

**ANALISIS KUALITAS BUTIR SOAL PADA UJI COBA EVALUASI
PEMBELAJARAN MATEMATIKA TINGKAT SMA**

ICCA WIDYA SURBAKTI

Universitas Negeri Surabaya

e-mail: icca.23127@mhs.unesa.ac.id

ABSTRAK

Pendidikan merupakan salah satu komponen penting dalam kehidupan. Salah satu komponen pendidikan yang memiliki peran penting untuk mendukung proses pendidikan yaitu evaluasi. Penelitian ini bertujuan untuk menganalisis kualitas butir soal evaluasi pembelajaran pada mata pelajaran matematika melalui uji coba instrumen dengan jumlah subjek sebanyak 30 orang siswa Sekolah Menengah Atas (SMA) yang diambil secara acak. Metode penelitian yang digunakan kuantitatif deskriptif yang bertujuan untuk menganalisis kualitas butir soal tes evaluasi pembelajaran. Analisis melibatkan beberapa aspek, yaitu reliabilitas, daya pembeda, tingkat kesukaran, dan validitas butir soal. Data diperoleh dari skor jawaban peserta pada lima butir soal dengan bobot skor jawaban benar sebesar 20 dan salah sebesar 0. Dari hasil uji realibilitas menunjukkan nilai koefisien sebesar 0.79 mengindikasikan bahwa tes ini memiliki tingkat konsistensi yang tinggi, sehingga dapat diandalkan dalam evaluasi pembelajaran. Berdasarkan rangkuman hasil analisis, dua butir soal direkomendasikan untuk digunakan, sementara tiga butir lainnya perlu direvisi untuk meningkatkan kualitas instrumen evaluasi. Penelitian ini menegaskan pentingnya analisis butir soal guna memastikan instrumen evaluasi memiliki validitas, reliabilitas, daya pembeda, dan tingkat kesukaran yang memadai.

Kata Kunci: Analisis, Kualitas Butir Soal, Evaluasi Pembelajaran.

ABSTRACT

Education is one of the important components in life. One of the components of education that has an important role in supporting the educational process is evaluation. This study aims to analyze the quality of learning evaluation questions in mathematics subjects through instrument trials with a total of 30 subjects of Senior High School (SMA) students who were taken randomly. The research method used is quantitative descriptive which aims to analyze the quality of the learning evaluation test questions. The analysis involves several aspects, namely reliability, discriminatory power, level of difficulty, and validity of the test items. The data was obtained from the participants' answer scores on five questions with a score weighting of 20 for correct answers and 0 for incorrect answers. The results of the reliability test show a coefficient value of 0.79 indicating that this test has a high level of consistency, so it can be relied on in learning evaluation. Based on the summary of the analysis results, two questions are recommended for use, while the other three questions need to be revised to improve the quality of the evaluation instrument. This study emphasizes the importance of item analysis to ensure that evaluation instruments have adequate validity, reliability, discriminatory power, and level of difficulty.

Keywords: Analysis, Quality of Question Items , Learning Evaluation.

PENDAHULUAN

Pendidikan merupakan alah satu elemen penting dalam kehidupan. Pendidikan dapat berjalan dengan lancar dan optimal jika semua komponennya berfungsi dengan baik. Salah satu komponen yang mendukung proses pendidikan adalah evaluasi, yang bahkan menjadi hal yang sangat penting di dalamnya. Evaluasi dibutuhkan dalam pendidikan sebagai sarana untuk mengendalikan, menjamin, dan menetapkan mutu pendidikan. Evaluasi dilakukan secara



menyeluruh di setiap jalur atau jenjang pendidikan, menjadikannya sebagai inti dari pendidikan itu sendiri. Dengan kata lain, ketika kita membicarakan pendidikan, kita tidak bisa lepas dari evaluasi Pendidikan (Sari, Utomo, & Sumarwati, 2022).

Dalam proses pembelajaran, evaluasi sangat penting bagi guru untuk mengukur sejauh mana pemahaman siswa dalam menyelesaikan masalah matematika. Menurut Undang-Undang tahun 2005, pasal 10 ayat 1 mengenai guru dan dosen, seorang guru diharapkan menguasai empat kompetensi, salah satunya adalah kompetensi pedagogik. Dalam kompetensi ini, guru diharuskan melakukan evaluasi terhadap hasil belajar siswa. Tujuannya adalah untuk menilai perkembangan belajar siswa, memperbaiki program pengajaran yang telah dilaksanakan, serta memastikan apakah siswa telah menguasai materi yang diajarkan atau belum (Yusuf, 2020).

