

## Analisis Butir Soal Penilaian Tengah Semester (PTS) Matematika Kelas XI Berdasarkan Teori Klasik

Gunawan<sup>1a)</sup>, Lailatul Asria<sup>2b)</sup>

<sup>1</sup>Program Studi Magister Pendidikan Matematika, FMIPA, Universitas Negeri Yogyakarta, Jalan Colombo No. 1 Yogyakarta, DIY, Indonesia

<sup>2</sup>Program Studi Pendidikan Matematika, FKIP, Universitas Tidar, Jalan Kapten Suparman No. 39 Magelang, Jawa Tengah, Indonesia

e-mail: <sup>a)</sup>gunawan.2023@student.uny.ac.id, <sup>b)</sup>lailatulasriya112@gmail.com

### Abstrak

Penelitian ini bertujuan untuk menganalisis kualitas butir soal Penilaian Tengah Semester (PTS) mata pelajaran matematika SMAN 1 Kota Mungkid. Penelitian ini merupakan penelitian deskriptif kuantitatif. Subjek penelitian ini adalah siswa kelas XI MIPA 4 dengan instrumen tes berupa soal pilihan ganda. Butir soal dianalisis berdasarkan teori klasik yaitu indeks kesukaran, daya pembeda, reliabilitas, dan efektivitas pengecoh. Hasil analisis indeks kesukaran terdapat 3 soal kategori sukar, 12 soal kategori sedang, dan 5 soal kategori mudah. Daya pembeda soal menunjukkan 16 soal kategori sangat baik, 1 soal kategori cukup baik, 2 soal kategori sedang, dan 1 soal kategori buruk. Analisis reliabilitas menunjukkan soal memenuhi kriteria reliabilitas dengan kategori baik. Efektivitas pengecoh pada soal menunjukkan 3 soal kategori sangat baik, 6 soal kategori baik, 8 soal kategori kurang baik, 2 soal kategori tidak baik, dan 1 soal kategori sangat tidak baik. Secara keseluruhan terdapat 8 butir soal memenuhi kriteria soal sangat baik dan dapat digunakan kembali.

**Kata Kunci:** analisis butir soal, PTS, teori klasik

## *Item Analysis of Mathematics Mid-Semester Assessment for Class XI Based on Classical Theory*

### Abstract

*This study aims to analyze the quality of the Mid Semester (PTS) items in mathematics at SMAN 1 Kota Mungkid. This research was quantitative and descriptive. The subjects of this study were students of class XI MIPA 4, with the test instrument being multiple choice questions. The items were analyzed based on classical theory, namely the index of difficulty, discriminated power, reliability, and effectiveness of the distractor. The analysis results of the difficulty index are 3 questions in the difficult category, 12 questions in the medium category, and 5 questions in the easy category. The discriminated power of the items showed 16 items in the very good category, 1 item in the fairly good category, 2 items in the moderate category, and 1 item in the bad category. The reliability analysis showed that the questions met the reliability criteria in the good category. The effectiveness of the distractor on the questions showed 3 items in the very good category, 6 items in the good category, 8 items in the poor category, 2 items in the bad category, and 1 item in the very bad category. Overall, there were 8 items that met the criteria of very good questions and could be reused.*

**Keywords:** item analysis, PTS, classical theory

### PENDAHULUAN

Pendidikan merupakan usaha sadar yang dilakukan oleh keluarga, masyarakat dan pemerintah melalui bimbingan,

pembelajaran dan/atau latihan yang berlangsung di sekolah dan luar sekolah (Sagala, 2014). Menurut Purwati dkk. (2021), pendidikan dianggap sebagai salah

satu penentu kemajuan suatu bangsa sehingga layak mendapat perhatian oleh seluruh pihak mulai dari dalam pelaksanaan kebijakan hingga dalam tahapan evaluasi. Salah satu bentuk evaluasi dalam pendidikan sekolah menengah adalah penilaian tengah semester.

Berdasarkan Anggraena dkk. (2022) pada Panduan Pembelajaran dan Asesmen Pendidikan Anak Usia Dini, Pendidikan Dasar, dan Menengah, penilaian tengah semester merupakan bagian dari asesmen yang digunakan untuk mencari bukti ataupun dasar pertimbangan tentang ketercapaian tujuan pembelajaran. Asesmen atau sistem penilaian yang berkualitas berpengaruh pada peningkatan kualitas pendidikan. Oleh karena itu, peningkatan asesmen perlu juga untuk selalu dilaksanakan.

