

Analisis Butir Soal Pilihan Berganda Pada Training Teknisi**Level 0 - Apprentice Tahun 2023**Fery Irawan¹, Sutiman²^{1,2}Pendidikan Teknik Otomotif, Universitas Negeri Yogyakarta

E-mail: feryirawan.2020@student.uny.ac.id, sutiman@uny.ac.id

Abstract

Item analysis is a process used to evaluate the quality of each test item to ensure that the questions effectively measure participants' abilities. Based on observations at PT SGMW Motor Indonesia, item analysis for the Level 0 – Apprentice Technician Certification in 2023 had not been implemented, as indicated by the lack of attention to the difficulty level of the test items. This study employed a descriptive method with a quantitative approach, involving certification participants as the research subjects and the Level 0 – Apprentice certification test items as the objects. Data were collected through documentation and analyzed using Microsoft Excel to assess validity, reliability, item difficulty, discrimination index, and distractor effectiveness. The results show that out of 250 items analyzed, 34% (84 items) were rejected, 32% (80 items) required revision, and 34% (86 items) were accepted. These findings highlight the importance of conducting systematic item analysis to ensure the availability of a standardized, valid, and reliable item bank. Implementing such procedures can support more accurate and objective assessment processes in technician certification, ultimately improving the quality of competency measurement.

Keywords: *certification exam; item analysis; Microsoft Excel; technician training*

Abstrak

Analisis butir soal merupakan proses untuk mengevaluasi kualitas setiap item dalam suatu tes agar mampu mengukur kemampuan peserta secara akurat. Berdasarkan observasi di PT SGMW Motor Indonesia, analisis butir soal pada sertifikasi teknisi *Level 0 – Apprentice* 2023 belum dilaksanakan, yang terlihat dari minimnya perhatian terhadap tingkat kesulitan soal. Penelitian ini menggunakan metode deskriptif dengan pendekatan kuantitatif, dengan subjek peserta sertifikasi dan objek berupa soal ujian teknisi *Level 0 – Apprentice*. Data dikumpulkan melalui dokumentasi dan dianalisis menggunakan Microsoft Excel untuk menilai validitas, reliabilitas, tingkat kesulitan, daya pembeda, serta efektivitas pengecoh. Hasil penelitian menunjukkan bahwa dari 250 soal yang dikaji, 34% (84 soal) masuk kategori ditolak, 32% (80 soal) perlu direvisi, dan 34% (86 soal) dapat disimpan sebagai soal berkualitas. Temuan ini menegaskan pentingnya pelaksanaan analisis butir soal secara berkala untuk memastikan tersedianya bank soal yang valid, reliabel, serta sesuai standar evaluasi pembelajaran, sehingga proses sertifikasi mampu menghasilkan pengukuran kompetensi teknisi yang lebih objektif dan tepat sasaran.

Kata kunci: *analisis butir; Microsoft Excel; soal sertifikasi; training teknisi*

PENDAHULUAN

Kemajuan sains dan teknologi menimbulkan tantangan baru yang menuntut pengembangan ilmu berkelanjutan. Ketika batas komunikasi kian menyempit, kompetisi global pun meningkat. Karena itu, peningkatan kapabilitas sumber daya manusia menjadi strategis, dengan pendidikan dan pendidik berperan sentral menjaga efektivitas pembelajaran.

Peran pendidik semakin kompleks seiring kemajuan teknologi sehingga menuntut kemampuan adaptif dalam membentuk generasi berkompetensi tinggi dan siap bersaing global. Pencapaian kompetensi peserta didik sangat ditentukan oleh perancangan pembelajaran yang relevan dengan dunia kerja dan kurikulum berbasis kompetensi, terbukti mampu meningkatkan hasil belajar vokasi (Akhyar, Ihsan, Suharno, Tamrin, & Purwanto, 2024). Selain itu, kompetensi pedagogik, profesional, kepribadian, dan sosial pendidik turut menentukan daya saing lulusan (Wahyudi & Arifin, 2023).

