

## Penggunaan ANATES 4.0.9 dalam menguji kelayakan soal pilihan ganda sumatif matematika

### Use of ANATES 4.0.9 in testing the feasibility of mathematics summative multiple choice questions

Dito Risky Karindi<sup>1)\*</sup>, Ruffi'i<sup>2)</sup>

<sup>1)2)</sup> Sekolah Pascasarjana, Universitas PGRI Adi Buana, Surabaya, Indonesia

\*ditos2tepunipasby@gmail.com

Diterima: 16 Oktober 2024 | Direvisi: 4 November 2024 | Dipublikasi: 24 Desember 2024

#### Abstrak

Guru harus mampu mempersiapkan dan melaksanakan kegiatan penilaian dengan benar, serta menggunakan hasil penilaian sesuai dengan keperluan. Tujuan penelitian ini adalah mengevaluasi kelayakan soal sumatif matematika pilihan ganda materi fungsi komposisi dan invers fungsi berbasis HOTS dan kontekstual. Penelitian ini menggunakan pendekatan deskriptif kuantitatif. Sebanyak 72 siswa kelas XI dari salah satu SMA Negeri di Kota Surabaya menjadi subjek penelitian ini pada semester ganjil tahun ajaran 2024/2025. Data dikumpulkan melalui tes pilihan ganda dengan 25 soal. Data dianalisis dengan *software* ANATES 4.0.9. Hasil penelitian terkait butir soal yaitu sebanyak 56% valid dan 44% tidak valid; reliabilitas tes mencapai 0,71, yang memiliki kategori tinggi; 16% soal memiliki daya pembeda sangat bagus, 48% bagus, 24% cukup bagus, dan 12% buruk; 68% soal memiliki tingkat kesukaran sedang, 20% mudah, dan 12% sangat mudah. Hasil penelitian menyatakan bahwa 40% memiliki kualitas soal yang sangat baik, sehingga layak digunakan pada tes sumatif yang akan datang atau dimasukkan ke dalam bank soal. Sementara 60% harus dievaluasi terlebih dahulu oleh peneliti. Ini akan membuat pengembangan bank soal lebih baik dan akan dapat digunakan untuk melakukan berbagai evaluasi pembelajaran.

**Kata Kunci:** ANATES, Butir Soal, Kelayakan Soal

#### Abstract

*A Teachers must be able to prepare and carry out assessment activities correctly, and use the results of the assessment as needed. The purpose of this study was to evaluate the feasibility of multiple-choice summative math questions on HOTS and contextual-based composition and inverse function material. This research used a quantitative descriptive approach. A total of 72 grade XI students from one of the public high schools in Surabaya City were the subjects of this study in the odd semester of the 2024/2025 academic year. Data were collected through a multiple choice test with 25 questions. The data were analyzed with ANATES 4.0.9 software. The results of research related to question items are 56% valid and 44% invalid; test reliability reached 0.71, which has a high category; 16% of questions have very good differentiating power, 48% are good, 24% are quite good, and 12% are bad; 68% of questions have moderate difficulty, 20% are easy, and 12% are very easy. The results of the study state that 40% have very good question quality, so they are suitable for use in future summative tests or are included in the question bank. While 60% must be evaluated first by the researcher. This will make the development of the question bank better and will be able to be used to conduct various learning evaluations.*

**Keywords:** ANATES, Question Items, Question Worthiness

Published by [Linear: Journal of Mathematics Education](#)

This is an open access article under the [CC BY SA](#) license



## **PENDAHULUAN**

Matematika sangat penting untuk perkembangan IPTEK, dan berkat bantuan matematika, IPTEK berkembang pesat. Matematika adalah ilmu yang sangat dekat dengan aktivitas manusia dan berasal dari peristiwa yang benar-benar terjadi pada manusia. Ilmu ini kemudian menerapkan untuk membuat peristiwa tersebut lebih mudah dipelajari dan dipahami oleh orang lain. Selain itu, logika manusia membentuk matematika karena ide-ide matematika dapat diperoleh dari aktivitas berpikir (Usman et al., 2022). Matematika adalah salah satu mata pelajaran yang sangat penting dalam pendidikan, seperti yang ditunjukkan oleh jumlah waktu yang dihabiskan guru untuk mengajar matematika dibandingkan dengan mata pelajaran lain. Matematika diajarkan di semua jenjang pendidikan, dari pendidikan dasar hingga menengah (Sardia et al., 2021).

Keberhasilan proses pembelajaran matematika dipengaruhi oleh faktor-faktor seperti siswa, guru, fasilitas, metode dan sistem penilaian yang saling bergantung satu sama lain. Proses pembelajaran matematika terkait materi dianggap berhasil apabila hasil yang diperoleh memenuhi tujuan pembelajaran matematika (Munauwarah et al., 2022). Tingkat keberhasilan proses belajar siswa dinilai melalui kegiatan tes atau ujian. Guru harus mampu mempersiapkan dan melaksanakan kegiatan penilaian dengan tepat dan menggunakan hasil penilaian sesuai dengan keperluannya. Guru dapat melihat tingkat penguasaan materi pembelajaran yang diberikan serta kesulitan yang dihadapi siswa melalui kegiatan penilaian. Kemampuan yang harus dikuasai oleh pendidik yaitu kemampuan pedagogik. Kompetensi pedagogik adalah kemampuan untuk memahami siswa, merencanakan dan melaksanakan pembelajaran, menilai hasil belajar, dan mengembangkan siswa untuk memaksimalkan potensi mereka. Pendidik juga memiliki kemampuan untuk mengevaluasi apakah tes atau ujian yang mereka buat cukup untuk meningkatkan kemampuan siswa (Santosa & Badawi, 2022).

Hasil belajar matematika diperoleh setelah serangkaian kegiatan proses belajar mengajar selesai dilaksanakan. Kemampuan seorang siswa dalam belajar matematika sangat bergantung pada kualitas tes yang diambil. Sebab, tes dianggap baik pasti menghasilkan informasi baik juga dan sebaliknya, informasi tentang kemampuan matematika siswa akan menjadi biasa-biasa saja jika tes tersebut tidak dilaksanakan dengan baik. Untuk dapat mengerjakan ujian dengan baik, tentunya harus mematuhi prinsip-prinsip pelaksanaan ujian, khususnya mulai dari menyusun kumpulan soal hingga menganalisis soal (Lukman, 2023).