Menurut Irmayta, Rudibyani, dan Efkar (2018), asesmen adalah sebuah proses yang dilakukan oleh guru untuk mengambil suatu keputusan terhadap siswa. Semakin berkualitasnya asesmen, maka akan semakin mudah guru memahami kelebihan maupun kelemahan dari siswa. Karena dengan mengetahui kelemahan siswa, maka guru dapat melakukan evaluasi proses pembelajaran yang telah dilakukan dengan memanfaatkan kelebihan dari siswa.

Ragam bentuk asesmen pada jenjang pendidikan meliputi banyak hal, salah satunya ialah tes. Menurut Verschoor (2007), tes adalah sebuah alat untuk mengukur kemampuan peserta ujian. Sedangkan Suwanto (2022) berpendapat bahwa tes prestasi hasil belajar adalah tes yang digunakan untuk mengukur pengetahuan dan pemahaman siswa terhadap materi pelajaran. Untuk mengukur kemampuan siswa dalam menguasai materi matematika, SMAN 1 Kota Mungkid menerapkan suatu bentuk tes atau evaluasi

melalui ujian Penilaian Tengah Semester (PTS). SMAN 1 Kota Mungkid mengembangkan tipe soal PTS berbentuk tes objektif (pilihan ganda). Tipe butir soal pilihan ganda adalah suatu tipe butir soal yang memiliki alternatif jawaban lebih dari dua pilihan (Ratnaningsih, Isfarudi, & Soleiman, 2011).

*Multiple choice test* terdiri atas bagian keterangan (*stem*) dan bagian kemungkinan jawaban atau alternatif (*options*). Kemungkinan jawaban (*option*) terdiri atas satu jawaban yang benar yaitu kunci jawaban dan beberapa pengecoh (*distractor*). Pembentuk soal pilihan ganda (*multiple choice*) dibagi menjadi dua bagian, yaitu bagian keterangan (*stem*) dan jawaban alternatif (*option*). *Stem* soal bisa berupa pernyataan atau pertanyaan, sedangkan *option* soal terdiri atas beberapa alternatif, satu jawaban benar dan yang lain berupa jawaban pengecoh (*distractor*) (Mulyadi, 2010).

Jumlah alternatif jawaban dalam pilihan ganda yang digunakan di SMAN 1 Kota Mungkid adalah 5 (lima) pilihan. Ratnaningsih dan Isfarudi (2013) menyatakan bahwa soal yang baik harus memiliki pengecoh yang relatif homogen, sehingga sulit ditebak oleh siswa. Pembentuk soal yang memiliki *option* pengecoh yang heterogen membuat soal kurang berarti atau menjadi lemah, karena siswa cenderung lebih mudah menebak jawaban yang benar (Ratnaningsih, dkk., 2011). Hal tersebut dapat dimaknai, bahwasanya soal pada akhirnya tidak dapat membedakan antara siswa yang pandai dan kurang pandai dari hasil belajarnya.

Soal tes yang bermutu dapat membantu siswa dalam meningkatkan kemampuan dalam memahami materi matematika. Semakin tinggi kemampuan siswa dalam memahami materi matematika,

semakin besar pula kesempatan siswa dalam menjawab benar suatu soal dan mencapai tujuan kompetensi yang ditetapkan. Begitupun sebaliknya semakin rendah kemampuan siswa dalam memahami materi matematika, semakin kecil pula kesempatan siswa dalam menjawab benar suatu soal dan mencapai tujuan kompetensi yang ditetapkan. Sehingga untuk dapat melihat soal tersebut berguna atau tidak, maka perlu dilaksanakan analisis butir soal dalam pembuatan soal tes penilaian tengah semester.

Pengukuran kualitas sebuah perangkat tes dapat dilalui dengan melaksanakan analisis kualitatif dan kuantitatif. Analisis kualitatif adalah analisis yang dilaksanakan sebelum perangkat tes diberikan kepada siswa dengan mempertimbangkan kesesuaiannya dengan aspek materi, konstruksi dan bahasa, sedangkan analisis kuantitatif dapat dilaksanakan dengan menggunakan pendekatan yaitu teknik teori uji klasik (*classical test theory*) (Mariati, 2009).

