

Analisis Butir Soal UAS Matematika di SDN 072 Sibaungbaung Kecamatan Panyabungan Utara

Yulia Rahmi^{1*}, Fitri Agustina², Muhammad Aspar Nasution³

^{1,2,3} Tadris Matematika, Sekolah Tinggi Agama Islam Negeri Mandailing Natal, Indonesia

Email: yuliarahmi801@gmail.com

Abstract

This study aims to analyze the quality of Mathematics Final Semester Examination (UAS) items at SDN 072 Sibaungbaung, Panyabungan Utara District, in the 2024/2025 academic year based on five main aspects: validity, reliability, difficulty level, discrimination index, and distractor effectiveness. The research employed a descriptive quantitative approach using answer sheets from 20 fifth-grade students who completed 20 multiple-choice items. The analysis was carried out with the assistance of Microsoft Excel. The results revealed that 65% of the items were valid, while 35% were invalid. The instrument demonstrated good reliability with a coefficient of $r_{11} = 0.760$, exceeding the critical value of $r_{tabel} = 0.700$. Most items (75%) were categorized as easy, while 25% were classified as moderate, with no items falling into the difficult category. The discrimination analysis showed that 90% of the items were classified as weak, indicating limited ability to distinguish between high- and low-ability students. Meanwhile, the distractor analysis revealed that 71.67% of the distractors functioned effectively, whereas 28.33% were non-functional.

Keywords: item analysis, validity, reliability, difficulty level, discrimination index, distractors

Abstrak

Penelitian ini bertujuan untuk menganalisis kualitas butir soal Ujian Akhir Semester (UAS) Matematika di SDN 072 Sibaungbaung Kecamatan Panyabungan Utara tahun ajaran 2024/2025 berdasarkan lima aspek utama, yaitu validitas, reliabilitas, tingkat kesukaran, daya beda, dan efektivitas distraktor. Penelitian menggunakan pendekatan kuantitatif deskriptif dengan data yaitu lembar jawaban 20 siswa kelas V yang mengerjakan 20 butir soal pilihan ganda. Analisis dilakukan dengan bantuan aplikasi Microsoft Excel. Hasil penelitian menunjukkan bahwa 65% butir soal valid dan 35% tidak valid. Reliabilitas instrumen tergolong baik dengan koefisien $r_{11} = 0,760$ lebih besar dari $r_{tabel} = 0,700$. Sebagian besar soal (75%) termasuk kategori mudah, sementara 25% berada pada kategori sedang, tanpa ada soal kategori sulit. Analisis daya beda memperlihatkan dominasi butir soal dalam kategori lemah (90%), sehingga kurang mampu membedakan siswa berkemampuan tinggi dan rendah. Analisis distraktor menunjukkan bahwa 71,67% pengecoh berfungsi efektif, sedangkan 28,33% tidak berfungsi.

Kata Kunci: analisis butir soal, validitas, reliabilitas, tingkat kesukaran, daya beda, distraktor

How to Cite: Rahmi, Y., Agustina, F., & Nasution, M.A. (2024). Analisis Butir Soal UAS Matematika di SDN 072 Sibaungbaung Kecamatan Panyabungan Utara. *Journal of Mathematics in Teaching and Learning*, 2 (2), 145-155.

PENDAHULUAN

Evaluasi pembelajaran menjadi salah satu komponen penting dalam sistem pendidikan yang bertujuan untuk mengetahui sejauh mana tujuan pembelajaran tercapai (Arikunto, 2010). Evaluasi membuat guru memperoleh informasi objektif mengenai penguasaan siswa terhadap materi yang telah dipelajari (Hamimi et al., 2020; Sudjana, 1995). Salah satu bentuk evaluasi yang banyak digunakan yaitu tes hasil belajar dalam bentuk ujian akhir semester (UAS). UAS beberapa tujuan, yaitu sebagai alat ukur pencapaian belajar siswa dan sebagai umpan balik guru untuk menilai efektivitas proses pembelajaran (Sudjana, 1995). UAS menjadi instrumen yang penting dalam pembelajaran matematika karena sifat matematika yang hierarkis dan konseptual, sehingga kualitas soal harus mencerminkan kemampuan siswa.