Evaluasi pembelajaran adalah proses mengukur nilai dan manfaat pembelajaran (Alti et al., 2023). Evaluasi juga merupakan cara untuk menilai seberapa baik siswa mencapai tujuan

program. Dianggap sebagai sarana penting untuk mencapai tujuan pembelajaran adalah penilaian. Proses evaluasi membantu guru menilai kemampuan dan pemahaman siswa tentang materi pelajaran (Usman et al., 2022). Proses evaluasi juga membantu guru mengevaluasi kesesuaian metode pengajaran yang digunakan dan mengukur keberhasilan siswa dalam mencapai tujuan pembelajaran (Ashayidatina & Alber, 2023). Evaluasi adalah giat yang terencana bertujuan untuk mencari tahu tentang kondisi suatu objek menggunakan instrumen yang kemudian digunakan untuk membandingkan hasil tersebut sebagai tolak ukur membuat kesimpulan (Zein & Darto, 2012).

Kesuksesan belajar dapat diukur dari ketuntasan belajar siswa secara klasikal di dalam kelas. Sehingga dapat dikatakan bahwa semakin besar persentase ketuntasan belajar di dalam kelas, maka pembelajaran guru dapat dikatakan berhasil. Sebaliknya semakin kecil persentase ketuntasan belajar di dalam kelas, maka pembelajaran yang dilaksanakan guru dalam kelas berarti gagal. Ketuntasan belajar sering dilihat dari hasil tes atau sumatif yang dilakukan guru setelah menyelesaikan serangkaian materi yang diajar di dalam kelas. Namun, seringkali guru memiliki kekeliruan terhadap penilaian tersebut sehingga menyebabkan peserta didik mendapatkan nilai yang rendah atau bahkan tinggi. Nilai rendah atau tinggi yang diperoleh peserta didik dipengaruhi oleh kualitas soal yang dibuat guru dalam melaksanakan tes atau sumatif dalam kelas. Semakin sulit soal, semakin banyak siswa yang mendapatkan nilai rendah. Sebaliknya, semakin mudah soal, semakin banyak siswa yang mendapatkan nilai tinggi.

Evaluasi analitis terhadap soal tes sangat penting dalam menentukan kualitas soal. Sehingga dapat dikatakan kualitas soal menjadi tolak ukur dalam layak atau tidaknya soal tes digunakan untuk kegiatan sumatif selanjutnya (Sardia et al., 2021). Faktor-faktor seperti reliabilitas, validitas, daya pembeda, tingkat kesukaran, dan kualitas pengecoh dapat digunakan untuk menilai kualitas butir soal yang layak atau tidak digunakan. Ada kemungkinan bahwa pengujian ini akan digunakan untuk membuat kesimpulan apakah soal tersebut layak digunakan atau direvisi dan akan digunakan kembali serta dibuang atau tidak digunakan dalam tes sumatif berikutnya. Dalam melakukan pengukuran, pendidik membutuhkan alat uji seperti tes. Untuk memastikan hasil pengukuran sesuai dengan yang diharapkan, maka sangat penting untuk menggunakan tes yang tepat. (Hodiyanto, 2017). Perhatian dasar pengembang tes ketika membuat tes adalah sifat dan kualitas soal tes serta bagaimana peserta ujian merespons soal tersebut. Validitas dan reliabilitas tes apa pun pada akhirnya bergantung pada karakteristik item soal. Karakteristik tersebut adalah tingkat kesulitan item dan item daya beda. Teori tes

memprediksikan hasil tes dengan mengidentifikasi parameter kesukaran soal, daya beda soal, dan kemampuan peserta tes (O. A. & E. R. I., 2016).

Guru dapat menggunakan dua jenis tes yaitu tes formatif dan tes sumatif. Tes formatif diberikan selama proses pembelajaran dan mengevaluasi apa yang telah dipelajari siswa. Tes sumatif diberikan pada akhir pelajaran dan mencakup lebih dari satu bab atau unit materi. Siswa akan menjalani ujian akhir setelah menyelesaikan kursus mereka (Nurfiqah et al., 2015). Analisis butir soal sangat penting untuk menilai kualitas setiap aspek dari seperangkat soal. Analisis butir soal yang dapat dilakukan secara kualitatif atau kuantitatif memiliki tujuan utama untuk mendapatkan informasi tentang karakteristik setiap butir soal melalui telaah dan analisis empiris. Melalui analisis hasil ujian, hasil dimanfaatkan untuk menilai kualitas soal dan hasil belajar siswa (Elviana, 2020). Guru jarang melakukan kegiatan analisis butir soal karena beberapa faktor. Misalnya, soal tidak sepenuhnya menunjukkan tingkat kompetensi yang dimiliki siswa. Faktor lain adalah bahwa kegiatan ini dianggap rumit dan memakan waktu yang cukup lama (Hidayati & Nisa', 2023).

Alat penilaian berupa soal tes harus melalui tahap analisis bentuk soal untuk menghasilkan soal tes yang layak digunakan sebagai alat penilaian (Munauwarah et al., 2022). Tes yang berkualitas tinggi adalah tes yang mengukur kemampuan dan keberhasilan siswa secara memadai dan efektif. Kualitas soal ujian dapat diukur dengan melakukan kegiatan analisis butir soal (AR et al., 2024). Dalam melakukan evaluasi tes sumatif dapat berupa soal pilihan ganda atau uraian. Namun, dalam kenyataannya dalam melakukan penilaian soal tes pilihan ganda lebih mudah dilakukan karena penilai cukup melihat kunci jawaban yang telah tersedia. Sedangkan untuk soal uraian, penilai terkadang masih terpengaruh subjektivitas dalam melakukan penilaian. Dalam menganalisis butir soal pun soal tes pilihan ganda lebih mudah dilakukan. Setiap soal pilihan ganda yang dibuat oleh guru untuk tes sumatif harus dianalisis sesuai dengan pedoman pembuatan soal yang baik agar soal tersebut dapat mengukur tingkat pengetahuan siswa sesuai dengan standar kompetensi yang telah ditetapkan. Hal ini dilakukan untuk memastikan apakah soal yang dibuat oleh guru layak dan baik, dan untuk memberikan hasil yang optimal untuk mengukur dan meningkatkan pemahaman siswa (Zahiroh & Soleman Ritonga, 2021).

ANATES adalah *software* yang digunakan untuk menilai soal pilihan ganda berdasarkan daya pembeda, kualitas pengecoh, validitas, tingkat kesukaran, dan reliabilitas. Anates adalah salah satu *software* komputer yang dirancang untuk menganalisis pertanyaan. Program ini sangat bermanfaat untuk guru dan pemerhati evaluasi pendidikan (Cahyaningrum et al., 2023).