Dalam kompetensi pedagogik, pendidik wajib melakukan evaluasi pembelajaran untuk menilai keberhasilan program, mengidentifikasi kelebihan dan kekurangan proses belajar, serta menyediakan data bagi pengambilan keputusan. Hasil belajar menjadi indikator utama efektivitas pembelajaran (Anas & Budiman, 2020). Evaluasi terstruktur mencakup perencanaan, pelaksanaan, dan penilaian yang meningkatkan mutu pembelajaran (Suyitno, 2022). Ketentuan ini selaras dengan Permendiknas No. 41/2007. Partisipasi aktif dalam pembelajaran berbasis proyek juga meningkatkan berpikir kritis (Indasah & Sutiman, 2023).

Evaluasi melibatkan proses pengukuran dan penilaian; pengukuran sebagai tahap awal mencakup pemberian nilai numerik terhadap kinerja akademis, sedangkan penilaian menghimpun data kemajuan belajar untuk pengambilan keputusan. Penelitian menunjukkan bahwa instrumen tes yang valid dan reliabel menjadi dasar pengukuran kompetensi siswa (Ardhani, 2020). Analisis Rasch juga menegaskan pentingnya informasi karakteristik butir untuk perbaikan pembelajaran (Ngadi, 2023). Selain itu, instrumen non-tes valid dan reliabel menilai ranah afektif-psikomotor secara komprehensif, melampaui tes tertulis secara akurat berkelanjutan (Rahman, 2020).

Instrumen evaluasi harus memenuhi unsur validitas, reliabilitas, relevansi, dan kemampuan membedakan agar menghasilkan data yang akurat. Dalam evaluasi

kognitif, instrumen soal dianalisis secara kualitatif melalui telaah materi, konstruksi, dan bahasa sebelum digunakan. Sementara itu, analisis kuantitatif dilakukan setelah pengerjaan soal dengan bantuan perangkat lunak untuk menilai kualitas butir soal. Kedua metode ini memastikan instrumen layak digunakan dan mencerminkan kemampuan peserta didik secara objektif.

Berdasarkan observasi di PT SGMW Motor Indonesia, terlihat bahwa analisis butir soal pada sertifikasi teknisi Level 0–Apprentice belum dilaksanakan, sehingga tingkat kesulitan soal kurang terpantau karena pemilihan item hanya dilakukan secara acak. Kondisi ini membuat karakteristik penilaian dan kualitas soal tidak terdefinisi jelas. Oleh sebab itu, analisis butir soal diperlukan agar instrumen sertifikasi menghasilkan data yang akurat dan mendukung keputusan instruktur.

METODE

Studi deskriptif kuantitatif ini mengevaluasi kualitas item ujian Training Teknisi Level 0–Apprentice PT SGMW Motor Indonesia tahun 2023 dengan menelaah daya pembeda, tingkat kesulitan, dan efektivitas distraktor. Analisis dilakukan melalui pengolahan butir soal, kunci jawaban, dan respons peserta menggunakan Microsoft Excel.

1. Validitas

Validitas menunjukkan sejauh mana instrumen mampu mengukur variabel secara akurat. Menurut Yadnyawati (2019), validitas mencakup isi, konstruk, prediksi, dan kesamaan, dengan dua jenis pertama dapat ditetapkan tanpa analisis statistik. Validitas prediktif dan kesamaan memerlukan pengujian statistik untuk menilai keterkaitan tes. Suatu tes dinyatakan valid jika nilai korelasi melampaui standar product moment, sehingga item dapat dikode 1 atau 0 untuk evaluasi kualitas.

