Inggris secara fungsional dalam konteks kejuruan, bukan sekadar pemahaman linguistik umum. Dengan demikian, instrumen penilaian ESP perlu dirancang dengan mempertimbangkan relevansi kontekstual dan level kognitif yang sesuai dengan tuntutan pekerjaan.

2.2. Taksonomi Bloom sebagai Kerangka Pengembangan Soal

Taksonomi Bloom yang direvisi oleh Anderson & Krathwohl[10] sebagaimana dikembangkan oleh Saputra et al.[11] mengklasifikasikan kemampuan kognitif ke dalam enam level: mengingat (C1), memahami (C2), menerapkan (C3), menganalisis (C4), mengevaluasi (C5), dan mencipta (C6). Untuk tingkat SMK, distribusi soal yang umum direkomendasikan mencakup C1–C4 dengan proporsi yang seimbang antara kemampuan tingkat rendah (LOTS) dan tingkat tinggi (HOTS).

Brookhart & Nitko[12] menegaskan bahwa instrumen penilaian yang berkualitas harus memiliki keselarasan antara tujuan pembelajaran, level kognitif yang ditargetkan, dan konstruksi soal. Peta kognitif (cognitive map) melalui kisi-kisi soal menjadi alat penting untuk memastikan keselarasan tersebut terwujud.

2.3. Analisis Butir Soal: Konsep dan Parameter

Analisis butir soal adalah proses pengumpulan, peringkasan, dan penggunaan informasi dari respons peserta didik untuk membuat keputusan tentang setiap butir soal [12]. Menurut Retnawati[8], parameter utama yang dikaji dalam analisis butir soal pilihan ganda meliputi: (a) tingkat kesukaran (difficulty index), (b) daya pembeda (discrimination index), (c) kualitas pengecoh (distractor analysis), dan (d) reliabilitas tes.

Tingkat kesukaran dinyatakan dalam indeks P, dengan kategori: mudah ($P > 0,70$), sedang ($0,30 \leq P \leq 0,70$), dan sukar ($P < 0,30$). Daya pembeda mengacu pada kemampuan butir soal untuk membedakan peserta didik berkemampuan tinggi dari yang rendah, dengan indeks $D \geq 0,40$ dianggap sangat baik [8]. Reliabilitas tes menggunakan formula Spearman-Brown melalui metode belah dua (split-half), dengan nilai $\geq 0,70$ dianggap memenuhi standar keterpercayaan [13].

3. METODE PENELITIAN

3.1. Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan jenis penelitian evaluatif. Creswell & Creswell[14] menjelaskan bahwa penelitian deskriptif-evaluatif bertujuan mengumpulkan dan menganalisis data untuk mendeskripsikan karakteristik suatu fenomena secara sistematis. Dalam konteks ini, fenomena yang dideskripsikan adalah kualitas butir soal ESP berbasis Taksonomi Bloom.

3.2. Subjek dan Instrumen

Subjek penelitian adalah 30 peserta didik kelas XI Program Keahlian Teknik Otomotif SMK pada tahun pelajaran 2024/2025. Instrumen yang dianalisis berupa tes pilihan ganda sebanyak 25 butir soal ESP yang dikembangkan berbasis Taksonomi Bloom level C1 hingga C4. Soal mencakup topik-topik teknis otomotif seperti: engine components, lubrication system, cooling system, electrical system, brake system, exhaust system, fuel system, transmission, steering, suspension, diesel system, safety system, dan engine management.

Distribusi level kognitif soal yang dikembangkan adalah sebagai berikut: C1 (Remembering) sebanyak 6 soal (24%), C2 (Understanding) sebanyak 7 soal (28%), C3 (Applying) sebanyak 7 soal (28%), dan C4 (Analyzing) sebanyak 5 soal (20%). Distribusi ini dirancang untuk menjamin cakupan kemampuan kognitif yang proporsional sesuai karakteristik peserta didik SMK [5].