Ini adalah aplikasi untuk menganalisis soal pilihan ganda dan uraian. Anates mempunyai kemampuan menganalisis tes untuk: melatih keterampilan secara sistematis, terukur, berkesinambungan, dan menyeluruh pada setiap item pertanyaan yang dimasukkan. Ini termasuk menghitung skor (asli atau tertimbang); menghitung tingkat kesulitan soal, mengidentifikasi pengidentifikasi kinerja, membagi subjek ke dalam kelompok yang lebih tinggi atau lebih rendah, dan menghitung korelasi antara skor butir soal dan skor keseluruhan (Merlin Purwati et al., 2021).

Penelitian ini dilaksanakan dalam pembelajaran matematika pada kelas XI. Evaluasi sumatif dilaksanakan pada materi fungsi komposisi dan invers fungsi dengan menggunakan soal pilihan ganda sebanyak 25 soal dan sebanyak 5 butir soal tersebut mengarah ke soal *Higher Order Thinking Skills* (HOTS) yang dapat mengukur kemampuan berfikir tingkat tinggi peserta didik. Selain itu, *novelty* dalam penelitian ini memberikan temuan terkait kemudahan peneliti dalam melakukan pengujian kelayakan soal sumatif pilihan ganda pada materi fungsi komposisi dan invers fungsi dapat menggunakan *software* ANATES 4.0.9 yang terdiri dari aspek reliabilitas, validitas, kualitas pengecoh, daya pembeda dan tingkat kesukaran yang sebelumnya belum ada penelitian terkait hal tersebut.

Berdasarkan latar belakang tersebut, tujuan penelitian ini adalah mengevaluasi kelayakan soal sumatif matematika pilihan ganda materi fungsi komposisi dan invers fungsi berbasis HOTS dan kontekstual yang diberikan kepada siswa kelas XI dari salah satu SMA Negeri di kota Surabaya pada semester ganjil tahun ajaran 2024/2025. Selain itu, untuk mengetahui apakah bagian soal yang digunakan dalam sumatif matematika tersebut layak digunakan untuk sumatif berikutnya, sehingga peneliti dapat membuat kesimpulan setelah menganalisis butir soal.

## **METODE PENELITIAN**

Jenis Penelitian ini adalah penelitian deskriptif kuantitatif. Tujuan penelitian deskriptif adalah untuk menjelaskan fakta, gejala, fenomena, dan peristiwa yang sedang terjadi. Data yang digunakan dalam penelitian kuantitatif ini terdiri dari angka yang telah dianalisis oleh peneliti menggunakan kriteria metode ilmiah yang konkrit, logis, terukur, objektif, dan sistematis (Alti et al., 2023).

Penelitian ini dilakukan saat semester ganjil tahun pelajaran 2024/2025 pada 72 siswa kelas XI pada salah satu SMA Negeri di Kota Surabaya yang terdiri dari 36 siswa kelas XI-10 dan 36 siswa kelas XI-11. Untuk mengumpulkan data, peneliti menggunakan soal matematika

sumatif tentang materi fungsi komposisi dan invers fungsi. Soal ini terdiri dari 25 butir soal pilihan ganda dengan 5 pilihan jawaban, yaitu A, B, C, D, dan E. Peneliti menganalisis validitas, daya pembeda, kualitas pengecoh, tingkat kesukaran, dan reliabilitas masing-masing butir soal menggunakan program ANATES 4.0.9.

## **HASIL DAN PEMBAHASAN**

Berdasarkan analisis kuantitatif yang dilakukan saat semester ganjil tahun pelajaran 2024/2025 di kelas XI pada salah satu SMA Negeri di Kota Surabaya yang melibatkan analisis soal sumatif matematika berbentuk pilihan ganda materi fungsi komposisi dan invers fungsi dengan 72 siswa yang menjadi responden serta 25 butir soal, diperoleh kesimpulan tentang validitas, daya pembeda, kualitas pengecoh, tingkat kesukaran, reliabilitas.

### **1. Validitas dan Reliabilitas Tes**

Validitas yang tinggi terjadi ketika penilaian setiap elemen pertanyaan sesuai dengan skor keseluruhan. Menganalisis reliabilitas item kuesioner sangatlah penting. Hal ini karena analisis ini menentukan tingkat konsistensi pertanyaan, yang memungkinkan Anda menentukan apakah pertanyaan tersebut dapat diandalkan (AR et al., 2024).

Dalam menentukan validitas dan reliabilitas suatu tes, peneliti menggunakan kriteria validitas butir soal yakni.

**Tabel 1. Kriteria Validitas Butir Soal**

<b>Nilai r</b>	<b>Keterangan</b>
0,91 – 1,00	Sangat Valid
0,71 – 0,90	Valid
0,41 – 0,70	Cukup Valid
0,21 – 0,40	Kurang Valid
Negatif – 0,20	Sangat Kurang Valid

Sumber: (Aquatika et al., 2022)

Sedangkan dalam menentukan reliabilitas tes, peneliti menggunakan kriteria reliabilitas tes yakni.

**Tabel 2. Kriteria Reliabilitas**

<b>Nilai r</b>	<b>Keterangan</b>
$\geq 0,70$	Tinggi
$< 0,70$	Rendah

Sumber: (Sardia et al., 2021)

Item soal dianggap valid jika skornya berkorelasi positif dengan skor total. Software ANATES dapat digunakan untuk mengevaluasi validitas item berdasarkan hasil korelasi antara skor item dan skor total. Butir soal dianggap valid jika hasilnya signifikan dan sebaliknya. Semua soal yang dinyatakan valid harus dipelihara dalam database soal (Khasanah et al., 2023).

Pembiasaan guru dengan teknik persiapan pertanyaan akan membantu memperbaiki pertanyaan yang tidak valid. Pertanyaan disusun dengan baik dan memiliki konten yang benar-benar mewakili tujuan pengukuran, yang meningkatkan kemungkinan validitasnya (Muluki et al., 2020).

Hasil analisis validitas dan reliabilitas pada pengujian butir soal menggunakan ANATES 4.0.9 didistribusikan sebagai berikut.

