

## Analisis Kualitas Butir Soal Penilaian Akhir Tahun Mata Pelajaran Matematika di Sekolah Menengah Pertama

Aam Choitrotul Cholidiyah<sup>1\*</sup>

<sup>1</sup> Magister Pendidikan Matematika, UIN Maulana Malik Ibrahim Malang, Indonesia

Email: [aamcholidiah27@gmail.com](mailto:aamcholidiah27@gmail.com)

### Abstract

The purpose of this study was to measure the level of quality of End of Year Assessment (PAT) items in mathematics subjects. A quantitative approach with a descriptive type was used in this study. The participants involved were 30 Budi Mulia Junior High School students. The data collected were in the form of PAT question sheets, answer sheets, and answer keys. Data analysis was conducted using Microsoft Excel application on 30 items. The results showed that 36.67% of the questions were included in the valid category, the items tested were reliable, the difficulty level of the items was in the poor category, the differentiating power was in the weak category, and the distractors were functioning properly. Thus, the tested math items need to be improved and reviewed. This research contributes to teachers to revise questions that are not yet feasible and store good category questions in the question bank.

**Keywords:** item analysis; end-of-year assessment; mathematics; junior high school

### Abstrak

Tujuan penelitian ini yaitu untuk mengukur tingkat kualitas butir soal Penilaian Akhir Tahun (PAT) pada mata pelajaran matematika. Pendekatan kuantitatif dengan jenis deskriptif digunakan dalam penelitian ini. Partisipan yang dilibatkan yaitu 30 siswa SMP Budi Mulia. Data yang dikumpulkan berupa lembar soal PAT, lembar jawaban, dan kunci jawaban. Analisis data dilakukan berbantuan aplikasi Microsoft Excel pada 30 butir soal. Hasil penelitian menunjukkan bahwa 36,67% soal termasuk dalam kategori valid, butir soal yang diujikan telah reliabel, tingkat kesukaran butir soal termasuk dalam kategori kurang baik, daya pembeda termasuk dalam kategori lemah, dan distractor yang berfungsi dengan baik. Dengan demikian, butir-butir soal matematika yang diujikan perlu diperbaiki dan dikaji kembali. Penelitian ini berkontribusi bagi guru untuk merevisi soal yang belum layak dan menyimpan soal berkategori baik ke dalam bank soal.

**Kata Kunci:** analisis butir soal; penilaian akhir tahun; matematika; sekolah menengah pertama (SMP)

**How to Cite:** Cholidiyah, A.C. (2023). Analisis Kualitas Butir Soal Penilaian Akhir Tahun Mata Pelajaran Matematika di Sekolah Menengah Pertama. *Journal of Mathematics in Teaching and Learning*, 1 (2), 43-52.

## PENDAHULUAN

Pendidikan nasional bertujuan untuk membangun kualitas manusia serta mengembangkan potensi dan karakter individu (Indonesia, 2003). Kualitas pendidikan yang baik tidak hanya diukur melalui tujuan dan proses pembelajaran, tetapi juga dari hasil belajar yang didapatkan siswa (Arikunto, 2010). Aktivitas untuk menilai hasil belajar dikenal dengan istilah evaluasi atau penilaian hasil belajar (*asesment*) (Patriana et al., 2021; Sa'dijah et al., 2015). Dengan kata lain, penilaian hasil belajar berperan sebagai pilar yang membangun kualitas pendidikan ke arah yang lebih baik.

Penilaian hasil belajar menjadi aspek yang diperhitungkan dalam proses pendidikan (Sudjiono, 2005). Penilaian berperan dalam mengukur ketercapaian suatu tujuan pembelajaran (Anwar et al., 2022; Arikunto, 2010; Gronlund, 1998). Selain itu, penilaian hasil belajar menjadi tolak ukur dalam pengambilan keputusan bagi para pemangku kebijakan di dunia pendidikan (Ernest, 1991). Misalnya di tingkat sekolah, asesmen sumatif seperti Penilaian Tengah Semester (PTS) dan Penilaian Akhir Tahun (PAT) menjadi salah satu tahapan evaluasi yang ditujukan guna mengukur kompetensi siswa

Received April 10, 2023; Revised May 12, 2023; Accepted June 2, 2023

setelah melalui serangkaian proses pembelajaran selama satu tahun ajaran (Ardiansyah, 2021; Mania et al., 2020; Sa'dijah et al., 2015).

