Journal of Innovative and Creativity
Journal Homepage: https://joecy.org/index.php/joecy
e-mail: galungwaknook@gmail.co

Faculty of Education
University of Pahlawan Tuanku Tambusai
ISSN 2775-771X (Online)
ISSN 2962-570X (Printed)
5 (3) 2025, Page: 33362-33372

Analisis Butir Soal Elemen Ekonomi Bisnis dan Administrasi Umum Fase F dengan Menggunakan Aplikasi ANATES Versi 4

Niken Lilisanti Putri Kinasih[1], Erin Erindra Mei Rahma[2], Ardini Dhaniah Arum[3], Unzilla Nafisah Saputri [4], Shalwa Rizqita[5], Aisyah Nur Indriyani Saleh[6], Luqman Hakim [7], Vivi Pratiwi [8]
[1,2,3,4,5,6,7,8]Fakultas Ekonomika dan Bisnis, Universitas Negeri Surabaya
Email: 24080304115mhs.unesa.ac.id, 24080304105@mhs.unesa.ac.id, 24080304153@mhs.unesa.ac.id, 24080304040@mhs.unesa.ac.id, 24080304152@mhs.unesa.ac.id, 24080304113@mhs.unesa.ac.id, luqmanhakim@unesa.ac.id, vivipratiwi@unesa.ac.id

Published: November , 2025

ABSTRAK

Penelitian ini menerapkan metode klasik guna menilai kualitas butir soal pada topik elemen ekonomi bisnis dan administrasi umum, yang memuat aspek validitas, daya pembeda, tingkat kesukaran, reliabilitas, serta efektivitas pengecoh. Pendekatan penelitian yang diterapkan yaitu kuantitatif deskriptif. Pengumpulan data dilakukan lewat peredaran tes kepada peserta didik SMK/MAK yang mengkaji materi ekonomi bisnis dan administrasi umum melibatkan responden sejumlah 20 orang. Data hasil tes dihasilkan melalui observasi dan ditelaah menggunakan aplikasi Anates versi 4. Hasil analisis menyajikan butir soal pada materi tersebut memiliki tingkat validitas yang tinggi, daya pembeda dan tingkat kesukaran yang tergolong cukup baik, reliabilitas yang dinilai telah sangat memadai, serta efektivitas pengecoh yang berada pada kategori kurang baik.

Kata Kunci: Analisis Butir Soal, Anates, Evaluasi, Validitas, SMK Akuntansi

ABSTRACT

This study applies classical methods to assess the quality of test items on the topics of business economics and general administration, which include aspects of validity, discriminating power, difficulty level, reliability, and effectiveness of distractors. The research approach applied is descriptive quantitative. Data collection was conducted by distributing tests to vocational school/MAK students studying business economics and general administration, involving 20 respondents. The test results were obtained through observation and analyzed using the Anates version 4 application. The analysis results show that the test items on this material have a high level of validity, fairly good discriminating power and difficulty level, adequate reliability, and poor distractor effectiveness.

Keyword: *Item Analysis, Anates, Evaluation, Validity, Vocational High School Accounting*

© The Author(s). 2021 Open Access This article is licensed under a Creative Commons Atibution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0

PENDAHULUAN

Evaluasi yaitu tahap krusial dalam kegiatan belajar guna menilai pencapaian tujuan pembelajaran. Menurut (Zuhri, 2024), evaluasi pembelajaran berperan sebagai alat untuk memperoleh informasi yang objektif mengenai kemampuan peserta didik, efektivitas strategi pembelajaran, serta kualitas instrumen penilaian yang digunakan. Evaluasi merupakan tahap penutup selama kegiatan belajar yang dimaksudkan untuk menilai tingkat pencapaian tujuan pembelajaran (Zahiroh & Ritonga, 2021). Evaluasi menjadi komponen integral dalam sistem pendidikan karena tidak hanya menilai hasil belajar, tetapi juga memberikan masukan yang membantu pendidik meningkatkan efektivitas pembelajaran. Menurut (Anshari et al., 2024), kegiatan evaluasi yang baik harus memenuhi prinsip sistematis, objektif, dan berkesinambungan. Artinya, proses penilaian perlu dilakukan secara terencana dan berdasarkan data yang valid agar hasilnya dapat dipercaya. Informasi yang dihasilkan dari evaluasi tidak hanya menggambarkan prestasi peserta didik, tetapi juga menjadi dasar pengambilan keputusan dalam peningkatan mutu pendidikan. Evaluasi berfungsi sebagai alat

kontrol sekaligus pendorong perbaikan dalam dunia pendidikan, dengan demikian perannya sangat penting.

Saat melakukan evaluasi, guru perlu berpedoman pada kurikulum yang digunakan di sekolah. Dalam pelaksanaannya, evaluasi pendidikan menggunakan berbagai instrumen. Tes adalah salah satu produk pembelajaran yang dapat dijadikan alat evaluasi (Arifien & Wijayanto, 2022). Tes merupakan instrumen penilaian yang dimanfaatkan guna mengukur pemahaman peserta didik mengenai materi pelajaran. Menurut (Saputri et al., 2023), tes yang efektif perlu memenuhi standar validitas, reliabilitas, daya pembeda, serta tingkat kesukaran yang proporsional. Tes menjadi metode untuk menilai kompetensi siswa setelah mengikuti proses pembelajaran (Umami & Rusdi, 2021). Maka dari itu, tes digunakan sebagai media penilaian untuk mengukur tingkat pencapaian pembelajaran, yang hasilnya dapat menjadi acuan dalam memperbaiki serta meningkatkan mutu soal (Rahmawati & Rahman, 2025). Salah satu jenis tes yang paling sering diterapkan yaitu tes pilihan ganda, karena efisien dalam pelaksanaan dan penilaiannya. Namun demikian, penyusunan tes pilihan ganda tidak dapat dilakukan secara sembarangan. Menurut (Zuhri, 2024), kualitas tes sangat dipengaruhi oleh mutu butir soal yang disusun. Jadi, setiap butir soal harus dianalisis sebelumnya guna mengerti tingkat kesukaran, daya pembeda, validitas, reliabilitas, serta efektivitas pengecohnya.