Received March 14, 2024; Revised April 02, 2024; Accepted May 15, 2024

Kualitas tes tidak hanya ditentukan oleh cakupan materi dan kesesuaian kurikulum, tetapi juga oleh karakteristik psikometris pada butir soal (Arifin, 2015). Analisis butir soal menjadi penting untuk mengetahui sejauh mana instrumen tes tersebut valid, reliabel, memiliki tingkat kesulitan yang sesuai, daya pembeda yang pas, dan distraktor yang optimal (Hamimi et al., 2020; Tilaar et al., 2020). Menurut Arikunto (2010), analisis butir soal merupakan proses untuk menilai mutu soal yang digunakan dalam tes sehingga memberikan gambaran yang lebih tepat mengenai kemampuan siswa. Tanpa analisis ini, soal yang digunakan mungkin kurang representatif dan tidak sesuai tujuan pembelajaran. Oleh karena itu, analisis butir soal menjadi bagian penting dalam penyusunan maupun penggunaan tes hasil belajar.

Sejumlah penelitian menunjukkan bahwa kualitas soal UAS yang diberikan ke siswa masih sering bermasalah. Cholidiyah (2023) dalam penelitiannya menemukan bahwa sebagian soal UAS matematika sekolah dasar tidak memenuhi kriteria validitas dan reliabilitas. Demikian pula, penelitian Fatimah & Alfath (2019) menunjukkan bahwa banyak distraktor pada soal pilihan ganda tidak berfungsi, sehingga tidak membedakan siswa berkemampuan tinggi dengan siswa berkemampuan rendah. Hal ini menunjukkan kelemahan pada instrumen evaluasi yang digunakan di sekolah. Apabila kondisi tersebut terus dibiarkan, hasil evaluasi tidak akan mencerminkan kemampuan siswa yang sebenarnya. Hal ini sekaligus menegaskan bahwa penelitian analisis butir soal penting sebagai sarana memperbaiki kualitas evaluasi pembelajaran.

Penelitian lainnya dalam lingkup mancanegara juga menunjukkan pentingnya analisis butir soal dalam meningkatkan kualitas evaluasi. DeVellis & Thorpe (2021) menegaskan bahwa kualitas instrumen tes bergantung pada kemampuan setiap item untuk mengukur dan mengklasifikasikan siswa dengan kemampuan berbeda. Sementara itu, Haladyna et al. (2019) menyoroti pentingnya desain distraktor yang berfungsi, karena distraktor yang tidak dipilih siswa berarti tidak efektif dalam mengukur kompetensi. Selain itu, literatur global di bidang asesmen juga menekankan pentingnya validitas dan reliabilitas instrumen sebagai syarat utama tes (Ciftci & Karadag, 2016; Hesamamiri et al., 2013). Dengan demikian, dapat disimpulkan bahwa praktik analisis butir soal merupakan praktik yang berstandar internasional untuk menjamin kualitas instrumen evaluasi.

Sayangnya, guru di sekolah dasar dalam praktiknya masih jarang melakukan analisis butir soal secara sistematis. Banyak guru yang menyusun atau menggunakan soal UAS tanpa memeriksa validitas dan reliabilitasnya terlebih dahulu (Cholidiyah, 2023). Hal ini dapat menyebabkan hasil evaluasi yang diperoleh tidak mencerminkan kemampuan siswa secara akurat. Keterbatasan waktu, pengetahuan, dan keterampilan teknis sering kali menjadi alasan utama mengapa analisis butir soal jarang dilakukan (Basri et al., 2021). Padahal dengan analisis butir soal, guru dapat memperbaiki dan menyimpan soal berkualitas untuk digunakan kembali pada ujian berikutnya. Dengan kata lain, analisis butir soal bukan hanya bermanfaat bagi evaluasi saat ini, tetapi juga untuk perbaikan instrumen tes di masa depan.

Konteks lokal juga memengaruhi pentingnya analisis instrumen tes. Misalnya SDN 072 Sibaungbaung Kecamatan Panyabungan Utara, Kabupaten Mandailing Natal, yang merupakan salah satu sekolah dasar yang rutin melaksanakan UAS pada setiap akhir semester. Namun hingga kini, belum

ada artikel ilmiah yang khusus menelaah kualitas soal UAS Matematika di sekolah. Ketiadaan penelitian ini membuat pihak sekolah tidak memiliki data empiris yang dapat dijadikan dasar perbaikan soal. Padahal, informasi mengenai validitas, reliabilitas, tingkat kesulitan, daya pembeda, dan efektivitas distraktor akan membantu guru dalam meningkatkan kualitas evaluasi. Sehingga dilakukannya analisis butir soal di sekolah tersebut menjadi hal relevan untuk dilakukan.