**Tabel 3. Hasil Validitas Butir Soal**

No	Koefisien Validitas	Validitas	Nomor Soal	Banyak Soal	Prosentase
1	0,55	Valid	1, 5, 6, 7, 9, 11, 12, 13, 14, 16, 17, 18, 19, 22	14	56%
2		Tidak Valid	2, 3, 4, 8, 10, 15, 20, 21, 23, 24, 25	11	44%

Sumber: Diolah oleh Peneliti

**Tabel 4. Hasil Reliabilitas Soal**

Komponen	Hasil ANATES	Keterangan
Rata-rata	15,81	-
Simpangan Baku	4,90	-
KorelasiXY	0,55	Cukup Valid
Reliabilitas Tes	0,71	Reliabilitas Sangat Tinggi

Sumber: Software ANATES 4.0.9

Hasil dari penelitian terkait validitas soal sumatif matematika berbentuk pilihan ganda materi fungsi komposisi dan invers fungsi di kelas XI pada salah satu SMA Negeri di Kota Surabaya menunjukkan 56% dari jumlah soal dinyatakan valid serta 44% dari jumlah soal dinyatakan tidak valid. Hasil tersebut hampir sama dengan penelitian Khasanah et al. yaitu Validitas soal cukup baik karena terdapat 8 butir soal (53%) dalam kategori valid, dan 7 butir soal (47%) tidak valid (Khasanah et al., 2023). Soal valid tersebut merupakan hal yang tepat dilakukan karena dapat dikatakan mengukur apa yang telah diukur. Validitas adalah tingkat ketepatan instrumen untuk mengukur objek yang ingin diukur (Antari et al., 2021). Butir soal yang dikatakan valid yaitu 1, 5, 6, 7, 9, 11, 12, 13, 14, 16, 17, 18, 19, 22 dapat dipertahankan untuk tes sumatif berikutnya. Sedangkan yang tidak valid 2, 3, 4, 8, 10, 15, 20, 21, 23, 24, 25 dapat diubah.

Reliabilitas didefinisikan sebagai sejauh mana suatu item pertanyaan dapat menghasilkan hasil yang sama dalam berbagai percobaan. Dengan kata lain, reliabilitas mengacu pada konsistensi atau stabilitas hasil dari waktu ke waktu atau antar penilai (Warju et al., 2020). Hasil penelitian menunjukkan soal mempunyai nilai reliabilitas sangat tinggi, yang berarti bahwa jika soal tersebut diujikan pada subjek lain, nilainya akan tetap sama. Dengan demikian, dapat

disimpulkan bahwa soal sumatif matematika berbentuk pilihan ganda materi fungsi komposisi dan invers fungsi memiliki reliabilitas yang tinggi. Hal tersebut hampir sama dengan hasil penelitian Khasanah et al. yaitu reliabilitas soal dalam kategori tinggi karena koefisien reliabilitas mencapai 0,72 yang berarti lebih dari sama dengan 0,70 (Khasanah et al., 2023). Namun, hal tersebut berbanding terbalik dengan hasil penelitian Friatma et al. yaitu reliabilitas hasil tes yang diperoleh adalah 0,46. Angka reliabilitas ini menunjukkan bahwa soal tersebut termasuk dalam kategori cukup, dan soal tersebut reliabel. Kriteria reliabilitas yang cukup berarti butir soal yang digunakan cukup konsisten dalam mengukur pengetahuan siswa (Friatma & Anhar, 2019).

## 2. Daya Pembeda

Tujuan dari analisis karakteristik item pertanyaan adalah untuk memperjelas kemampuan item pertanyaan yang diajukan dan membedakan siswa dengan kemampuan rendah dan tinggi. Adapun kriteria daya pembeda (D) yakni

**Tabel 5. Kriteria Daya Pembeda**

Nilai D	Keterangan
$0,70 < D \leq 1,00$	Sangat Bagus
$0,40 < D \leq 0,70$	Bagus
$0,20 < D \leq 0,40$	Cukup Bagus
$0,00 < D \leq 0,20$	Buruk
$D \leq 0,00$	Sangat Buruk

Sumber: (AR et al., 2024)

Hasil analisis daya pembeda dengan program ANATES 4.0.9 didistribusikan yakni.

**Tabel 6. Hasil Daya Pembeda Butir Soal**

No	Kategori Daya Pembeda	Nomor Soal	Banyak Soal	Prosentase
1	Sangat Bagus	5, 11, 12, 19	4	16%
2	Bagus	1, 6, 7, 9, 13, 14, 16, 17, 18, 21, 22, 23	12	48%
3	Cukup Bagus	2, 4, 8, 15, 20, 24	6	24%
4	Buruk	3, 10, 25	3	12%

Sumber: Diolah oleh Peneliti

Kemampuan untuk membedakan siswa yang memiliki kemampuan yang tinggi dibanding yang rendah dikenal sebagai distraktor (Muluki et al., 2020). Koefisien daya distraktor setiap butir pertanyaan berkorelasi positif dengan kualitasnya dalam membedakan siswa berkemampuan tinggi dibanding rendah (Sardia et al., 2021). Pada ANATES dalam pembagian kelompok tinggi (unggul) dan kelompok rendah (asor) berlaku pengambilan sampel 27% dari masing-masing kelompok tersebut. Sehingga pendistribusiannya pada kelompok unggul sebanyak 19 siswa dan kelompok asor sebanyak 19 siswa.



Hasil penelitian terkait daya pembeda soal sumatif matematika berbentuk pilihan ganda materi fungsi komposisi dan invers fungsi di kelas XI pada salah satu SMA Negeri di Kota Surabaya menunjukkan Soal ini memiliki daya pembeda yang dikategorikan 16% sangat bagus, 48% bagus, 24% cukup bagus, dan 12% buruk. Hal tersebut hampir sama dengan hasil penelitian AR et al. yaitu daya pembeda butir soal ulangan akhir semester (UAS) ganjil mata pelajaran matematika siswa kelas X SMAN 1 Majene tahun ajaran 2021/2022 diperoleh bahwa sebanyak 10 butir soal (50%) berada pada kategori baik, sebanyak 8 butir soal (40%) berada pada kategori cukup baik dan terdapat 2 butir soal (10%) berada pada kategori jelek (AR et al., 2024).

Soal-soal dengan daya pembeda yang sangat bagus termasuk nomor 5, 11, 12, 19 dan yang bagus termasuk nomor 1, 6, 7, 9, 13, 14, 16, 17, 18, 21, 22, 23 menunjukkan kemampuan untuk membedakan tingkat kecerdasan siswa dengan baik, sehingga soal-soal tersebut dapat dikumpulkan dalam kumpulan soal dan digunakan untuk sumatif selanjutnya. Soal-soal dengan daya pembeda yang cukup bagus termasuk nomor 2, 4, 8, 15, 20, 24 dan cukup menunjukkan kemampuan untuk membedakan tingkat kemampuan siswa; namun, soal-soal ini harus diperiksa ulang dan diperbaiki agar memiliki daya pembeda yang baik untuk sumatif berikutnya. Soal-soal dengan daya pembeda yang buruk termasuk nomor 3, 10, 25 dan tidak menunjukkan kemampuan untuk membedakan tingkat kemampuan siswa.

