



ANALISIS KUALITAS SOAL PILIHAN GANDA DALAM ASESMEN SUMATIF MATEMATIKA MENGGUNAKAN TEORI TES KLASIK

Annisa Nur Azizah¹, Fuad Luky Atmaja²

^{1,2}Tadris Matematika, UIN Raden Mas Said Surakarta, Indonesia

Email korespondensi: nurazizahannisa421@gmail.com

Riwayat Artikel:		
Diajukan: Juni 2026	Diterima: Juni 2026	Diterbitkan: Juli 2026

Abstract

The purpose of this study is to describe the quality of objective test items used in summative assessments for mathematics based on classical test theory. This study will examine the aspects of difficulty, discriminative power, and the effectiveness of distractor options. The method used in this study is a descriptive quantitative approach involving 57 students from classes VIII A1 and VIII B as participants. The data used in this study were taken from 25 multiple-choice items on the summative assessment for the first semester of the 2025/2026 academic year, which were then analyzed using Microsoft Excel. The analysis results showed that the majority of items (56%) fell into the “moderate difficulty” category, while 40% were categorized as easy and 4% as difficult. In terms of discriminative power, 56% of the items fell into the “good to very good” category; meanwhile, 16% of the items were found to have low discriminative power. The effectiveness of distractors indicated that 70.67% of distractors functioned effectively, and 80% of the exam questions had good to very good distractor quality. Overall, this assessment tool is of good quality; however, some questions need to be revised to ensure more accurate and appropriate measurement of student skills.

Keywords: item analysis, classical test theory, difficulty level, discriminative power, effectiveness of distractors.

Abstrak

Tujuan dari penelitian ini adalah guna menggambarkan kualitas butir-butir soal objektif yang diterapkan dalam penilaian sumatif untuk mata pelajaran matematika berdasarkan teori tes klasik. Penelitian ini akan meneliti aspek-aspek kesulitan, daya pembeda, dan efektivitas pilihan jawaban yang bersifat pengalih perhatian. Metode yang dimanfaatkan pada penelitian ini ialah dengan memakai pendekatan kuantitatif deskriptif yang di dalamnya terdapat 57 siswa yang dilibatkan dari kelas VIII A1 serta VIII B sebagai partisipan. Data-data yang dimanfaatkan pada penelitian ini diambil dari 25 butir soal yang berupa pilihan ganda pada penilaian sumatif untuk semester pertama tahun ajaran 2025/2026 yang kemudian dianalisis menggunakan Microsoft Excel. Hasil analisis memperlihatkan bahwasanya mayoritas soal (56%) tergolong pada kategori “kesulitan sedang”, sementara 40% soal dikategorikan sebagai mudah dan 4% sebagai sulit. Dalam hal daya diskriminasi, 56% soal termasuk dalam kategori “baik hingga sangat baik”; sementara itu, 16% soal ditemukan memiliki daya diskriminasi yang rendah. Efektivitas pengalih perhatian menunjukkan bahwa 70,67% pengalih perhatian berfungsi secara efektif, dan 80% soal ujian memiliki kualitas pengalih perhatian yang baik hingga sangat baik. Secara keseluruhan, alat penilaian ini memiliki kualitas yang baik; meskipun demikian, beberapa soal perlu direvisi agar pengukuran keterampilan siswa menjadi lebih akurat dan sesuai.

Kata kunci: analisis butir soal, teori tes klasik, tingkat kesukaran, daya pembeda, efektivitas pengecoh.

PENDAHULUAN

Pendidikan mempunyai peranan yang krusial guna memajukan kualitas sumber daya manusia. Di antara berbagai disiplin ilmu, matematika menjadi salah satu bidang yang sangat berkontribusi dalam membangun kehidupan manusia yang lebih baik, karena dipandang sebagai fondasi utama untuk menguasai, memperoleh, serta mengembangkan ilmu pengetahuan dan teknologi (Nurazizah et al., 2022). Di lingkungan sekolah, matematika turut berperan penting dalam mencetak siswa yang berkualitas, yaitu dengan melatih mereka agar mampu berpikir secara logis dan teratur (Rofiq et al., 2024).

Agar efektivitas pembelajaran matematika dapat terukur secara objektif, dalam proses pendidikan, evaluasi pembelajaran menjadi bagian penting untuk mengetahui ketercapaian tujuan pembelajaran serta kemampuan peserta didik. Melalui evaluasi, guru dapat memperoleh informasi yang digunakan sebagai dasar dalam memperbaiki proses pembelajaran selanjutnya (Magdalena et al., 2023; Pratama et al., 2024). Menurut Kasiyani & Isnaniah (2026), evaluasi merupakan suatu proses yang bersifat komprehensif, meliputi berbagai kegiatan yang dilakukan secara sistematis mulai dari pengumpulan, pengolahan, analisis, hingga penafsiran informasi yang berkaitan dengan pembelajaran, guna mengukur sejauh mana efektivitas proses pembelajaran dan ketercapaian tujuan serta hasil belajar siswa. Pengukuran kemampuan matematika peserta didik secara akurat memerlukan instrument tes yang telah teruji kualitasnya, karena kemampuan berpikir siswa yang kompleks hanya dapat terukur apabila soal yang diberikan mampu merepresentasikan tingkat kemampuan tersebut secara proporsional (Rahman & Nursyam, 2026).