2. Tingkat Kesukaran

Berlandaskan Arifin (2010), tingkat kesulitan soal ditentukan dari perbandingan jumlah peserta yang menjawab benar dengan total peserta, sehingga butir dengan sedikit jawaban benar tergolong sulit. Pertanyaan yang terlalu mudah atau sukar dapat mengurangi keadilan penilaian, sehingga proporsi 25% mudah, 50% sedang, dan 25% sukar direkomendasikan (Santoso, 2013). Butir dinilai baik bila berkategori sedang, sehingga kategori sedang diberi nilai 1, sedangkan mudah dan sukar bernilai 0.

3. Daya Pembeda

Berlandaskan Amirono & Daryanto (2016), daya pembeda merupakan kemampuan butir soal dalam membedakan siswa berkemampuan tinggi dan rendah. Sudijono (2017) juga menegaskan bahwa daya pembeda berfungsi mengelompokkan siswa sesuai tingkat kapabilitasnya, sehingga dapat disimpulkan sebagai kapasitas soal untuk mengklasifikasikan siswa secara tepat. Dalam praktiknya, kategori sangat baik, baik, dan cukup diberi nilai 1, sedangkan jelek dan negatif diberi nilai 0 untuk mempermudah penilaian mutu butir.

4. Efektivitas Pengecoh

Menurut Arifin (2010), pada tes pilihan ganda, distraktor disajikan untuk membingungkan peserta tes, sejalan dengan (Sudijono, 2017) yang menegaskan bahwa distraktor efektif apabila minimal 5% peserta memilihnya. Dengan demikian, distraktor berfungsi sebagai opsi pengecoh yang efektif ketika mencapai batas keterpilihan tersebut. Efektivitasnya dinilai melalui kategori kualitas, dan butir diberi skor 1 jika minimal cukup, atau 0 bila tergolong buruk.

5. Tindak Lanjut Hasil Analisis

Kriteria kualitas butir soal ditetapkan berdasarkan pemenuhan empat aspek analisis. Butir dengan skor 4 dikategorikan sangat baik; skor 3 baik; skor 2 cukup; skor 1 buruk; dan skor 0 sangat buruk. Penetapan skor mencerminkan terpenuhinya atau tidak terpenuhinya aspek validitas, tingkat kesulitan, daya pembeda, serta efektivitas pengecoh dalam evaluasi kualitas butir soal.

HASIL DAN PEMBAHASAN

HASIL

Hasil analisis butir soal menggunakan Microsoft Excel menunjukkan temuan terkait kualitas item, meliputi validitas, reliabilitas, tingkat kesukaran, dan daya pembeda dalam pengukuran sebagai berikut:

1. Validitas

Rumus korelasi biserial titik dimanfaatkan guna menentukan validitas. Hasil perhitungan dibandingkan dengan tabel product moment pada taraf signifikansi 5% oleh peserta tes yang menganalisis setiap soal. Temuan analisis soal pada 10 materi pilihan ganda yang berjumlah 250 soal adalah sebagai berikut:

Tabel 1 Distribusi Soal Berdasarkan Validitas

No.	Kategori	Jumlah	Persentase	Ket
1.	Valid	88 butir	35%	Diterima
2.	Tidak Valid	162 butir	65%	Direvisi

Studi terhadap 250 soal pilihan ganda mengungkapkan bahwa 88 soal (35%) dianggap valid, sedangkan 162 soal (65%) dianggap tidak valid. Dengan demikian, dapat disimpulkan bahwa sebagian besar soal pilihan ganda tidak berkontribusi terhadap validitas soal. Pertanyaan yang memiliki kemanjuran item yang buruk tidak boleh digunakan kembali dan harus dihilangkan atau diubah.

2. Reliabilitas

Kajian reliabilitas pertanyaan secara kuantitatif menunjukkan bahwa item pertanyaan memiliki nilai reliabilitas sebesar 0.9179 yang menunjukkan adanya hubungan yang kuat. Hal ini menunjukkan bahwa penyelidikan dapat diandalkan atau temuan penilaian cukup konsisten dan dapat dipercaya.