Selain itu, evaluasi pembelajaran juga memiliki peran penting dalam menilai kualitas proses pembelajaran serta mengukur sejauh mana tujuan pembelajaran telah tercapai. Salah satu instrumen yang digunakan dalam evaluasi adalah tes, yang terdiri dari sejumlah butir soal yang harus memenuhi kriteria kualitas tertentu agar hasil evaluasi dapat dipertanggungjawabkan.

Menurut Salmina dan Fadillah dalam (Ulfahyana, 2022) untuk menilai kualitas tes yang telah disusun dan diberikan kepada siswa, seharusnya dilakukan analisis soal dengan tujuan untuk mengidentifikasi butir-butir soal yang baik dan yang kurang baik. "Evaluasi melalui analisis butir soal sangat berguna untuk menilai soal-soal yang berkualitas sehingga layak dijadikan sebagai tolok ukur keberhasilan pembelajaran siswa. Analisis butir soal dapat dihitung melalui beberapa unsur, seperti Validitas, Reliabilitas, Tingkat Kesukaran, Daya Pembeda, dan Fungsi Pengecoh. Dengan analisis ini, dapat diidentifikasi soal-soal yang baik, yang perlu direvisi, atau yang sebaiknya dibuang, serta soal mana yang layak dimasukkan ke dalam bank soal".

Perangkat soal yang digunakan di sekolah selama ini diterapkan tanpa adanya jaminan kualitas soal, baik dari segi rasional maupun empirik. Padahal, menurut teori evaluasi, hal ini seharusnya dilakukan sebelum soal diberikan kepada peserta tes. Secara rasional, suatu perangkat soal seharusnya memenuhi kualitas isi atau konten materi yang akan diuji. Hal tersebut dapat dicapai dengan melakukan analisis yang melibatkan penilaian dari para ahli yang dianggap kompeten dalam substansi materi yang akan diuji. Secara empirik, perangkat soal yang akan diujikan seharusnya memenuhi kualitas karakteristik butir soal, seperti pengukuran tingkat kesulitan, daya pembeda, efektivitas pengecoh, serta sejauh mana butir soal tersebut valid secara internal. Selain itu, perangkat soal yang akan digunakan juga harus terjamin keandalannya (*reliable*) untuk diterapkan (Ulfahyana, 2022). Karena perangkat soal bisa dijadikan tolok ukur keberhasilan suatu pembelajaran, maka berangkat dari hal tersebut penulis ingin melakukan analisis kualitas butir soal pada uji coba evaluasi pembelajaran matematika tingkat Sekolah Menengah Atas (SMA). Melalui analisis ini, dapat diketahui apakah butir soal tersebut layak digunakan atau memerlukan revisi.

METODE PENELITIAN

Penelitian ini menggunakan metode kuantitatif deskriptif yang bertujuan untuk menganalisis kualitas butir soal tes evaluasi pembelajaran berdasarkan beberapa kriteria, yakni reliabilitas, validitas, daya pembeda, dan tingkat kesukaran soal. Subjek dalam penelitian ini diambil secara acak melibatkan 30 orang siswa Sekolah Menengah Atas (SMA) dengan 5 butir soal sebagai instrumen tes. Data yang diperoleh dari hasil tes peserta mencakup jumlah jawaban benar, jawaban salah, dan skor total setiap peserta akan dianalisis menggunakan program Anates.

Analisis reliabilitas dilakukan untuk mengetahui sejauh mana instrumen tes dapat menghasilkan hasil yang konsisten dengan menggunakan rumus korelasi antar skor ganjil-genap. Reliabilitas berkaitan dengan kepercayaan. Dimana suatu tes memiliki tingkat



kepercayaan yang tinggi, jika tes yang digunakan memberikan hasil yang tetap. Uji reliabilitas bertujuan untuk mengetahui tingkat konsistensi dari sebuah tes yang diujikan, untuk menentukan nilai reliabilitas suatu tes bisa menggunakan formula Kuder-Richardson KR-20 (Arikunto, 2009).