Tabel 1. Distribusi Kisi-Kisi Soal Berdasarkan Level Kognitif Taksonomi Bloom

Level Kognitif	Materi Pokok	Jumlah Soal	Persentase	Nomor Soal
C1 (Remembering)	Engine, Lubrication, Cooling, Electrical, Brake, Exhaust	6	24%	1, 2, 3, 4, 5, 6
C2 (Understanding)	Fuel, Transmission, Steering, Suspension, Diesel, Safety, Engine Management	7	28%	7–13
C3 (Applying)	Engine, Brake, Electrical, Fuel, Cooling, Transmission, Steering	7	28%	14–20
C4 (Analyzing)	Engine Mgmt, Brake, Electrical, Exhaust, Suspension	5	20%	21–25
Total		25	100%	1–25

Sumber: Dokumen kisi-kisi soal (diolah, 2025)

3.3. Teknik Analisis Data

Data dianalisis menggunakan program ANATES Versi 4.0 yang dikembangkan oleh Karno To dan Yudi Wibisono. Analisis mencakup: (a) reliabilitas tes dengan metode belah dua Spearman-Brown; (b) tingkat kesukaran setiap butir soal; (c) daya pembeda berdasarkan kelompok unggul dan asor (masing-masing 27% dari subjek); dan (d) kualitas pengecoh serta korelasi skor butir dengan skor total [1], [8].

Catatan: Tabel distribusi kisi-kisi soal berdasarkan level kognitif Taksonomi Bloom dapat disisipkan di sini sesuai format tabel JPTI.

4. HASIL DAN PEMBAHASAN

4.1. Deskripsi Data Skor Peserta Didik

Tes ESP yang diujikan kepada 30 peserta didik menghasilkan rata-rata skor sebesar 13,73 dengan simpangan baku 4,27. Skor tertinggi diperoleh oleh dua peserta didik dengan nilai sempurna (25), sementara skor terendah adalah 8. Distribusi skor ini menunjukkan sebaran yang relatif normal dengan nilai rata-rata berada pada kategori sedang jika merujuk pada Kriteria Ketuntasan Minimum (KKM) sebesar 75 dari skor maksimal 100 (atau 18,75 dari 25).

Tabel 2. Distribusi Skor Peserta Didik Kelompok Unggul dan Asor

No	Nama Siswa	Skor	Kelompok	Keterangan
1	Akhmad Ramadhani	25	Unggul	Sempurna
2	Dimas Oktaviano	25	Unggul	Sempurna
3	Adji Imas Rahman	19	Unggul	Tinggi
4	Adit Tiya Firmansyah	17	Unggul	Tinggi
5	Adrian Dani Andika	17	Unggul	Tinggi
...	...	...	...	...
27	Kevin Putra Harianto	8	Asor	Rendah
28	Muhammad Yoga	8	Asor	Rendah

Sumber: Output ANATES V4 (diolah, 2025)

4.2. Reliabilitas Tes

Hasil analisis ANATES menunjukkan koefisien korelasi antara skor ganjil dan genap (KorelasiXY) sebesar 0,59. Setelah dikoreksi menggunakan formula Spearman-Brown, diperoleh koefisien reliabilitas tes sebesar 0,75. Nilai ini berada pada kategori tinggi menurut kriteria Guilford (dalam Sumintono & Widhiarso, 2021) yang menetapkan koefisien $\geq 0,70$ sebagai reliabel.

Reliabilitas sebesar 0,75 mengindikasikan bahwa instrumen soal ESP yang dikembangkan memiliki tingkat konsistensi pengukuran yang memadai. Azwar[13] menyatakan bahwa untuk keperluan penelitian dan evaluasi formatif, reliabilitas $\geq 0,70$ sudah dianggap mencukupi. Temuan ini sejalan dengan penelitian Yuwono[7] yang menemukan bahwa instrumen berbasis ESP pada tingkat SMK umumnya menghasilkan koefisien reliabilitas antara 0,68–0,82 ketika dikembangkan secara sistematis.

