

Kualitas Butir Soal Informatika Materi Kecerdasan Artifisial Kelas XI: Analisis ANATES dan Peta Revisi Instrumen

Amilia Khasanah Marzuki^{*1}, Ruffi², Sabariah³

^{1,2,3}Program Studi Teknologi Pendidikan, Sekolah Pascasarjana, Universitas PGRI Adi Buana Surabaya, Indonesia

Email: ¹250020069@student.unipasby.ac.id, ²ruffi.adibuana@gmail.com, ³sabariah@unipasby.ac.id

Abstrak

Penelitian ini bertujuan menganalisis kualitas 25 butir soal pilihan ganda Informatika materi kecerdasan artifisial (AI) kelas XI serta menyusun peta revisi instrumen berbasis indikator psikometrik. Penelitian menggunakan desain deskriptif kuantitatif evaluatif dengan sumber data berupa kisi-kisi soal, kunci jawaban, dan 35 lembar jawaban peserta didik SMA Al Muslim. Data dianalisis menggunakan ANATES versi 4.09 berdasarkan teori tes klasik yang mencakup validitas butir, reliabilitas tes, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Hasil penelitian menunjukkan koefisien reliabilitas sebesar 0,73 sehingga instrumen tergolong memadai untuk evaluasi kelas. Dari sisi validitas, 13 butir (52%) berkategori cukup, 7 butir (28%) rendah, 4 butir (16%) sangat rendah, dan 1 butir (4%) berkorelasi negatif. Tingkat kesukaran terdiri atas 1 butir sangat mudah, 12 butir mudah, dan 12 butir sedang, sehingga belum tersedia butir sukar. Daya pembeda menunjukkan 15 butir (60%) berkategori baik sampai sangat baik, tetapi efektivitas pengecoh masih menjadi kelemahan karena hanya 4 butir (16%) yang seluruh distraktornya berfungsi baik. Sintesis indikator menunjukkan 5 butir dapat dipertahankan, 17 butir perlu direvisi, dan 3 butir harus ditulis ulang. Instrumen layak digunakan secara terbatas, tetapi perlu revisi terarah pada distraktor, redaksi stem, dan proporsi butir berpikir tingkat tinggi sebelum dijadikan bank soal final.

Kata kunci: analisis butir soal, ANATES, evaluasi pembelajaran, informatika, literasi AI.

Quality of Grade XI Informatics Test Items on Artificial Intelligence: ANATES Analysis and Instrument Revision Map

Abstract

This study aimed to analyze the quality of 25 multiple-choice Informatics items on artificial intelligence (AI) for Grade XI students and to develop an instrument revision map based on psychometric indicators. The study employed an evaluative quantitative descriptive design using a test blueprint, answer key, and 35 student response sheets from SMA Al Muslim as data sources. The data were analyzed using ANATES version 4.09 within the classical test theory framework, covering item validity, test reliability, difficulty level, discrimination power, and distractor effectiveness. The results showed a reliability coefficient of 0.73, indicating that the instrument was adequate for classroom evaluation. In terms of validity, 13 items (52%) were categorized as adequate, 7 items (28%) as low, 4 items (16%) as very low, and 1 item (4%) showed a negative correlation. Difficulty analysis revealed 1 very easy item, 12 easy items, and 12 moderate items, indicating the absence of difficult items. Discrimination analysis showed that 15 items (60%) were good to very good; however, distractor effectiveness remained the main weakness because only 4 items (16%) had all distractors functioning properly. The synthesis of all indicators showed that 5 items could be retained, 17 items required revision, and 3 items should be rewritten. The instrument is usable on a limited basis, but targeted revision of distractors, item stems, and the proportion of higher-order thinking items is required before it can serve as a final school item bank.

Keywords: item analysis, ANATES, learning evaluation, informatics, AI literacy

1. PENDAHULUAN

Pembelajaran Informatika pada fase F menuntut peserta didik tidak hanya mengenali teknologi digital, tetapi juga mampu menganalisis strategi algoritmik secara kritis, menilai efisiensi solusi, serta memahami keterbatasan teknologi yang digunakan [1]. Tuntutan tersebut semakin relevan ketika materi kecerdasan artifisial (AI) mulai masuk ke ruang kelas. Pada konteks ini, asesmen tidak cukup hanya menguji ingatan terhadap istilah

AI, tetapi juga perlu mengukur kemampuan siswa dalam menafsirkan data, mengenali bias, mengevaluasi keluaran AI generatif, dan mengambil keputusan secara bertanggung jawab.