Analisis butir soal yaitu aktivitas yang dilaksanakan oleh guru guna mengevaluasi standar setiap soal yang digunakan (Augustia & Agustia, 2025). Melalui proses ini, guru dapat mengidentifikasi soal yang memenuhi standar kualitas dan layak digunakan kembali, serta soal yang perlu diperbaiki atau dihilangkan. Menurut (Anshari et al., 2024), analisis butir soal yang dilakukan dengan metode klasik dapat membantu guru dalam menilai kualitas setiap butir soal, mengidentifikasi kelemahan instrumen tes, dan melakukan perbaikan terhadap soal yang tidak memenuhi kriteria. Hasil evaluasi akan lebih valid dan reliabel, dengan demikian mendukung pengambilan keputusan yang tepat.Dalam perkembangan teknologi saat ini, kegiatan analisis butir soal semakin mudah dilakukan melalui berbagai perangkat lunak, salah satunya Anates. Aplikasi Anates yaitu perangkat lunak yang dimanfaatkan dalam melaksanakan analisis butir soal berlandaskan model klasik, mencakup aspek validitas, reliabilitas, tingkat kesulitan, daya pembeda, serta efektivitas pengecoh (Alista et al., 2023). Menurut (Saputri et al., 2023), penggunaan aplikasi seperti Anates membantu guru memperoleh hasil analisis yang lebih efisien serta meminimalkan kesalahan perhitungan manual. Penelitian ini berfokus pada analisis butir soal pilihan ganda mata pelajaran Ekonomi Bisnis dan Administrasi Umum bagi peserta didik SMK/MAK yang mengkaji materi ekonomi bisnis dan administrasi umum. Melewati penerapan aplikasi Anates, penelitian ini bertujuan guna menilai kualitas soal mengacu pada aspek validitas, reliabilitas, tingkat kesukaran, daya pembeda, serta efektivitas pengecoh. Dari hasil analisis dimaksudkan guna menyajikan gambaran tentang sejauh mana soal-soal tersebut tergolong valid sebagai instrumen evaluasi serta menjadi dasar perbaikan dalam penyusunan soal pada periode berikutnya.

METODE PENELITIAN

Penelitian ini dilaksanakan di Program Studi Pendidikan Akuntansi, Jurusan Pendidikan Ekonomi, Fakultas Ekonomika dan Bisnis Universitas Negeri Surabaya pada semester ganjil tahun akademik 2024/2025. Subjeknya yaitu mahasiswa Program S1 Pendidikan Akuntansi yang sedang menempuh mata kuliah Evaluasi Belajar dan Pembelajaran, serta siswa SMK/MAK yang mengkaji materi ekonomi bisnis dan administrasi umum melewati soal yang disusun oleh mahasiswa tersebut. Fokus studi adalah analisis butir soal, meliputi aspek validitas, daya pembeda, tingkat kesukaran, reliabilitas, serta efektivitas pengecoh. Studi tersebut dilaksanakan dengan metode kuantitatif yang berorientasi pada pendekatan deskriptif. Pendekatan ini dipilih guna mengumpulkan data kuantitatif yang mampu menggambarkan kualitas setiap butir soal. Metode kuantitatif dipilih karena data yang dikumpulkan berbentuk numerik yang akan dianalisis menggunakan perangkat lunak Anates V.4.09 dan dianalisis berdasarkan kriteria tertentu sebelum ditarik kesimpulan (Riani et al., 2020).

Soal yang dianalisis dalam penelitian ini merupakan hasil karya mahasiswa yang mengikuti mata kuliah Evaluasi Belajar dan Pembelajaran. Sampel penelitian terdiri atas 20 peserta didik kelas

XI SMK/MAK yang mempelajari materi ekonomi bisnis dan administrasi umum, yang dipilih menggunakan teknik *simple random sampling*. Analisis dilakukan terhadap 54 butir soal pilihan ganda yang dikembangkan berdasarkan indikator kompetensi dalam Kurikulum Merdeka, dengan memanfaatkan perangkat lunak Anates versi 4. Proses analisis diawali dengan tahap persiapan, yang meliputi penyusunan butir soal untuk keperluan analisis, perancangan dan instalasi perangkat lunak Anates versi 4, serta pengaturan pembagian tugas dan materi kerja dalam kelompok. Selanjutnya, pada tahap pelaksanaan, soal yang telah disusun dibagikan kepada siswa kelas XI SMK/MAK dan juga diuji coba kepada mahasiswa peserta mata kuliah. Setelah itu, jawaban siswa dikoreksi dan dinilai, kemudian data hasil tes dimasukkan ke dalam program Anates versi 4 untuk dianalisis lebih lanjut.

Validitas

Validitas menjadi indikator dasar guna mengonfirmasi bahwa soal yang disusun sepenuhnya selaras dengan tujuan pembelajaran. Apabila suatu soal dinyatakan valid, berarti isi soal tersebut telah sesuai dengan keterampilan yang ingin diukur. Hasil uji validitas biasanya diklasifikasikan menjadi valid atau tidak valid berlandaskan nilai koefisien korelasi. Semakin tinggi nilai korelasinya, semakin valid butir soal tersebut. Validitas dapat dinilai salah satunya melalui membandingkan skor total hasil tes siswa mengacu pada skor tiap butir soal. Hubungan atau korelasi antara keduanya menunjukkan tingkat validitas suatu butir soal. Setiap soal dapat dinilai apakah layak digunakan atau perlu direvisi, dengan demikian proses evaluasi menjadi lebih terarah.

Daya Pembeda

Indikator ini menunjukkan perbedaan tingkat capaian dalam memilih jawaban antara kedua kelompok tersebut. Indeks diskriminasi, yang juga disebut daya pembeda (D), merupakan nilai yang menggambarkan perbedaan tingkat kemampuan peserta (Alpusari, 2014). Indeks ini berperan penting dalam analisis butir soal untuk memastikan keadilan pengukuran kemampuan siswa, mengoptimalkan dalam menentukan pilihan terkait perbaikan soal, sekaligus memastikan hasil evaluasi pembelajaran memiliki tingkat kepercayaan dan akurasi yang tinggi. Apabila nilai indeks daya pembeda menunjukkan angka 0, hal tersebut terjadi karena jawaban benar dari kelompok atas serta kelompok bawah menunjukkan proporsi yang setara. Nilai daya pembeda (D) rentang nilainya dari -1 sampai +1, namun hanya nilai positif yang bermakna. Semakin mengarah menuju angka 1, semakin baik potensi butir soal tersebut dalam mengategorikan peserta didik berkemampuan tinggi serta rendah.