Analisis butir soal dengan teori uji klasik memperhatikan beberapa aspek meliputi tingkat kesukaran butir, daya pembeda butir, penyebaran *option* (pilihan) jawaban, dan reliabilitas skor tes (Safari, 2005). Menurut Mariati (2009), analisis butir soal dengan menggunakan teori uji klasik adalah analisis yang paling mudah walaupun memiliki keterbatasan dalam pelaksanaannya. Berdasarkan penjelasan tersebut, dapat diambil kesimpulan bahwa menganalisis kualitas butir soal dianggap memiliki kontribusi penting dalam kegiatan evaluasi hasil pembelajaran matematika.

Seorang guru harus mengetahui kualitas butir soal matematika yang telah disusun tersebut apakah sudah memiliki kualitas yang baik atau belum melalui analisis butir soal. Analisis butir soal dilakukan untuk menstandarisasikan butir soal yang telah disusun agar diperoleh informasi mengenai butir soal yang bermutu baik (Tobari, 2015).

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis soal Penilaian Tengah Semester (PTS) menggunakan teori uji klasik. Hal tersebut berguna untuk mengukur apakah item-item soal tersebut mempunyai tingkat kesulitan yang cukup dengan kemampuan siswa, mempunyai daya beda yang tinggi, kemudian faktor tebakan yang paling kecil, sehingga dapat memberikan informasi pengukuran yang akurat.

## **METODE**

Metode penelitian yang digunakan dalam penelitian ini adalah deskriptif kuantitatif. Subjek dalam penelitian adalah siswa kelas XI MIPA 4 SMA Negeri 1 Kota Mungkid Tahun Ajaran 2022/2023. Data yang dianalisis merupakan hasil jawaban PTS. Ukuran sampel yang diambil sebanyak 36 siswa. Instrumen tes yang digunakan adalah soal PTS dengan jumlah butir soal sebanyak 20 soal pilihan ganda, dimana setiap soal memiliki lima pilihan jawaban yaitu A, B, C, D, dan E. Soal terdiri dari dua bagian materi, yaitu nomor 1-10 adalah persamaan trigonometri, dan nomor 11-20 adalah soal jumlah dan selisih sinus cosinus. Lebih lanjut, rincian materi soal PTS berdasarkan kisi-kisi matematika kelas XI disajikan pada Tabel 1 berikut.

Tabel 1. Rincian Materi Soal PTS

Materi	No. Soal
Persamaan trigonometri fungsi sinus	1, 4, 7
Persamaan trigonometri fungsi cosinus	2, 5
Persamaan trigonometri fungsi tangen	3, 6
Persamaan trigonometri berbentuk persamaan kuadrat	8, 9, 10
Jumlah dan selisih dua sudut (Sinus)	11, 12, 15, 16
Jumlah dan selisih dua sudut (Cosinus)	13, 17, 18
Jumlah dan selisih dua sudut (Tangen)	14, 19, 20

Tahapan-tahapan yang dilakukan pada penelitian ini sebagai berikut.

### Menghitung Indeks Kesukaran Soal

Menentukan besarnya indeks kesukaran secara matematis dirumuskan oleh Saifuddin (2003) sebagai berikut.

$$P = \frac{n_1}{N} \quad (1)$$

Keterangan:

$P$  = indeks kesukaran butir

$n_1$  = jumlah peserta tes yang menjawab benar

$N$  = banyaknya siswa yang menjawab butir soal tersebut

Setelah data besarnya indeks kesukaran butir soal dihitung, kemudian data tersebut dikategorikan sebagai berikut (Allen & Yen, 1979).