Literasi AI dipahami sebagai seperangkat kompetensi untuk memahami konsep dasar AI, menggunakan aplikasi berbasis AI secara tepat, dan mengevaluasi dampak sosial-etisnya [2], [3]. Kajian terbaru pada pendidikan K-12 menunjukkan bahwa pembelajaran AI yang bermakna cenderung menekankan keterlibatan siswa dalam aktivitas berbasis proyek, kolaborasi manusia-komputer, serta refleksi kritis terhadap hasil yang dihasilkan sistem AI [4]. Hal ini sejalan dengan kebutuhan penguatan berpikir kritis dan prompt engineering dalam pendidikan modern [5]. UNESCO juga menegaskan bahwa pemanfaatan AI dalam pendidikan harus berpusat pada manusia, menjaga privasi, dan mendorong validasi kritis terhadap keluaran teknologi [6]. Dengan demikian, instrumen evaluasi pada materi AI perlu selaras dengan karakter materi yang konseptual, aplikatif, dan etis.

Kualitas instrumen merupakan syarat penting agar skor tes benar-benar merepresentasikan capaian belajar. Instrumen yang reliabel tetapi berisi butir lemah tetap dapat menghasilkan keputusan pembelajaran yang kurang tepat. Oleh karena itu, analisis butir perlu dilakukan untuk menelaah validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh [8]. Pada praktik sekolah, ANATES banyak digunakan karena menyediakan prosedur analisis yang relatif mudah diakses guru dan menghasilkan informasi praktis untuk merevisi butir [9].

Sejumlah penelitian sebelumnya memperlihatkan bahwa soal buatan guru masih sering bermasalah pada distribusi tingkat kesukaran, daya pembeda, atau kualitas pengecoh [10]-[12]. Temuan mutakhir pada instrumen komputer akuntansi fase F juga menunjukkan pola serupa, yaitu reliabilitas dapat berada pada kategori memadai sementara sebagian butir masih perlu diperbaiki pada tingkat validitas dan distraktor [13]. Secara khusus, efektivitas pengecoh penting karena pengecoh yang tidak dipilih, terlalu jelas salah, atau tidak setara dengan kunci dapat menurunkan daya diagnostik butir [14].

Meskipun analisis butir telah banyak dilakukan, kajian yang secara khusus menelaah instrumen Informatika materi AI di tingkat SMA masih terbatas. Selain itu, banyak studi berhenti pada pelaporan indeks statistik tanpa menyajikan peta tindak lanjut yang langsung dapat digunakan guru. Kebaruan penelitian ini terletak pada penggabungan analisis psikometrik berbasis ANATES dengan peta revisi butir yang memisahkan item menjadi kategori dipertahankan, direvisi, dan ditulis ulang. Penelitian ini bertujuan menganalisis kualitas butir soal Informatika materi AI kelas XI dan merumuskan rekomendasi revisi instrumen untuk pengembangan bank soal sekolah yang lebih valid, reliabel, dan diagnostik.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan deskriptif kuantitatif dengan orientasi evaluatif. Pendekatan ini dipilih karena fokus penelitian bukan menguji perlakuan pembelajaran, melainkan menilai mutu instrumen tes berdasarkan data empirik jawaban peserta didik. Subjek data adalah 35 lembar jawaban peserta didik kelas XI SMA AI Muslim, sedangkan objek analisis adalah 25 butir soal pilihan ganda materi kecerdasan artifisial beserta kunci jawabannya. Data pendukung berupa kisi-kisi soal, naskah soal, kunci jawaban, dan keluaran program ANATES.

Kisi-kisi instrumen memuat indikator literasi AI, antara lain definisi AI, hubungan AI dan machine learning, peran data, bias algoritma, etika AI, supervised learning, unsupervised learning, train-test split, overfitting, evaluasi akurasi model, natural language processing, computer vision, sistem rekomendasi, privasi data, verifikasi keluaran AI, dan peran manusia dalam pengambilan keputusan. Sebaran ranah kognitif diklasifikasikan menurut taksonomi revisi Anderson dan Krathwohl [15], mulai dari C1 sampai C5.