Salah satu elemen penting pada PAT yaitu kualitas butir soal yang digunakan (Mania et al., 2020). Butir soal yang baik perlu menggambarkan kemampuan siswa secara akurat (Arikunto, 2010). Keakuratan butir soal yang digunakan dikaji melalui analisis butir soal (Anas, 2006). Karenanya, analisis butir soal menjadi langkah penting untuk memastikan bahwa soal-soal yang digunakan dalam asesmen telah berkualitas dengan analisis uji validitas, uji reliabilitas, analisis daya pembeda, dan analisis tingkat kesulitan (Susanto et al., 2015; Tilaar et al., 2020).

Menurut Suzana (2018), pengabaian terhadap analisis butir soal mengakibatkan soal memiliki banyak kelemahan apabila digunakan untuk menilai hasil belajar. Kelemahan tersebut meliputi ketidaksesuaian dengan cakupan materi, opsi jawaban yang mudah ditebak, dan tingkat kesukaran yang tidak seimbang (Susanto et al., 2015; Suzana, 2018; Tilaar et al., 2020). Kelemahan tersebut menyebabkan hasil penilaian tidak mencerminkan kompetensi siswa secara objektif (Sa'dijah et al., 2015). Oleh karena itu, analisis butir soal berperan dalam mengetahui butir soal yang 'tidak layak' diujikan, sehingga asesmen dapat mengukur kompetensi siswa secara adil dan akurat.

Dalam lingkup pendidikan matematika, analisis butir soal memiliki peran yang esensial (Anwar et al., 2022; Suzana, 2018; Utomo, 2018). Dalam mempelajari matematika dibutuhkan pemahaman konsep serta keterampilan dalam berpikir dan pemecahan masalah (Alghar, 2022; Huinchahue et al., 2021; Kilpatrick et al., 2001). Karenanya, butir soal matematika yang diujikan perlu dirancang guna mengukur berbagai aspek kemampuan siswa, mulai dari pemahaman dan penerapan konsep hingga analisis dan evaluasi (Anwar et al., 2022; Hardianti, 2019). Tanpa asesmen yang baik, hasil penilaian tidak mencerminkan kemampuan siswa secara objektif (Sa'dijah et al., 2015).

Observasi awal dilakukan peneliti untuk melihat penerapan analisis butir soal matematika secara nyata. Peneliti mewawancarai satu guru matematika di sebuah SMP di Kabupaten Malang. Hasil wawancara menunjukkan guru jarang melakukan analisis butir soal dalam merancang butir soal ujian. Guru gemar memakai kembali soal ujian tahun sebelumnya. Tidak jarang guru menyadur dan 'memodifikasi' butir soal yang diperoleh dari lembar kerja siswa (LKS) maupun buku ajar tanpa kajian yang mendalam. Hal ini menunjukkan bahwa soal-soal ujian yang digunakan belum melalui proses analisis butir soal.

Hasil observasi awal selaras dengan berbagai penelitian sebelumnya. Seperti yang dipaparkan Susanto et al. (2015) bahwa beberapa guru SMA di Bandar Lampung tidak melakukan analisis butir soal matematika untuk ujian semester. Hamimi et al. (2020) melaporkan bahwa soal yang dibuat oleh beberapa guru matematika SMP di Susoh tidak dihitung analisis butir soalnya. Begitu juga yang dipaparkan Khasanah et al. (2023) bahwa beberapa guru Madrasah Aliyah di Singosari mengabaikan analisis butir soal dalam asesmen formatif dan sumatif. Temuan-temuan ini menunjukkan bahwa analisis butir soal sering diabaikan dalam pembelajaran matematika.

Berdasarkan uraian tersebut, penelitian ini bermaksud untuk melakukan analisis butir soal pada

soal matematika yang diterapkan di penilaian akhir tahun salah satu SMP di Kabupaten Malang. Penelitian ini akan mengevaluasi kualitas soal berdasarkan kriteria validitas, reliabilitas, daya beda, serta tingkat kesukaran. Dengan demikian, judul pada penelitian ini yaitu analisis kualitas butir soal penilaian akhir tahun mata pelajaran matematika di sekolah menengah pertama.