Tingkat Kesukaran

Tingkat kesukaran ditetapkan berdasarkan persentase peserta didik yang memberikan jawaban tepat (Hendrayadi et al., 2024). Soal yang sesuai mempunyai taraf kesukaran yang proporsional dan tidak terlalu rendah tingkat kesukarannya sehingga semua peserta dapat menjawab, serta tidak terlalu tinggi, hanya sebagian kecil yang berhasil menjawabnya (Bano et al., 2022). Maka dari itu, soal yang baik mempunyai komposisi yang setara antara taraf kesulitan mudah, sedang, dan sulit dengan perbedaan ideal 3:5:2 dan 3:4:3. Keseimbangan dalam penyusunan soal menjadi kunci untuk memperoleh hasil evaluasi yang adil (Fiska et al., 2021). Indeks tingkat kesulitan soal berada pada rentang 0,00 hingga 1,00. Nilai yang hampir 0 menggambarkan bahwa soal tergolong sangat sukar, adapun nilai yang hampir mencapai 1 mengidikasikan bahwa soal tersebut sangat mudah.

Reliabilitas

Reliabilitas berfungsi guna menilai konsistensi suatu tes dalam mengukur aspek yang relevan dievaluasi. Sementara itu, (Farida & Musyarofah, 2021) berpendapat jika *output* pengukuran dianggap dinyatakan reliabel apabila menghasilkan nilai yang konsisten melalui pengukuran berulang pada kelompok subjek yang sama selama karakteristik yang dinilai tidak mengalami perubahan. Sebuah

alat ukur dikatakan andal apabila digunakan berulang terhadap gejala yang sama yang menghasilkan hasil yang stabil. Artinya, reliabilitas menekankan pada stabilitas hasil pengukuran terhadap objek yang identik.

Tabel 1. Kriteria koefisien Reliabilitas Butir Soal

Indeks Reliabilitas (r)	Tingkat Reliabilitas
$0{,}80 < r \leq 1{,}00$	Sangat tinggi
$0{,}60 < r \leq 0{,}80$	Tinggi
$0{,}40 < r \leq 0{,}60$	Sedang
$0{,}20 < r \leq 0{,}40$	Rendah
$0{,}00 < r \leq 0{,}20$	Sangat rendah

Sumber :(YUSUF, 2024)

Instrumen Cronbach Alpha digunakan untuk menilai tingkat reliabilitas suatu alat ukur. Berdasarkan Sugiyono (2017), suatu alat ukur dinyatakan reliabel apabila koefisien Cronbach Alpha ≥ 0,6. Dalam penelitian ini, batas koefisien reliabilitas ditetapkan sebesar 0,6 dengan 2 standar, ketika nilai reliabilitas berada > 0,6, alat ukur dinilai cukup stabil serta layak digunakan. Apabila nilai reliabilitas < 0,6, maka instrumen tersebut belum memenuhi standar konsistensi yang diperlukan..

Efektivitas Pengecoh

Melaksanakan pengolahan data efektivitas pengecoh (*distractor*) atau distribusi pilihan jawaban yang dipilih oleh peserta pada tiap butir soal.. Menurut (Qadir et al., 2024) pengecoh (*distractor*), yang juga disebut sebagai alternatif jawaban penyesat, merupakan pilihan jawaban yang disajikan dalam suatu tes namun tidak berfungsi sebagai kunci jawaban yang benar. Ketika efektivitas pengecoh dalam suatu butir soal rendah, maka indeks daya pembeda cenderung rendah serta taraf kesukaran soal menurun. Namun, jika pengecoh berfungsi secara optimal, maka daya pembeda meningkat dan soal menjadi lebih sulit (Akhmadi, 2021). Penggunaan pengecoh dalam butir soal bertujuan guna mengidentifikasi perbedaan terhadap tingkat kemampuan siswa. Apabila suatu pilihan jawaban diberi kode (-) untuk kurang baik, (+) untuk baik, serta (++) untuk sangat baik, berarti pengecoh tersebut dinilai berfungsi secara optimal. Namun, jika diberi tanda (--) yang menunjukkan buruk, atau (---) yang menunjukkan sangat buruk, berarti pengecoh tersebut dinilai tidak berfungsi. Dalam hal ini, akan muncul penanda khusus (blok) yang mengindikasikan bahwa pengecoh tersebut tidak bekerja sebagaimana mestinya (R. M. Sari & Surtani, 2020).

HASIL DAN PEMBAHASAN

Pendidik memiliki peran penting dalam memastikan kualitas soal yang digunakan dalam pembelajaran. Untuk itu, analisis menyeluruh terhadap setiap komponen soal perlu dilakukan, mulai dari tahap penyusunan, penggunaan, revisi, hingga evaluasi. Telaah soal bertujuan menilai secara mendalam aspek-aspek butir soal sebelum digunakan dalam ujian, seperti kesesuaian isi, struktur, dan bahasa (Nuradinafika et al., 2025).

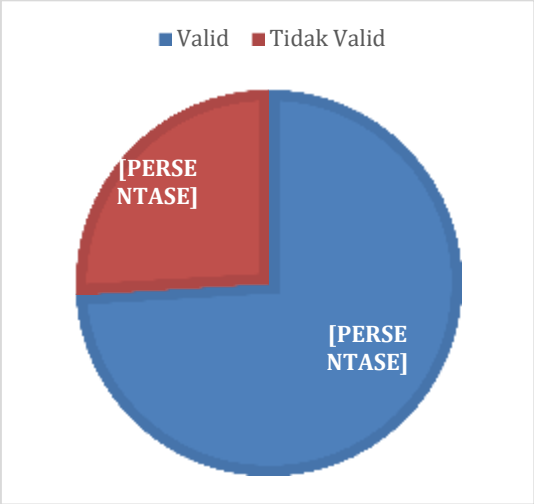
Melalui pendekatan kuantitatif deskriptif, penelitian ini membuktikan bahwa pendidik mampu menganalisis soal secara menyeluruh, baik dari aspek kualitatif maupun kuantitatif. Analisis ini mencakup pengujian validitas, daya pembeda, tingkat kesukaran, efektivitas pengecoh, serta reliabilitas. Dengan analisis yang tepat, soal mampu menjadi instrumen pengukuran yang lebih presisi dan mendukung pencapaian tujuan pembelajaran.

