



ANALISIS BUTIR SOAL *HIGH ORDER THINKING SKILLS* (HOTS) BERBANTUAN *ARTIFICIAL INTELLIGENCE* (AI) UNTUK PEMBELAJARAN PERAWATAN DAN PERBAIKAN SASIS SEPEDA MOTOR DI SMK

Rizki Kasprianto¹, Wahid Munawar², Sriyono³

Universitas Pendidikan Indonesia, Fakultas Pendidikan Teknik dan Industri
Jl. Dr. Setiabudhi No. 229 Bandung 40154

rizkikasprianto@upi.edu; wahidmunawar@upi.edu; sriyono@upi.edu

ABSTRACT/ABSTRAK

This study aims to analyze the quality of High Order Thinking Skills (HOTS)-based test items generated by Artificial Intelligence (AI), specifically ChatGPT, in the subject of Motorcycle Chassis Maintenance and Repair in vocational high schools (SMK). The analysis includes validity, reliability, difficulty level, discriminating power, and distractor effectiveness. The research method used is a survey with a descriptive approach, with item analysis techniques utilizing Anates V4 software. The research subjects consisted of 153 twelfth-grade students majoring in Motorcycle Engineering at SMKN 8 Bandung and SMKN 7 Baleendah. Data were collected through a written test comprising 60 HOTS-based questions generated by ChatGPT. The results show that the tested questions have good validity and high reliability. In terms of difficulty level, 13.33% of the questions were classified as easy, 58.33% as moderate, and 28.33% as difficult. The discriminating power analysis indicates that most questions effectively differentiate between students with high and low abilities. The distractor effectiveness was also optimal in guiding students' answer choices. This study concludes that AI-generated HOTS-based questions have the potential to enhance the quality of learning evaluation in vocational high schools, with good validity and reliability, AI can serve as an innovative solution for educators in developing more efficient and high-quality assessment instruments.

Penelitian ini bertujuan untuk menganalisis kualitas butir soal berbasis *High Order Thinking Skills* (HOTS) yang dihasilkan oleh *Artificial Intelligence* (AI), khususnya ChatGPT, dalam mata pelajaran Perawatan dan Perbaikan Sasis Sepeda Motor di SMK. Analisis mencakup validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas distraktor. Metode penelitian yang digunakan adalah survei dengan pendekatan deskriptif, dengan teknik analisis butir soal menggunakan perangkat lunak Anates V4. Subjek penelitian terdiri dari 153 siswa kelas XII jurusan Teknik Sepeda Motor di SMKN 8 Bandung dan SMKN 7 Baleendah. Data dikumpulkan melalui tes tertulis yang terdiri dari 60 butir soal HOTS yang dibuat menggunakan ChatGPT. Hasil penelitian menunjukkan bahwa soal-soal yang diuji memiliki tingkat validitas yang baik dan reliabilitas yang

ARTICLE INFO

Article History:

Submitted/Received
04 Jun 2025

First Revised
13 Jun 2025

Accepted
06 Jul 2025

Online Date
14 Jul 2025

Publication Date
14 Jul 2025

Keywords:

High Order Thinking Skills (HOTS); Artificial Intelligence (AI); item analysis; vocational high school.

Kata kunci:

High Order Thinking Skills (HOTS); Artificial Intelligence (AI); analisis butir soal; SMK.

tinggi. Dari segi tingkat kesukaran, 13,33% soal tergolong mudah, 58,33% sedang, dan 28,33% sukar. Daya pembeda menunjukkan sebagian besar soal mampu membedakan siswa dengan kemampuan tinggi dan rendah secara efektif. Efektivitas distraktor juga cukup optimal dalam mengarahkan pilihan jawaban siswa. Kesimpulan penelitian ini menegaskan bahwa soal HOTS berbasis AI berpotensi meningkatkan kualitas evaluasi pembelajaran di SMK, dengan validitas dan reliabilitas yang baik, AI dapat menjadi solusi inovatif bagi pendidik dalam menyusun instrumen evaluasi yang lebih efisien dan berkualitas.