Satu di antara beberapa alat untuk mengevaluasi yang secara umum diterapkan pada kegiatan belajar mengajar ialah tes. Agar tes dapat mengukur kompetensi siswa secara akurat dan adil, instrumen tersebut harus memenuhi standar kualitas tertentu (Asri et al., 2024). Analisis butir soal menjadi satu diantara beberapa langkah yang bisa dimanfaatkan guna meningkatkan mutu ujian. Sebuah item dapat dianggap berkualitas tinggi jika memiliki tingkat kesulitan yang seimbang, daya beda yang cukup, dan distraktor yang berfungsi dengan baik (Fitriani, 2021). Ketiga kriteria ini adalah indikator utama dari Teori Ujian Klasik (CTT), yang digunakan untuk mengevaluasi kualitas item berdasarkan data empiris (Istiyova, 2025). Teori Tes Klasik adalah pendekatan untuk menganalisis kualitas item yang menekankan hubungan antara nilai siswa dan karakteristik item yang diuji (Sumaryanta, 2021).

Berdasarkan keterangan yang didapat dari wawancara bersama guru matematika MTs N 2 Surakarta, diketahui bahwa penyusunan soal umumnya mengacu pada ATP atau silabus serta kesepakatan dalam forum Musyawarah Guru Mata Pelajaran (MGMP). Namun, analisis kualitas butir soal belum dilakukan secara rutin dan sistematis setelah

pelaksanaan tes. Akibatnya, kualitas instrument yang digunakan belum diketahui secara empiris, sehingga hasil evaluasi pembelajaran berpotensi belum sepenuhnya mencerminkan kemampuan peserta didik secara akurat. Hasil asesmen sumatif juga menunjukkan bahwa capaian peserta didik dalam mata pelajaran matematika masih bervariasi.

Hasil dari penelitian-penelitian sebelumnya menunjukkan bahwa banyak butir soal belum memenuhi standar kualitas yang diharapkan. Temuan penelitian menunjukkan adanya sejumlah butir soal yang tidak valid dan memiliki daya beda yang rendah, sehingga butir-butir tersebut perlu diperbaiki. Namun, penelitian ini belum secara menyeluruh mengevaluasi efektivitas pilihan pengalih perhatian sebagai faktor penting dalam menentukan kualitas butir soal (Amanda et al., 2025). Sejalan dengan itu, penelitian oleh Novriyanti et al. (2024) juga menemukan bahwa meskipun tingkat kesukaran soal relatif proporsional, masih terdapat kelemahan pada aspek daya beda dan keberfungsian pengecoh yang belum optimal. Temuan ini mengindikasikan bahwa diperlukan analisis menyeluruh terhadap kualitas butir soal agar instrumen penilaian yang digunakan dapat memberikan informasi yang akurat tentang kemampuan peserta. Pemilihan MTs N 2 Surakarta sebagai lokasi penelitian didasarkan pada adanya kebutuhan evaluasi kualitas instrument asesmen matematika yang belum dianalisis secara empiris dan rutin. Selain itu, keberagaman karakteristik siswa yang mencakup kelas reguler dan pendidikan khusus, memungkinkan terjadinya gambaran yang lebih komprehensif dan representatif mengenai kualitas butir-butir

Dengan demikian, tujuan dilakukannya penelitian ini ialah guna menilai kualitas dari soal-soal pada penilaian somatik dalam mata pelajaran matematika dengan menggunakan pendekatan teori tes klasik, melalui analisis tiga aspek: tingkat kesulitan, daya pembeda, serta efektivitas pengecoh. Diharapkan hasil dari dilakukannya penelitian ini bisa dijadikan sebagai panduan dalam mengembangkan alat penilaian yang lebih berkualitas serta mendukung optimasi evaluasi proses pembelajaran.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif deskriptif. Pendekatan ini diterapkan guna mendapatkan gambaran objektif tentang kualitas dari butir-butir soal yang didasarkan dari hasil menganalisis secara statistik (Sugiyono, 2024). Melalui pendekatan ini, karakteristik butir soal dapat diidentifikasi berdasarkan tingkat kesukaran, daya pembeda, dan efektivitas pengecoh sehingga diperoleh informasi mengenai kualitas asesmen yang digunakan (Nababan et al., 2025).

Populasi dalam penelitian ini meliputi semua siswa kelas VIII yang terdiri dari program reguler dan program khusus. Metode pengambilan sampel yang diterapkan

adalah *cluster sampling*, yaitu pengambilan sampel berdasarkan kelompok yang telah terbentuk (Sugiyono, 2024). Sampel dalam penelitian ini terdiri dari 57 siswa yang berasal kelas VIII A1 program khusus dan VIII B program regular. Pemilihan kedua kelas tersebut dilakukan untuk memperoleh gambaran kualitas butir soal yang lebih luas berdasarkan karakteristik peserta didik pada dua program pembelajaran yang terdapat di MTs N 2 Surakarta.