Validitas

Tes dikatakan valid apabila hasil pengukurannya benar-benar mencerminkan kondisi atau karakteristik sebenarnya dari objek yang diukur, selaras dengan sasaran pengukuran yang ditentukan

(Ramadhan et al., 2024). Butir dengan korelasi tinggi menunjukkan kontribusi yang baik terhadap tes secara keseluruhan, sedangkan butir dengan korelasi rendah mungkin perlu diperbaiki atau dihapus agar kualitas tes tetap terjaga.

Gambar 1. Diagram Validitas



Sumber: data diolah (2025)

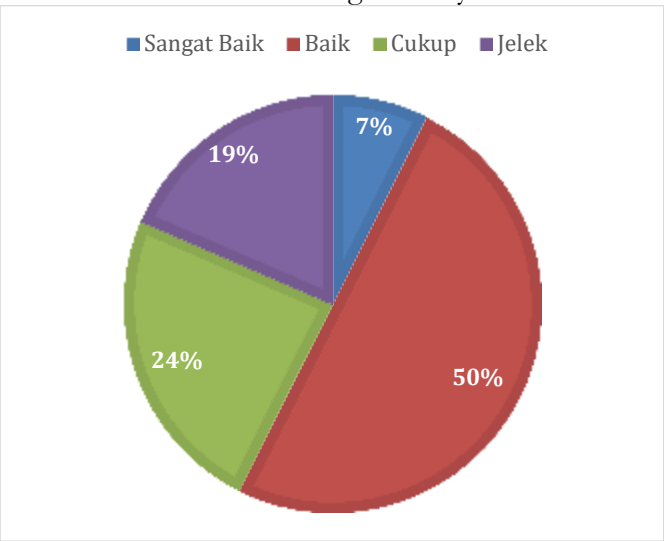
Berdasarkan hasil analisis terhadap 54 butir soal, ditemukan bahwa sebanyak 33 butir termasuk dalam klasifikasi sangat signifikan, 7 butir signifikan, 9 butir tidak signifikan, dan 5 butir berada pada klasifikasi NAN (Not a Number). Kategori NAN menunjukkan bahwa koefisien korelasi untuk butir tersebut tidak dapat dihitung karena nilai korelasinya tidak terbaca atau gagal diproses oleh sistem, sehingga ditampilkan sebagai NAN. Jika diklasifikasikan berdasarkan kategori validitas, maka diperoleh distribusi sebagai berikut: sangat signifikan sebanyak 33 soal (61%), signifikan sebanyak 7 soal (13%), tidak signifikan sebanyak 9 soal (17%), dan NAN sebanyak 5 soal (9%). Temuan ini menunjukkan bahwa mayoritas butir soal memiliki korelasi yang signifikan terhadap skor keseluruhan, sehingga dapat dianggap valid dan mampu mengukur konstruk yang dimaksud. Butir-butir yang tergolong tidak signifikan maupun NAN perlu ditinjau kembali, baik dari segi isi materi, konstruksi soal, maupun pilihan jawabannya, agar kualitas instrumen dapat ditingkatkan secara menyeluruh.

Daya Pembeda

Soal yang memiliki daya pembeda tinggi menunjukkan efektivitasnya dalam mengidentifikasi perbedaan kemampuan peserta didik, sehingga dapat meningkatkan validitas dan reliabilitas keseluruhan tes (Saputri et al., 2023). Berdasarkan hasil analisis menggunakan aplikasi ANATES terhadap 54 butir soal, diperoleh empat kategori daya pembeda, yakni sangat baik, baik, cukup, serta jelek. Klasifikasi tersebut membantu peneliti maupun pendidik dalam mengevaluasi kualitas butir soal serta menentukan butir mana yang layak digunakan kembali, direvisi, atau dibuang (Pradita et al., 2023).

Butir dengan daya pembeda tinggi menandakan bahwa soal tersebut mampu mengukur kemampuan siswa secara adil, adapun butir dengan daya pembeda rendah bahkan negatif sebaiknya diperbaiki karena tidak dapat membedakan siswa dengan baik (Rismawati et al., 2024). Maka dari itu, analisis daya pembeda menjadi langkah krusial guna mengonfirmasi bahwa instrumen evaluasi yang diterapkan sanggup mengukur kemampuan peserta didik dengan objektif serta akurat.

Gambar 2. Diagram Daya Pembeda



Sumber: data diolah (2025)

Hasil analisis mencerminkan bahwa kebanyakan soal mempunyai daya pembeda dalam kategori baik sampai sangat baik, yang artinya soal-soal ini mampu berkerja dengan baik dalam membedakan kompetensi siswa. Secara rinci, terdapat 4 butir soal dengan klasifikasi sangat baik (DP ≥ 0,70), yakni nomor 32, 40, 45, dan 53. Kemudian terdapat 27 butir soal dalam kategori baik (0,40 ≤ DP < 0,70), di antaranya nomor 4, 5, 6, 7, 8, 9, 24, 26, 28, 29, 30, 33, 34, 35, 36, 37, 38, 39, 41, 42, 43, 44, 47, 48, 49, 50, 51 dan 52. Sementara itu, 13 butir soal termasuk kategori cukup (0,20 ≤ DP < 0,40), ada di nomor 3, 11, 12, 14, 15, 20, 21, 22, 23, 25, 31, dan 54. Adapun 10 butir soal masuk kategori jelek (DP < 0,20), yaitu nomor 1, 2, 10,13, 16, 17, 18, 19, 27 dan 46, dengan satu di antaranya (nomor 13) memiliki nilai daya pembeda negatif dianggap tidak valid serta dan sebaiknya direvisi atau dibuang.

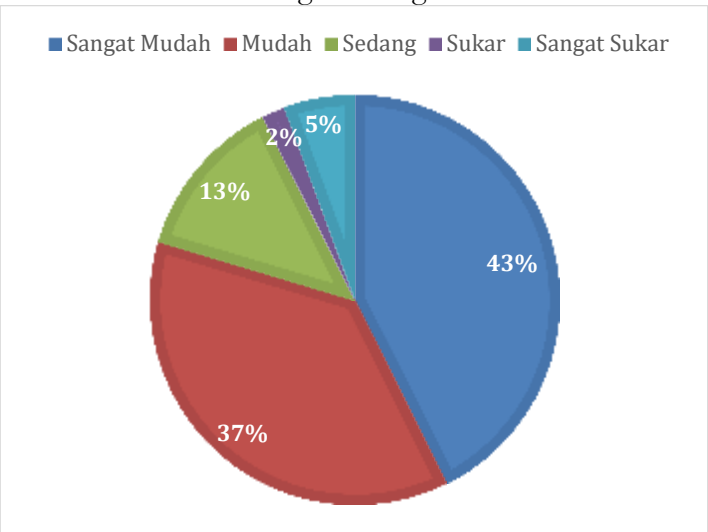
Dilihat dari semua hasil, terlihat bahwa mayoritas butir soal sudah sesuai standar yang baik, karena mampu membedakan secara efektif antara kelompok siswa berkemampuan tinggi serta rendah. Akan tetapi, sebagian butir soal dengan daya pembeda rendah dan negatif sebaiknya mendapat perhatian agar dilakukan revisi sehingga kualitas tes dapat meningkat. Dapat disimpulkan bahwa instrumen tes tersebut tergolong baik serta layak digunakan, dengan beberapa perbaikan kecil pada instrumen yang memiliki daya pembeda kurang optimal.