Tabel 2. Kriteria Indeks Kesukaran Soal

Nilai $p$	Kriteria
$p = 0,0$	Terlalu sukar
$0,0 < p < 0,3$	Sukar
$0,3 \leq p \leq 0,7$	Sedang
$0,7 < p < 1,0$	Mudah
$p = 1,0$	Terlalu mudah

### 1. Menghitung Daya Pembeda

Teknik untuk menentukan nilai daya beda pada penelitian ini, dengan menggunakan rumus sebagai berikut.

$$DP = \frac{\sum A_K}{\sum A} - \frac{\sum B_K}{\sum B} \quad (2)$$

Keterangan:

$DP$  = Daya pembeda

$\sum A_K$  = Jumlah siswa yang menjawab benar pada kelompok atas

$\sum B_K$  = Jumlah siswa yang menjawab benar pada kelompok bawah

$\sum A$  = Jumlah siswa pada kelompok atas

$\sum B$  = Jumlah siswa pada kelompok bawah

Setelah ditentukan nilai daya beda, kemudian data tersebut diberikan patokan indeks daya beda pada Tabel 3 sebagai berikut (Ebel & Frisbie, 1986).

Tabel 3. Kriteria Indeks Daya Pembeda

Indeks Daya	Kriteria
$\geq 0,4$	Butir soal sangat baik, dapat diterima
$0,3 - 0,39$	Butir soal cukup baik, dapat diterima dengan revisi
$0,2 - 0,29$	Butir soal sedang, perlu direvisi dan menjadi sasaran perbaikan
$\leq 0,19$	Butir soal buruk, ditolak atau dibuang dan digantikan butir soal yang lain

### 2. Menghitung Reliabilitas Soal

Menghitung reliabilitas skor tes bertujuan untuk mengetahui tingkat ketepatan (*precision*) dan kekonsistenan skor tes. Indeks reliabilitas berkisar antara 0-1. Semakin tinggi koefisien reliabilitas suatu tes (mendekati 1), semakin tinggi pula ketepatannya.

Reliabilitas dapat dihitung menggunakan koefisien alfa dalam Crocker dan Algina (1986), yang dirumuskan sebagai berikut.

$$\alpha = \frac{n}{n-1} \left( 1 - \frac{\sum_{i=1}^n \sigma_i^2}{\sigma_x^2} \right) \quad (3)$$

Keterangan:

$n$  = jumlah butir soal

$\sigma_x^2$  = ragam skor soal

$\sigma_i^2$  = ragam skor per butir soal

Setelah ditentukan nilai reliabilitas, kemudian data tersebut diberikan patokan kriteria reliabilitas pada Tabel 4 berikut (Zein & Darto, 2012).

Tabel 4. Kriteria Reliabilitas Soal

Koefisien Reliabilitas	Kriteria
$0,80 < \alpha \leq 1,00$	Sangat baik
$0,60 < \alpha \leq 0,80$	Baik
$0,40 < \alpha \leq 0,60$	Cukup baik
$0,20 < \alpha \leq 0,40$	Buruk
$\alpha \leq 0,20$	Sangat buruk

### 3. Efektivitas Pengecoh

Dasar dalam menganalisis soal dapat dengan melaksanakan penyebaran pilihan jawaban. Hal tersebut bertujuan untuk mengetahui apakah jawaban yang tersedia berfungsi atau tidak. Menurut Safari (2005), suatu pilihan jawaban dapat dinyatakan berfungsi apabila tidak dipilih oleh 5% dari jumlah siswa dan dipilih oleh lebih banyak siswa yang belum memahami materi. Adapun kriteria efektivitas pengecoh untuk soal pilihan ganda dengan 5 pilihan jawaban adalah sebagai berikut (Rahayu & Djazari, 2016).

Tabel 5. Kriteria Efektivitas Pengecoh

Banyak Pengecoh yang Berfungsi	Kriteria
4	Sangat baik
3	Baik
2	Kurang baik
1	Tidak baik
0	Sangat tidak baik

## HASIL DAN PEMBAHASAN

Analisis butir soal ini didasarkan pada jawaban siswa kelas XI MIPA 4 pada Penilaian Tengah Semester (PTS) Gasal Tahun Ajaran 2022/2023 Mata Pelajaran Matematika. Adapun hasil PTS siswa adalah sebagai berikut.