1. PENDAHULUAN

Seiring perkembangan teknologi, kecerdasan buatan (Artificial Intelligence/AI) mulai banyak dimanfaatkan dalam dunia pendidikan. AI memiliki potensi untuk mendukung guru dalam merancang soal yang lebih efisien, berkualitas dan bervariasi. Salah satu implementasinya adalah penggunaan ChatGPT sebagai alat bantu dalam menyusun soal. Trust dalam Pardomuan (2024) menyatakan bahwa AI dapat mempercepat proses pembelajaran, mendorong kreativitas, serta menyediakan sumber belajar yang lebih luas. (Direktorat Pembelajaran & Kemahasiswaan, 2024) menyatakan bahwa *generative AI* dapat membantu membuat materi ajar baru, seperti pertanyaan untuk kuis dan latihan, penjelasan konsep, serta rangkuman materi ajar.

Kemampuan berpikir tingkat tinggi atau *High Order Thinking Skills* (HOTS) menjadi aspek penting dalam dunia pendidikan untuk mempersiapkan siswa menghadapi tantangan global. Konteks pembelajaran vokasional seperti di Sekolah Menengah Kejuruan (SMK), penerapan HOTS sangat diperlukan, terutama dalam mata pelajaran Perawatan dan Perbaikan Sasis Sepeda Motor yang menuntut keterampilan analitis, berpikir kritis, dan pemecahan masalah. Hal ini sejalan dengan (Pusat Penilaian Pendidikan, 2019) menyatakan bahwa setiap peserta didik perlu memiliki keterampilan berpikir tingkat tinggi agar mampu berperan secara maksimal, baik sebagai individu yang mandiri dan kritis maupun sebagai anggota masyarakat yang produktif.

Berdasarkan kenyataan di sekolah, penyusunan soal HOTS masih sering dilakukan secara konvensional, yaitu secara manual tanpa bantuan teknologi atau alat digital pendukung. Guru menyusun soal berdasarkan pengalaman pribadi atau referensi terbatas dengan rujukan yang bersumber kepada buku konvensional dengan capaian pembelajaran beserta kompetensi inti dan kompetensi dasar dengan keterbatasan waktu. (Ilmi, 2021) menyatakan bahwa guru memiliki berbagai tanggung jawab, termasuk administrasi, mengajar, mengoreksi tugas, serta menyusun perangkat evaluasi. Hal ini berdampak pada kualitas instrumen evaluasi yang digunakan di kelas, yang terkadang dirasa kurang optimal dalam mengukur kemampuan berpikir tingkat tinggi siswa. Dibutuhkan alternatif solusi yang mampu membantu pendidik dalam menyusun instrumen penilaian yang lebih bervariasi dan berkualitas.

ChatGPT merupakan salah satu bentuk *Generative AI* yang mampu menghasilkan soal secara otomatis berdasarkan perintah pengguna (prompt). Beberapa penelitian terdahulu seperti yang dilakukan oleh Zuhri dkk. (2024) menunjukkan bahwa soal yang dihasilkan

AI melalui platform *QuestionWell* memiliki validitas dan reliabilitas yang cukup baik. Namun, masih diperlukan analisis mendalam untuk mengetahui kualitas butir soal secara menyeluruh, termasuk tingkat kesukaran, daya pembeda, dan efektivitas distraktor dalam soal HOTS yang dihasilkan AI. Penelitian yang dilakukan oleh Ristalia dkk. (2025) menekankan pada analisis butir soal pilihan ganda menggunakan aplikasi Anates pada siswa SMK, dengan fokus pada pengukuran teknis seperti tingkat kesulitan, daya pembeda, efektivitas pengecoh, dan reliabilitas soal. Hasilnya menunjukkan bahwa Anates efektif dalam meningkatkan kualitas evaluasi. Meskipun penelitian tersebut relevan dalam konteks penggunaan Anates, penelitian ini belum menyentuh aspek pemanfaatan *Artificial Intelligence* (AI) dalam pembuatan soal, khususnya yang berbasis ChatGPT. Selain itu, penelitian oleh Dewi dkk. (2020) mengembangkan instrumen penilaian HOTS untuk materi hereditas di SMA dan menganalisis validitas serta reliabilitasnya menggunakan Anates. Hasilnya menunjukkan bahwa instrumen yang dikembangkan sangat valid dan reliabel. Instrumen tersebut disusun secara manual oleh peneliti, bukan dihasilkan oleh teknologi AI.