Analisis data dilakukan menggunakan ANATES versi 4.09 berdasarkan teori tes klasik. Indikator yang dianalisis meliputi: (1) validitas butir melalui korelasi skor butir dengan skor total; (2) reliabilitas tes melalui koefisien belah dua Spearman-Brown; (3) tingkat kesukaran; (4) daya pembeda; dan (5) efektivitas pengecoh. Interpretasi validitas dan reliabilitas mengacu pada prinsip bahwa instrumen harus menghasilkan data yang tepat dan konsisten sebelum digunakan untuk pengambilan keputusan [8].

Agar hasil analisis tidak berhenti pada deskripsi statistik, setiap butir disintesis menjadi rekomendasi tindak lanjut. Butir dikategorikan dipertahankan apabila memiliki validitas memadai, daya pembeda baik, dan pengecoh relatif berfungsi. Butir dikategorikan direvisi apabila masih memiliki kelemahan yang dapat diperbaiki melalui perbaikan redaksi stem, penataan opsi, atau peningkatan level kognitif. Butir dikategorikan ditulis ulang apabila memiliki validitas sangat rendah atau negatif disertai daya pembeda buruk, sehingga berisiko menghasilkan kesalahan pengukuran. Seluruh data dianalisis secara agregat tanpa menampilkan identitas peserta didik.

3. HASIL DAN PEMBAHASAN

3.1. Deskripsi Instrumen dan Sebaran Kognitif

Hasil ANATES menunjukkan skor rata-rata peserta didik sebesar 17,71 dari skor maksimum 25 dengan simpangan baku 4,11. Skor terendah adalah 10 dan skor tertinggi adalah 25. Rentang ini menunjukkan bahwa instrumen dapat dikerjakan oleh sebagian besar peserta didik, tetapi belum cukup kuat untuk membedakan kelompok kemampuan atas secara tajam. Kondisi tersebut berkaitan dengan distribusi level kognitif yang masih didominasi pemahaman dan penerapan dasar.

Tabel 1. Distribusi Level Kognitif pada Instrumen yang Dianalisis

Level kognitif	Nomor butir	Jumlah	Persentase
C1	1	1	4%
C2	2, 3, 4, 5, 6, 9, 10, 11, 16, 17	10	40%
C3	7, 8, 12, 13, 18, 19, 21	7	28%
C4	14, 15, 20, 22, 23, 24	6	24%
C5	25	1	4%
Total	-	25	100%

Tabel 1 memperlihatkan bahwa level C2 mendominasi instrumen sebesar 40%, diikuti C3 sebesar 28% dan C4 sebesar 24%. Hanya satu butir berada pada level C5. Komposisi ini menunjukkan bahwa tes lebih banyak mengukur pemahaman konseptual dan penerapan sederhana daripada evaluasi kritis terhadap kasus AI. Padahal, materi AI menuntut kemampuan untuk menilai risiko bias, mengecek kebenaran keluaran AI, dan mempertimbangkan peran manusia dalam pengambilan keputusan. Oleh karena itu, pengembangan bank soal perlu menambah butir C4 dan C5 berbasis skenario agar instrumen tidak hanya mudah dikerjakan, tetapi juga mampu mengungkap kedalaman berpikir siswa.

3.2. Validitas dan Reliabilitas Butir Soal

Koefisien reliabilitas tes sebesar 0,73 dengan korelasi belah dua 0,57. Nilai tersebut menunjukkan bahwa instrumen memiliki konsistensi internal yang memadai untuk evaluasi kelas. Namun, reliabilitas total tidak otomatis menjamin semua butir berkualitas. Evaluasi pada tingkat item tetap diperlukan karena butir yang lemah dapat tersembunyi dalam instrumen yang secara keseluruhan tampak reliabel.

Tabel 2. Distribusi Kategori Validitas Butir Soal

Kategori	Rentang koefisien	Nomor butir	n (%)
Cukup	0,41-0,70	3, 6, 10, 11, 13, 14, 15, 18, 19, 20, 21, 22, 23	13 (52%)
Rendah	0,21-0,40	1, 2, 4, 8, 9, 17, 25	7 (28%)
Sangat rendah	0,00-0,20	5, 7, 12, 24	4 (16%)
Negatif	< 0,00	16	1 (4%)

Sebanyak 13 butir atau 52% berada pada kategori validitas cukup, sedangkan 12 butir lainnya masih berada pada kategori rendah, sangat rendah, atau negatif. Butir nomor 16 menjadi prioritas perbaikan karena memiliki korelasi negatif. Dalam analisis butir, korelasi negatif dapat mengindikasikan kunci jawaban yang perlu diperiksa ulang, stem yang ambigu, atau opsi jawaban yang tidak berfungsi sesuai tujuan. Temuan ini menguatkan bahwa pengambilan keputusan terhadap bank soal perlu dilakukan pada tingkat butir, bukan hanya berdasarkan reliabilitas total.