Dengan demikian, tujuan penelitian ini yaitu untuk mendeskripsikan kualitas soal UAS Matematika di SDN 072 Sibaungbaung Kecamatan Panyabungan Utara berdasarkan lima aspek utama analisis butir soal. Penelitian ini tidak hanya berkontribusi pada penyajian data kuantitatif tentang kualitas instrumen tes, tetapi juga memberikan rekomendasi praktis bagi sekolah. Harapannya, penelitian ini dapat memperkaya literatur analisis butir soal pada jenjang sekolah dasar, khususnya di daerah yang masih minim melakukan penelitian analisis butir soal. Diharapkan sekolah dapat memperbaiki instrumen evaluasi sehingga mutu pembelajaran matematika semakin berkembang.

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan jenis penelitian deskriptif. Pendekatan kuantitatif dipilih karena analisis difokuskan pada perhitungan statistik sederhana guna menilai kualitas butir soal (Arikunto, 2010). Jenis penelitian deskriptif digunakan untuk menggambarkan kondisi soal UAS Matematika sebagaimana adanya berdasarkan data empiris yang diperoleh dari siswa. Data penelitian berupa hasil tes UAS Matematika yang terdiri atas 20 butir soal pilihan ganda. Penelitian ini dilaksanakan di SDN 072 Sibaungbaung Kecamatan Panyabungan Utara, Kabupaten Mandailing Natal. Subjek penelitian merupakan siswa kelas V yang berjumlah 20 orang dan mengikuti UAS Matematika semester ganjil tahun ajaran 2024/2025.

Teknik pengumpulan data dilakukan dengan menghimpun lembar jawaban siswa pada UAS Matematika. Data yang diperoleh kemudian dianalisis menggunakan analisis butir soal. Analisis mencakup lima aspek, yaitu validitas, reliabilitas, tingkat kesulitan, daya pembeda, dan distraktor. Validitas soal diuji dengan korelasi *product moment* Pearson, dengan kategori sangat rendah ($r \leq 0,20$), rendah ($0,21-0,40$), sedang ($0,41-0,60$), tinggi ($0,61-0,80$), dan sangat tinggi ($0,81-1,00$). Reliabilitas tes dihitung dengan rumus *KR-20*, dengan kategori tidak reliabel ($r_{11} < 0,20$), reliabilitas rendah ($0,20-0,40$), sedang ($0,41-0,60$), tinggi ($0,61-0,80$), dan sangat tinggi ($0,81-1,00$). Tingkat kesulitan soal ditentukan berdasarkan proporsi siswa yang menjawab benar, dengan kategori mudah ($p > 0,70$), sedang ($0,30 \leq p \leq 0,70$), dan sulit ($p < 0,30$). Daya pembeda soal dihitung dengan membandingkan proporsi jawaban benar antara kelompok atas dan kelompok bawah, dengan kategori sangat jelek ($D < 0,20$), cukup ($0,20-0,40$), baik ($0,40-0,70$), dan sangat baik ($D \geq 0,70$). Efektivitas distraktor dianalisis dengan melihat distribusi pilihan jawaban. Suatu distraktor dinyatakan berfungsi apabila dipilih minimal 5% siswa, sedangkan jika tidak dipilih sama sekali dinyatakan tidak berfungsi. Prosedur penelitian dilakukan mulai dari pengumpulan data, pengolahan data dengan SPSS, dan interpretasi hasil sesuai kriteria standar analisis butir soal.

HASIL DAN PEMBAHASAN

Hasil Uji Validitas

Validitas suatu butir soal merupakan sejauh mana soal tersebut mampu mengukur aspek yang hendak diukur. Suatu soal dinyatakan valid apabila menunjukkan hubungan yang signifikan dengan skor keseluruhan tes. Skor pada setiap butir memiliki keselarasan dengan skor total, baik dalam peningkatan maupun penurunannya. Adapun rumus korelasi *product moment* yaitu

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

Keterangan

r_{xy} : koefisien korelasi antara skor item dan skor total

N : banyak responden

X : skor item

Y : skor total

Temuan uji validitas butir soal UAS mata pelajaran matematika di SDN 072 Sibaungbaung pada tahun ajaran 2024/2025, yang dianalisis menggunakan aplikasi Microsoft Excel, dapat dilihat pada Tabel 1.