3. Tingkat Kesukaran

Tingkat kesulitan suatu soal mengacu pada kemungkinan jawaban yang benar dengan tingkat kompetensi siswa tertentu. Kerumitan setiap soal dikategorikan menjadi lima tingkatan: sangat mudah, sedang, sedang, sulit, dan sangat sulit. Alokasi pertanyaan menurut tingkat kerumitannya ditunjukkan sebagai berikut:

Tabel 2 Distribusi Butir Soal Berdasarkan Tingkat Kesukaran

No.	Kategori	Jumlah	Persentase	Ket
1.	Sangat Mudah	9 butir	4%	Ditolak
2.	Mudah	133 butir	53%	Direvisi
3.	Sedang	105 butir	42%	Diterima
4.	Sukar	3 butir	1%	Direvisi
5.	Sangat Sukar	0 butir	0%	Ditolak

Analisis perhitungan menunjukkan bahwa dari total 250 soal yang dihasilkan, 9 soal tergolong sangat mudah, 133 soal mudah, 105 soal sedang, 3 soal menantang, dan sisanya tergolong sangat sulit.

4. Daya Pembeda

Indeks diferensiasi item dapat digunakan sebagai kriteria untuk memastikan kualitas suatu item. Baik mengacu pada kemampuan membedakan secara akurat siswa yang berkemampuan berbeda-beda, sedangkan tidak baik menunjukkan ketidakmampuan mengidentifikasi siswa berketerampilan tinggi dan rendah.

Selain itu, hasil identifikasi atribut berbagai tingkat pertanyaan ditunjukkan sebagai berikut:

Tabel 3 Distribusi Butir Soal Berdasarkan Daya Beda

No.	Kategori	Jumlah	Persentase	Ket
1.	Baik Sekali	1 butir	0%	Diterima
2.	Baik	6 butir	2%	Diterima
3.	Cukup	32 butir	13%	Direvisi
4.	Jelek	116 butir	46%	Ditolak
5.	Buang	95 butir	38%	Ditolak

Hasil perhitungan menunjukkan bahwa 2% dari 250 soal memiliki daya pembeda yang memuaskan. Selain itu, 13% item menunjukkan tingkat daya pembeda yang cukup sehingga memerlukan revisi. Sedangkan item sisanya (84%) ditolak karena indeks db < 0,1.

5. Efektivitas pengecoh

Kemanjuran pengecoh menggambarkan seberapa efektif opsi jawaban salah mengelabui siswa yang belum menguasai materi, sehingga pola sebaran respons dapat dianalisis secara tepat dan menghasilkan pengelompokan item berdasarkan variasi kekuatan pengecohnya.

Tabel 4 Distribusi Butir Soal Berdasarkan Efektivitas Pengecoh

No.	Kategori	Jumlah	Persentase	Ket
1.	Sangat Baik	2 butir	1%	Diterima
2.	Baik	21 butir	8%	Diterima
3.	Cukup	113 butir	45%	Direvisi
4.	Tidak Baik	114 butir	46%	Ditolak

Berdasarkan pemeriksaan terhadap 250 soal pilihan ganda, dapat disimpulkan bahwa 9% soal termasuk dalam kelompok yang dapat dipertahankan. Sedangkan 45% soal termasuk kelompok memadai dan memerlukan revisi. Sedangkan sisanya tergolong kurang baik kemudian dibuang.

PEMBAHASAN

1. Validitas

Dari 250 soal yang dianalisis, sebanyak 88 soal atau 35% dinyatakan valid, sedangkan 162 soal atau 65% dinyatakan tidak valid. Berdasarkan data tersebut, kualitas soal ditinjau dari aspek validitas masih perlu diperbaiki karena sebagian besar butir belum menunjukkan hubungan yang memadai dengan skor total. Butir yang tidak valid sebaiknya tidak langsung digunakan kembali sebelum diperiksa ulang dari aspek materi, konstruksi, bahasa, dan kesesuaian kunci jawaban.