3.3. Contoh Sub-Bab Kedua

Tabel 3. Distribusi Tingkat Kesukaran Butir Soal

Kategori	Nomor butir	Jumlah	Persentase
Sangat mudah	4	1	4%
Mudah	2, 5, 6, 7, 8, 9, 10, 11, 12, 18, 20, 24	12	48%
Sedang	1, 3, 13, 14, 15, 16, 17, 19, 21, 22, 23, 25	12	48%
Sukar/Sangat sukar	-	0	0%

Distribusi tingkat kesukaran menunjukkan tidak ada butir yang berada pada kategori sukar atau sangat sukar. Walaupun proporsi butir mudah dan sedang tampak seimbang, ketiadaan butir sukar menyebabkan instrumen kurang optimal untuk memetakan siswa berkemampuan tinggi. Hal ini diduga berkaitan dengan dominasi butir C2 dan konteks soal yang dekat dengan pengalaman sehari-hari, misalnya pengenalan contoh AI atau pilihan tindakan etis yang sangat jelas. Untuk bank soal, diperlukan variasi konteks yang lebih kompleks dan jawaban yang menuntut alasan konseptual, bukan sekadar pengenalan istilah.

3.4. Daya Pembeda dan Efektivitas Pengecoh

Tabel 4. Distribusi Daya Pembeda Butir Soal

Kategori	Nomor butir	Jumlah	Persentase
Sangat baik	14	1	4%
Baik	3, 6, 8, 10, 11, 13, 15, 18, 19, 20, 21, 22, 23, 25	14	56%
Cukup	1, 2, 4, 5, 7, 9, 17	7	28%
Buruk	12, 24	2	8%
Sangat buruk	16	1	4%
Sangat baik	14	1	4%

Daya pembeda menunjukkan 15 butir atau 60% berada pada kategori baik sampai sangat baik. Artinya, sebagian besar soal telah mampu membedakan peserta didik dengan kemampuan relatif tinggi dan rendah. Namun, butir nomor 12, 16, dan 24 perlu menjadi prioritas karena memiliki daya pembeda buruk atau sangat buruk. Pada butir 16, daya pembeda negatif menunjukkan bahwa kelompok bawah justru lebih banyak menjawab benar dibandingkan kelompok atas. Butir seperti ini tidak layak dimasukkan ke bank soal sebelum diperiksa dan ditulis ulang.

Tabel 5. Ringkasan Efektivitas Pengecoh Berdasarkan Keluaran ANATES

Kondisi pengecoh	Nomor butir	Jumlah	Persentase
Semua pengecoh berfungsi baik	3, 4, 5, 13	4	16%
Satu pengecoh kurang baik	6, 7, 10, 15	4	16%
Ada pengecoh buruk/sangat buruk	1, 2, 8, 9, 11, 12, 14, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25	17	68%

Efektivitas pengecoh merupakan kelemahan paling menonjol. Hanya 4 butir atau 16% yang seluruh pengecohnya berfungsi baik, sedangkan 17 butir atau 68% masih memiliki pengecoh buruk atau sangat buruk. Pengecoh yang tidak dipilih siswa, terlalu mudah dieliminasi, atau tidak sejajar dengan kunci dapat meningkatkan peluang jawaban benar karena tebakan. Oleh karena itu, revisi tidak cukup dilakukan pada kunci atau stem, tetapi juga harus mengembangkan pengecoh yang mewakili miskonsepsi umum siswa terhadap AI, misalnya salah membedakan supervised dan unsupervised learning, menganggap akurasi selalu cukup untuk menilai model, atau tidak memahami risiko bias data.