Tabel 6. Hasil PTS Siswa

Data Statistik	Hasil
Banyaknya siswa	36
Nilai tertinggi	95
Nilai terendah	15
Rata-rata nilai	59,31

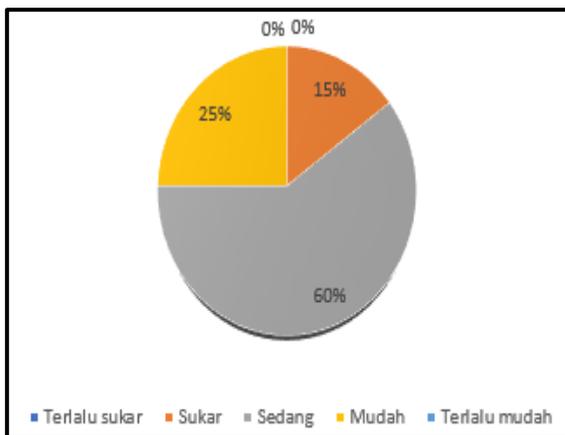
Kemudian, dari jawaban siswa dilakukan analisis butir soal dengan menggunakan teori klasik. Analisis yang dilakukan yaitu menentukan indeks kesukaran soal, daya pembeda soal, reliabilitas soal, dan penyebaran pilihan jawaban.

### 1. Indeks Kesukaran Soal

Setelah dilakukan analisis indeks kesukaran soal pada hasil jawaban siswa, didapatkan hasil berdasarkan Tabel 7 dan Gambar 1 berikut.

Tabel 7. Hasil Analisis Indeks Kesukaran Soal

Kategori	Butir Soal	Jumlah
Sukar	10, 18, 20	3
Sedang	5, 6, 7, 8, 9, 11, 12, 13, 14, 16, 17, 19	12
Mudah	1, 2, 3, 4, 15	5



Gambar 1. Persentase Indeks Kesukaran Soal

Berdasarkan Tabel 7 dan Gambar 1, diperoleh bahwa sebanyak 3 butir soal (15%) berada pada kategori sukar, 12 butir soal (60%) berada pada kategori sedang, 5 butir soal (25%) berada pada kategori mudah, dan tidak ada butir soal yang berada pada kategori sangat sukar dan sangat mudah. Proporsi tingkat kesukaran soal ini relevan dengan penelitian Son (2019) yang menyatakan bahwa 25% soal berada pada kategori sukar, 50% soal berada pada kategori sedang, dan 15% soal berada pada kategori mudah.

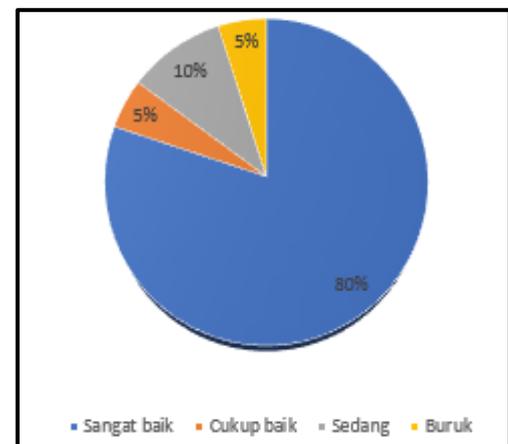
Kriteria tingkat kesukaran soal dapat digunakan ketika butir soal tidak terlalu sukar dan tidak terlalu mudah (Lestari & Yudhanegara, 2017). Berdasarkan hasil yang diperoleh, semua butir soal (100%) memenuhi kriteria tersebut yaitu soal berada pada kategori mudah, sedang, dan sukar.

## 2. Daya Pembeda Soal

Setelah dilakukan analisis daya pembeda soal pada hasil jawaban siswa, didapatkan hasil berdasarkan Tabel 8 dan Gambar 2 berikut.

Tabel 8. Hasil Analisis Daya Pembeda Soal

Kategori	Butir Soal	Jumlah
Sangat baik	2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 17, 18, 19	16
Cukup baik	3	1
Sedang	1, 15	2
Buruk	20	1



Gambar 2. Persentase Daya Pembeda Soal

Berdasarkan Tabel 8 dan Gambar 2, diperoleh bahwa sebanyak 16 butir soal (80%) memiliki daya pembeda sangat baik, 1 butir soal (5%) memiliki daya pembeda cukup baik, 2 butir soal (10%) memiliki daya pembeda sedang, dan 1 butir soal (5%) memiliki daya pembeda buruk.