Penelitian ini memiliki kebaruan pada penggunaan ChatGPT untuk menghasilkan butir soal HOTS di bidang otomotif SMK, serta menganalisis kualitas butir soal tersebut menggunakan perangkat lunak Anates V4. Penggunaan ChatGPT dalam konteks ini belum banyak diteliti, terutama pada mata pelajaran vokasional seperti perawatan dan perbaikan sasis sepeda motor.

Dampak dari penelitian ini diharapkan dapat memberikan kontribusi nyata terhadap dunia pendidikan, khususnya bagi guru SMK dalam menyusun instrumen evaluasi pembelajaran yang berkualitas dan efisien. Selain itu, hasil penelitian ini dapat menjadi rujukan dalam pengembangan kebijakan pemanfaatan AI secara bijak di lingkungan pendidikan.

Tujuan dari penelitian ini adalah untuk menganalisis kualitas butir soal HOTS yang dihasilkan oleh ChatGPT pada mata pelajaran Perawatan dan Perbaikan Sasis Sepeda Motor di SMK. Instrumen analisis yang digunakan adalah Anates V4, yaitu perangkat lunak yang umum digunakan untuk mengevaluasi kualitas soal secara kuantitatif berdasarkan parameter validitas, tingkat kesukaran, daya pembeda, efektivitas pengecoh (distraktor), dan reliabilitas. Penelitian ini diharapkan dapat memberikan gambaran lebih lengkap mengenai potensi pemanfaatan AI dalam menyusun instrumen evaluasi pembelajaran yang berkualitas, sesuai kebutuhan pendidikan abad 21.

2. METODE PENELITIAN

Penelitian ini menggunakan metode survei dengan pendekatan deskriptif kuantitatif. Metode ini dipilih karena sesuai dengan tujuan penelitian, yaitu untuk mendeskripsikan kualitas butir soal HOTS yang dihasilkan oleh *Artificial Intelligence* (AI) berbasis ChatGPT, dalam mata pelajaran Perawatan dan Perbaikan Sasis Sepeda Motor di SMK.

Sugiyono (2019) menjelaskan bahwa metode survei merupakan jenis penelitian yang dapat diterapkan pada populasi dalam skala besar maupun kecil, namun analisisnya difokuskan pada data yang diperoleh dari sampel sebagai representasi populasi tersebut. Tujuan dari metode ini adalah untuk mengidentifikasi kecenderungan umum, persebaran karakteristik, serta hubungan antara variabel-variabel yang bersifat sosiologis maupun psikologis.

Penelitian ini menggunakan metode survei deskriptif, karena fokusnya adalah menganalisis data dari sampel siswa SMK untuk menggambarkan kualitas butir soal HOTS yang dihasilkan oleh ChatGPT. Melalui analisis sampel ini, penelitian bertujuan untuk menganalisis kualitas butir soal seperti validitas, tingkat kesukaran, daya pembeda, efektivitas distraktor, dan reliabilitas, menggunakan perangkat lunak Anates V4, tanpa menguji hubungan antar variabel. Pendekatan survei deskriptif ini tepat digunakan untuk memperoleh gambaran objektif tentang kualitas butir soal berbasis AI dalam konteks pembelajaran vokasional di SMK.

Sugiyono (2019) menyatakan bahwa penelitian survei dapat bersifat deskriptif, komparatif, asosiatif, gabungan keduanya, maupun menggunakan analisis jalur dan model persamaan struktural (SEM)".