4.3. Tingkat Kesukaran Butir Soal

Tabel 3. Distribusi Tingkat Kesukaran Butir Soal

Kategori	Jumlah Butir	Persentase	Nomor Butir
Mudah ($>70\%$)	6	24%	2, 3, 7, 11, 14, 23
Sedang (30–70%)	13	52%	1, 4, 6, 8, 9, 10, 12, 15, 16, 17, 19, 22, 24
Sukar ($<30\%$)	6	24%	5, 13, 18, 20, 21, 25
Total	25	100%	—
Mudah ($>70\%$)	6	24%	2, 3, 7, 11, 14, 23
Sedang (30–70%)	13	52%	1, 4, 6, 8, 9, 10, 12, 15, 16, 17, 19, 22, 24
Sukar ($<30\%$)	6	24%	5, 13, 18, 20, 21, 25
Total	25	100%	—

Sumber: Output ANATES V4 (diolah, 2025)

Distribusi tingkat kesukaran menunjukkan komposisi yang relatif ideal. Sebanyak 6 butir (24%) berkategori mudah, 13 butir (52%) sedang, dan 6 butir (24%) sukar. Retnawati[8] merekomendasikan distribusi ideal untuk tes sumatif adalah 25% mudah, 50% sedang, dan 25% sukar, sehingga hasil penelitian ini mendekati distribusi ideal tersebut.

Butir soal yang berkategori sukar (nomor 5, 13, 18, 20, 21, dan 25) umumnya berasal dari level kognitif C4 (Analyzing) dan level C2–C3 yang memerlukan pemahaman mendalam tentang sistem teknis kendaraan. Misalnya, butir nomor 5 tentang jenis minyak rem (DOT 3/DOT 4) hanya dijawab benar oleh 7 dari 30 siswa (23,33%), mengindikasikan bahwa terminologi teknis ini belum sepenuhnya dikuasai. Sementara itu, butir soal mudah seperti nomor 2 (oil filter) dan 7 (perbedaan MPI-GDI) dijawab benar oleh 24 siswa (80%).

4.4. Daya Pembeda Butir Soal

Tabel 4. Distribusi Daya Pembeda Butir Soal

Kategori DP	Jumlah	Persentase	Nomor Butir
Sangat Baik ($\geq 62,50\%$)	5	20%	5, 12, 13, 18, 20, 25
Baik (37,50–50,00%)	8	32%	1, 2, 10, 11, 14, 15, 16, 19, 21, 22
Cukup (25,00–37,50%)	8	32%	6, 8, 9, 23, 24
Perlu Revisi ($<12,50\%$)	4	16%	3, 4, 7, 17
Total	25	100%	—
Sangat Baik ($\geq 62,50\%$)	5	20%	5, 12, 13, 18, 20, 25
Baik (37,50–50,00%)	8	32%	1, 2, 10, 11, 14, 15, 16, 19, 21, 22
Cukup (25,00–37,50%)	8	32%	6, 8, 9, 23, 24

Sumber: Output ANATES V4 (diolah, 2025)

Analisis daya pembeda dilakukan berdasarkan kelompok unggul (8 siswa dengan skor tertinggi) dan kelompok asor (8 siswa dengan skor terendah). Hasil menunjukkan bahwa 5 butir soal (20%) memiliki daya pembeda sangat baik ($\geq 62,50\%$), yaitu butir nomor 5, 12, 13, 18, 20, dan 25 dengan indeks daya pembeda

62,50–75,00%. Butir-butir ini terbukti efektif membedakan peserta didik yang menguasai materi dari yang belum.

Empat butir soal yang memerlukan revisi (nomor 3, 4, 7, dan 17) menunjukkan daya pembeda rendah, bahkan butir nomor 17 memiliki indeks daya pembeda 0,00% yang berarti butir tersebut tidak mampu membedakan kelompok unggul dari kelompok asor sama sekali. Analisis lebih lanjut pada butir nomor 17 (tentang injektor kotor) menunjukkan bahwa jawaban peserta didik dari kedua kelompok tersebar merata di semua pilihan, mengindikasikan kemungkinan adanya kelemahan pada konstruksi soal atau kunci jawaban yang perlu diverifikasi ulang.

4.5. Korelasi Skor Butir dengan Skor Total

Tabel 5. Distribusi Signifikansi Korelasi Skor Butir dengan Skor Total

Signifikansi	Jumlah	Persentase	Nomor Butir
Sangat Signifikan	5	20%	5, 13, 18, 20, 25
Signifikan	4	16%	12, 19, 21, 22
Tidak Signifikan	16	64%	1,2,3,4,6,7,8,9,10,11,14,15,16,17,23,24
Total	25	100%	—
Sangat Signifikan	5	20%	5, 13, 18, 20, 25
Signifikan	4	16%	12, 19, 21, 22
Tidak Signifikan	16	64%	1,2,3,4,6,7,8,9,10,11,14,15,16,17,23,24
Total	25	100%	—