Sejalan dengan teori bahwa koefisien daya pembeda yang lebih tinggi menunjukkan kemampuan suatu soal untuk membedakan siswa yang berkemampuan tinggi dengan rendah. Dengan kata lain, jika semua atau sebagian besar siswa yang cerdas dapat menjawab soal dengan benar, maka koefisien daya pembeda soal tersebut akan lebih tinggi (AR et al., 2024).

### **3. Tingkat Kesukaran**

Simbol P singkatan dari kata "Proporsi," digunakan untuk menggambar indeks kesukaran ini untuk evaluasi (Rezki Amalia et al., 2021). Tingkat kesulitan yang diberikan oleh software ANATES 4.0.9 diklasifikasikan dalam lima kategori: sangat mudah, mudah, sedang, sukar, dan sangat sukar (Khasanah et al., 2023).

Tingkat kesulitan terkait dengan berapa banyak atau setidaknya siswa menjawab pertanyaan menjawab soal dengan benar. Tingkat kesukaran soal dalam tes dilihat dari kemampuan siswa dalam menjawab soal tersebut, tidak berdasarkan asumsi guru yang menyusun soal, karena soal yang sulit atau mudah bagi guru belum tentu sulit atau mudah bagi siswa (Manfaat et al., 2021).

Adapun kriteria tingkat kesulitan untuk masing-masing item soal yakni.

**Tabel 7. Kriteria Tingkat Kesukaran**

Nilai P	Keterangan
$0,00 < P \leq 0,30$	Sulit
$0,30 < P \leq 0,70$	Sedang
$0,70 < P \leq 1,00$	Mudah

Sumber: (Zahiroh & Soleman Ritonga, 2021)

Hasil analisis tingkat kesukaran dengan program software ANATES 4.0.9 didistribusikan yakni.

**Tabel 8. Hasil Tingkat Kesukaran Butir Soal**

No	Kategori Tingkat Kesukaran	Nomor Soal	Banyak Soal	Prosentase
1	Sedang	3, 6, 7, 11, 12, 13, 14, 15, 16, 18, 19, 20, 21, 22, 23, 24, 25	17	68%
2	Mudah	1, 2, 5, 9, 17	5	20%
3	Sangat Mudah	4, 8, 10	3	12%

Sumber: Diolah oleh Peneliti

Tingkat kesukaran setiap butir tes hasil belajar menentukan kualitasnya. Soal dapat dianggap baik jika tidak terlalu sulit atau terlalu mudah (Muluki et al., 2020). Hasil penelitian terkait daya pembeda soal sumatif matematika berbentuk pilihan ganda materi fungsi komposisi dan invers fungsi di kelas XI pada salah satu SMA Negeri di Kota Surabaya menunjukkan soal sumatif dominan dengan tingkat kesukaran sedang sebanyak 68%, yaitu nomor butir 3, 6, 7, 11, 12, 13, 14, 15, 16, 18, 19, 20, 21, 22, 23, 24, 25, dan dianggap memiliki kualitas yang baik. Soal dengan tingkat kesukaran mudah sebanyak 20% dan 12% memiliki tingkat kesukaran yang sangat mudah. Hal tersebut berbanding terbalik dengan hasil penelitian dari Indah Sari et al. bahwa sejumlah 50% butir soal berkategori sangat mudah, 40% berkategori mudah, dan 10% berkategori sedang. Hasil analisis tersebut juga menggambarkan bahwa tidak ada soal dalam kategori sulit (Indah Sari et al., 2022). Soal-soal yang lebih mudah memiliki lebih banyak siswa yang mampu menjawab dengan benar, sementara soal-soal yang lebih sulit memiliki lebih sedikit siswa yang mampu menjawab dengan benar (Mira et al., 2021). Kumpulan soal yang baik memiliki rasio kesulitan mudah:sedang:sulit yang seimbang dan memiliki rasio 3:5:2 (Fiska et al., 2021).

#### 4. Kualitas Pengecoh

Distraktor efektif karena mereka dapat menipu siswa yang tidak tahu kunci jawaban dengan memberi mereka pilihan yang salah. Semakin banyak siswa yang memilih distraktor, semakin baik fungsinya (Rezki Amalia et al., 2021). Suatu distraktor dikatakan berhasil jika dipilih oleh siswa dari kelompok terendah atau yang berkemampuan kurang (Khasanah et al., 2023). Interpretasi efektivitas distraktor didasarkan pada kriteria berikut pada skala Likert. (1)

Untuk pertanyaan yang keempat distraktornya berfungsi, maka potensi pengecohnya tergolong sangat baik. (2) Untuk pertanyaan yang ketiga distraktornya berfungsi, maka potensi pengecohnya tergolong baik. (3) Untuk pertanyaan yang kedua distraktornya berfungsi, maka potensi pengecohnya tergolong cukup. (4) untuk pertanyaan yang satu distraktornya berfungsi, maka potensi pengecohnya tergolong kurang baik. (5) untuk pertanyaan yang tidak memiliki distraktornya berfungsi, maka potensi pengecohnya tergolong sangat tidak baik (Widayanti et al., 2021).

Peneliti menentukan kualitas pengecoh (distraktor) berdasarkan hasil ANATES 4.0.9. Hasil analisis butir soal didistribusikan yakni.

**Tabel 9. Hasil Kualitas Pengecoh Butir Soal**

No	Kategori Kualitas Pengecoh	Nomor Soal	Banyak Soal	Prosentase
1	Sangat Baik	12, 16	2	8%
2	Baik	7, 9, 11, 15, 17, 18, 19, 23, 24, 25	10	40%
3	Cukup Baik	2, 4, 5, 6, 10, 14, 20, 21, 22	9	36%
4	Kurang Baik	1, 3, 8, 13	4	16%

Sumber: Diolah oleh Peneliti

Soal pilihan ganda biasanya memiliki pilihan yang mengganggu atau pilihan alternatif. Siswa yang menjawab salah secara merata memilih pertanyaan dengan pengecoh yang baik, sedangkan siswa yang menjawab salah memilih pertanyaan dengan pengecoh yang kurang baik secara tidak merata. Ketika setidaknya 5% peserta tes memilih jawaban tersebut, pengecoh berfungsi dengan baik (Toifur & Setyowati, 2022). Soal-soal pada aspek kompetensi yang berbeda mempunyai kualitas yang baik jika siswa yang lebih cerdas menjawab pertanyaan tersebut dibandingkan siswa yang kurang cerdas. Soal dikatakan berkualitas rendah apabila siswa yang cerdas dan kurang cerdas dapat menjawab soal tersebut, begitu pula sebaliknya (Fiska et al., 2021). Dalam penelitian ini, hanya 4 soal yang memiliki daya pembeda sangat tinggi, yang berarti mereka dapat membedakan siswa dengan kemampuan rendah dan tinggi. Soal yang mempunyai daya pembeda baik, cukup baik, dan buruk, memerlukan revisi terkait konstruk atau materi yang sedang diujikan, serta variasi pilihan jawaban yang lebih sedikit untuk mengganggu jawaban peserta didik.