3. HASIL PENELITIAN

Analisis butir tes secara kuantitatif dilakukan berdasarkan teori tes klasik pada data empirik tentang jawaban responden pada tiap-tiap butir tes. Analisis menghasilkan karakteristik butir tes berupa statistik butir yang meliputi tingkat kesukaran butir tes, daya pembeda butir tes, efektivitas distraktor, reliabilitas, dan validitas.

Hasil analisis butir secara kuantitatif dengan bantuan program Anates v4. Tingkat kesukaran yang diperoleh dari hasil pengujian soal dapat dilihat pada Tabel 1.

Tabel 1. Tingkat Kesukaran

Kategori	Nomor Butir Tes	Jumlah (%)
Sukar	14, 18, 19, 20, 22, 25, 26, 27, 28, 45, 46, 47, 48, 49, 50, 55, 56	17 (28,33%)
Sedang	7, 9, 10, 12, 13, 15, 16, 17, 21, 23, 24, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 51, 52, 53, 54, 57, 58, 59, 60	35 (58,33%)
Mudah	1, 2, 3, 4, 5, 6, 8, 11	8 (13,33%)

Tingkat kesukaran menunjukkan sejauh mana soal itu mudah, sedang atau sulit bagi peserta tes tertentu. Soal yang terlalu mudah atau terlalu sukar kurang memberikan informasi bermakna mengenai kemampuan peserta. Soal HOTS berbasis ChatGPT menunjukkan variasi tingkat kesukaran yang ideal, sehingga sesuai digunakan untuk mengukur kemampuan berpikir tingkat tinggi siswa SMK.

Tahapan selanjutnya adalah daya beda. Daya beda butir tes adalah kemampuan butir yang dengan skornya dapat membedakan peserta dari kelompok berkemampuan tinggi dan peserta dari kelompok berkemampuan rendah. Makin tinggi daya beda butir, maka makin besar perbedaan skor yang dihasilkan oleh kelompok berkemampuan tinggi dan kelompok berkemampuan rendah. Semakin tinggi daya beda butir, semakin banyak peserta dari kelompok berkemampuan tinggi yang dapat menjawab butir itu dengan benar serta makin sedikit peserta dari kelompok berkemampuan rendah yang dapat menjawab dengan benar. Daya beda butir tes melibatkan pembagian peserta ke kelompok berkemampuan tinggi dan kelompok berkemampuan rendah.

Daya pembeda soal yang dianalisis menggunakan program komputer Anates versi 4 digambarkan pada Tabel 2.

Tabel 2. Daya Pembeda

Kategori	Nomor Butir Tes	Jumlah (%)
Cukup memuaskan	7, 9, 12, 15, 21, 24, 26, 27, 29, 31, 32, 33, 36, 37, 38, 39, 40, 41, 42, 43, 44, 51, 52, 53, 54, 58, 60	27 (51,92%)
Sedikit atau tanpa revisi	13, 14, 16, 17, 20, 34, 55, 56, 59	9 (17,31%)
Perbatasan atau perlu revisi	18, 25, 30, 35, 46, 50	6 (11,54%)
Dibuang atau revisi total	10, 19, 22, 23, 28, 45, 47, 48, 49	9 (17,31%)

Tahapan selanjutnya adalah Efektivitas Distraktor. Efektifitas distraktor perlu diperhatikan guna mengetahui banyaknya peserta didik menjawab benar sesuai dengan kunci jawaban dan berapa banyak peserta didik yang memilih distraktor atau pengecoh. Pengecoh berfungsi guna mengidentifikasi peserta tes yang berkemampuan tinggi. Pengecoh dikatakan efektif jika dipilih oleh peserta tes yang berasal dari kelompok bawah yaitu peserta didik yang mempunyai kemampuan rendah. Dan jika dipilih oleh peserta didik yang berkemampuan tinggi maka pengecoh tersebut tidak berfungsi dengan baik.