Kriteria soal dapat digunakan apabila soal memiliki kriteria daya pembeda minimal pada kategori sedang (Dewi, Hariastuti, & Utami, 2019). Berdasarkan hasil yang diperoleh, 19 soal memenuhi kriteria tersebut yaitu berada pada rentang

kategori daya pembeda sedang, cukup baik, dan sangat baik. Sehingga soal tersebut dapat digunakan dengan catatan untuk soal pada kategori sedang dan cukup baik diperlukan revisi lebih lanjut. Sedangkan, terdapat 1 soal yang tidak memenuhi kriteria tersebut yaitu soal nomor 20 karena berada pada kategori buruk dan soal tersebut tidak dapat digunakan kembali dan perlu diganti soal yang lain. Hal ini relevan dengan penelitian Angriani dkk. (2021) yang menyatakan bahwa terdapat 1 soal memiliki daya beda kurang baik dan tidak dapat digunakan.

### 3. Reliabilitas Soal

Setelah dilakukan analisis reliabilitas soal pada hasil jawaban siswa, didapatkan hasil berdasarkan Tabel 9 berikut.

Tabel 9. Hasil Analisis Reliabilitas Soal

Koefisien Reliabilitas	Interpretasi
0,762	Baik

Berdasarkan Tabel 9, diperoleh hasil koefisien reliabilitas dari soal PTS adalah 0,762 dengan kategori baik. Instrumen soal layak digunakan apabila hasil koefisien reliabilitas minimal berada pada kategori cukup baik (Lestari & Yudhanegara, 2017). Sehingga berdasarkan hasil tersebut, soal PTS memenuhi batas minimal reliabilitas dan layak untuk digunakan. Hal ini relevan dengan penelitian Usman, Wiwid, dan Sulisti (2022). yang menyatakan bahwa soal ulangan matematika kelas XI memiliki tingkat reliabilitas sebesar 0,771 atau kategori baik.

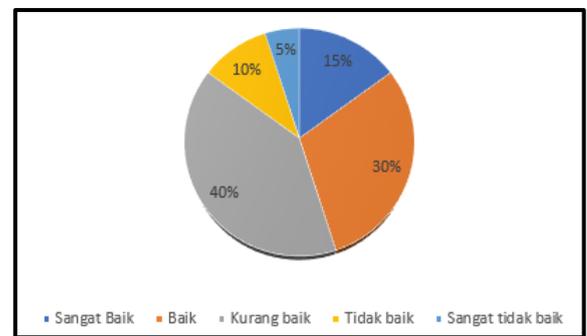
### 4. Efektivitas Pengecoh

Setelah dilakukan analisis efektivitas pengecoh pada hasil jawaban

matematika siswa kelas XI, didapatkan hasil berdasarkan Tabel 10 dan Gambar 3 berikut.

Tabel 10. Hasil Analisis Efektivitas Pengecoh

Kategori	Butir Soal	Jumlah
Sangat baik	6, 10, 20	3
Baik	3, 5, 9, 13, 16, 18	6
Kurang baik	2, 7, 8, 11, 12, 14, 17, 19	8
Tidak baik	1, 4	2
Sangat tidak baik	15	1



Gambar 3. Persentase Efektivitas Pengecoh

Berdasarkan Tabel 10 dan Gambar 3, didapat bahwa sebanyak 3 butir soal (15%) memiliki efektivitas pengecoh pada kategori sangat baik, 6 butir soal (30%) berada pada kategori baik, 8 butir soal (40%) berada pada kategori kurang baik, 2 butir soal (10%) berada pada kategori tidak baik, dan 1 butir soal (5%) berada pada kategori sangat tidak baik.

Soal dapat digunakan kembali apabila efektivitas pengecoh minimal berada pada kategori baik (Toifur & Setyowati, 2022). Soal dengan kategori kurang baik dapat digunakan kembali pada tes berikutnya dengan catatan soal tersebut diperbaiki agar meningkat menjadi baik dan dapat digunakan. Berdasarkan hasil yang

diperoleh, 17 butir soal (85%) memenuhi kriteria efektivitas pengecoh dan dapat digunakan kembali. Sedangkan, 3 butir soal yaitu nomor 1, 4, dan 15 tidak dapat digunakan kembali. Hal ini relevan dengan penelitian Budiman dan Jailani (2014) yang menyatakan bahwa efektivitas pengecoh yang berfungsi dengan baik sebesar 83,33% dari total semua soal.