Hasil penelitian terkait kualitas pengecoh soal sumatif matematika berbentuk pilihan ganda materi fungsi komposisi dan invers fungsi di kelas XI pada salah satu SMA Negeri di Kota Surabaya menunjukkan 8% memiliki pengecoh sangat baik, 40% memiliki pengecoh baik, 36% memiliki pengecoh cukup baik, dan 16% memiliki pengecoh kurang baik. Hal tersebut berbanding terbalik dengan hasil penelitian AR et al. yaitu efektivitas pengecoh butir soal

ulangan akhir semester (UAS) ganjil mata pelajaran matematika siswa kelas X SMAN 1 Majene tahun ajaran 2021/2022 diperoleh bahwa sebanyak masing-masing 6 butir soal (30%) memiliki efektivitas pengecoh yang sangat baik dan kurang baik, sebanyak 3 butir soal (15%) memiliki efektivitas pengecoh yang baik dan sebanyak 5 butir soal (25%) memiliki efektivitas pengecoh yang tidak baik (AR et al., 2024).

Soal dengan pengecoh sangat baik, seperti nomor 12, 16, dan pengecoh baik, seperti nomor 7, 9, 11, 15, 17, 18, 19, 23, 24, 25, memiliki pilihan pengecoh yang berfungsi baik dan dipilih oleh siswa secara merata. Ini memastikan bahwa soal-soal tersebut dapat diterima, dikumpulkan dalam kumpulan soal, dan digunakan untuk sumatif selanjutnya. Pilihan pengecoh cukup baik, yaitu nomor 2, 4, 5, 6, 10, 14, 20, 21, 22 dan pilihan pengecoh kurang baik, yaitu nomor 1, 3, 8, 13, harus diteliti untuk mengetahui alasan mengapa pengecoh tidak berfungsi baik, dan kemudian direvisi atau bahkan diubah dengan soal lain untuk sumatif berikutnya. Faktor pendukung pemilihan distraktor karena opsi jawaban yang tidak seragam serta susunan kata pada soal termasuk kunci jawaban yang kurang tepat. Distraktor tidak akan berhasil dengan baik kecuali jika dipilih oleh siswa yang tidak terlalu memahami materi (Khasanah et al., 2023).

## **5. Kualitas Butir Soal**

Beberapa kriteria (validitas, tingkat kesulitan, daya pembeda, dan efektivitas pengecoh) digunakan untuk mengevaluasi kualitas soal. Menurut Skala Likert berikut: (1) Soal dengan empat kriteria adalah sangat baik dan dapat dimasukkan ke dalam bank soal; (2) Soal dengan tiga kriteria adalah baik tetapi tidak dapat dimasukkan ke dalam bank soal; (3) Soal dengan dua kriteria adalah baik tetapi perlu direvisi; dan (4) Soal dengan hanya satu kriteria adalah buruk dan harus dibuang (Widayanti et al., 2021).

Hasil analisis kualitas butir soal berdasarkan hasil analisis butir soal dari rekap analisis butir menggunakan program ANATES 4.0.9 didistribusikan yakni.

**Tabel 10. Hasil Kualitas Butir Soal**

No	Kategori Kualitas Butir Soal	Nomor Soal	Jumlah Soal	Persentase
1	Sangat Baik dan Layak Digunakan	5, 6, 7, 12, 14, 16, 17, 18, 19, 22	10	40%
2	Baik dan Perlu Direvisi	1, 2, 9, 13, 15, 20, 21, 23, 24	9	36%
3	Cukup Baik dan Perlu Direvisi	3, 4, 8, 10, 11, 25	6	24%

Sumber: Diolah oleh Peneliti

Soal dapat dikategorikan menjadi lima kategori berdasarkan skala Likert: sangat baik, baik, sedang, kurang baik, dan sangat kurang. Kategori ini didasarkan pada kualitas pengecoh,

validitas, tingkat kesukaran, daya pembeda, dan reliabilitas. Peneliti menyatakan bahwa sebanyak 40% dari total butir soal memiliki kualitas soal yang sangat baik diantaranya butir soal nomor 5, 6, 7, 12, 14, 16, 17, 18, 19, 22. Soal-soal tersebut memenuhi kriteria kualitas pengecoh, validitas, tingkat kesukaran, daya pembeda, dan reliabilitas sehingga layak digunakan pada tes sumatif berikutnya atau juga dapat dikumpulkan pada bank soal yang dimiliki peneliti. Selain itu, ada soal dengan kualitas yang baik sebanyak 36% di nomor 1, 2, 9, 13, 15, 20, 21, 23, 24 dan soal dengan kualitas yang cukup baik sebanyak 24% di nomor 3, 4, 8, 10, 11, 25. Agar soal tersebut dapat digunakan, kualitas pengecoh, validitas, tingkat kesukaran, daya pembeda, dan reliabilitas harus dievaluasi terlebih dahulu. Hal tersebut berbanding terbalik dengan hasil penelitian Usman et al. yaitu terdapat 17 butir (85%) soal yang sudah memenuhi kriteria yang ditentukan sehingga tergolong dalam kriteria layak digunakan. Sedangkan untuk soal yang tidak memenuhi kriteria yang ditentukan sehingga tergolong ke dalam kriteria soal perlu direvisi terdapat sebanyak 3 soal (15%) (Usman et al., 2022).

## **KESIMPULAN**

Peneliti dapat membuat kesimpulan diantaranya soal sumatif matematika berbentuk pilihan ganda materi fungsi komposisi dan invers fungsi di kelas XI pada salah satu SMA Negeri di Kota Surabaya menunjukkan 56% atau 14 soal valid dan 44% atau 11 butir soal tidak valid. Nilai reliabilitas sebesar 0,71 yang artinya soal mempunyai nilai reliabilitas tinggi. Soal sumatif menunjukkan bahwa secara keseluruhan, soal tersebut memiliki daya pembeda yang baik: 16% atau 4 soal daya pembedanya sangat bagus, 48% atau 12 soal daya pembedanya bagus, 24% atau 6 soal daya pembedanya cukup bagus, dan 12% atau 3 soal daya pembedanya buruk.