Instrumen penelitian ini terdiri atas 25 butir soal yang berupa pilihan ganda pada assessmen sumatif akhir semester gasal mata pelajaran matematika. Setiap soal memiliki 4 opsi jawaban yakni opsi A, opsi B, opsi C, serta opsi D. Soal tersebut mencakup beberapa materi dan submateri berdasarkan kompetensi yang diujikan. Adapun materi yang diujikan meliputi Operasi Aljabar, Sistem Persamaan Linear dengan Dua Variabel (SPLDV), Gradien dan Persamaan Garis Lurus, serta Fungsi Linear. Data penelitian berupa hasil jawaban siswa terhadap 25 butir soal tersebut.

Proses mengumpulkan data-data untuk penelitian ini dilaksanakan dengan memanfaatkan jenis teknik dokumentasi, yaitu melalui pengumpulan naskah soal beserta lembar jawaban peserta didik. Data yang telah dikumpulkan kemudian dianalisis memakai jenis pendekatan yakni teoritis klasik guna menilai kualitas dari masing-masing butiran soal. Tahapan-tahapan yang dilaksanakan pada penelitian ini ialah berikut.

1. Menghitung Tingkat Kesukaran Soal

Tingkatan kesukaran suatu butir soal mempunyai fungsi ialah guna mengelompokkan soal-soal ke dalam kategori sukar, mudah, maupun sedang menurut perspektif peserta didik. Soal dianggap semakin mudah apabila semakin banyak siswa yang mampu menjawabnya dengan benar, dan berlaku sebaliknya (Khotimah & Hidayat, 2023). Menentukan besarnya Tingkat kesukaran dirumuskan oleh Arikunto (2018) sebagai berikut

$$P = \frac{B}{JS}$$

Keterangan:

B : Jumlah siswa yang menjawab benar

P : Indeks Tingkat Kesukaran

JS : Jumlah seluruh siswa yang mengikuti tes

Tabel 1. Klasifikasi Tingkat Kesukaran

Indeks Tingkat Kesukaran	Interpretasi
0 - 0,30	Sukar
0,31 – 0,70	Sedang
0,71 – 1,00	Mudah

Sumber: (Arikunto, 2018)

2. Menghitung Daya Pembeda

Daya pembeda pada suatu butiran soal mencerminkan sebesar apa kemampuan soal tersebut dalam menjadi pembeda para peserta didik yang memiliki kompetensi tinggi dan juga para peserta didik yang memiliki kompetensi rendah (Shofiah et al., 2023). Menurut Arikunto (2018), daya beda butir soal dapat dihitung melalui rumus sebagai berikut.

$$D = \frac{Ba}{Ja} - \frac{Bb}{Jb}$$

Keterangan :

D : Indeks daya pembeda

Ba : Jumlah siswa pada kelompok atas yang menjawab benar

Bb : Jumlah siswa pada kelompok bawah yang menjawab benar

Ja : Jumlah seluruh siswa pada kelompok atas

Jb : Jumlah seluruh siswa pada kelompok bawah

Tabel 2. Klasifikasi Daya Pembeda

Indeks	Interpretasi
0 - 0,20	Jelek
0,21 – 0,40	Cukup
0,41 – 0,70	Baik
0,71 – 1,00	Baik sekali
Negatif	Semuanya tidak baik

Sumber: (Arikunto, 2018)

3. Menghitung Efektivitas Pengecoh

Pengecoh merupakan pilihan jawaban pada soal pilihan ganda selain kunci jawaban yang benar. Menurut Arikunto (2018) pengecoh dinyatakan mempunyai fungsi secara baik apabila dipilih setidaknya 5% dari total peserta yang melaksanakan tes. Proses menilai mengenai efektifitas dari pengecoh pada tiap-tiap butir soal dilaksanakan dengan cara menerapkan kriteria yang tercantum dalam tabel di bawah ini.

Tabel 3. Klasifikasi Efektivitas Pengecoh

Pengecoh yang Berfungsi	Klasifikasi
3	Sangat baik
2	Baik
1	Kurang baik
0	Tidak baik

Sumber: (Nafs et al., 2023)

4. Menarik Kesimpulan Hasil Analisis

Tahap terakhir dilakukan dengan menarik kesimpulan berdasarkan hasil analisis tingkat kesukaran, daya pembeda, dan efektivitas pengecoh pada setiap butir soal. Kesimpulan dirumuskan berdasarkan hasil perhitungan masing-masing indikator

analisis butir soal yang kemudian diinterpretasikan sesuai dengan kriteria klasifikasi yang telah ditetapkan. Selain itu, hasil analisis juga diperkuat dengan merujuk temuan-temuan dari penelitian terdahulu yang relevan sebagai bahan pembandingan dan pendukung dalam proses penarikan kesimpulan. Dengan demikian, diperoleh gambaran mengenai kualitas butir soal asesmen sumatif matematika secara keseluruhan.

HASIL DAN PEMBAHASAN

Proses menganalisis butir-butir soal ini dilakukan berdasarkan hasil dari jawaban para siswa terhadap soal berbentuk pilihan ganda pada siswa kelas VIII A1 dan VIII B dalam penilaian Asesmen Sumatif Semester Gasal Tahun Ajaran 2025/2026 pada mata pelajaran Matematika. Adapun hasil dari asesmen sumatif siswa disajikan pada tabel berikut berikut.