Dengan proporsi butir valid sebesar 35%, kualitas soal pada aspek validitas belum dapat dikategorikan baik. Perbaikan perlu diarahkan pada penajaman indikator soal, kesesuaian tingkat berpikir, dan pemeriksaan kembali alternatif jawaban agar setiap butir benar-benar mengukur kompetensi teknisi level 0-apprentice.

2. Reliabilitas

Ketergantungan pertanyaan menunjukkan konsistensi pengukuran yang terpercaya, dengan keandalan dihitung melalui koefisien r_{11} menggunakan data Excel. Hasil analisis menunjukkan nilai reliabilitas 0.9179, yang menempatkan soal pilihan ganda pada kategori sangat tinggi, sehingga temuan diprediksi tetap stabil ketika diujikan pada kelompok serupa.

3. Tingkat Kesukaran

Tingkat kesulitan tiap aspek diukur melalui tes yang mengelompokkan soal ke lima kategori: sangat mudah hingga sangat sulit. Hasil analisis menunjukkan distribusi kerumitan yang beragam, yakni 9 sangat mudah, 133 mudah, 105 sedang, 3 sulit, dan sisanya sangat sulit, sehingga evaluasi berlangsung lebih terstruktur. Soal dengan tingkat kesukaran sedang menunjukkan kualitas yang optimal karena tidak terlalu mudah maupun terlalu menantang. Pertanyaan yang terlalu sulit dapat menurunkan motivasi, sedangkan yang terlalu mudah gagal mendorong keterlibatan siswa. Hal ini sejalan dengan temuan bahwa soal terbaik berada pada kategori kesukaran sedang (Amalia & Widayati, 2012).

4. Daya Pembeda

Indeks diskriminasi item mengukur kemampuan item untuk membedakan antara siswa dengan bakat tinggi dan siswa dengan bakat rendah. Penelitian ini mengelompokkan fitur pertanyaan ke dalam tiga kategori berbeda: (1) sangat baik, (2) cukup baik, dan (3) buruk. Berdasarkan hasil tes identifikasi item, saya dapat mengategorikan temuan analisis dan menentukan proporsi setiap item yang dapat dikenali. Distribusi temuan indeks daya pembeda untuk item-item menunjukkan bahwa 2% item menunjukkan daya pembeda yang memuaskan dan dapat diterima. Selain itu, 13% item perlu diperbarui karena tingkat kekuatan diferensiasinya yang memuaskan. Sedangkan item sisanya (84%) ditolak karena indeks db di bawah 0,1. Menurut Azwar (2016), objek yang memiliki indeks db kurang dari 0,1 sebaiknya dihilangkan atau diabaikan.

5. Efektivitas pengecoh

Efektivitas fungsi pengecoh dapat dinilai dengan menganalisis pola distribusi jawaban. Distraktor efektif bila dipilih oleh minimal 5% dari seluruh individu yang mengikuti ujian. Menurut Kartowagiran (2012) dan Mardapi (2017), pengecoh dianggap berfungsi jika dipilih oleh minimal 5% peserta tes. Pemeriksaan atas 250 soal menunjukkan hanya 22 soal (9%) memiliki pengecoh efektif, sementara 113 soal (45%) cukup baik dan 114 soal (46%) kurang berfungsi. Temuan ini menegaskan bahwa kualitas pengecoh secara keseluruhan belum memadai karena mayoritas tidak bekerja efektif.

6. Tindak Lanjut Hasil Analisis Butir Soal

Butir soal dinilai layak digunakan apabila memenuhi validitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh; bila ada kriteria tidak terpenuhi, soal wajib direvisi terlebih dahulu sehingga hasil analisis dapat menghasilkan instrumen evaluasi yang lebih akurat dan andal.