Validitas butir soal diukur melalui korelasi antara skor butir dengan skor total, yang digunakan untuk mengetahui sejauh mana setiap butir soal mampu mengukur apa yang seharusnya diukur. Hasil korelasi menunjukkan beberapa butir soal memiliki validitas sangat signifikan, signifikan, maupun cukup. Sedangkan daya pembeda butir soal dianalisis dengan membandingkan jumlah jawaban benar antara kelompok atas dan kelompok bawah, yang masing-masing terdiri dari 8 subjek. Perbedaan ini dinyatakan dalam indeks daya pembeda yang dikategorikan sebagai sangat baik, baik, cukup, atau kurang. Sedangkan tingkat kesukaran dihitung berdasarkan persentase jumlah peserta yang menjawab benar pada setiap butir soal, yang diklasifikasikan ke dalam kategori sangat mudah, mudah, dan cukup sulit.

Hasil analisis dari semua kriteria digunakan sebagai dasar untuk menentukan kualitas butir soal, apakah digunakan, direvisi, atau dibuang. Melalui pendekatan ini, penelitian diharapkan dapat memberikan gambaran yang jelas mengenai kualitas soal serta langkah perbaikan yang diperlukan untuk meningkatkan efektivitas evaluasi pembelajaran.

HASIL DAN PEMBAHASAN

Hasil

A. Skor Data Dibobot

Uji coba ini dilakukan terhadap 30 subyek untuk mengukur tingkat keberhasilan peserta dalam menjawab soal dengan jumlah total sebanyak 5 butir. Penilaian dilakukan dengan memberikan bobot tertentu pada setiap jawaban, yaitu:

1. Jawaban benar diberi bobot 20 poin.
2. Jawaban salah atau kosong tidak mendapatkan bobot (0 poin).

Skor asli dihitung berdasarkan jumlah jawaban benar, sedangkan skor berbobot diperoleh dengan mengalikan skor asli dengan bobot 20. Oleh karena itu, skor maksimum yang dapat diperoleh setiap peserta adalah 100 (5 jawaban benar x 20). Pada tabel 1. Menunjukkan adalah hasil perhitungan untuk setiap peserta skor data dibobot

Tabel 1. Hasil Skor Data Dibobot

No Urt	No Subyek	Benar	Salah	Kosong	Skor Asli	Skor Bobot
1	1	5	0	0	5	100
2	2	4	1	0	4	80
3	3	3	2	0	3	60
4	4	5	0	0	5	100
5	5	2	3	0	2	40
6	6	5	0	0	5	100
7	7	4	1	0	4	80
8	8	3	2	0	3	60
9	9	5	0	0	5	100
10	10	4	1	0	4	80
11	11	5	0	0	5	100
12	12	2	3	0	2	40
13	13	3	2	0	3	60
14	14	5	0	0	5	100

15	15	4	1	0	4	80
16	16	3	2	0	3	60
17	17	5	0	0	5	100
18	18	4	1	0	4	80
19	19	5	0	0	5	100
20	20	2	3	0	2	40
21	21	4	1	0	4	80
22	22	3	2	0	3	60
23	23	5	0	0	5	100
24	24	5	0	0	5	100
25	25	4	1	0	4	80
26	26	3	2	0	3	60
27	27	5	0	0	5	100
28	28	4	1	0	4	80
29	29	3	2	0	3	60
30	30	5	0	0	5	100

- Jumlah Subyek : 30
- Butir Soal : 5
- Bobot untuk jawaban benar : 20
- Bobot untuk jawaban salah : 0

B. Reliabilitas Tes

Reliabilitas adalah sejauh mana suatu tes memberikan hasil yang konsisten ketika digunakan berulang kali pada subjek yang sama atau dalam kondisi yang sebanding. Reliabilitas sangat penting dalam pengukuran karena menunjukkan stabilitas dan konsistensi suatu alat ukur.