### 5. Analisis Butir Soal secara Keseluruhan

Berdasarkan hasil analisis indeks kesukaran, daya pembeda, dan efektivitas pengecoh, dilakukan analisis secara keseluruhan dengan membagi menjadi 3 kriteria kualitas soal yang ditunjukkan pada Tabel 11 berikut (Sugiyono, 2010)

Tabel 11. Kriteria Kualitas Butir Soal

Kriteria Terpenuhi	Kualitas Butir Soal	Revisi	Bank Soal
3	Sangat baik	Tidak	Ya
2	Baik	Minor	Belum
1	Sedang	Sedang	Belum
0	Tidak baik	Dibuang	Tidak

Berdasarkan hasil analisis, didapatkan hasil secara keseluruhan sebagai berikut.

Tabel 12. Analisis Butir Soal Keseluruhan

Kriteria Butir Soal	Jumlah Soal	Nomor Soal
Sangat baik	8	3, 5, 6, 9, 10, 13, 16, 18
Baik	12	1, 2, 4, 7, 8, 11, 12, 14, 15, 17, 19, 20
Sedang	0	-
Tidak baik	0	-

Berdasarkan Tabel 12, diperoleh bahwa 8 butir soal (40%) berada pada kriteria butir soal sangat baik dan soal dapat digunakan kembali tanpa revisi. Kemudian, 12 butir soal (60%) berada pada kriteria butir soal baik dan belum bisa digunakan dan harus dilakukan revisi agar memenuhi tiga kriteria tersebut. Hal ini relevan dengan hasil penelitian Toifur dan Setyowati (2022) yang menyatakan bahwa terdapat 7 soal yang dapat digunakan kembali dan soal lainnya belum dapat digunakan karena masih diperlukan revisi dan beberapa ditolak karena tidak memenuhi kriteria.

### SIMPULAN

Berdasarkan uraian pada hasil dan pembahasan, dapat ditarik kesimpulan bahwa hasil analisis kriteria tingkat indeks kesukaran menunjukkan 3 butir soal (15%) berada pada kategori sukar, 12 butir soal (60%) kategori sedang, 5 butir soal (25%) kategori mudah. Hasil analisis daya pembeda menunjukkan 16 butir soal (80%) memiliki daya pembeda sangat baik, 1 butir soal (5%) cukup baik, 2 butir soal (10%) sedang, dan 1 butir soal (5%) buruk. Analisis reliabilitas soal berada pada kategori baik. Hasil analisis efektivitas pengecoh menunjukkan 3 butir soal (15%) memiliki efektivitas pengecoh pada kategori sangat baik, 6 butir soal (30%) baik, 8 butir soal (40%) kurang baik, 2 butir soal (10%) tidak baik, dan 1 butir soal (5%) sangat tidak baik. Analisis secara keseluruhan menunjukkan bahwa 8 butir soal (40%) berada pada kriteria butir soal sangat baik dan soal dapat digunakan kembali tanpa revisi. Kemudian, 12 butir soal (60%) berada pada kriteria butir soal baik dan harus dilakukan revisi agar memenuhi kriteria yang ada.