3.5. Sintesis Tindak Lanjut dan Implikasi Pengembangan Bank Soal

Tabel 6. Sintesis Tindak Lanjut Butir Soal

Status	Nomor butir	Jumlah	Persentase
Dipertahankan	3, 6, 10, 13, 15	5	20%
Direvisi	1, 2, 4, 5, 7, 8, 9, 11, 14, 17, 18, 19, 20, 21, 22, 23, 25	17	68%
Ditulis ulang	12, 16, 24	3	12%

Tabel 6 menunjukkan bahwa hanya lima butir yang dapat dipertahankan tanpa revisi substansial, yaitu nomor 3, 6, 10, 13, dan 15. Sebanyak 17 butir perlu direvisi, terutama pada redaksi stem, kesetaraan opsi, dan kualitas pengecoh. Tiga butir, yaitu nomor 12, 16, dan 24, direkomendasikan untuk ditulis ulang karena memiliki kombinasi kelemahan pada lebih dari satu indikator. Pendekatan sintesis ini penting karena kualitas butir tidak seharusnya diputuskan hanya dari satu indeks. Butir yang validitasnya rendah tetapi masih memiliki daya pembeda cukup dapat direvisi, sedangkan butir dengan validitas negatif dan daya pembeda buruk lebih aman ditulis ulang.

Tabel 7. Arah Revisi untuk Butir Prioritas

Butir	Temuan utama	Arah perbaikan yang disarankan
12	Validitas sangat rendah (0,174), daya pembeda buruk (0,11), dan tingkat kesukaran mudah.	Ubah stem menjadi skenario klasifikasi data AI yang menuntut alasan, misalnya membedakan klasifikasi, clustering, dan regresi pada kasus data yang lebih kompleks. Opsi jawaban dibuat setara panjang dan kedekatan maknanya.
16	Validitas negatif (-0,110), daya pembeda sangat buruk (-0,11), serta terdapat pengecoh tidak berfungsi.	Periksa ulang kunci dan redaksi. Butir lebih aman ditulis ulang dengan konteks natural language processing yang eksplisit, satu jawaban benar, dan pengecoh yang merepresentasikan miskonsepsi umum siswa.
24	Validitas sangat rendah (0,120), daya pembeda buruk (0,11), dan tingkat kesukaran mudah.	Naikkan tuntutan kognitif melalui skenario verifikasi bertingkat, misalnya membandingkan keluaran AI dengan dua sumber berbeda dan meminta justifikasi pilihan tindakan.

Implikasi praktis penelitian ini adalah perlunya siklus pengembangan bank soal yang berulang: menyusun kisi-kisi, menulis butir, menguji coba, menganalisis, merevisi, dan menguji kembali. Guru juga perlu menyusun pengecoh berdasarkan miskonsepsi siswa, bukan sekadar membuat opsi yang tampak berbeda. Pada materi AI, butir berkualitas sebaiknya menghadirkan konteks autentik seperti verifikasi jawaban AI generatif, pemilihan data latih dan data uji, penilaian bias, atau penentuan kapan keputusan harus tetap dikendalikan manusia. Strategi ini sejalan dengan orientasi literasi AI yang menempatkan pemahaman teknis, pertimbangan etis, dan human oversight sebagai bagian dari kompetensi belajar.

Keterbatasan penelitian ini terletak pada ukuran sampel yang masih terbatas pada 35 peserta didik dari satu sekolah dan penggunaan teori tes klasik sebagai satu-satunya pendekatan analisis. Karena itu, hasil penelitian lebih tepat digunakan untuk perbaikan instrumen lokal daripada generalisasi luas. Penelitian berikutnya disarankan menguji butir hasil revisi pada sampel yang lebih besar, membandingkan hasil dengan model Rasch atau item response theory, serta menelaah proses respons siswa saat mengerjakan soal AI.

4. KESIMPULAN

Instrumen tes Informatika materi kecerdasan artifisial kelas XI yang dianalisis dalam penelitian ini memiliki reliabilitas memadai dengan koefisien 0,73 sehingga dapat digunakan secara terbatas untuk evaluasi kelas. Namun, kualitas butir pada tingkat item masih bervariasi. Dari 25 butir, 13 butir berkategori validitas cukup, sedangkan sisanya masih rendah, sangat rendah, atau negatif. Tingkat kesukaran didominasi butir mudah dan sedang tanpa adanya butir sukar, sehingga kemampuan instrumen untuk memetakan siswa berkemampuan tinggi belum optimal.

Daya pembeda secara umum cukup baik karena 60% butir berada pada kategori baik sampai sangat baik, tetapi efektivitas pengecoh menjadi kelemahan utama. Sintesis seluruh indikator menunjukkan 5 butir dapat dipertahankan, 17 butir perlu direvisi, dan 3 butir harus ditulis ulang, terutama nomor 12, 16, dan 24. Dengan demikian, instrumen belum ideal sebagai bank soal final. Revisi terarah pada kualitas pengecoh, kejelasan stem, pemeriksaan kunci, dan peningkatan proporsi butir C4-C5 perlu dilakukan sebelum instrumen digunakan secara lebih luas.