Tingkat Kesukaran

Semakin tinggi persentase peserta yang memilih opsi benar, semakin mudah suatu soal, di sisi lain, semakin rendah persentasenya, semakin sulit pula soal tersebut (Dia & Syah, 2022). Berdasarkan analisis terhadap 54 butir soal menggunakan aplikasi Anates, diperoleh distribusi tingkat kesukaran yang beragam. Aplikasi ini mengelompokkan soal menjadi 5 kelompok: sangat mudah, mudah, sedang, sukar, serta sangat sukar, yang membantu guru atau peneliti dalam menentukan butir yang layak dipertahankan, direvisi, atau dibuang (Damayanti et al., 2020).

Berdasarkan hasil analisis menggunakan aplikasi Anates, diperoleh data mengenai tingkat kesukaran dari 54 butir soal pada mata pelajaran Ekonomi Bisnis dan Administrasi Umum. Taraf kesukaran soal mencerminkan seberapa besar kemungkinan siswa dapat memberikan jawaban yang benar. Soal dengan nilai kesukaran tinggi cenderung lebih mudah diselesaikan, sedangkan nilai yang rendah menunjukkan soal semakin sulit.

Gambar 3. Diagram Tingkat Kesukaran



Sumber: data diolah (2025)

Dari analisis, diketahui bahwa mayoritas butir soal masuk dalam kategori sangat mudah dan mudah. Sebanyak 23 butir soal (42,59%) termasuk dalam kategori sangat mudah, di antaranya nomor 1, 2, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 25, 26, 28, dan 54. Selain itu, terdapat 20 butir soal (37,04%) yang diklasifikasikan sebagai soal mudah, yaitu nomor 3, 23, 29, 30, 31, 32, 33, 35, 37, 38, 39, 40, 41, 42, 43, 44, 47, 48, 49 dan 52. Sementara itu, 7 butir soal (12,96%) berada dalam kategori sedang, mencakup nomor 8, 24, 36, 45, 50, 51 dan 53. Hanya 1 butir soal (1,85%), yaitu nomor 4, yang tergolong sukar, dan 3 butir soal (5,56%) termasuk dalam kategori sangat sukar, yaitu nomor 27, 34, dan 46.

Sebanyak 41 butir soal, atau sekitar 79,68% dari total soal yang dianalisis, termasuk dalam kategori sangat mudah dan mudah. Hal ini menunjukkan bahwa sebagian besar soal tergolong ringan bagi peserta didik dan tidak menimbulkan kesulitan berarti dalam proses pengerjaan. Karakteristik soal yang demikian dapat mencerminkan tingkat kompleksitas yang rendah atau kesesuaian yang tinggi dengan kemampuan siswa. Total tersebut mengindikasikan bahwa perlu dilakukan peninjauan ulang terhadap tingkat kesukaran soal agar distribusinya lebih seimbang dan mampu mengukur kemampuan peserta didik secara lebih komprehensif. Meskipun hal ini mencerminkan pemahaman siswa yang baik, dominasi soal yang terlalu mudah mampu mengurangi keterampilan tes dalam membedakan tingkat kemampuan peserta. Agar hasil evaluasi mencerminkan kemampuan secara lebih menyeluruh, instrumen ke depan perlu disusun secara lebih seimbang, dengan perbaikan pada soal-soal yang terlalu mudah dan penambahan soal yang lebih menantang.

Reliabilitas

Analisis terhadap butir soal pada Elemen Ekonomi Bisnis dan Administrasi Umum menggunakan program Anates menunjukkan nilai reliabilitas sebesar 0,97, yang tergolong dalam golongan sangat tinggi berdasarkan kisaran 0,80 hingga 1,00.

Tabel 2. Hasil Reliabilitas

Rata-rata	41,85
Simpangan Baku	8,19
Korelasi XY	0,94
Reliabilitas tes	0,97

Sumber : data diolah (2025)

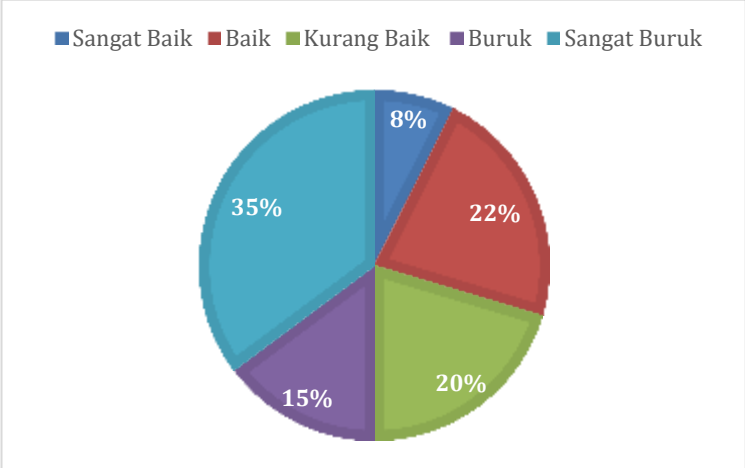
Dari hasil Anates versi 4 diperoleh nilai reliabilitas sebesar 0,97, yang menandakan tes memiliki tingkat konsistensi internal yang sangat tinggi. Nilai reliabilitas yang hampir mendekati angka satu ini mengindikasikan bahwa butir soal dalam instrumen mempunyai keterkaitan yang kuat dalam mengukur kemampuan yang sama secara konsisten. Menurut (Karindi & Rufi'i, 2024), nilai reliabilitas di atas 0,80 termasuk dalam kategori sangat tinggi dan mengungkapkan bahwa alat ukur tersebut mampu memberikan hasil yang akurat dalam menilai kompetensi siswa. Hasil ini juga sejalan dengan temuan (Subhaktiyasa, 2024) yang menyatakan bahwa instrumen dengan reliabilitas tinggi menggambarkan stabilitas hasil pengukuran terhadap konstruk yang diuji. Tes ini dapat dikatakan mencerminkan mutu yang memadai serta dapat dimanfaatkan secara efektif dalam proses evaluasi pembelajaran.