Berdasarkan tingkat kesulitan, peneliti menemukan bahwa 68% atau 17 soal tingkat kesulitannya sedang, 20% atau 5 soal tingkat kesulitannya mudah, dan 12% atau 3 soal tingkat kesulitannya sangat mudah. Oleh karena itu, soal sumatif ini dominan oleh soal dengan tingkat kesulitannya sedang dan dapat dianggap memiliki kualitas yang baik. Dalam penelitian, kualitas pengecoh menunjukkan bahwa 8% atau 2 soal pengecohnya sangat baik, 40% atau 10 soal pengecohnya baik, 36% atau 9 soal pengecohnya cukup baik, dan 16% atau 4 soal pengecohnya kurang baik.

Hasil analisis soal sumatif matematika berbentuk pilihan ganda yang dilakukan pada materi fungsi komposisi dan invers fungsi di kelas XI pada salah satu SMA Negeri di Kota Surabaya menunjukkan bahwa 40%, atau setidaknya 10 butir, dari soal tersebut berkualitas

sangat baik. Hal tersebut memenuhi kriteria daya pembeda, validitas, kualitas pengecoh, dan tingkat kesukaran. Oleh karena itu, soal tersebut dapat digunakan untuk tes sumatif berikutnya atau digabungkan ke dalam bank soal yang dimiliki peneliti. Selain itu, terdapat 36% atau sebanyak 9 butir soal berkualitas baik, dan 24% atau sebanyak 6 butir soal berkualitas cukup baik. Soal tersebut perlu direvisi untuk daya pembeda, validitas, kualitas pengecoh, dan tingkat kesukarannya sebelum digunakan untuk sumatif selanjutnya.

Peneliti dapat melakukan revisi pada soal sumatif matematika berbentuk pilihan ganda materi fungsi komposisi dan invers fungsi di kelas XI pada salah satu SMA Negeri di Kota Surabaya yang memiliki kualitas yang cukup dan baik terkait : (1) tingkat kesulitan yang masih dalam kategori mudah serta sangat mudah dan tidak ada kategori sulit, (2) kevalidan butir soal yang hampir mendekati dan tidak menimbulkan ketimpangan persentase yang signifikan, (3) daya pembeda yang sangat baik masih sangat minimal jumlah butir soalnya, dan (4) kualitas pengecoh didominasi oleh jenis pengecoh yang buruk.

Penggunaan software ANATES 4.0.9 dapat digunakan untuk guru sebagai alat yang digunakan untuk mengetahui kelayakan butir soal sehingga menghasilkan soal-soal tes yang berkualitas dan dimanfaatkan dalam kumpulan soal yang dimiliki oleh guru guna untuk tes selanjutnya. Penggunaan software ANATES 4.0.9 juga dapat digunakan untuk peneliti selanjutnya guna menguji kelayakan butir soal dari seperangkat tes yang diberikan kepada siswa dalam kegiatan penilaian formatif atau sumatif.

## **DAFTAR PUSTAKA**

- Alti, R. P., Zulyuri, Z., & Violita. (2023). Analisis Kualitas Butir Soal Ujian Tengah Semester Mata Pelajaran Biologi Kelas X di MAN 1 Solok Selatan. *Jurnal Metaedukasi: Jurnal Ilmiah Pendidikan*, 4(2), 70–75. <https://doi.org/10.37058/metaedukasi.v4i2.4089>
- Antari, B., Susanta, A., & Siagian, T. A. (2021). Analisis Soal Penilaian Akhir Semester Matematika SMA Negeri 8 Kota Bengkulu Tahun Ajaran 2020/2021. *JEMS (Jurnal Edukasi Matematika Dan Sains)*, 9(2), 299–312. <https://doi.org/10.25273/jems.v9i2.10421>
- Aquatika, F., Rofi'i, & Bayu Pratama, Moch. H. (2022). Penggunaan Aplikasi Anates pada Analisis Butir Soal Penilaian Tengah Semester Bahasa Indonesia Kelas X MIPA 8 SMA Muhammadiyah 2 Surabaya. *Jurnal Metalanguage*, 4(4), 1–7. <https://doi.org/10.56707>
- AR, A., Suharti, Mania, S., & Hasrianti, A. (2024). Analisis Kualitas Butir Soal Ulangan Akhir Semester Ganjil Pada Mata Pelajaran Matematika Siswa Kelas X SMA Negeri. *Jurnal Theorems (The Original Research Of Mathematics)*, 8(2), 284–296. <https://doi.org/10.31949/th.v8i2.7442>