Tabel 4. Hasil Asesmen Sumatif Siswa

Data Statistik	Hasil
Jumlah siswa	57
Nilai tertinggi	96
Nilai terendah	20
Rentang	76
Nilai rata-rata	62,1754386

Berdasarkan data statistik dari Asesmen Sumatif Semester Gasal tersebut, diketahui bahwa rentang nilai mencapai 76, yang memperlihatkan bahwasanya terdapat variasi tingkat kemampuan yang cukup besar antarpeserta didik. Nilai rata-rata yang tercatat sebesar 62,18 mengindikasikan bahwa secara keseluruhan, hasil belajar siswa tergolong dalam kategori sedang. Kondisi ini menegaskan perlunya dilakukan analisis terhadap butir-butir soal, terkhusus dilihat atau ditinjau dari daya pembedanya, tingkatan kesukarannya, serta aktivitas pengecohnya, agar kualitas instrumen evaluasi yang digunakan dapat terjamin.

1. Tingkat Kesukaran Soal

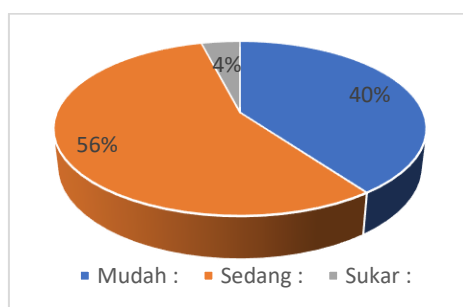
Kemudahan maupun kesulitan suatu butir soal bisa dipahami dari banyaknya siswa yang memberikan jawaban benar. Hasil dari dilakukannya analisis pada tingkat kesukaran soal berdasarkan jawaban siswa tersaji pada tabel 5 serta gambar 1 di bawah ini.

Tabel 5. Distribusi Tingkat Kesukaran Soal

Kategori	Butir Soal	Jumlah
Sukar	21	1
Sedang	2, 5, 10, 11, 12, 13, 16, 17, 18, 19, 20, 23, 24, 25	14

Mudah	1, 3, 4, 6, 7, 8, 9, 14, 15, 22	10
-------	---------------------------------	----

Berdasarkan hasil analisis terhadap 25 butir soal, diperoleh gambaran distribusi tingkat kesukaran sebagai berikut: sebanyak 10 butir soal (40%) masuk pada kategori mudah, sebanyak 14 butir soal (56%) berada pada kategori sedang, serta hanya terdapat 1 butir soal (4%) yang masuk pada kategori sukar. Butir soal dengan kategori mudah meliputi nomor 1, 3, 4, 6, 7, 8, 9, 14, 15, dan 22, sedangkan butir soal berkategori sedang mencakup nomor 2, 5, 10, 11, 12, 13, 16, 17, 18, 19, 20, 23, 24, dan 25. Adapun satu-satunya butir soal yang tergolong sukar adalah butir nomor 21.



Gambar 1. Persentase Tingkat Kesukaran Soal

Merujuk pada standar proporsionalitas tingkat kesukaran, suatu tes dinilai baik apabila terdapat soal-soal dengan tingkat kesukaran yang seimbang. Proporsi ideal tingkat kesulitan soal ialah 30% soal dengan kategori mudah, 50% soal dengan kategori sedang, serta 20% soal dengan kategori sulit, yang dapat diwakili dengan perbandingan 3:5:2 (Pradita et al., 2023). Jika dibandingkan dengan standar proporsionalitas tingkat kesukaran, distribusi soal pada penelitian ini belum sepenuhnya ideal. Persentase soal sedang sebesar 56% sudah mendekati proporsi yang dianjurkan, namun persentase soal mudah yang mencapai 40% jauh melampaui batas ideal 30%, sementara soal sukar hanya sebesar 4% sehingga masih jauh di bawah proporsi yang disarankan. Kondisi ini mengindikasikan bahwa perangkat soal yang dimanfaatkan belum bisa merepresentasikan kemampuan dari para peserta didik dengan proporsional.

Dominasi soal berkategori sedang pada dasarnya merupakan hal yang positif, sebab soal dengan tingkat kesukaran sedang memiliki indeks diskriminasi yang cenderung lebih optimal dibandingkan soal yang terlalu mudah atau terlalu sukar (Arikunto, 2018). Soal dengan tingkat kesukaran sedang mampu membuat tes berfungsi lebih adil karena peserta didik dengan penguasaan konsep rendah masih memiliki peluang untuk menjawab, sementara peserta didik dengan penguasaan tinggi tetap perlu mengerahkan kemampuan berpikirnya. Namun demikian, minimnya proporsi soal sukar menjadi catatan penting yang perlu diperhatikan oleh penyusun soal ke depannya.

Temuan dari penelitian ini juga sejalan dengan hasil dari penelitian yang pernah dilaksanakan oleh Auliya et al. (2025), yang menunjukkan bahwasanya kualitas perangkat tes seringkali didominasi oleh butir soal berkategori sedang (67,68%) dengan proporsi soal sukar yang sangat rendah (7,41%). Temuan serupa juga dikemukakan oleh Habiba Ulfahyana (2022) yang menyatakan bahwa tingkat kesukaran soal asesment masih belum mencapai standar ideal sebab mayoritas soal-soal terdiri dari soal yang mudah, sehingga belum dianggap menantang untuk para peserta didik yang melaksanakan tes.