## METODE

Pendekatan kuantitatif dengan jenis deskriptif digunakan dalam penelitian ini. Pendekatan kuantitatif deskriptif bertujuan untuk menggambarkan karakteristik suatu populasi atau fenomena tertentu tanpa menguji hubungan antar variabel (Creswell & Creswell, 2017). Data yang diperoleh tergolong ke dalam data kuantitatif, yaitu nilai Penilaian Akhir Tahun (PAT) mata pelajaran matematika. Data yang dikumpulkan berupa lembar soal PAT, lembar jawaban, serta kunci jawaban. Partisipan dalam penelitian ini yaitu 30 siswa di SMP Budi Mulia tahun ajaran 2022/2023 di Kabupaten Malang. Evaluasi dilakukan terhadap 30 butir soal PAT matematika. Analisis data dilakukan dengan aplikasi Ms. Excel. Teknik analisis yang diterapkan terdiri atas uji validitas, uji reabilitas, tingkat kesukaran, analisis daya beda soal, dan analisis *distractor* (pengecoh). Adapun hasil analisis direpresentasikan dalam tabel dan yang diinterpretasikan secara deksriptif.

## HASIL DAN PEMBAHASAN

### Hasil Uji Validitas

Validitas butir soal diartikan sebagai ukuran sejauh mana suatu soal dapat mengukur apa yang seharusnya diukur. Kevalidan butir soal terpenuhi apabila memiliki korelasi signifikan dengan skor total. Skor pada butir soal mendukung peningkatan atau penurunan skor total. Dengan kata lain suatu butir soal memperoleh nilai validitas yang baik jika skor butir soal memiliki kesejajaran dengan skor total. Rumus korelasi *product moment* digunakan untuk menghitung validitas (Sudjiono, 2005).

$$r_{hitung} = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{\{N\sum X^2 - (\sum X)^2\}\{N\sum Y^2 - (\sum Y)^2\}}}$$

Keterangan

$r_{hitung}$ : koefisien korelasi

$N$ : banyak siswa

$X$ : skor pada butir soal

$Y$ : total skor

Hasil uji validitas butir soal PAT pada mata pelajaran matematika di SMP Budi Mulia tahun ajaran 2022/2023, yang dianalisis dengan aplikasi Ms. Excel, ditampilkan Tabel 1.

Tabel 1. Hasil uji validitas butir soal

No	Validitas	Nomor Butir Soal	Banyaknya	Persentase
1.	$\geq 0,361$ (Valid)	2, 3, 5, 9, 12, 15, 16, 21, 25, 26, 28	11	36,67%
2.	$< 0,361$ (Tidak Valid)	1, 4, 6, 7, 8, 10, 11, 13, 14, 17, 18, 19, 20, 22, 23, 24, 27, 29, 30	19	63,33%

Tabel 1 menunjukkan bahwa dari 30 butir soal, hanya 11 soal (36,67%) yang memenuhi kriteria validitas ( $r_{hitung} \geq 0,361$ ), sementara 19 soal (63,33%) dinyatakan tidak valid ( $r_{hitung} < 0,361$ ). Soal yang valid dapat digunakan sebagai bahan evaluasi karena telah memenuhi kriteria dan tidak menyimpang dari tujuan pengukuran. Sebaliknya, soal yang tidak valid perlu direvisi atau bahkan dihindari penggunaannya karena tidak memenuhi kriteria pengukuran yang sesuai.

Temuan ini selaras dengan Tilaar et al. (2020) yang menunjukkan 40% butir soal pada ujian akhir semester berada dalam kategori tidak valid. Penelitian Khasanah et al. (2023) menunjukkan 47% butir soal ujian harian berada dalam kelompok tidak valid. Di sisi lain, Arifin (2017) memaparkan tiga aspek yang mempengaruhi validitas butir soal, diantaranya aspek instrumen tes, aspek penskoran, dan aspek jawaban siswa. Selain itu, validitas soal dapat ditingkatkan dengan memastikan siswa membaca soal secara teliti, menjawab dengan cermat, dan memeriksa kembali jawaban mereka sebelum menyerahkan lembar jawaban (Gronlund, 1998).