- Ashayidatina, L., & Alber. (2023). Analisis Butir Soal Ujian Semester Ganjil Bahasa Indonesia Kelas X SMK Negeri 1 Mempura. *Jurnal Pendidikan Bahasa Dan Sastra Indonesia*, 8(2), 131–137. <https://doi.org/10.26737/jp-bsi.v8i2.5039>
- Cahyaningrum, I. Y., Fuady, A., & Sunismi. (2023). Analisis Butir Soal Sumatif Akhir Semester Ganjil Mata Pelajaran Matematika Kelas VII dengan Berbantuan Aplikasi Software Anates. *Mathema Journal*, 5(2), 67–81. <https://doi.org/10.33365/jm.v5i2.2601>
- Elviana. (2020). Analisis Butir Soal Evaluasi Pembelajaran Pendidikan Agama Islam Menggunakan Program Anates. *Jurnal Mudarrisuna*, 10(2), 58–74. <https://doi.org/10.22373/jm.v10i2.7839>
- Fiska, J. M., Hidayati, Y., Qomaria, N., & Hadi, W. P. (2021). Analisis Butir Soal Ulangan Harian IPA Menggunakan Software Anates Pada Pendekatan Teori Tes Klasik. *Jurnal Natural Science Educational Research*, 4(1), 65–76. <https://doi.org/10.21107/nser.v4i1.8133>
- Friatma, A., & Anhar, A. (2019). Analysis of validity, reliability, discrimination, difficulty and distraction effectiveness in learning assessment. *Journal of Physics: Conference Series*, 1387(1). <https://doi.org/10.1088/1742-6596/1387/1/012063>
- Hidayati, K., & Nisa', Z. E. K. (2023). Analisis Butir Soal Penilaian Akhir Semester Matematika. *Aksioma : Jurnal Program Studi Pendidikan Matematika*, 12(3), 3516. <https://doi.org/10.24127/ajpm.v12i3.7575>
- Hodiyanto. (2017). Analisis Butir Soal Pilihan Ganda Matematika Sekolah Menengah Pertama. *Buana Matematika*, 7, 53–60. <https://doi.org/10.36456/buanamatematika.v7i2.:1044>
- Indah Sari, V. N., Yudi Utomo, A. P., & Sumarwati. (2022). Kualitas Soal Bahasa Indonesia di SMP Muhammadiyah 1 Pontianak: Analisis Butir Soal. *Jurnal Pendidikan Bahasa Dan Sastra Indonesia*, 2, 112–119. <https://doi.org/https://doi.org/10.15294/jpbsi.v11i2.58091>
- Khasanah, I., Fuady, A., & Sunismi. (2023). Analisis Soal Ulangan Harian Matematika Bentuk Pilihan Ganda. *Mathema Journal*, 5(2), 2023. <https://doi.org/10.33365/jm.v5i2.2818>
- Lukman, A. (2023). Karakteristik Butir Soal Pas Matematika SMA Kelas X. *Jurnal\_Ilmiah\_Matematika*, 4, 25–34. <https://doi.org/10.5281/zenodo.8146055>
- Manfaat, B., Nurazizah, A., & Misri, M. A. (2021). Analysis of Mathematics Test Items Quality for High School. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 25(1), 108–117. <https://doi.org/10.21831/pep.v25i1.39174>
- Merlin Purwati, L., Arianty, R., Malaka Syakilah, D., Ridlo, S., & Susilaningsih, E. (2021). Analisis Soal Tes Pilihan Ganda Berbasis Higher Order Thinking Skill menggunakan Aplikasi Anates Windows Versi 4.0.9 For Windows. *JurnalPendidikanUniversitasGarut*, 15(2), 460–473. <https://doi.org/http://dx.doi.org/10.52434/jp.v15i2.1287>
- Mira, K., Hayon, V. H. B., & Tinenti, Y. R. (2021). Analisis Butir Soal Pilihan Ganda Pada Materi Pokok Larutan Elektrolit Dan Nonelektrolit Kelas X IPA SMA Negeri 1 Lamba Leda Manggarai Timur. *Wahana: Tridarma Perguruan Tinggi*, 73, 2. <https://doi.org/https://doi.org/10.36456/wahana.v73i2.4201>

- Muluki, A., Bundu, P., & Sukmawati. (2020). Analisis Kualitas Butir Tes Semester Ganjil Mata Pelajaran IPA Kelas IV MI Radhiatul Adawiyah. *Jurnal Ilmiah Sekolah Dasar*, 4(1), 86–96.
- Munauwarah, R., Suhartati, & Khairul Umam. (2022). Analisis Butir Soal Ujian Semester Genap Mata Pelajaran Matematika pada Kelas X SMA di Kota Banda Aceh Tahun Ajaran 2020/2021. *Jurnal Ilmiah Mahasiswa Pendidikan Matematika*, 7(2), 278–284.
- Nurfiqah, S., Supardi, I., & Novita, D. (2015). The Analysis On The Items Of The English Test Made By The Teacher. *Jurnal Pendidikan Dan Pembelajaran Khatulistiwa*, 4(9). <https://doi.org/10.26418/jppk.v4i9.11247>
- O. A., A., & E. R. I., A. (2016). Comparative Analysis of Classical Test Theory and Item Response Theory Based Item Parameter Estimates of Senior School Certificate Mathematics Examination. *European Scientific Journal, ESJ*, 12(28), 263. <https://doi.org/10.19044/esj.2016.v12n28p263>
- Rezki Amalia, N., Halik, A., & Mukhlisa, N. (2021). Analisis Butir Soal Matematika Pada Siswa Sekolah Dasar. *Pinisi Journal Of Education*, 219–230.
- Santosa, S., & Badawi, J. A. (2022). Analisis Butir Soal Pilihan Ganda Tema Pertumbuhan dan Perkembangan Makhhluk Hidup Kelas III Madrasah Ibtidaiyah. *Jurnal Basicedu*, 6(2), 1678–1686. <https://doi.org/10.31004/basicedu.v6i2.2206>
- Sardia, Y., Fitriati, A., & Salmina, M. (2021). Analisis Kualitas Butir Soal Try Out Bidang Studi Matematika SMA N 1 Simeulue Tengah. *Jurnal Ilmiah Mahasiswa*, 2(1).
- Toifur, I., & Setyowati, E. (2022). Analisis Butir Soal Simulasi Ujian Sekolah Mata Pelajaran Matematika SMA Negeri 1 Salaman. *Math Locus: Jurnal Riset Dan Inovasi Pendidikan Matematika*, 3(1), 28–39.
- Usman, Wiwid, & Sulisti, H. (2022). Analisis Butir Soal Ulangan Matematika Semester Genap Kelas XI SMA Negeri 2 Pulau Maya. *Jurnal Al 'Adad: Jurnal Tadris Matematika*, 1(2), 34–43. <https://doi.org/10.24260/add.v1i2.1133>
- Warju, Rizki Ariyanto, S., Soeryanto, & Adi Trisna, R. (2020). Analisis Kualitas Butir Soal Tipe HOTS Pada Kompetensi Sistem Rem Siswa Di Sekolah Menengah Kejuruan. *Jurnal Pendidikan Teknologi Dan Kejuruan*, 17(1), 95–104.
- Widayanti, W., Bistari, & Suparjan. (2021). Analisis Butir Soal Pilihan Ganda Penilaian Tengah Semester Pada Pembelajaran Tematik Kelas V Sekolah Dasar Negeri 39 Pontianak Kota. *Jurnal Didika: Wahana Ilmiah Pendidikan Dasar*, 7(2), 279–296. <https://doi.org/10.29408/didika.v7i2.4370>
- Zahiroh, U., & Soleman Ritonga, P. (2021). Analisis Kualitas Butir Soal Pilihan Ganda Mata Pelajaran Kimia Pada Ujian Akhir Semester (UAS) Kelas XI MAN 2 Kepulauan Meranti. *Jedchem (Journal Education and Chemistry)*, 3(1), 11–20. <https://doi.org/10.36378/jedchem.v3i1.780>
- Zein, M., & Darto. (2012). *Evaluasi Pembelajaran Matematika* (1st ed.). Daulat Riau.