Tabel 2. Hasil Korelasi antara Skor Ganjil dan Skor Genap

NO. URUT	NO. SUBYEK	SKOR GANJIL	SKOR GENAP	SKOR TOTAL
1	1	3	2	5
2	2	3	1	4
3	3	2	1	3
4	4	3	2	5
5	5	0	2	2
6	6	3	2	5
7	7	3	1	4
8	8	2	1	3
9	9	3	2	5
10	10	3	1	4
11	11	3	2	5
12	12	1	1	2
13	13	2	1	3
14	14	3	2	5
15	15	2	2	4
16	16	2	1	3
17	17	3	2	5
18	18	3	1	4



19	19	3	2	5
20	20	1	1	2
21	21	3	1	4
22	22	2	1	3
23	23	3	2	5
24	24	3	2	5
25	25	2	2	4
26	26	2	1	3
27	27	3	2	5
28	28	2	2	4
29	29	2	1	3
30	30	3	2	5

- Rata-rata: 4,20
- Simpang Baku: 1,12
- KorelasiXY: 0,66
- Reliabilitas Tes: 0,79

Dalam tes ini, skor tertinggi adalah 5, yang dicapai oleh beberapa peserta seperti Ahmad, Citra, dan lainnya, sementara skor terendah adalah 2, yang dicapai oleh peserta seperti Dewi, Kartika, dan Santi.

C. Analisis Daya Pembeda

Dari analisis ini, dapat diidentifikasi bahwa kelompok unggul berhasil menjawab semua soal dengan benar, menunjukkan pemahaman yang baik terhadap materi yang diuji. Sementara itu, kelompok asor memerlukan perhatian khusus untuk meningkatkan kemampuan mereka, terutama melalui bimbingan dan pembelajaran tambahan untuk memahami materi yang dianggap sulit.

Tabel 3. Hasil Jawaban Benar Kelompok Atas

No.Urut	No Subyek	Skor	1	2	3	4	5
1	1	5	1	1	1	1	1
2	4	5	1	1	1	1	1
3	6	5	1	1	1	1	1
4	9	5	1	1	1	1	1
5	11	5	1	1	1	1	1
6	14	5	1	1	1	1	1
7	17	5	1	1	1	1	1
8	24	5	1	1	1	1	1

Jumlah Jawaban Benar: 8 | 8 | 8 | 8 | 8

Analisis terhadap kelompok unggul dan asor dalam tes menunjukkan adanya perbedaan performa peserta didik. Kelompok unggul terdiri dari subyek dengan skor sempurna, yaitu 5, pada setiap butir soal. Mereka menunjukkan kemampuan konsisten dalam menjawab seluruh soal dengan benar. Terdapat 8 peserta dalam kelompok unggul, seperti Ahmad, Citra, dan Wawan, yang masing-masing menjawab benar sebanyak 8 soal dari total 8.

Tabel 4. Hasil Jawaban Benar Kelompok Bawah

No.Urut	No Subyek	Skor	1	2	3	4	5
1	5	2	1	0	0	1	0
2	12	2	0	1	0	1	0
3	20	2	1	0	0	0	1
4	22	3	1	1	0	0	1
5	26	3	1	1	0	1	0
6	29	3	1	0	1	0	1
7	15	4	1	1	1	0	1
8	21	4	0	1	1	1	1

Jumlah Jawaban Benar: 6 | 5 | 3 | 4 | 5

Analisis terhadap kelompok asor mencakup peserta dengan skor lebih rendah, berkisar antara 2 hingga 4 pada setiap butir soal. Mereka cenderung memiliki jumlah jawaban benar yang lebih sedikit dibandingkan kelompok unggul. Contohnya, peserta seperti Dewi dan Kartika hanya memiliki skor 2 karena lebih banyak menjawab salah atau tidak menjawab. Berdasarkan data, jumlah jawaban benar pada kelompok ini lebih bervariasi, dengan jumlah jawaban benar masing-masing berkisar antara 3 hingga 6.

Tabel 5. Hasil Analisi Daya Pembeda

No Butir Baru	No Butir Asli	Kel. Atas	Kel. Bawah	Beda	Indeks DP (%)
1	1	8	6	2	25.00
2	2	8	5	3	37.50
3	3	7	2	5	62.50
4	4	6	1	5	62.50
5	5	8	6	2	25.00

Jumlah Subyek: 30

Kelompok atas/bawah (n): 8

Butir Soal: 5

D. Tingkat Kesukaran

Tabel 6. Hasil Analisis Tingkat Kesukaran

No Butir Baru	No Butir Asli	Jml Betul	Tkt. Kesukaran (%)	Tafsiran
1	1	26	86.67	Sangat Mudah
2	2	24	80.00	Mudah
3	3	18	60.00	Cukup Sulit
4	4	20	66.67	Mudah
5	5	25	83.33	Sangat Mudah

Tingkat kesukaran soal menunjukkan proporsi peserta didik yang menjawab benar pada setiap butir soal. Dari lima butir soal yang dianalisis, dua soal termasuk kategori sangat mudah (butir ke-1 dengan tingkat kesukaran 86,67% dan butir ke-5 dengan 83,33%), dua soal masuk kategori mudah (butir ke-2 dengan 80,00% dan butir ke-4 dengan 66,67%), serta satu soal berada di kategori cukup sulit (butir ke-3 dengan 60,00%).