Namun demikian, nilai reliabilitas yang sangat tinggi perlu ditafsirkan dengan hati-hati. (Mawardi et al., 2023) menjelaskan bahwa reliabilitas yang terlalu tinggi dapat menandakan adanya kesamaan atau pengulangan makna antarbutir (*redundansi item*) sehingga cakupan konten tes menjadi sempit. - Pemeriksaan ulang terhadap variasi tingkat kesukaran dan daya pembeda butir perlu dilakukan untuk memastikan bahwa seluruh indikator kompetensi telah terwakili secara proporsional. Selain itu, (Mardiyah, 2024) menambahkan bahwa faktor seperti jumlah butir soal, kualitas pengecoh, serta keragaman peserta tes juga mempengaruhi besarnya nilai reliabilitas.

Efektivitas Pengecoh

Melalui pemanfaatan aplikasi Anates versi 4, analisis terhadap butir soal dalam elemen Ekonomi Bisnis dan Administrasi Umum mengungkapkan bahwa pengecoh dikategorikan ke dalam lima tingkat efektivitas: sangat buruk, kurang efektif, cukup, baik, serta sangat baik. Melalui aplikasi ini, jumlah peserta didik yang memilih setiap opsi jawaban dapat langsung terlihat, namun informasi mengenai identitas individu maupun kelompok peserta yang memilih masing-masing pilihan tidak tersedia. Hasil analisis efektivitas pengecoh terhadap 54 butir soal memperlihatkan mayoritas *distraktor* berfungsi secara baik, meskipun terdapat beberapa yang kurang efektif karena jarang dipilih oleh peserta didik. Hasil tersebut menandakan bahwa pengecoh belum berhasil memengaruhi peserta didik dengan kompetensi yang rendah, sehingga diperlukan revisi agar dapat meningkatkan daya pembeda butir soal.

Gambar 4. Diagram Efektivitas Pengecoh



Sumber : data diolah (2025)

Dari hasil analisis memanfaatkan Anates *for Windows* menunjukkan bahwa kualitas pengecoh sangat bervariasi, mulai dari kategori sangat baik (++), baik (+), kurang baik (–), hingga sangat buruk (---). Sebagian besar pengecoh berfungsi secara baik, namun ada beberapa yang jarang dipilih peserta didik sehingga dinilai tidak efektif. Pengecoh yang kurang berfungsi ini tidak mampu

menarik peserta berkemampuan rendah, sehingga dapat menurunkan daya pembeda serta validitas butir soal.

Sebagaimana ditunjukkan oleh hasil analisis kualitas *distractor* butir soal, dapat diidentifikasi beberapa pola karakteristik pengecoh. Pertama, ditemukan sejumlah pengecoh dengan angka 0 yang menunjukkan bahwa opsi tersebut tidak ada siswa yang memilih, seperti pada soal nomor 1 opsi A, C, D, serta E, serta nomor 2 opsi B, C, D, dan E. Kedua, pengecoh dengan kategori "---" (Sangat Buruk) menunjukkan ketidakefektifan dalam menarik perhatian siswa, contohnya pada soal nomor 4 opsi E dan nomor 5 opsi A. Ketiga, *distractor* dengan kategori "++" (Sangat Baik) terbukti efektif mempengaruhi siswa yang tidak menguasai materi, seperti terlihat pada nomor 29 opsi A dan D serta 30 opsi B dan E. Di sisi lain, *distractor* dengan kategori "+" (Baik) juga berhasil menjalankan fungsinya dengan cukup baik dalam menarik pilihan siswa yang kurang memahami konsep, contohnya dapat ditemukan pada soal nomor 8 opsi A serta nomor 36 opsi B, D, dan E. Sebaliknya, *distractor* dengan kategori "--" (Buruk) menunjukkan kinerja yang tidak optimal karena gagal menjadi alternatif pilihan yang menantang, misalnya pada soal nomor 6 opsi B dan C. *Distractor* dengan kategori "-" (Kurang Baik) meskipun sedikit lebih baik dari kategori buruk, tetap belum berfungsi optimal, contohnya pada soal nomor 8 opsi C serta nomor 24 opsi B dan D

Menurut (Trianingsih, 2023), pengecoh yang efektif adalah alternatif jawaban yang berhasil menarik peserta didik yang kurang dalam penguasaan materi, sehingga mampu membedakan tingkat kemampuan peserta secara lebih jelas. Sementara itu, (Rismawati et al., 2024) menegaskan bahwa pengecoh yang tidak berfungsi akan menurunkan reliabilitas tes karena tidak mendukung proses pengukuran secara objektif. Sejalan dengan temuan tersebut, (Ayu & Realita, 2024) serta (O. P. D. Sari et al., 2023) mengungkapkan bahwa pengecoh yang tidak pernah dipilih menunjukkan rendahnya kualitas butir soal karena gagal menyesatkan peserta berkemampuan rendah. Revisi terhadap pengecoh yang tergolong "kurang baik" atau "sangat buruk" perlu dilakukan dengan memperhatikan kejelasan redaksi, kelogisan pilihan jawaban, serta keseimbangan panjang kalimat antar opsi.

KESIMPULAN

Dari hasil penelitian melalui aplikasi Anates versi 4, dapat disimpulkan bahwasanya kualitas Butir Soal Elemen Ekonomi Bisnis dan Administrasi Umum Fase F menunjukkan *output* yang beragam. Secara keseluruhan, instrumen tes ini memiliki validitas yang tergolong tinggi dengan 74,07% butir soal dinyatakan valid, serta reliabilitas yang sangat tinggi dengan nilai 0,97 yang menunjukkan konsistensi pengukuran yang *excellent*. Dari segi daya pembeda, lebih dari separuh soal telah memenuhi kriteria baik hingga sangat baik sehingga mampu membedakan kemampuan peserta didik secara efektif. Namun demikian, temuan penelitian mengungkapkan bahwa banyak soal memiliki taraf kesukaran yang terlalu mudah, dimana hampir 80% soal tergolong dalam kategori mudah dan sangat mudah, sehingga kurang optimal dalam mengukur variasi kemampuan peserta didik. Di sisi lain, efektivitas pengecoh pada banyak butir soal masih belum berfungsi optimal, dimana banyak pengecoh yang tidak berhasil mengecoh peserta didik yang kesulitan mengerti materi.