Sumber: Output ANATES V4 (diolah, 2025)

Korelasi skor butir dengan skor total (point-biserial correlation) menghasilkan 5 butir sangat signifikan (koefisien $\geq 0,449$; $p < 0,01$), yaitu butir nomor 5, 13, 18, 20, dan 25 dengan koefisien tertinggi 0,711 (butir 25) dan 0,693 (butir 13 dan 18). Empat butir signifikan (koefisien $\geq 0,349$; $p < 0,05$) meliputi nomor 12, 19, 21, dan 22.

Sebanyak 16 butir (64%) tidak mencapai signifikansi pada taraf $p < 0,05$ dengan $df = 28$. Hal ini berkorelasi dengan temuan daya pembeda yang rendah pada sebagian butir soal. Brown[16] menjelaskan bahwa butir soal dengan korelasi rendah terhadap skor total cenderung memiliki karakteristik psikometrik yang lemah dan perlu direvisi atau diganti. Namun, rendahnya korelasi pada beberapa butir juga dapat disebabkan oleh ukuran sampel yang terbatas ($N=30$), sehingga diperlukan replikasi dengan sampel yang lebih besar.

4.6. Kualitas Pengecoh

Analisis kualitas pengecoh (distractor analysis) menunjukkan hasil yang bervariasi. Pengecoh yang berfungsi sangat baik (dipilih oleh banyak siswa dari kelompok asor namun sedikit dari kelompok unggul) ditemukan pada butir-butir seperti nomor 5 (pengecoh B: DOT 3→SAE 30 dipilih 8 siswa; C: ATF dipilih 9 siswa) yang menunjukkan distraktor berfungsi efektif menggoda siswa yang belum paham.

Sebaliknya, beberapa pengecoh menunjukkan kinerja sangat buruk (ditandai tanda '---' pada output ANATES), seperti pengecoh D pada butir 7 (tidak ada siswa yang memilih), pengecoh C pada butir 14 (dipilih 5 siswa namun sebagian besar dari kelompok asor), dan pengecoh D pada butir 22 serta 23 yang sama sekali tidak berfungsi. Retnawati[8] merekomendasikan bahwa pengecoh yang tidak dipilih oleh minimal 5% peserta didik perlu diganti karena tidak berfungsi sebagai distractor yang efektif.

4.7. Rekomendasi Perbaikan Butir Soal

Berdasarkan hasil analisis multidimensi di atas, dapat diidentifikasi tiga kategori butir soal: (a) butir berkualitas baik yang dapat dipertahankan (nomor 1, 5, 8–16, 18–22, 25); (b) butir yang perlu direvisi pengecohnya (nomor 2, 3, 6, 23, 24); dan (c) butir yang perlu direvisi menyeluruh termasuk stem dan kunci jawaban (nomor 4, 7, 17). Revisi sebaiknya dilakukan melalui expert judgment dari pakar bahasa Inggris ESP dan mekanik otomotif untuk memastikan validitas isi tetap terjaga [1].

[Tabel 2. Distribusi Skor Peserta Didik Kelompok Unggul dan Asor — disisipkan sesuai format tabel JPTI]

[Tabel 3. Distribusi Tingkat Kesukaran Butir Soal — disisipkan sesuai format tabel JPTI]

[Tabel 4. Distribusi Daya Pembeda Butir Soal — disisipkan sesuai format tabel JPTI]

[Tabel 5. Distribusi Signifikansi Korelasi Skor Butir dengan Skor Total — disisipkan sesuai format tabel JPTI]

5. SIMPULAN

Penelitian ini menyimpulkan bahwa instrumen soal ESP berbasis Taksonomi Bloom untuk Program Keahlian Teknik Otomotif SMK yang dianalisis menggunakan program ANATES memiliki karakteristik psikometrik sebagai berikut: (1) Reliabilitas tes sebesar 0,75 termasuk kategori tinggi dan memenuhi standar instrumen yang dapat dipercaya; (2) Distribusi tingkat kesukaran menunjukkan komposisi yang mendekati ideal dengan 24% mudah, 52% sedang, dan 24% sukar; (3) Sebagian besar butir soal (80%) memiliki daya pembeda yang memadai (kategori cukup hingga sangat baik), namun 4 butir (20%) perlu direvisi; (4) Korelasi skor butir dengan skor total menunjukkan bahwa 9 butir (36%) signifikan atau sangat signifikan, sementara 16 butir (64%) belum mencapai signifikansi pada taraf $p < 0,05$; (5) Kualitas pengecoh bervariasi, dengan beberapa pengecoh perlu diganti karena tidak berfungsi efektif.