Tabel 5 Distribusi Kualitas Butir Soal

No.	Kategori	Jumlah	Persentase	Ket
1.	Baik Sekali	7 butir	3%	Diterima
2.	Baik	79 butir	32%	Diterima
3.	Cukup	80 butir	32%	Direvisi
4.	Tidak Baik	79 butir	32%	Ditolak
5.	Sangat Tidak Baik	5 butir	2%	Ditolak

Berdasarkan uraian di atas, maka soal yang tergolong “sangat baik” memenuhi keempat kriteria tersebut sehingga layak untuk digunakan kembali. Namun pada soal dengan kategori sangat baik dan cukup perlu dilakukan penyempurnaan terhadap syarat yang belum terpenuhi. Pada saat yang sama, pertanyaan yang termasuk dalam kategori kualitas buruk dan sangat buruk harus dihilangkan dan diganti dengan pertanyaan baru. Secara keseluruhan, hasil gabungan analisis validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh menunjukkan bahwa 86 butir soal dapat diterima, 80 butir perlu direvisi, dan 84 butir perlu ditolak. Tindak lanjut ini penting agar bank soal sertifikasi teknisi level 0 - apprentice memiliki kualitas yang lebih terstandar, objektif, dan dapat digunakan secara berkelanjutan.

KESIMPULAN

Berlandaskan analisis reliabilitas, validitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh menggunakan Microsoft Excel, diperoleh simpulan bahwa keseluruhan butir soal menunjukkan kualitas pengukuran sebagai berikut:

1. Dari total 250 pertanyaan, 162 pertanyaan ditemukan tidak validitas, yakni 65% dari total. 88 pertanyaan sisanya sah, juga mewakili 65% dari total pertanyaan. Dalam hal ketergantungan, r_{11} memperoleh nilai 0.9179 yang menempatkannya pada kategori sangat tinggi.
2. Dari 250 soal, tingkat kesukarannya sebagai berikut: 9 soal berkategori sangat mudah, 133 soal berkategori mudah, 105 soal berkategori sedang, 3 soal berkategori sukar, dan selebihnya berkategori sangat sulit.
3. Analisis pertanyaan memperlihatkan bahwasanya dari total 250 pertanyaan, 2% di antaranya menunjukkan kemampuan diskriminatif yang memuaskan dan memadai. Selain itu, 13% item menunjukkan tingkat daya pembeda yang cukup sehingga memerlukan revisi. Sedangkan item sisanya (84%) ditolak karena indeks db di bawah 0,1.
4. Efektivitas pengecoh dapat disimpulkan dari fakta bahwa 9% pertanyaan dapat diklasifikasikan sebagai sangat baik dan dapat dipertahankan. Sedangkan 45% soal termasuk kelompok memadai dan memerlukan revisi. Sedangkan sisanya tergolong kurang baik kemudian dibuang.
5. Kualitas pertanyaan ditentukan dengan menganalisis kategorisasinya. Hasil penelitian menunjukkan bahwa 34% (84 pertanyaan) tergolong ditolak, 32% (80 pertanyaan) diubah, dan 34% (86 pertanyaan) dipertahankan. Soal-soal yang diterima dapat digunakan sebagai bank soal yang terstandarisasi, sehingga menjadikannya berharga untuk evaluasi pembelajaran.

DAFTAR PUSTAKA

- Akhyar, M., Ihsan, F., Suharno, S., Tamrin, A., & Purwanto, P. (2024). Development Of Problem And Project-Based Learning Syntax To Improve Vocational Student Learning Outcomes. *Jurnal Pendidikan Teknologi dan Kejuruan*, 30(1), 01–19. <https://doi.org/10.21831/jptk.v30i1.53968>
- Amalia, A. N., & Widayati, A. (2012). Analisis Butir Soal Tes Kendali Mutu Kelas XII Sma Mata Pelajaran Ekonomi Akuntansi Di Kota Yogyakarta Tahun 2012. *Jurnal Pendidikan Akuntansi Indonesia*, X(1), 1–26. <https://doi.org/https://doi.org/10.21831/jpai.v10i1.919>