## DAFTAR PUSTAKA

- Allen, M. J. & Yen, W.M. (1979). *Introduction to measurement theory*. California: Brookd/Cole Publishing Company.
- Anggraena, Y., Ginanto, D., Felicia, N., Andiarti, A., Herutami, N., Alhapip, L., ... & Mahardika, R. L. (2022). *Panduan pembelajaran dan asesmen pendidikan anak usia dini, pendidikan dasar, dan menengah*. Jakarta: Badan Standar, Kurikulum, dan Asesmen Pendidikan Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia.
- Angriani, A. D., Mania, S., Alam. S., Rasyid, M. R., Kusumayanti, A. (2021). Analysis of final exam problems in mathematics semester of junior high school. *Alauddin Journal of Mathematics Education*, 3(1), 1-11.
- Budiman, A. & Jailani. (2014). Pengembangan instrumen asesmen Higher Order Thinking Skill (HOTS) pada mata pelajaran matematika SMP kelas VIII semester 1. *Jurnal Riset Pendidikan Matematika*, 1(2), 139-150.
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Forth Worth: Holt, Rinehart and Winston Inc.
- Dewi, S. S., Hariastuti, R. M., & Utami, A. U. (2019). Analisis tingkat kesukaran dan daya pembeda soal Olimpiade Matematika (OMI) tingkat SMP tahun 2018. *Transformasi-Jurnal Pendidikan Matematika & Matematika*, 3(1), 15-26.
- Ebel, R.L. & Frisbie, D.A. (1986). *Essentials of educational measurement*. New Jersey: Prentice Hall.
- Irmayta, E., Rudibyani, R. B., & Efkar, T. (2018). Pengembangan instrumen asesmen pengetahuan pada materi asam basa arrhenius. *Jurnal Pendidikan dan Pembelajaran Kimia*, 7(1), 63–76.
- Lestari, K. N. & Yudhanegara, M. R. (2017). *Penelitian pendidikan matematika*. Bandung: Refika Aditama.
- Mariati, I. (2009). Analisis butir soal dengan teori tes klasik (classical test theory) dan teori respons butir (item response theory) (Studi kasus: Soal ujian olimpiade sains provinsi bidang informatika 2009). *Pythagoras: Jurnal Matematika dan Pendidikan Matematika*, 5(2), 1-13.
- Mulyadi, M. (2010). *Evaluasi pendidikan: Pengembangan model evaluasi pendidikan agama di sekolah*. Malang: UIN Maliki Press.

- Purwati, H., Retnawati, H., Jailani, J., & Retnowati, T. H. (2021). Analisis karakteristik butir soal ujian nasional matematika SMP/MTs berdasarkan pendekatan teori tes klasik. *Jurnal Sains dan Edukasi Sains*, 4(2), 46-51.
- Rahayu, R. & Djazari, M. (2016). Analisis kualitas soal pra ujian nasional mata pelajaran ekonomi akuntansi. *Jurnal Pendidikan Akuntansi Indonesia*, 14(1), 84-94.
- Ratnaningsih, D. J., Isfarudi, I., & Soleiman, N. (2011). Analisis butir soal pilihan ganda ujian akhir semester mahasiswa di Universitas Terbuka dengan pendekatan teori tes klasik. *Jurnal Pendidikan Terbuka dan Jarak Jauh*, 12(2), 92-99.
- Ratnaningsih, D. J., & Isfarudi, I. (2013). Analisis butir tes objektif ujian akhir semester mahasiswa Universitas Terbuka berdasarkan teori tes modern. *Jurnal Pendidikan Terbuka dan Jarak Jauh*, 14(2), 98-109.
- Safari, M. A. (2005). *Teknik analisis butir soal instrumen tes dan non tes*. Jakarta: APSI.
- Sagala, S. (2014). *Konsep dan makna pembelajaran: Untuk membantu memecahkan problematika belajar dan mengajar*, Bandung: Alfabeta.
- Saifuddin. (2003). *Tes prestasi : Fungsi dan pengembangan pengukuran prestasi belajar*. Yogyakarta: Pustaka Pelajar.
- Son, A. L. (2019). Instrumentasi kemampuan pemecahan masalah matematis: Analisis reliabilitas, validitas, tingkat kesukaran dan daya beda butir soal. *Gema Wiralodra*, 10(1), 41-52.
- Sugiyono. (2010). *Metode penelitian pendidikan pendekatan kuantitatif, kualitatif, dan R&D*. Bandung: Alfabeta.
- Suwarto (2022). Karakteristik tes ilmu pengetahuan alam. *Jurnal Pendidikan*, 31(1), 109-120.
- Tobari, H. (2015). *Evaluasi soal-soal penerimaan pegawai baru dilengkapi dengan hasil penelitiannya*. Yogyakarta: Deepublish.
- 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.
- Verschoor, A. J. (2007). *A multiple objective test assembly approach for exposure control problems in computerized adaptive testing*. Measurement and Research Department Reports. Cito.

Zein, M. & Darto. (2012). *Evaluasi pembelajaran matematika*. Pekanbaru: Daulat Riau.