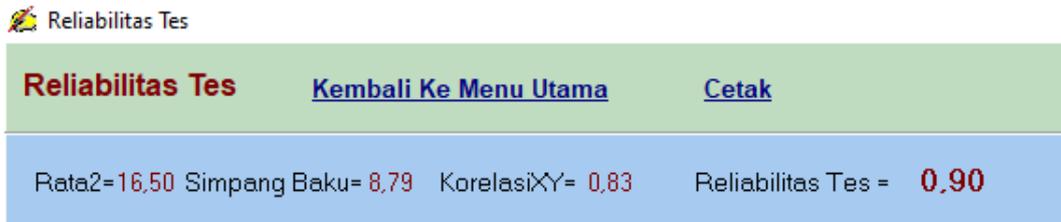
Efektivitas distraktor yang dianalisis menggunakan program komputer Anates versi 4 digambarkan pada Tabel 3.

Tabel 3. Efektivitas Distraktor

Kategori	Nomor Butir Tes dan Distraktor	Jumlah (%)
Berfungsi baik	12, 13, 14, 15, 16, 17, 18, 20, 21, 24, 25, 26, 27, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 46, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60	41 (95,34%)
Belum berfungsi baik	7, 9	1 (4,65%)

Tahapan selanjutnya adalah reliabilitas. Kualitas butir tes secara keseluruhan dapat dilihat dari hasil reliabilitas tes. Menggunakan bantuan program Anates V4, koefisien reliabilitas perangkat tes 0,90. Susetyo, B. (2015) menjelaskan bahwa sebuah instrumen dapat dikatakan reliabel apabila memiliki nilai koefisien korelasi minimal 0,50.

Reliabilitas yang dianalisis menggunakan program komputer Anates versi 4 digambarkan pada Gambar 1.



Gambar 1. Reliabilitas Tes

4. PEMBAHASAN

Berdasarkan hasil penelitian yang telah dianalisis menggunakan perangkat lunak Anates V4, dapat dilakukan pembahasan terkait validitas, dan reliabilitas dari soal-soal berbasis *High Order Thinking Skills* (HOTS) yang dihasilkan oleh ChatGPT dalam mata pelajaran Perawatan dan Perbaikan Sasis Sepeda Motor di SMK.

1. Validitas Konstruk

Validitas merupakan aspek penting dalam menentukan kualitas suatu instrumen evaluasi. Validitas yang dianalisis dalam penelitian ini adalah validitas konstruk, yaitu validitas yang menunjukkan sejauh mana butir-butir soal mampu mengukur aspek kognitif yang hendak diukur. Arikunto, S. (2016) mengemukakan bahwa sebuah tes dikatakan memiliki validitas konstruksi apabila butir-butir soal yang membangun tes tersebut mengukur setiap aspek berpikir seperti yang disebutkan dalam Tujuan Instruksional Khusus (TIK). Analisis terhadap tingkat kesukaran, daya pembeda, dan efektivitas distraktor dilakukan guna mengetahui kelayakan konstruksi setiap butir soal.

a. Tingkat Kesukaran

Tingkat kesukaran menggambarkan sejauh mana soal dapat dijawab dengan benar oleh peserta didik. Berdasarkan hasil analisis terhadap 60 butir soal, terdapat 35 butir soal (58,33%) yang termasuk kategori sedang, 17 butir soal (28,33%) tergolong sukar, dan 8 butir soal (13,33%) termasuk mudah. Sebaran ini menunjukkan bahwa mayoritas soal memiliki tingkat kesukaran yang berada dalam rentang ideal. Komposisi soal mudah dan

sukar juga tergolong proporsional, sehingga dapat membantu menggambarkan kemampuan peserta didik secara objektif.