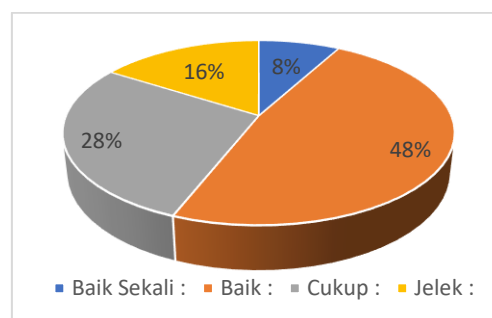
2. Daya Pembeda Soal

Setelah tingkat kesukaran dianalisis, tahap berikutnya adalah mengkaji kemampuan dari tiap-tiap butir soal guna menjadi pembeda dari para siswa yang memiliki tingkat penguasaan materi berbeda. Hasil analisis daya pembeda soal berdasarkan jawaban siswa tersaji pada tabel 6 serta gambar 2 di bawah ini.

Tabel 6. Distribusi Daya Pembeda Soal

Kategori	Butir Soal	Jumlah
Baik Sekali	10, 16	2
Baik	2, 3, 8, 9, 11, 12, 13, 17, 19, 21, 24, 25	12
Cukup	1, 5, 6, 7, 20, 22, 23	7
Jelek	4, 14, 15, 18	4
Negatif	-	-

Analisis terhadap daya beda dari 25 butir soal menunjukkan variasi kategori daya beda yang cukup beragam. Sebanyak 2 butir soal (8%) berada pada kategori baik sekali, yakni butir nomor 10 dan 16. Butir soal berkategori baik berjumlah 12 butir (48%), meliputi nomor 2, 3, 8, 9, 11, 12, 13, 17, 19, 21, 24, dan 25. Sementara itu, 7 butir soal (28%) masuk dalam kategori cukup, yaitu nomor 1, 5, 6, 7, 20, 22, dan 23. Sementara itu, 4 butir soal (16%) tergolong dalam kategori jelek, yakni nomor 4, 14, 15, dan 18. Tidak ditemukan butir soal yang termasuk kategori negatif, yang menunjukkan bahwa tidak ada soal yang dijawab dengan benar lebih sering oleh kelompok bawah dibandingkan dengan kelompok atas.



Gambar 2. Persentase Daya Pembeda Soal

Daya pembeda merupakan parameter penting dalam mengevaluasi kualitas butir soal karena menunjukkan sejauh mana soal tersebut dapat menjadi pembeda antara para peserta didik yang telah memahami materi serta yang belum paham pada materi (Arikunto, 2018). Butir soal dengan daya pembeda tinggi umumnya akan dijawab tepat atau benar oleh kelompok atas serta yang salah oleh kelompok bawah, sedangkan soal dengan daya pembeda rendah tidak mampu memberikan informasi yang jelas mengenai tingkat penguasaan peserta didik (Crocker & Algina, 1986, dalam Novriyanti et al., 2024). Rendahnya daya pembeda tidak hanya disebabkan oleh kelemahan konstruksi soal, tetapi juga dapat dipengaruhi oleh kompleksitas materi sehingga peserta didik menjawab tanpa melalui proses penalaran yang sistematis (Hestiwati & Lestari, 2021).

Secara keseluruhan, kualitas butir soal yang dianalisis menunjukkan hasil yang cukup baik. Sebanyak 56% soal berada pada kategori baik dan baik sekali, sehingga instrumen yang digunakan dinilai cukup efektif dalam mengidentifikasi perbedaan tingkat penguasaan materi antarsiswa. Temuan ini sejalan dengan hasil penelitian Hussain & Garba (2025) yang menganalisis daya pembeda butir soal matematika dan menemukan bahwa 60% butir soal memiliki daya pembeda pada kategori baik hingga baik sekali, sehingga instrumen dinilai memiliki kualitas psikometrik yang cukup memadai untuk digunakan dalam evaluasi pembelajaran. Meskipun demikian, keberadaan 4 butir soal berkategori jelek menunjukkan bahwa masih terdapat soal yang belum efektif dalam membedakan kemampuan peserta didik. Oleh karena itu butir soal dengan kategori jelek perlu direvisi.

Dalam penelitian ini, kelompok atas dan kelompok bawah ditentukan dengan mengambil 27% peserta tes pada masing-masing ujung distribusi skor (Kelley, 1939). Persentase tersebut dipilih karena dinilai mampu menghasilkan perbedaan kemampuan yang optimal antarkelompok sekaligus menjaga kestabilan hasil analisis (Ebel & Frisbie, 1991, dalam Novriyanti et al., 2024). Pengelompokan dilakukan berdasarkan total skor peserta tes tanpa membedakan asal program kelas untuk menjaga objektivitas analisis daya pembeda. Hasil pengelompokan menunjukkan bahwa peserta didik Program Khusus lebih banyak berada pada kelompok atas, sedangkan peserta didik Program Reguler cenderung mendominasi kelompok bawah. Kondisi tersebut mencerminkan adanya variasi kemampuan akademik yang menjadi karakteristik alami populasi penelitian dan sekaligus menunjukkan bahwa instrumen diujikan pada kelompok peserta didik dengan tingkat kemampuan yang beragam.