E. Validitas

Tabel 7. Hasil Korelasi Antara Skor Butir dengan Skor Total

No Butir Baru	No Butir Asli	Korelasi	Signifikansi
1	1	0.458	-
2	2	0.312	-
3	3	0.782	Sangat Signifikan
4	4	0.664	Signifikan
5	5	0.556	Signifikan

Dari hasil analisis, butir ke-3 memiliki korelasi tertinggi sebesar 0.782 dan dinyatakan sangat signifikan. Butir ke-4 dan ke-5 memiliki korelasi sebesar 0.664 dan 0.556 yang termasuk signifikan. Sementara itu, butir ke-1 dan ke-2 memiliki korelasi yang lebih rendah, yaitu 0.458 dan 0.312, tanpa signifikansi yang tercatat.

F. Rangkuman Hasil Analisis Butir Soal

Tabel 8. Hasil Analisis Butir Soal

No	Butir Soal	Kriteria Validitas	Kriteria Daya Beda	Kriteria Tingkat Kesukaran	Koefisien Reliabilitas	Keputusan
1	1	Cukup	Cukup	Sangat Mudah	0.79	Direvisi
2	2	Rendah	Cukup Baik	Mudah	0.79	Direvisi
3	3	Sangat Tinggi	Sangat Baik	Cukup Sulit	0.79	Digunakan
4	4	Tinggi	Baik	Mudah	0.79	Digunakan
5	5	Cukup	Cukup	Sangat Mudah	0.79	Direvisi

Pembahasan

Analisis butir soal merupakan salah satu cara yang bisa digunakan untuk mengetahui tingkat kesalahan dalam membuat butir soal pada soal ujian, jadi dengan melakukan analisis butir soal ini akan diperoleh tes yang baik (Masullah., et all, 2024). Untuk menilai kualitas butir soal didasarkan tiga kriteria utama: validitas, daya beda, dan tingkat kesukaran. Dari hasil uji realibilitas, berdasarkan data yang dianalisis, korelasi antara skor ganjil dan genap (KorelasiXY) sebesar 0,66 menunjukkan hubungan yang cukup kuat antara dua bagian skor tersebut. Menurut Arikunto (2009) hasil uji ini ada dua jenis yaitu, 1) koefisien reliabilitas $\geq 0,70$, maka reliable atau nilai reliabilitas yang tinggi dan 2) koefisien reliabilitas $< 0,70$, maka tidak reliable atau nilai reliabilitas rendah. Koefisien reliabilitas yang didapatkan sebesar 0,79 mengindikasikan bahwa tes ini memiliki tingkat konsistensi yang tinggi, sehingga dapat diandalkan dalam evaluasi pembelajaran.

Daya pembeda adalah indikator penting dalam evaluasi soal untuk menentukan sejauh mana suatu butir soal mampu membedakan peserta didik berkemampuan tinggi dan rendah. Analisis daya pembeda mengkaji butir-butir soal dengan tujuan untuk mengetahui kesanggupan soal dalam membedakan siswa yang tergolong mampu (tinggi prestasinya) dengan siswa yang tergolong kurang atau lemah prestasinya (Sudjana, 2016). Dari lima butir soal yang dianalisis, butir ke-3 dan ke-4 memiliki daya pembeda tertinggi sebesar 62,50%, sehingga tergolong sangat baik dalam membedakan kemampuan peserta didik. Sebaliknya, butir ke-1 dan ke-5 memiliki daya pembeda terendah sebesar 25,00%, yang menunjukkan kurang efektif



dan memerlukan revisi. Butir soal dengan daya pembeda tinggi ($\geq 40\%$) dapat dipertahankan, sedangkan soal dengan daya pembeda rendah ($< 40\%$) perlu diperbaiki untuk meningkatkan kualitas pengukuran. Dengan demikian, daya pembeda dapat menjadi alat penting untuk memperbaiki kualitas soal dalam evaluasi pembelajaran.