b. Daya Pembeda

Daya pembeda menunjukkan kemampuan soal dalam membedakan peserta didik yang berkemampuan tinggi dan rendah. Sebanyak 27 butir soal (51,92%) memiliki daya pembeda cukup memuaskan. Terdapat pula 9 butir soal (17,31%) yang termasuk dalam kategori baik atau hanya memerlukan sedikit revisi. Sebanyak 6 butir soal (11,54%) tergolong perlu revisi dan 9 butir soal (17,31%) berada pada kategori dibuang atau perlu revisi total. Proporsi soal dengan daya pembeda baik menunjukkan bahwa sebagian besar soal dapat digunakan untuk menilai perbedaan tingkat pemahaman siswa, meskipun masih terdapat beberapa soal yang perlu perbaikan agar kualitas instrumen semakin meningkat.

c. Efektivitas Distraktor

Distraktor atau pengecoh merupakan pilihan jawaban yang berfungsi untuk mengalihkan peserta didik yang kurang menguasai materi. Dari total 43 soal, sebanyak 41 butir (95,34%) memiliki distraktor yang berfungsi dengan baik. Hanya 2 butir soal (4,65%) yang distraktornya belum berfungsi secara optimal. Temuan ini memperlihatkan bahwa sebagian besar distraktor telah dirancang secara tepat dan mampu menjalankan fungsinya sebagai pengecoh yang valid.

2. Reliabilitas Soal

Reliabilitas mengacu pada konsistensi hasil yang diperoleh dari suatu tes ketika diujikan berulang kali. Berdasarkan analisis menggunakan Anates V4, reliabilitas soal yang dihasilkan oleh ChatGPT tergolong tinggi. Hal ini menunjukkan bahwa soal-soal yang dihasilkan mampu memberikan hasil yang reliabel. Arikunto, S. (2016) mengemukakan bahwa reliabilitas menunjuk pada satu pengertian bahwa sesuatu instrumen cukup dapat dipercaya untuk digunakan sebagai alat pengumpul data karena instrumen tersebut sudah baik. Instrumen yang baik tidak akan bersifat tendensius mengarahkan responden untuk memilih-memilih jawaban tertentu. Instrumen yang sudah dapat dipercaya, yang reliabel akan menghasilkan data yang dapat dipercaya juga. Reliabel dalam penilaian ini penting untuk memastikan bahwa hasil evaluasi benar-benar mencerminkan kemampuan siswa.

5. KESIMPULAN

Validitas butir soal HOTS yang dihasilkan oleh ChatGPT tergolong baik jika ditinjau dari tiga aspek utama, yaitu:

- a. Tingkat kesukaran soal menunjukkan bahwa sebagian besar butir berada dalam kategori sedang, yaitu sebesar 58,33%. Hal ini menunjukkan bahwa soal telah disusun secara proporsional dan mampu mengukur kemampuan siswa pada tingkat yang optimal.
- b. Daya pembeda soal menunjukkan bahwa sebanyak 51,92% butir soal memiliki kemampuan cukup memuaskan dalam membedakan peserta didik yang berkemampuan tinggi dan rendah. Meskipun terdapat sejumlah butir yang perlu direvisi atau dibuang, secara umum daya pembeda soal sudah memadai untuk digunakan dalam evaluasi pembelajaran.
- c. Efektivitas distraktor berada dalam kategori sangat baik, dengan 41 dari 43 (95,34%) butir soal memiliki distraktor yang berfungsi secara optimal. Sebagian besar opsi jawaban pengecoh berhasil menarik perhatian peserta yang kurang menguasai materi, sehingga mendukung validitas konstruk instrumen.

Hasil analisis reliabilitas menggunakan program Anates v4 menunjukkan bahwa butir soal HOTS yang dihasilkan oleh ChatGPT memiliki tingkat reliabilitas yang tinggi, dengan koefisien reliabilitas sebesar 0,90. Nilai ini menunjukkan bahwa soal yang dikembangkan memiliki konsistensi yang baik dalam mengukur kemampuan siswa dalam memahami materi Perawatan dan Perbaikan Sasis Sepeda Motor.