### **Hasil Uji Reliabilitas**

Reliabilitas mengacu pada tingkat keterandalan suatu tes. Artinya reabilitas mengukur sejauh mana tes tersebut memberikan hasil yang konsisten ketika digunakan berulang kali. Uji reliabilitas diukur dengan metode *Kuder Richardson-20 (KR-20)* (Ntumi et al., 2023). Metode ini cocok untuk menguji reliabilitas soal dengan format jawaban benar (point 1) atau salah (point 0). Adapun rumus KR-20 yaitu

$$r_{11} = \left( \frac{k}{k-1} \right) \left( \frac{s^2 - \sum pq}{s^2} \right)$$

Keterangan

$r_{11}$ : koefisien reliabilitas internal untuk semua butir soal

$p$ : proporsi subjek yang benar dalam menjawab butir soal

$q$ : proporsi subjek yang salah dalam menjawab butir soal ( $q = 1 - p$ )

$\sum pq$ : jumlah dari hasil  $p$  dikali  $q$

$k$ : banyaknya butir soal

$s$ : deviasi standar

Hasil uji reliabilitas soal PAT mata pelajaran matematika di SMP Budi Mulia tahun ajaran 2022/2023, yang dianalisis menggunakan aplikasi Microsoft Excel, ditampilkan oleh Tabel 2.

Tabel 2. Hasil uji reabilitas butir soal

Hasil $r_{11}$	Hasil $r_{tabel}$	Keterangan
0,430	0,361	Reliabel

Berdasarkan Tabel 2, indeks reliabilitas  $r_{11}$  soal PAT mata pelajaran matematika yaitu 0,430 lebih besar daripada nilai  $r_{tabel}$  dengan nilai 0,361 pada taraf signifikansi 5% dengan  $N = 30$ . Dengan demikian, soal-soal PAT matematika tersebut tergolong reliabel. Artinya, soal yang digunakan memiliki konsistensi dalam mengukur kompetensi siswa.

Hasil tes reabilitas ini selaras dengan Khasanah et al. (2023) yang memaparkan soal ulangan harian telah memenuhi kaidah reabilitas. Penelitian Hanan et al. (2023) melaporkan bahwa butir soal cerita pada materi KPK dan FPB telah memenuhi aspek reabilitas. Reliabilitas yang baik menunjukkan bahwa soal telah dirancang sesuai standar evaluasi dan mampu mendukung penilaian yang objektif (Arikunto, 2010; Gronlund, 1998). Hal-hal yang mempengaruhi reabilitas yaitu tingkat kesukaran, banyaknya soal, dan objektivitas penskoran (Khasanah et al., 2023).

#### *Hasil Analisis Tingkat Kesukaran*

Tujuan dari analisis tingkat kesukaran yaitu mengelompokkan soal ke dalam kategori sulit, sedang, atau mudah (Arikunto, 2010). Suatu butir soal dianggap layak apabila tingkat kesukarannya tidak terlalu sulit atau terlalu mudah. Artinya, tingkat kesukaran yang proporsional menggambarkan soal yang berkualitas. Rumus yang digunakan untuk mengukur tingkat kesukaran yaitu

$$P = \frac{\bar{S}}{S_{maks}}$$

keterangan:

$P$ : indeks kesukaran

$S$ : skor rata-rata pada butir soal

$S_{maks}$ : skor maksimum untuk butir soal tersebut

Tabel 3 menunjukkan hasil analisis tingkat kesukaran soal PAT matematika di SMP Budi Mulia tahun ajaran 2022/2023, yang telah dianalisis menggunakan aplikasi Microsoft Excel.

Tabel 3. Hasil analisis tingkat kesukaran butir soal

Tingkat Kesukaran	Nomor Soal	Banyaknya	Persentase
Sangat Sulit	-	0	0%
Sulit	3, 4, 6, 7, 9, 10, 11, 12, 13, 14, 17, 18, 19, 20, 21, 22, 23, 24, 26, 27, 28, 29, 30	23	76,67%
Sedang	1, 5, 8, 15, 16, 25	6	20%
Mudah	2	1	3,33%
Sangat Mudah	-	0	0%

Berdasarkan Tabel 3, 23 soal (76,67%) berada dalam kategori sulit, enam soal (20%) termasuk dalam kategori sedang, dan satu soal (3,33%) termasuk dalam kategori mudah. Tidak ditemukan soal yang tergolong sangat sulit maupun sangat mudah. Hasil ini menunjukkan bahwa soal yang disajikan didominasi pada tingkatan sulit dan sedang, sehingga belum layak untuk diujikan. Temuan ini selaras dengan penelitian Febriani (2016) yang menunjukkan 47,5% soal berada pada tingkat tidak layak. Sedangkan Mania et al. (2020) yang menunjukkan 20%, 28%, dan 22% berada pada tingkat sangat sukar, sukar, dan mudah sehingga kualitas soal yang disajikan kurang baik.