Tingkat kesukaran soal menunjukkan proporsi peserta didik yang menjawab benar pada setiap butir soal. Tingkat kesukaran adalah peluang siswa menjawab benar butir soal, dan jika semakin banyak siswa yang menjawab soal dengan benar maka soal tergolong mudah, sebaliknya jika semakin sedikit siswa yang menjawab soal dengan benar maka soal tergolong sukar, soal dikatakan baik adalah soal yang tidak terlalu sukar atau mudah, dikarenakan dengan soal yang terlalu mudah tidak mampu meningkatkan kemampuan siswa, sedangkan soal yang terlalu sukar menjadikan siswa putus asa mengerjakannya (Halik dalam Masullah et al., 2024). Dari lima butir soal yang dianalisis, dua soal termasuk kategori sangat mudah (butir ke-1 dengan tingkat kesukaran 86,67% dan butir ke-5 dengan 83,33%), dua soal masuk kategori mudah (butir ke-2 dengan 80,00% dan butir ke-4 dengan 66,67%), serta satu soal berada di kategori cukup sulit (butir ke-3 dengan 60,00%). Hasil analisis ini menunjukkan bahwa sebagian besar soal tergolong mudah hingga sangat mudah. Idealnya, tingkat kesukaran soal perlu disesuaikan agar distribusi soal mencakup kategori mudah, sedang, dan sulit untuk memperoleh hasil evaluasi yang lebih akurat dalam mengukur kemampuan peserta didik.

Menurut Sudijono dalam Arbiatin & Mulabbiyah (2020), tingkat validitas butir soal tes adalah ketepatan yang dimiliki oleh butir soal dalam mengukur apa yang hendak diukur menggunakan butir soal tersebut. Butir soal dikatakan valid apabila skor butir soal yang bersangkutan terbukti memiliki korelasi positif yang signifikan dengan skor totalnya. Korelasi skor butir dengan skor total menunjukkan hubungan antara skor pada tiap butir soal dengan skor total peserta didik. Dari hasil analisis, butir ke-3 memiliki korelasi tertinggi sebesar 0.782 dan dinyatakan sangat signifikan. Butir ke-4 dan ke-5 memiliki korelasi sebesar 0.664 dan 0.556 yang termasuk signifikan. Sementara itu, butir ke-1 dan ke-2 memiliki korelasi yang lebih rendah, yaitu 0.458 dan 0.312, tanpa signifikansi yang tercatat. Korelasi tinggi menunjukkan bahwa butir soal memiliki kontribusi yang baik terhadap total skor. Sebaliknya, butir dengan korelasi rendah memerlukan revisi untuk memastikan keterkaitan yang lebih kuat dengan tujuan pengukuran. Analisis ini penting untuk menentukan validitas butir soal dalam evaluasi pembelajaran.

Analisis ini menunjukkan bahwa variasi skor antar peserta cukup kecil, dengan simpang baku yang rendah, sehingga instrumen tes dinilai memadai untuk mengevaluasi kemampuan subjek. Namun, terdapat beberapa skor rendah yang memerlukan perhatian lebih untuk perbaikan instrumen atau metode pembelajaran yang digunakan. Hasil analisis ini menegaskan pentingnya reliabilitas dalam memastikan bahwa hasil evaluasi pembelajaran dapat diinterpretasikan secara valid dan konsisten.

Berdasarkan hasil analisis butir soal yang telah dilakukan, berikut adalah temuan untuk setiap butir soal:

1. Butir soal 1 memiliki validitas yang cukup, daya beda yang cukup, dan tingkat kesukaran yang sangat mudah. Koefisien reliabilitasnya sebesar 0.79, yang menunjukkan bahwa soal ini masih memerlukan revisi agar lebih efektif.