Melalui temuan penelitian yang diperoleh, peneliti mengajukan beberapa saran konstruktif. Bagi pendidik dan penyusun soal, sangat disarankan untuk melakukan revisi mendalam terhadap instrumen tes, khususnya pada butir-butir soal yang tidak valid, mempunyai daya pembeda rendah, dan efektivitas pengecoh yang tidak optimal. Perbaikan dapat difokuskan pada penyusunan pengecoh yang lebih logis dan menantang, serta menyesuaikan tingkat kesukaran soal agar lebih seimbang dan proporsional, tidak didominasi oleh kategori mudah. Bagi peneliti selanjutnya, hasil ini dapat dijadikan pijakan untuk kajian yang lebih mendalam dengan memperluas cakupan sampel, mengombinasikan analisis kuantitatif dengan telaah kualitatif, atau bahkan mengaplikasikan pendekatan teori tes modern seperti *Item Response Theory* (IRT) untuk mendapatkan pemahaman yang lebih komprehensif tentang kualitas butir soal. Terakhir, bagi lembaga pendidikan, disarankan untuk secara berkelanjutan memfasilitasi pelatihan dan sosialisasi penggunaan *software* analisis butir soal,

contohnya ANATES, guna mengoptimalkan kapasitas pendidik dalam menyusun instrumen evaluasi yang valid, reliabel, dan berkualitas tinggi, yang pada akhirnya akan berkontribusi signifikan terhadap peningkatan mutu proses dan hasil pembelajaran.

DAFTAR PUSTAKA

Akhmadi, M. N. (2021). Analisis Butir Soal Evaluasi Tema 1 Kelas 4 SDN Plumbungan Menggunakan Program Anates. *Ed-Humanistics: Jurnal Ilmu Pendidikan*, *6*(1), 799–806.

Alista, Y. F., Syahzanani, R. A., Alista, Y. F., Syahzanani, R. A., Maret, U. S., & Maret, U. S. (2023). Analisis Butir Soal Ulangan Harian Fisika. *SIPTEK : Seminar Nasional Inovasi dan Pengembangan Teknologi Pendidikan*, *1*(1), 1–11.

Alpusari, M. (2014). 106 Analisis Butir Soal, Program Komputer Anates Versi 4.0 For Windows Mahmud Alpusari. *Jurnal Primary Program Studi Pendidikan Guru Sekolah Dasar Fakultas Keguruan dan Ilmu Pendidikan Universitas Riau*, *3*, 106–115.

Anshari, M. I., Nasution, R., Irsyad, M., Alifa, A. Z., & Zuhriyah, I. A. (2024). Analisis Validitas dan Reliabilitas Butir Soal Sumatif Akhir Semester Ganjil Mata Pelajaran PAI. *Edukatif : Jurnal Ilmu Pendidikan Vol*, *6*(1), 965–977. https://doi.org/https://doi.org/10.31004/edukatif.v6i1.5931

Arifien, M., & Wijayanto, P. A. (2022). Analisis Kualitas Butir Soal Tengah Semester Geografi Sma Islam Al Azhar 14 Kota Semarang. *Geodika: Jurnal Kajian Ilmu dan Pendidikan Geografi*, *5*(2), 206–215. https://doi.org/10.29408/geodika.v5i2.3696

Augustia, A. D., & Agustia, C. N. (2025). Analisis Validitas dan Reliabilitas Soal Pilihan Ganda dengan Menggunakan Software Anates pada Mata Pelajaran Perpajakan. *AKADEMIK Jurnal Mahasiswa Ekonomi & Bisnis*, *5*(1), 250–265. https://doi.org/https://doi.org/10.37481/jmeb.v5i1.1165

Ayu, D., & Realita. (2024). Analisis Butir Soal pada Mata Pelajaran Pendidikan Agama Islam di Kelas V SDN 3 Tanjung Pura. *Alacrity : Journal Of Education*, *4*(3), 140–151. https://doi.org/https://doi.org/10.52121/alacrity.v4i3.418

Bano, V. O., Marambaawang, D. N., & Njoeroemana, Y. (2022). Analisis Kriteria Butir Soal Ujian Sekolah Mata Pelajaran Ipa Di Smp Negeri 1 Waingapu. *JURNAL IDEAS Pendidikan, Sosial, dan Busaya*, *8*(1), 145–152. https://doi.org/10.32884/ideas.v8i1.660

Damayanti, W. W., Halidjah, S., & Pranata, R. (2020). Analisis Tingkat Kesukaran Butir Soal Pilihan Ganda pada Penilaian Tengah Semester Kelas IV. *JURNAL PENDIDIKAN DAN PEMBELAJARAN KHATULISTIWA*, *10*(11). https://doi.org/https://doi.org/10.26418/jppk.v10i11.50458

Dia, E. E., & Syah, M. A. S. (2022). Tingkat Kesukaran Soal Penilaian Tengah Semester Mata Pelajaran Bahasa Indonesia di Madrasah. *Fourth Conference on Research and Community Services STKIP PGRI Jombang*.

Farida, & Musyarofah, A. (2021). Validitas dan Reliabilitas dalam Analisis Butir Soal. *Al-Mu'arrib: Jurnal Pendidikan Bahasa Arab*, *I*(1), 34–44. https://doi.org/https://doi.org/10.32923/al-muarrib.v1i1.2100

Fiska, J. M., Hidayati, Y., Qomaria, N., Hadi, W. P., Pendidikan, F. I., Trunojoyo, U., & Bangkalan, M. (2021). Analisis Butir Soal Ulangan Harian Ipa Menggunakan Software Anates pada Pendekatan Teori Tes Klasik. *Jurnal Natural Science Educational Research*, *4*(1).

Hendrayadi, H., Kustati, M., & Amelia, R. (2024). Analisis Ulangan Harian Mata Pelajaran PAI di SMA Negeri 10 Padang Tahun Pelajaran 2023/2024 (Telaah Terhadap Reliabilitas, Daya Beda Dan Tingkat Kesukaran Menggunakan Software Anates). *Jurnal Review Pendidikan dan Pengajaran (JRPP)*, *7*(3), 6954–6961.