Soal dalam kategori sulit berjumlah 23 butir (76,67%) dan mencakup nomor 3, 4, 6, 7, 9, 10, 11,

12, 13, 14, 17, 18, 19, 20, 21, 22, 23, 24, 26, 27, 28, 29, dan 30. Soal dengan tingkat kesukaran sulit sebaiknya dikurangi kuantitasnya dan tidak diperbanyak untuk ujian selanjutnya (Mania et al., 2020). Hal ini karena persentase tingkat kesukaran tidak sesuai proporsi 3-4-3 atau 30% sulit, 40% sedang, dan 30% mudah (Arikunto, 2010). Soal dalam kategori sedang berjumlah 6 butir (20%) dengan nomor 1, 5, 8, 15, 16, dan 25. Soal dengan tingkat kesukaran sedang sebaiknya diperbanyak, setidaknya sampai 40%. Karena soal dengan tingkat kesukaran sedang bersifat menantang bagi siswa dan melatih pemahaman konsep, sehingga tidak mematahkan semangat siswa dalam menyelesaikan soal (Halik et al., 2019; Khasanah et al., 2023).

Kemudian terdapat 1 soal (3.33%) dalam kategori mudah, yaitu nomor 2. Angka ini jauh di bawah standar, yaitu 20%-30% untuk soal mudah. Soal dengan tingkat kesukaran mudah sebaiknya diperbanyak. Hal ini untuk menambah motivasi siswa dalam mengerjakan soal (Arikunto, 2010). Soal berkategori mudah juga membantu siswa berkemampuan rendah dalam memahami konsep matematika yang telah dipelajarinya (Anas, 2006).

Sementara soal dengan tingkat kesukaran sangat sulit dan sangat mudah tidak ditemukan. Peneliti menyarankan agar setidaknya 5%-10% soal berada di kategori ini. Soal yang sangat sulit memberikan tantangan tersendiri bagi siswa berkemampuan tinggi (Arikunto, 2010). Soal yang sangat mudah memberikan motivasi bagi siswa berkemampuan rendah (Arifin, 2017). Dapat disimpulkan bahwa butir soal dalam PAT tersebut berada pada kategori yang kurang baik, karena tidak sesuai dengan proporsi tingkat kesulitan yang normal, yaitu 3-4-3 atau 3-5-2 (Arikunto, 2010).

#### **Hasil Analisis Daya Beda**

Kemampuan suatu butir soal untuk membedakan siswa yang memiliki kemampuan tinggi dan rendah disebut sebagai daya beda. Daya beda suatu soal yang tinggi menunjukkan semakin baik kemampuannya dalam membedakan siswa berdasarkan tingkat kemampuan mereka. Untuk menganalisis daya beda, digunakan rumus berikut

$$DB = \frac{BaB}{Ba} - \frac{BbS}{Bb}$$

Keterangan:

*DB* sebagai indeks daya beda

*BaB* sebagai banyak siswa yang menjawab benar di kelompok atas

*Ba* sebagai banyak siswa di kelompok atas

*BbS* sebagai banyak siswa yang menjawab benar di kelompok bawah

*Bb* sebagai banyak siswa di kelompok bawah.

Tabel 4 menunjukkan hasil analisis daya beda soal PAT matematika di SMP Budi Mulia tahun ajaran 2022/2023, yang telah dianalisis menggunakan aplikasi Microsoft Excel.

Tabel 4. Hasil analisis daya beda pada butir soal

Tingkat Daya Beda	Kategori	Nomor Soal	Banyaknya	Persentase
0,71-100	Sangat Kuat	-	0	0%
0,41-0,71	Kuat	25	1	3,34%
0,21-0,40	Sedang	1, 12, 15, 16, 21, 22, 26	7	23,33%
0-0,20	Lemah	2, 3, 5, 6, 7, 8, 9, 10, 11, 17, 18, 19, 23, 27, 28, 30	16	53,33%
< 0	Sangat Lemah	4, 13, 14, 20, 24, 29	6	20%

Hasil pada Tabel 4 menunjukkan persentase bahwa 3,34% soal tergolong dalam kategori kuat, 23,33% soal d tergolong dalam kategori sedang, 53,33% soal tergolong dalam kategori lemah, 20% soal tergolong dalam kategori sangat lemah, dan 0% soal tergolong dalam kategori sangat kuat. Banyaknya soal berkategori lemah menunjukkan butir soal yang digunakan tidak efektif dalam mengukur perbedaan kemampuan siswa. Hal ini mungkin disebabkan karena dua hal. Pertama, soal yang terlalu sulit, sehingga siswa berkemampuan tinggi pun kesulitan menjawab. Kedua, siswa di kelas tersebut mungkin banyak yang berkemampuan rendah, sehingga mereka belum mampu menyelesaikan soal-soal yang membutuhkan kemampuan analitis tinggi.