3. Efektivitas Pengecoh

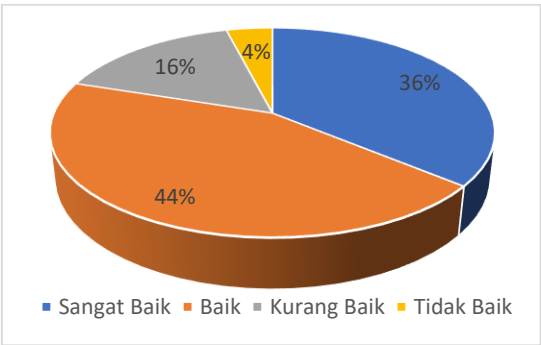
Untuk mengetahui sejauh mana alternatif jawaban pada setiap butir soal pilihan ganda berfungsi sebagai pengecoh yang efektif, dilakukan analisis terhadap distribusi

pilihan jawaban siswa. Hasil analisis efektivitas pengecoh berdasarkan jawaban siswa disajikan pada Tabel 7 dan Gambar 3 berikut.

Tabel 7. Distribusi Efektivitas Pengecoh

Kategori	Pengecoh yang efektif	Butir Soal	Jumlah
Sangat baik	3	10, 13, 17, 18, 19, 21, 22, 23, 25	9
Baik	2	1, 2, 4, 5, 7, 8, 9, 11, 12, 20, 24	11
Kurang baik	1	3, 6, 14, 16	4
Tidak baik	0	15	1

Berdasarkan hasil analisis terhadap 25 butir soal, sebanyak 53 dari 75 pengecoh (70,67%) dinyatakan berfungsi efektif, sementara 22 pengecoh lainnya (29,33%) tidak efektif. Jika ditinjau per butir soal, 9 butir (36%) termasuk di dalam kategori sangat baik, 11 butir soal (44%) masuk pada kategori baik, 4 butir soal (16%) masuk pada kategori kurang baik, dan 1 butir soal (4%) masuk pada kategori tidak baik. Sehingga dengan demikian, secara keseluruhan sebanyak 80% butiran soal mempunyai pengecoh yang berfungsi dalam kategori baik hingga sangat baik.



Gambar 3. Persentase Efektivitas Pengecoh

Pengecoh dinyatakan berfungsi secara efektif ketika terdapat paling sedikit 5% dari seluruh peserta tes memilih jawaban tersebut (Arikunto, 2018). Pengecoh yang tidak berfungsi secara langsung menurunkan kualitas soal karena memperbesar peluang peserta menjawab benar secara tebakan, sehingga soal kehilangan kemampuannya dalam mengukur penguasaan konsep secara akurat (Novriyanti et al., 2024). Butir soal nomor 15 yang seluruh pengecohnya tidak berfungsi menjadi perhatian khusus, karena secara psikometrik butir ini hanya beroperasi layaknya soal benar-salah yang menyamar sebagai pilihan ganda, sehingga perlu direvisi secara menyeluruh. Adapun butir nomor 3, 6, 14, dan 16 yang berkategori kurang baik juga memerlukan perbaikan pada konstruksi alternatif jawabannya.

Oleh karena itu, perbaikan perlu difokuskan pada butir soal nomor 3, 6, 14, 15, dan 16 dengan menganalisis pola kesalahan peserta didik guna mengidentifikasi

miskonsepsi yang paling sering muncul pada materi terkait. Miskonsepsi tersebut selanjutnya dapat dijadikan dasar dalam menyusun pengecoh baru agar instrumen tes mampu mengukur pemahaman konseptual peserta didik secara lebih optimal.

Secara umum, hasil penelitian ini menunjukkan kualitas pengecoh yang relatif baik. Temuan ini sejalan dengan penelitian Hestiwati & Lestari (2021) yang menemukan bahwa 80% pengecoh pada instrumen soal pilihan ganda matematika berfungsi secara efektif. Penelitian lain yang dilakukan oleh Gunawan & Asria (2023) pada mata pelajaran matematika juga menunjukkan pola distribusi efektivitas pengecoh yang bervariasi, yaitu 15% berkategori sangat baik, 30% baik, 40% kurang baik, dan 15% tidak baik. Kesamaan tersebut menunjukkan bahwa kualitas pengecoh sangat dipengaruhi oleh kemampuan penyusun soal dalam merancang alternatif jawaban yang *plausibel* dan sesuai dengan kesalahan umum peserta didik.