Karindi, D. R., & Rufi'i. (2024). Penggunaan Anates 4.0.9 Dalam Menguji Kelayakan Soal Pilihan Ganda Sumatif Matematika. *Linear: Journal of Mathematics Education*, *5*(2). https://doi.org/https://doi.org/10.32332/2zg4f753

Mardiyah, H. (2024). Penggunaan Aplikasi Anates pada Analisis Butir Soal Penilaian Akhir Jenjang. *WIDYALOKA*, *11*(2). https://jurnal.ikipwidyadarmasurabaya.ac.id/index.php/widyaloka/article/view/180

Mawardi, M. S., Fuady, A., & Sunismi. (2023). Analisis Butir Soal Pilihan Ganda Menggunakan Anates pada Penilaian Tengah Semester Kelas VII D SMP Negeri 1 Ngajum Kabupaten Malang. *Wahana : Tridarma Perguruan Tingg*, *75*(1). https://doi.org/https://doi.org/10.36456/wahana.v75i1.6820

Nuradinafika, A., Afriyanti, N. A., Hakim, L., & Pratiwi, V. (2025). Analisis Butir Soal Elemen Ekonomi Bisnis dan Administrasi Umum Menggunakan Aplikasi Anates Versi 4. *Educational Journal: General and Specific Research*, *5*(1), 45–63.

Pradita, E., Megawanti, P., & Yulianingsih. (2023). Analisis Tingkat Kesukaran , Daya Pembeda , dan Fungsi Distraktor PTS Matematika SMPN Jakarta. *HIMPUNAN: JURNAL ILMIAH MAHASISWA PENDIDIKAN MATEMATIKA*, *3*(80), 109–118.

Qadir, A., Huda, N., & Hermina, D. (2024). Analisis Butir Tes : Tingkat Kesukaran, Daya Pembeda dan

Efektivitas Pengecoh. *Al Furqan : Jurnal Agama, Sosial, dan Budaya*, *3*(3), 1450–1467. https://publisherqu.com/index.php/Al-Furqan

Rahmawati, N. S., & Rahman, M. F. (2025). Analisis Butir Soal Objektif Penilaian Sumatif Akhir Semester Gasal Mata Pelajaran Ekonomi. *JIIP (Jurnal Ilmiah Ilmu Pendidikan)*, *8*(8), 8863–8873. https://doi.org/https://doi.org/10.54371/jiip.v8i8.8837

Ramadhan, M. F., Siroj, R. A., & Afgani, M. W. (2024). Validitas and Reliabilitas. *Journal on Education*, *06*(02), 10967–10975.

Riani, D., Almujab, S., Dina, A., & Budiarto, R. (2020). Analisis Butir Soal dan Kemampuan Siswa dalam Menjawab Soal Ujian Nasional pada Mata Pelajaran Ekonomi. *Oikos: Jurnal Kajian Pendidikan Ekonomi dan Ilmu Ekonomi*, *IV*(1), 70–79. https://doi.org/https://doi.org/10.23969/oikos.v4i1.2425

Rismawati, Mukhlisa, N., & Lukman. (2024). Analisis Butir Soal Penilaian Akhir Semester (Pas) Ganjil Tahun Ajaran 2023/2024 pada Mata Pelajaran Matematika Upt Sd Negeri 228 Pinrang. *Pendas : Jurnal Ilmiah Pendidikan Dasar*, *09*(3).

Saputri, H. A., Zulhijrah, Larasati, N. J., & Shaleh. (2023). Analisis Instrumen Assesmen : Validitas, Reliabilitas, Tingkat Kesukaran dan Daya Beda Butir Soal. *Didaktik : Jurnal Ilmiah PGSD FKIP Universitas Mandiri*, *09*(5), 2986–2995. https://doi.org/https://doi.org/10.36989/didaktik.v9i5.2268

Sari, O. P. D., Sunismi, & Fuady, A. (2023). Analisis Instrumen Soal Ujian Akhir Semester (Uas) Mata Pelajaran Matematika Peminatan pada Soal Pilihan Ganda. *JUMLAHKU: Jurnal Matematika Ilmiah Universitas Muhammadiyah Kuningan*, *9*(1), 1–12. https://doi.org/https://doi.org/10.33222/jumlahku.v9i1.2718

Sari, R. M., & Surtani. (2020). Analisis Kualitas Butir Soal Ujian Akhir Semester Genap pada Mata Pelajaran Geografi Kelas X IIS Tahun Ajaran 2018/2019 Di Sma Pertiwi 1 Padang. *JURNAL BUANA JURUSAN GEOGRAFI FAKULTAS ILMU SOSIAL – UNP*, *4*(2). https://doi.org/https://doi.org/10.24036/buana.v4i2.830

Subhaktiyasa, P. G. (2024). Evaluasi Validitas dan Reliabilitas Instrumen Penelitian Kuantitatif : Sebuah Studi Pustaka. *Journal of Education Research*, *5*(4), 5599–5609.

Trianingsih, L. (2023). Analisis Kualitas Tes dan Butir Soal Integral pada Evaluasi Formatif Matematika Teknik. *IJCEE Vol.9*, *9*(2), 44–57.

Umami, R., & Rusdi, M. (2021). Pengembangan Instrumen Tes untuk Mengukur Higher Order Thinking Skills (HOTS) Berorientasi Programme For International Student Asessment (PISA) pada Peserta Didik. *Jurnal Penelitian Pendidikan dan Pengajaran Matematika*, *7*(1), 57–68. https://doi.org/10.37058/jp3m.v7i1.2069

YUSUF, F. W. (2024). Analisis Butir Soal Asesmen Sumatif Biologi Materi Perubahan Lingkungan Dengan Menggunakan Anates Pada Kelas X SMA. *LEARNING : Jurnal Inovasi Penelitian Pendidikan dan Pembelajaran*, *4*(1), 126–135. https://doi.org/https://doi.org/10.51878/learning.v4i1.2777

Zahiroh, U., & Ritonga, P. S. (2021). *Analisis Kualitas Butir Soal Pilihan Ganda Mata Pelajaran Kimia pada Ujian Akhir Semester (UAS) Kelas XI MAN 2 Kepulauan Meranti* (Vol. 3, Nomor 1). http://repository.uin-suska.ac.id/id/eprint/28622

Zuhri, N. Z. (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*, 693–704.