Hasil analisis daya beda ini diperkuat oleh Magdalena et al. (2021) yang memaparkan soal-soal dengan daya beda lemah sering kali menunjukkan ketidaksesuaian antara kompetensi siswa dengan tingkat kesulitan soal. Febriani (2016) menyatakan bahwa soal dengan daya beda terlalu lemah tidak layak digunakan dalam evaluasi pembelajaran. Menurut Anas (2006), daya beda sangat penting untuk memastikan bahwa soal dapat mengukur kompetensi yang sesuai dengan tujuan pembelajaran. Soal dengan daya beda tinggi menunjukkan bahwa soal tersebut mampu membedakan kemampuan siswa secara efektif dan relevan (Anas, 2006; Arifin, 2017).

#### **Hasil Analisis Distraktor (Pengecoh)**

Tujuan analisis distraktor atau pengecoh yaitu untuk menilai efektivitas setiap opsi jawaban dalam membedakan tingkat kemampuan siswa. Pengecoh dianggap efektif apabila lebih banyak dipilih oleh peserta tes dari kelompok bawah daripada oleh peserta tes dari kelompok atas. Sebaliknya, jika pengecoh lebih sering dipilih oleh peserta tes kelompok atas, maka pengecoh tersebut tidak berfungsi dengan baik. Distraktor dianggap berkualitas apabila paling sedikit dipilih oleh 5% dari peserta tes.

Efektivitas pengecoh diukur dengan patokan berdasarkan kriteria jumlah peserta tes kelompok atas dan kelompok bawah yang memilih opsi pengecoh tersebut, yang dijelaskan sebagai berikut.

- 1) Pengecoh berfungsi apabila  $NA < NB$
- 2) Pengecoh tidak berfungsi apabila  $NA = NB$
- 3) Pengecoh menyesatkan apabila  $NA > NB$

Keterangan:

*NA* : Banyaknya peserta kelompok atas yang memilih pengecoh

*NB* : Banyaknya peserta kelompok bawah yang memilih pengecoh.

Tabel 5 menyajikan hasil analisis distraktor untuk soal PAT matematika di SMP Budi Mulia tahun ajaran 2022/2023, yang dianalisis menggunakan Microsoft Excel.

Tabel 5. Hasil analisis distractor (pengecoh) pada butir soal

Kategori	Banyaknya	Persentase opsi jawaban	Persentase pengecoh
Kunci Jawaban	30	25%	0%
Pengecoh Berfungsi	86	71,66%	95,55%
Pengecoh tidak berfungsi	4	3,34%	4,45%
Pengecoh Menyesatkan	0	0%	0%

Berdasarkan Tabel 5, sebagian besar pengecoh (71.66%) berfungsi efektif. Hal ini menunjukkan opsi jawaban selain kunci jawaban dapat mengalihkan perhatian siswa berkemampuan rendah, sesuai dengan tujuan utama dari pengecoh. Kemudian sebanyak 4 pengecoh (3.34%) tidak berfungsi efektif. Hal ini menandakan pengecoh tersebut lebih sering dipilih oleh siswa berkemampuan tinggi. Selain itu, tidak ditemukan pengecoh yang tergolong menyesatkan. Artinya, semua pengecoh yang digunakan tidak menimbulkan kesalahan konseptual bagi siswa berkemampuan tinggi.

Hasil analisis ini selaras dengan penelitian Febriani (2016) bahwa pengecoh yang berfungsi dengan baik menandakan siswa yang belum cermat dalam menjawab soal. Beberapa hal yang menyebabkan pengecoh berfungsi dengan baik yaitu ketidaksiapan siswa untuk mengikuti ujian dengan maksimal, keterbatasan waktu yang membuat siswa tidak teliti dalam memahami soal, pilihan jawaban yang mirip sehingga membingungkan siswa, dan tingkat soal yang dirasa sulit bagi siswa (Anas, 2006; Arikunto, 2010; Khasanah et al., 2023; Suzana, 2018).