KESIMPULAN

Berdasarkan hasil analisis terhadap kualitas butir soal pada asesmen sumatif matematika yang dilakukan dengan pendekatan Teori Tes Klasik, dapat disimpulkan bahwa secara umum kualitas instrumen tergolong cukup baik, kendati masih dijumpai beberapa aspek yang memerlukan perbaikan. Dari perspektif tingkat kesukaran, mayoritas butir soal berada pada kategori sedang, yakni sebanyak 14 butir (56%), sementara 10 butir (40%) termasuk dalam kategori mudah, dan 1 butir (4%) tergolong sukar. Temuan ini mengindikasikan bahwa distribusi tingkat kesukaran belum sepenuhnya ideal, mengingat proporsi soal mudah masih cukup dominan sedangkan jumlah soal sukar tergolong sangat sedikit. Pada aspek daya pembeda, diperoleh hasil bahwa 14 butir soal (56%) termasuk dalam kategori baik hingga sangat baik, 7 butir (28%) dalam kategori cukup, dan 4 butir (16%) dalam kategori kurang baik. Sementara itu, dari aspek efektivitas pengecoh, diketahui bahwa 53 dari 75 pengecoh (70,67%) dinyatakan berfungsi secara efektif, serta 20 butir soal (80%) memiliki tingkat efektivitas pengecoh dalam kategori baik hingga sangat baik. Temuan-temuan tersebut mengindikasikan bahwa sebagian besar butir soal telah mampu melaksanakan fungsi pengukurannya dengan cukup baik, terutama dalam hal membedakan kemampuan peserta didik serta menyediakan alternatif jawaban yang efektif.

Meskipun demikian, masih terdapat sejumlah butir soal yang perlu direvisi, khususnya pada aspek tingkat kesukaran, daya pembeda, dan efektivitas pengecoh yang belum optimal. Oleh karena itu, penyusun soal perlu memperhatikan proporsi tingkat kesukaran, kualitas konstruksi soal, serta penyusunan pengecoh yang lebih *plausibel* agar instrumen evaluasi mampu mengukur kemampuan peserta didik secara lebih akurat dan proporsional. Penelitian selanjutnya disarankan untuk melakukan analisis butir soal

dengan cakupan sampel yang lebih luas serta meninjau aspek kualitas instrumen lainnya, seperti validitas dan reliabilitas, agar diperoleh gambaran kualitas tes yang lebih komprehensif.

UCAPAN TERIMA KASIH (PILIHAN)

Peneliti berterima kasih pada pihak sekolah atas dukungan dan izin yang diberikan selama proses penelitian ini berlangsung. Ucapan terima kasih khusus diberikan pada para guru matematika yang sudah membantu seluruh orang yang proses mengumpulkan data-data serta memberikan akses terhadap hasil penilaian kinerja sumatif siswa. Selain itu, peneliti juga berterima kasih pada siswa dan siswi kelas VIII A1 dan VIII B yang sudah turut serta memberikan partisipasinya pada penelitian ini, serta pada seluruh pihak yang melalui bantuan, antusiasme, atau bentuk dukungan lainnya, telah memberikan kontribusi penting bagi kelancaran penyelesaian penelitian dan penulisan artikel ini.

DAFTAR PUSTAKA

- Amanda, P., Bistari, & Kresnadi, H. (2025). Analisis Kualitas Butir Soal Asesmen Sumatif Akhir Semester Gasal Matematika Kelas IV Sekolah Dasar. *Jurnal Pengabdian Masyarakat Dan Riset Pendidikan*, 4(2), 7203–7210. <https://jerk.in.org/index.php/jerk.in/article/view/2973>
- Arikunto, S. (2018). *Dasar Dasar Evaluasi Pendidikan - Arikunto Suharsimi* (R. Damayanti (ed.); 3rd ed., Vol. 3). Bumi Aksara. <https://ipusnas2.perpusnas.go.id/read-book>
- Asri, S. A., Syahrul, & Ishak, S. (2024). Analisis Kualita Butir Soal Ulangan Akhir Semester Genap Matematika Kelas VIII SMPN SINJAI. *Pedagogy*, 10(1), 120–133.
- Auliya, A. S., Hikmah, N., Tyaningsih, R. Y., & Turmuzi, M. (2025). Analisis Kualitas Butir Soal Asesmen Tengah Semester Genap Mata Pelajaran Matematika Kurikulum Merdeka Kelas VII SMP Negeri 2 Kuripan Tahun Ajaran 2024/2025. *Pendas : Jurnal Ilmiah Pendidikan Dasar*, 10(2), 210–224.
- Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Rinehart and Winston.
- Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of Educational Measurement. In Educational Researcher (Fifth Edit)*. Rajkamal Electirc Press. <https://doi.org/https://doi.org/10.2307/1175572>
- Fitriani, N. (2021). Penilaian Hasil Proses Belajar Mengajar Cetakan ketujuh belas. *Paedagoria : Jurnal Kajian, Penelitian Dan Pengembangan Kependidikan*, 12(2), 199–205.
- Gunawan, G., & Asria, L. (2023). Analisis Butir Soal Penilaian Tengah Semester (PTS) Matematika Kelas XI Berdasarkan Teori Klasik. *MATH LOCUS: Jurnal Riset Dan Inovasi Pendidikan Matematika*, 4(1), 1–11.