Secara keseluruhan, instrumen ini layak digunakan sebagai alat evaluasi dengan tetap melakukan revisi pada butir-butir yang teridentifikasi bermasalah. Penelitian lanjutan dengan ukuran sampel yang lebih besar dan analisis menggunakan Item Response Theory (IRT) disarankan untuk memberikan gambaran psikometrik yang lebih komprehensif.

DAFTAR PUSTAKA

- [1] D. Mardapi, "Pengukuran, Penilaian, dan Evaluasi Pendidikan," Edisi 3. Yogyakarta: Nuha Medika, 2021.
- [2] E. P. Widoyoko, "Evaluasi Program Pembelajaran: Panduan Praktis bagi Pendidik dan Calon Pendidik," Edisi Revisi. Yogyakarta: Pustaka Pelajar, 2021.
- [3] T. Hutchinson and A. Waters, "English for Specific Purposes: A Learning-Centred Approach." Cambridge: Cambridge University Press, 1987.
- [4] B. Paltridge and S. Starfield, Eds., "The Handbook of English for Specific Purposes," 2nd ed. Hoboken, NJ: Wiley-Blackwell, 2023.
- [5] Kemendikbudristek, "Capaian Pembelajaran SMK/MAK Fase F." Jakarta: Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, 2022.
- [6] Z. Prasetyo and R. Sari, "Analisis kualitas butir soal ujian akhir semester Bahasa Inggris di SMK: Kajian psikometrik berbasis ANATES," "Jurnal Pendidikan Vokasi," vol. 12, no. 1, pp. 45–58, 2022, doi: 10.21831/jpv.v12i1.44302.
- [7] G. C. Yuwono, "Penggunaan program ANATES V4 dalam analisis butir soal pilihan ganda," "Jurnal Kependidikan: Penelitian Inovasi Pembelajaran," vol. 5, no. 1, pp. 14–27, 2021, doi: 10.21831/jk.v5i1.35658.
- [8] H. Retnawati, "Analisis Kuantitatif Instrumen Penelitian: Panduan Peneliti, Mahasiswa, dan Psikometrian." Yogyakarta: Parama Publishing, 2021.
- [9] M. Rahardjo and A. Santoso, "Item analysis of English proficiency tests in Indonesian vocational schools: A psychometric study," "Journal of English Teaching," vol. 9, no. 2, pp. 115–130, 2023, doi: 10.33541/jet.v9i2.4117.
- [10] L. W. Anderson and D. R. Krathwohl, Eds., "A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Educational Objectives." New York: Longman, 2001.
- [11] H. Saputra, E. Febriyanti, and S. Ramadhan, "Pengembangan instrumen penilaian HOTS berbasis Taksonomi Bloom revisi pada pembelajaran Bahasa Inggris SMK," "Jurnal Pendidikan Bahasa dan Sastra," vol. 22, no. 1, pp. 89–102, 2022, doi: 10.17509/bs_jpbsp.v22i1.47001.
- [12] S. M. Brookhart and A. J. Nitko, "Educational Assessment of Students," 9th ed. Hoboken, NJ: Pearson, 2022.
- [13] S. Azwar, "Reliabilitas dan Validitas," Edisi 5. Yogyakarta: Pustaka Pelajar, 2022.
- [14] J. W. Creswell and J. D. Creswell, "Research Design: Qualitative, Quantitative, and Mixed Methods Approaches," 6th ed. Thousand Oaks, CA: SAGE Publications, 2023.
- [15] B. Sumintono and W. Widhiarso, "Aplikasi Pemodelan Rasch pada Assessment Pendidikan," Edisi 2. Bandung: Trim Komunikata, 2021.
- [16] J. D. Brown, "Language Assessment: Principles and Classroom Practices," 2nd ed. White Plains, NY: Pearson Education, 2021.