## KESIMPULAN

Berdasarkan temuan penelitian, peneliti menyimpulkan bahwa kualitas soal PAT matematika yang dikaji menunjukkan kualitas yang belum baik dan perlu diperbaiki. Hasil validitas menunjukkan 36.67% soal berada dalam kategori valid. Reabilitas butir soal menunjukkan bahwa butir soal memenuhi kriteria reliabel. Tingkat kesukaran butir soal memiliki kualitas yang kurang baik dengan 76.6% butir soal tergolong sukar. Daya beda yang digunakan pada soal perlu diperbaiki, karena persentasenya didominasi pada kategori lemah (53.3%) dan sangat lemah (20%). Sementara itu, analisis distraktor menunjukkan hasil yang cukup baik, di mana sebanyak 71.66% pengecoh berfungsi efektif dalam mengalihkan perhatian siswa berkemampuan rendah. Secara keseluruhan, butir soal yang digunakan dalam Penilaian Akhir Tahun masih perlu direvisi dan dikaji ulang agar lebih efektif dalam mengevaluasi kemampuan siswa.



Berdasarkan temuan dalam penelitian ini, diharapkan guru dan tenaga pendidik untuk memberikan upaya-upaya yang dapat meningkatkan kualitas soal yang diujikan dengan membiasakan mengidentifikasi soal dengan analisis butir soal sebelum dilakukan asesmen sumatif. Selain itu, penelitian di masa mendatang juga diharapkan dapat mengkaji kualitas soal dengan bentuk lainnya guna melengkapi hasil penelitian ini, seperti soal isian singkat, uraian, AKM, atau tugas berbasis proyek.

## REFERENSI

- Alghar, M. Z. (2022). *Proses berpikir analitis mahasiswa Matematika dalam menyelesaikan masalah Matematika berdasarkan teori APOS ditinjau dari jenis kelamin*. Universitas Islam Negeri Maulana Malik Ibrahim.
- Anas, S. (2006). *Pengantar Statistik Pendidikan*, Jakarta: PT. Raja Grafindo Persada.
- Anwar, L., Sa'dijah, C., & Murtafiah, W. (2022). *Asesmen berbasis HOTS dan Numerasi* (1st ed.). AE Media Grafika.
- Ardiansyah, M. (2021). Analisis Penilaian Tengah Semester Menggunakan Sistem CBT Pada Mata Pelajaran Matematika Di SMK Islam Perti Jakarta. *Research and Development Journal of Education*, 7(1), 29–38. <https://doi.org/10.30998/rdje.v7i1.7667>
- Arifin, Z. (2017). *Evaluasi Pembelajaran*. PT. Remaja Rosdakarya.
- Arikunto, S. (2010). *Dasar-dasar Evaluasi Pendidikan*. Bumi Aksara.
- Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage Publications.
- Ernest, P. (1991). *The Philosophy of Mathematics Education*. Taylor & Francis.
- Febriani, I. M. (2016). Analisis Butir Soal Ujian Akhir Semester (UAS) Bahasa Jerman Kelas X MIA 6 SMA Negeri 1 Maospati Tahun Pelajaran 2015/2016. *Laterne*, 5(2), 1–12.
- Gronlund, N. E. (1998). *Assessment of student achievement* (6th ed.). Allyn and Bacon.
- Halik, A. S., Mania, S., & Nur, F. (2019). Analisis butir soal ujian akhir sekolah (UAS) mata pelajaran matematika pada tahun ajaran 2015/2016 SMP Negeri 36 Makassar. *Al Asma : Journal of Islamic Education*, 1(1), 11–17. <https://doi.org/10.24252/asma.v1i1.11249>
- Hamimi, L., Zamharirah, R., & Rusydy, R. (2020). Analisis butir soal ujian Matematika kelas VII semester ganjil tahun pelajaran 2017/2018. *Mathema: Jurnal Pendidikan Matematika*, 2(1), 57–66. <https://doi.org/10.33365/jm.v2i1.459>
- Hanan, M. P., Jannah, R. R., & Alim, J. A. (2023). Analisis butir soal matematika materi KPK dan FPB berbasis cerita di SDN 111 Pekanbaru. *Journal of Educational Learning and Innovation (ELIa)*, 3(1), 92–106. <https://doi.org/10.46229/elia.v3i1.538>
- Hardianti, D. (2019). Analisis Butir Soal Tes Kemampuan Pemahaman Konsep Matematis Pada Materi Segiempat. *Metatika: Jurnal Pendidikan Matematika*, 1(2), 33–43.
- Huinchahue, J., Borromeo-Ferri, R., Reyes-Santander, P., & Garrido-Véliz, V. (2021). Mathematical