2. Butir soal 2 menunjukkan validitas yang rendah, daya beda yang cukup baik, dan tingkat kesukaran yang mudah. Dengan koefisien reliabilitas 0.79, soal ini juga perlu direvisi untuk meningkatkan kualitasnya.
3. Butir soal 3 memiliki validitas yang sangat tinggi, daya beda yang sangat baik, dan tingkat kesukaran yang cukup sulit. Soal ini memiliki koefisien reliabilitas 0.79 dan dapat digunakan tanpa revisi karena memenuhi kriteria kualitas yang baik.
4. Butir soal 4 memiliki validitas yang tinggi, daya beda yang baik, dan tingkat kesukaran yang mudah. Soal ini juga dapat digunakan tanpa revisi karena sudah memenuhi standar kualitas soal yang baik.
5. Butir soal 5 memiliki validitas yang cukup, daya beda yang cukup, dan tingkat kesukaran yang sangat mudah. Meskipun memiliki koefisien reliabilitas yang baik (0.79), soal ini perlu revisi untuk meningkatkan efektivitasnya.

Secara keseluruhan, beberapa butir soal membutuhkan revisi untuk meningkatkan kualitas dan efektivitas pengukuran, sementara beberapa lainnya sudah memenuhi kriteria untuk digunakan tanpa perubahan lebih lanjut.

KESIMPULAN

Berdasarkan hasil analisis terhadap 30 subjek dan 5 butir soal, diketahui bahwa tes memiliki reliabilitas sebesar 0,79, yang menunjukkan bahwa tes ini cukup konsisten dan dapat diandalkan. Dari segi daya beda, soal nomor 3 dan 4 memiliki daya pembeda yang sangat baik, validitas tinggi, dan tingkat kesukaran yang cukup ideal, sehingga dapat digunakan tanpa revisi. Namun, soal nomor 1, 2, dan 5 memiliki daya beda yang rendah, validitas yang kurang signifikan, serta tingkat kesukaran yang tergolong terlalu mudah, sehingga memerlukan revisi agar dapat lebih efektif dalam mengukur kemampuan peserta didik.

Tingkat kesukaran secara keseluruhan didominasi oleh soal yang mudah hingga sangat mudah, dengan pengecualian soal nomor 3 yang cukup sulit. Hal ini menunjukkan bahwa soal-soal perlu disesuaikan agar tingkat kesukaran lebih merata dan dapat mengevaluasi kemampuan peserta didik secara lebih menyeluruh. Dengan perbaikan pada butir soal yang lemah, instrumen ini diharapkan dapat menjadi alat evaluasi yang lebih akurat, adil, dan relevan dengan tujuan pembelajaran. Tes yang optimal akan mampu memberikan gambaran kemampuan peserta didik secara menyeluruh dan membantu dalam pengambilan keputusan pembelajaran yang lebih baik.

DAFTAR PUSTAKA

- Arbiatin, E., & Mulabbiyah (2020). Analisis Kelayakan Butir Soal Tes Penilaian Akhir Semester Mata Pelajaran Matematika Kelas VI Di SDN 19 Ampenan Tahun Pelajaran 2019/2020. *l-Midad : Jurnal PGMI*, 12(2), 146-171. Website <http://journal.uinmataram.ac.id/index.php/elmidad>
- Arikunto, S., & Cepi Abdul Jabar, S. (2009). buku cepi - Copy.pdf.
- Masullah, B. D., Zuhry, L.H., Usman, L. H., & Maulana, L. M. G. (2024). Analisis Butir Soal Ujian Akhir Semester Mata Pelajaran Matematika SMP Negeri 6 Praya Timur. *Elips: Jurnal Pendidikan Matematika*, 5(2), 152-161. DOI: <https://doi.org/10.47650/elips.v5i2.1219>
- Sari, V. N. I., Utomo, A. P. Y., & Sumarwati. (2024). Kualitas Soal Bahasa Indonesia di SMP Muhammadiyah 1 Pontianak: Analisis Butir Soal. *Jurnal Pendidikan Bahasa dan Sastra Indonesia*, 11(2), 112-119. <https://journal.unnes.ac.id/sju/index.php/jpbsi>
- Sudjana, N. (2016). Penilaian hasil proses belajar mengajar. Bandung: Remaja Rosdakarya.



Yusuf, R. (2020). Analisis Kualitas Butir Soal Ujian Nasional Mata Pelajaran Matematika Sekolah Menengah Pertama. *Journal of Didactic Mathematics*, 1(3), 158-164. Doi: 10.34007/jdm.v1i3.417.

Ulfahyana, H. (2022). Analisis Kualitas Butir Soal Ujian Sekolah Kelas IX Mata Pelajaran Matematika. *PRISMA (Jurnal Penalaran dan Riset Matematika)*, 1(2), 70-78.