<https://doi.org/10.31002/mathlocus.v4i1.3177>

- Habiba Ulfahyana. (2022). Analisis Kualitas Butir Soal Ujian Sekolah Kelas Ix Mata Pelajaran Matematika. *Jurnal Penalaran Dan Riset Matematika*, 1(2), 70–78. <https://doi.org/10.62388/prisma.v1i2.212>
- Hestiwati, C., & Lestari, F. (2021). Pengembangan Instrumen Soal Higher Order Thinking Berbasis Islam Pada Materi Perbandingan Kelas VII SMP Wahdiyah Kediri. *AL FARABI: Jurnal Matematika Dan Pendidikan Matematika*, 2(2), 63–71.
- Hussain, I. S., & Garba, B. M. (2025). Application of Classical Test Theory Model on 2021 Mathematics Objectives Items of Entrance Examination into Senior Science and Technical Colleges in Kano State, Nigeria. *Rima International Journal of Education (RIJE)*, 4(4), 365–373.
- Istiyova, L. R. (2025). Analisis butir soal pengukuran kemampuan berpikir kritis materi pencemaran lingkungan melalui teori tes klasik dan Rasch model. *Bio-Pedagogi*, 14(1), 36. <https://doi.org/10.20961/bio-pedagogi.v14i1.88792>
- Kasiyani, S., & Isnaniah, S. (2026). Analisis Kualitas Butir Soal Asesmen Sumatif Mata Pelajaran Bahasa Indonesia di Sekolah. *GHANCARAN: Jurnal Pendidikan Bahasa Dan Sastra Indonesia*, 7(2), 522–539. <https://doi.org/10.19105/ghancaran.v7i2.21239>
- Kelley, T. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology*, 30(1), 17–24. <https://doi.org/10.1037/h0057123>
- Khotimah, H., & Hidayat, N. (2023). *Evaluasi Pembelajaran: Konsep, Model, dan Pengembangan Tes Hasil Belajar (Dilengkapi dengan Evaluasi Pembelajaran Kurikulum Merdeka)*. Literasi Nusantara Abadi.
- Magdalena, I., Hidayati, N., Dewi, R. H., Septiara, S. W., & Maulida, Z. (2023). Pentingnya Evaluasi dalam Proses Pembelajaran dan Akibat Manipulasinya. *Jurnal Pendidikan Dan Sains*, 3(5), 810–823. <https://doi.org/https://doi.org/10.58578/masaliq.v3i5.1379>
- Nababan, E. D., Nurfahrunnisa, A., & Rachmaniar. (2025). Analisis Kualitas Butir Soal Tes Kemampuan Akademik (TKA) SMA di Kota Merauke. *Journal of Authentic Research*, 4(2), 2075–2083.
- Nafs, H., Nyoman Sridana, Hikmah, N., & Soeprianto, H. (2023). Analisis Kualitas Butir Soal Ulangan Akhir Semester Genap Mata Pelajaran Matematika Kelas VII SMPN 6 Mataram Tahun Ajaran 2022/2023. *Jurnal Ilmiah Profesi Pendidikan*, 8(4), 2502–7069. <https://doi.org/10.30605/pedagogy.v10i1.5672>
- Novriyanti, E., Sarifah, I., & Riyadi. (2024). Analisis Kualitas Butir Soal Ujian Tengah Semester Biologi Umum Menggunakan Model Rasch. *JagoMIPA: Jurnal Pendidikan Matematika Dan IPA*, 4(4), 718–733. <https://doi.org/10.53299/jagomipa.v4i4.927>
- Nurazizah, S., Damayanti, R., Janan, T., Sitaresmi, P. D. W., & Nuryami. (2022). Analisis Kesalahan Siswa dalam Menyelesaikan Soal Matematika pada Materi Relasi dan Fungsi. *AL JABAR: Jurnal Pendidikan Dan Pembelajaran Matematika*, 1(2), 9–17. <https://doi.org/10.51836/je.v9i2.622>
- Pradita, E., Megawanti, P., & Yulianingsih. (2023). Analisis Tingkat Kesukaran, Daya

- Pembeda, dan Fungsi Distraktor PTS Matematika SMPN Jakarta. *Original Research*, 3(80), 109–118.
- Pratama, P., Windianti, W., Susanti, I., & Syahrial, S. (2024). Peran Guru dalam Melaksanakan Evaluasi Pembelajaran di Sekolah. *SIMPATI: Jurnal Penelitian Pendidikan Dan Bahasa*, 2(3), 109–121. <https://doi.org/10.59024/simpati.v2i3.821>
- Rahman, N., & Nursyam, A. (2026). Literature Review: Analisis Kemampuan Shared Metacognition Siswa dalam Menyelesaikan Masalah Matematika. *AL JABAR: Jurnal Pendidikan Dan Pembelajaran Matematika*, 5(1), 34–56.
- Rofiq, A., Atmadi, Ayuni, Nurazizah, S., & Nuryami. (2024). Matematika Pada Materi Bilangan Bulat Dengan Strategi Problem Based Learning Menggunakan. *AL JABAR: Jurnal Pendidikan Dan Pembelajaran Matematika*, 2(2), 72–82.
- Shofiah, S., Bachtiar, E., KD, D. P., Syahropi, H., Zaman, N., Salam, Nurhemah, N., Djollong, A. F., N, D. A. W., Alinuridin, Putri, M. F. J. L., Rostiani, T., Haryani, Nurhemah, N., Ladisa, S., Jannah, M., & Hidayat, N. (2023). *Dasar-Dasar Evaluasi Pembelajaran*. PT. Mifandi Mandiri Digital Redaksi:
- Sugiyono. (2024). *Metode Penelitian: Kuantitatif, Kualitatif, dan R&D* (30th ed.). Alfabeta.
- Summaryanta. (2021). *Teori Tes Klasik dan Teori Respon Butir Konsep & Contoh Penerapannya* (Vol. 2). C.V. Convident.