6. REFERENSI

- Anas Sudijono. (2011). *Pengantar Evaluasi Pendidikan*, Jakarta: Raja Grafindo Persada.
- Asmawi Zainul & Neohi Nasution, (1997). *Penilaian Hasil Belajar*. Depdikbud: Universitas Terbuka.
- Direktorat Pembelajaran dan Kemahasiswaan, Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. (2024). *Panduan penggunaan generative artificial intelligence (GenAI) pada pembelajaran di perguruan tinggi*.
- Dewi, N. P., Rahmi, Y. L., Alberida, H., & Rahmawati, D. (2020). *Validitas dan reliabilitas instrumen penilaian kemampuan berpikir tingkat tinggi tentang materi*

- hereditas untuk peserta didik SMA/MA. Jurnal Eksakta Pendidikan (JEP)*, 4(2), 138–146. <https://doi.org/10.24036/jep/vol4-iss2/512>
- Engky Triyuana, R., Shofiyah, I. N., Hakim, L., & Pratiwi, V. (2025). *Analisis butir soal pilihan ganda dengan aplikasi Anates pada siswa SMK. Jurnal Manajemen dan Akuntansi WANARGI*, 2(2), 201–205.
- Faisal Pardomuan Siregar, Sri Wahyudi, Detri Amelia Chandra, & Ari Aprilia Dwiana. (2024). *ChatGPT Dalam Mendukung Pembelajaran di Sekolah. Jurnal Pendidikan Teknologi Informasi Dan Vokasional*, 6(1), 24–34. <http://jurnal.fkip.unila.ac.id/index.php/JPVTI/index>.
- Ilmi, D. (2021). *Problematika Guru dalam Pembelajaran Tahfiz dan Alternatif Solusinya (Studi Kasus di Pondok Tahfiz Yanbu'ul Qur'an Remaja Bejen Kudus)*. Repository.iainkudus.ac.id, 34.
- Pusat Penilaian Pendidikan, Badan Penelitian dan Pengembangan, Kementerian Pendidikan dan Kebudayaan. (2019). *Panduan penulisan soal HOTS–Higher Order Thinking Skills*.
- Ristalia Engky Triyuana, Irda Nur Shofiyah, Luqman Hakim, & Vivi Pratiwi. (2025). Analisis Butir Soal Pilihan Ganda dengan Aplikasi Anates pada Siswa SMK. WANARGI : Jurnal Manajemen Dan Akuntansi, 2(2), 201-205. <https://doi.org/10.62017/wanargi.v2i2.3733>
- Sapitri, A., Kurniati, T., & Yuliawati, A. 2022. *Analisis Kualitas Soal UAS Biologi SMA Kelas X dan XI MIA*. Bioeduca: Journal of Biology Education, 4(1), 45–56.
- Suharsimi Arikunto. (2016). *Dasar-dasar Evaluasi Pendidikan*. Jakarta: PT Bumi Aksara.
- Sugiyono. (2019). *Metodelogi Penelitian Kuantitatif dan Kualitatif Dan R&D*. Bandung: ALFABETA.
- Susetyo, Budi. (2015). *Prosedur Penyusunan dan Analisis Tes untuk Penilaian Hasil Belajar Bidang Kognitif*. Bandung: Refika Aditama.
- Wandri, R., Malahati, F., Romadhon, K., & Ramadhan, S. 2023. *Analisis Butir Soal Tipe Multiple Choice Questions pada Penilaian Harian Sekolah Dasar*. 10(2), 93–105.
- Zuhri, N. Z., Syihabuddin, S., & Tatang, T. (2024). *Analisis Validitas, Reliabilitas, dan Tingkat Kesukaran Soal Bahasa Arab Tingkat SMP Berbasis Artificial Intelligence (AI) melalui Platform QuestionWell*. Jurnal Pendidikan Dan Pembelajaran Indonesia (JPPI), 4(2), 693–704. <https://doi.org/10.53299/jppi.v4i2.576>