- Thinking Styles—The Advantage of Analytic Thinkers When Learning Mathematics. *Education Sciences*, 11(6), 289–297. <https://doi.org/10.3390/educsci11060289>
- Indonesia, P. R. (2003). *Undang-Undang Republik Indonesia No. 20 Tahun 2003*.
- Khasanah, I., Fuady, A., & Sunismi, S. (2023). Analisis Soal Ulangan Harian Matematika Bentuk Pilihan Ganda. *Mathema: Jurnal Pendidikan Matematika*, 5(2), 110–125. <https://doi.org/10.33365/jm.v5i2.2818>
- Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics*. National Academies Press.
- Magdalena, I., Fauziah, S. N., Faziah, S. N., & Nupus, F. S. (2021). Analisis Validitas, Reliabilitas, Tingkat Kesulitan dan Daya Beda Butir Soal Ujian Akhir Semester Tema 7 Kelas III SDN Karet 1 Sepatan. *Bintang*, 3(2), 198–214.
- Mania, S., Fitriani, F., Majid, A. F., Ichiana, N. N., & Abrar, A. I. P. (2020). Analisis butir soal ujian akhir sekolah. *Al Asma: Journal of Islamic Education*, 2(2), 274–284. <https://doi.org/10.24252/asma.v2i2.16569>
- Ntumi, S., Agbenyo, S., & Bulala, T. (2023). Estimating the Psychometric Properties (“Item Difficulty, Discrimination and Reliability Indices”) of Test Items Using Kuder-Richardson Approach (KR-20). *Shanlax International Journal of Education*, 11(3), 18–28. <https://doi.org/10.34293/education.v11i3.6081>
- Patriana, W. D., Utama, S., & Wulandari, M. D. (2021). Pembudayaan Literasi Numerasi untuk Asesmen Kompetensi Minimum dalam Kegiatan Kurikuler pada Sekolah Dasar Muhammadiyah. *Jurnal Basicedu*, 5(5), 3413–3430.
- Sa’dijah, C., Qohar, A., & Anwar, L. (2015). *Asesmen dalam Pembelajaran Matematika yang Mendukung Sikap, Pengetahuan, dan Keterampilan Matematika*. UM Penerbit & Percetakan.
- Sudjiono, A. (2005). *Pengantar Evaluasi Belajar*. Raja Grafindo Persada.
- Susanto, H., Rinaldi, A., & Novalia, N. (2015). Analisis Validitas Reliabilitas Tingkat Kesukaran dan Daya Beda pada Butir Soal Ujian Akhir Semester Ganjil Mata Pelajaran Matematika Kelas XII IPS di SMA Negeri 12 Bandar Lampung Tahun Ajaran 2014/2015. *Al-Jabar: Jurnal Pendidikan Matematika*, 6(2), 203–218. <https://doi.org/10.24042/ajpm.v6i2.50>
- Suzana, A. (2018). Analisis Tingkat Kesukaran dan Daya Beda Butir-Butir Soal Penilaian Akhir Tahun Matematika Kelas X di SMA Negeri 1 Purbalingga. *Jurnal MathGram Matematika*, 2(2), 1–8.
- Tilaar, A. L. F., Sulangi, V. R., & Pelealu, E. (2020). Analisis Kualitas Butir Soal Tes Buatan Guru Mata Pelajaran Matematika Kelas X SMA Negeri 1 Airmadidi Tahun Ajaran 2019/2020. *Jurnal Sains, Matematika & Edukasi (JSME)*, 8(2), 190–196. <https://doi.org/10.15408/jp3i.v8i1.13068>
- Utomo, B. (2018). Analisis validitas isi butir soal sebagai salah satu upaya peningkatan kualitas pembelajaran di madrasah berbasis nilai-nilai Islam. *Jurnal Pendidikan Matematika (Kudus)*, 2(1), 155–175. <https://doi.org/10.21043/jpm.v1i2.4883>