DAFTAR PUSTAKA

- [1] Kementerian Pendidikan Dasar dan Menengah, "CP & ATP - Informatika Fase F." Accessed: May 12, 2026. [Online]. Available: <https://guru.kemendikdasmen.go.id/kurikulum/referensi-penerapan/capaian-pembelajaran/sd-sma/informatika/fase-f/>
- [2] D. Long and B. Magerko, "What is AI literacy? Competencies and design considerations," in Proc. 2020 6CHI Conf. Human Factors in Computing Systems, 2020, pp. 1-16, doi: 10.1145/3313831.3376727.
- [3] D. T. K. Ng, J. K. L. Leung, S. K. W. Chu, and M. S. Qiao, "Conceptualizing AI literacy: An exploratory review," *Computers and Education: Artificial Intelligence*, vol. 2, p. 100041, 2021, doi: 10.1016/j.caeai.2021.100041.
- [4] I. H. Y. Yim and J. Su, "Artificial intelligence (AI) learning tools in K-12 education: A scoping review," *Journal of Computers in Education*, vol. 12, pp. 93-131, 2025, doi: 10.1007/s40692-023-00304-9.

-
- [5] Y. Walter, "Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education," *International Journal of Educational Technology in Higher Education*, vol. 21, no. 15, 2024, doi: 10.1186/s41239-024-00448-3.
- [6] F. Miao and W. Holmes, *Guidance for Generative AI in Education and Research*. Paris: UNESCO, 2023, doi: 10.54675/EWZM9535.
- [7] D. Ifenthaler, R. Majumdar, P. Gorissen, M. Judge, S. Mishra, J. Raffaghelli, and A. Shimada, "Artificial intelligence in education: Implications for policymakers, researchers, and practitioners," *Technology, Knowledge and Learning*, vol. 29, no. 4, pp. 1693-1710, 2024, doi: 10.1007/s10758-024-09747-0.
- [8] F. Yusup, "Uji validitas dan reliabilitas instrumen penelitian kuantitatif," *Tarbiyah: Jurnal Ilmiah Kependidikan*, vol. 7, no. 1, pp. 17-23, 2018, doi: 10.18592/tarbiyah.v7i1.2100.
- [9] R. L. Ariany and A. Al-Ghifari, "Penggunaan software ANATES untuk validasi instrumen tes," *Al-Khidmat: Jurnal Ilmiah Pengabdian Kepada Masyarakat*, vol. 1, no. 1, pp. 73-78, 2018, doi: 10.15575/jak.v1i1.3327.
- [10] L. Hamimi, R. Zamharirah, and R. Rusydy, "Analisis butir soal ujian matematika kelas VII semester ganjil tahun pelajaran 2017/2018," *Mathema: Jurnal Pendidikan Matematika*, vol. 2, no. 1, pp. 57-66, 2020, doi: 10.33365/jm.v2i1.459.
- [11] I. Y. Cahyaningrum, A. Fuady, and Sunismi, "Analisis butir soal sumatif akhir semester ganjil mata pelajaran matematika kelas VII dengan berbantuan aplikasi software ANATES," *Mathema: Jurnal Pendidikan Matematika*, vol. 5, no. 2, pp. 67-81, 2023, doi: 10.33365/jm.v5i2.2601.
- [12] S. Mania, F. Fitriani, A. F. Majid, N. N. Ichiana, and A. I. P. Abrar, "Analisis butir soal ujian akhir sekolah," *Al Asma: Journal of Islamic Education*, vol. 2, no. 2, pp. 274-284, 2020, doi: 10.24252/asma.v2i2.16569.
- [13] G. W. S. Pidekso, G. P. S. Ramadhan, M. R. A. Sampurno, M. N. Hafizh, V. Pratiwi, and L. Hakim, "Analisis karakteristik butir soal elemen komputer akuntansi fase F menggunakan software ANATES," *Jurnal Pendidikan Akuntansi Indonesia*, vol. 23, no. 2, pp. 84-95, 2025, doi: 10.21831/jpai.v23i2.91595.
- [14] A. A. Rezigalla et al., "Item analysis: The impact of distractor efficiency on the difficulty index and discrimination power of multiple-choice items," *BMC Medical Education*, vol. 24, no. 445, 2024, doi: 10.1186/s12909-024-05433-y.
- [15] L. W. Anderson and D. R. Krathwohl, Eds., *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman, 2001.