- Amirono, & Daryanto. (2016). *Evaluasi dan Penilaian Pembelajaran Kurikulum 2013*. Gava Media.
- Anas, Z., & Budiman, A. (2020). EVALUASI PROGRAM PEMBELAJARAN PEKERJAAN DASAR TEKNIK OTOMOTIF KELAS X TKRO DI SMK NEGERI 1 SEDAYU. *Jurnal Pendidikan Vokasi Otomotif*, 2(2), 1–22. <https://doi.org/10.21831/jpvo.v2i2.33561>
- Ardhani, Y. (2020). KUALITAS BUTIR SOAL PENILAIAN AKHIR TAHUN MATA PELAJARAN TEKNOLOGI DASAR OTOMOTIF KELAS X TEKNIK KENDARAAN RINGAN OTOMOTIF DI SMK MUHAMMADIYAH GAMPING. *Jurnal Pendidikan Vokasi Otomotif*, 3(1), 85–94. <https://doi.org/10.21831/jpvo.v3i1.34917>
- Arifin. (2010). *Metode penelitian kualitatif, kuantitatif, dan R & D*. Alfabeta.
- Azwar, S. (2016). *Metode Penelitian*. Pustaka Pelajar.
- Rahman, M. H. (2020). ANALISIS RANAH PSIKOMOTOR KOMPETENSI DASAR TEKNIK PENGUKURAN TANAH KURIKULUM SMK TEKNIK KONSTRUKSI DAN PROPERTI. *Jurnal Pendidikan Teknologi dan Kejuruan*, 17(1), 53. <https://doi.org/10.23887/jptk-undiksha.v17i1.23022>
- Indasah, S. S., & Sutiman. (2023). Pengaruh Tingkat Partisipasi Dalam Model Pembelajaran Project-Basedlearning Terhadap Kemampuan Berpikir Kritis Anggota Tim Garuda UNY. *Jurnal Pendidikan Vokasi Otomotif*, 6(1), 85–97. <https://doi.org/https://doi.org/10.21831/jpvo.v6i1.67433>
- Kartowagiran. (2012). *Penulisan Butir Soal*. Universitas Negeri Yogyakarta.
- Mardapi, D. (2017). *Pengukuran Penilaian dan Evaluasi Pendidikan* (2 ed.). Parama Publishing.
- Ngadi. (2023). Analisis Model Rasch Untuk Mengukur Kompetensi Pengetahuan Siswa SMKN 1 Kalianget Pada Mata Pelajaran Perawatan Sistem Kelistrikan Sepeda Motor. *Jurnal Pendidikan Vokasi Otomotif*, 6(1), 1–19. <https://doi.org/https://doi.org/10.21831/jpvo.v6i1.63479>
- Santoso, S. (2013). *Statistik Ekonomi Plus Aplikasi SPSS*. Umpo Press.
- Sigit Wahyudi, K., & Arifin, Z. (2023). Profil Ideal Guru Sekolah Menengah Kejuruan Teknik Kendaraan Ringan Otomotif. *Jurnal Pendidikan Vokasi Otomotif*, 5(2), 51–65. <https://doi.org/https://doi.org/10.21831/jpvo.v5i2.60198>
- Sudijono, A. (2017). *Pengantar Evaluasi Pendidikan*. PT. RajaGrafindo Persada.
- Suyitno. (2022). Evaluasi Peran Guru Produktif Otomotif dalam Proses Pembelajaran Jarak Jauh (PJJ) Melalui Google Classroom. *Jurnal Pendidikan Vokasi Otomotif*, 5(1). <https://doi.org/https://doi.org/10.21831/jpvo.v5i1.55662>
- Yadnyawati, I. A. G. (2019). *Evaluasi Pembelajaran* (I. K. Suda, Ed.). UNHI Press.