Tabel 1. Hasil uji validitas butir soal

No	Validitas	Nomor Butir Soal	Banyaknya	Persentase
1.	$\geq 0,3598$ (Valid)	2, 8, 9,10, 11, 12, 13, 14, 15, 16, 17, 18, 20	13	65%
2.	$< 0,3598$ (Tidak Valid)	1, 3, 4, 5, 6, 7, 19	7	35%

Berdasarkan Tabel 1, dari total 20 butir soal UAS Matematika, terdapat 13 soal (65%) yang dinyatakan valid karena memiliki nilai $r_{hitung} \geq 0,3598$. Sementara itu, sebanyak 7 soal (35%) dinyatakan tidak valid karena nilai r_{hitung} lebih kecil dari batas kritis tersebut. Temuan ini menunjukkan bahwa sebagian besar soal yang digunakan telah sesuai untuk mengukur kemampuan siswa sesuai tujuan pembelajaran. Namun, masih terdapat sepertiga soal yang tidak valid sehingga berpotensi menurunkan kualitas instrumen secara keseluruhan. Oleh karena itu, butir soal yang tidak valid sebaiknya diperbaiki melalui perbaikan redaksi, kesesuaian indikator, maupun tingkat kesulitan agar lebih tepat digunakan dalam evaluasi pembelajaran.

Soal valid pada penelitian ini menunjukkan kualitas instrumen yang cukup baik, meskipun masih perlu perbaikan pada soal tidak valid. Hasil ini sejalan dengan penelitian Khasanah et al. (2023) yang menemukan bahwa sekitar 60% soal ulangan harian termasuk dalam kategori valid. Penelitian Gronlund (1998) juga menegaskan pentingnya validitas sebagai indikator utama kualitas evaluasi pembelajaran karena instrumen yang valid memberikan informasi yang akurat bagi guru. Namun, Tilaar et al. (2020) mencatat bahwa masih banyak soal ujian akhir semester di sekolah dasar yang belum memenuhi aspek validitas, terutama akibat kurangnya kesesuaian indikator dengan butir soal. Dengan demikian, guru perlu memperkuat keterampilan penyusunan soal, telaah butir soal, dan menganalisis kuantitatif dan kualitatif pada instrumen tes.

Hasil Uji Reliabilitas

Reliabilitas berkaitan dengan sejauh mana suatu tes dapat dipercaya dalam memberikan hasil yang konsisten ketika diujikan secara berulang. Dengan kata lain, reliabilitas menilai kestabilan instrumen dalam mengukur kemampuan peserta didik. Pengujian reliabilitas pada penelitian ini menggunakan rumus Kuder Richardson-20 (KR-20) (Ntumi et al., 2023), yang tepat diterapkan pada instrumen berbentuk tes dengan pilihan jawaban benar (skor 1) dan salah (skor 0). Adapun formula KR-20 dapat dituliskan sebagai berikut.

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

Keterangan

r_{11} : koefisien reliabilitas

p : Proporsi benar

q : $1 - p$

σ_t^2 : varians total

k : jumlah soal

Analisis reliabilitas soal PAT mata pelajaran matematika di SDN 072 Sibaungbaung tahun pelajaran 2024/2025 dengan bantuan Microsoft Excel tersaji pada Tabel 2.

Tabel 2. Hasil uji reabilitas butir soal

Hasil r_{11}	Hasil r_{tabel}	Keterangan
0,760	0,700	Reliabel

Berdasarkan Tabel 2, diperoleh nilai koefisien reliabilitas $r_{11} = 0,760$, yang lebih besar dari nilai $r_{\text{tabel}} = 0,700$ pada taraf signifikansi 5%. Hal ini menunjukkan bahwa instrumen tes memiliki reliabilitas yang baik, sehingga hasil pengukuran yang diperoleh dari tes ini dapat dipercaya dan konsisten. Dengan kata lain, soal UAS Matematika ini dapat digunakan sebagai alat evaluasi karena menghasilkan data yang stabil meskipun digunakan pada waktu yang berbeda. Nilai reliabilitas yang tinggi juga mengindikasikan bahwa soal-soal tersebut memiliki homogenitas internal yang memadai, sehingga dapat mengukur kompetensi siswa secara berkesinambungan.

Hasil penelitian ini memperlihatkan bahwa instrumen soal UAS Matematika tergolong reliabel, sesuai dengan kriteria reliabilitas instrumen yang dikemukakan oleh Arikunto (2010). Penelitian serupa oleh Hanan et al. (2023) menunjukkan bahwa instrumen tes matematika dengan nilai reliabilitas di atas 0,70 dapat dikategorikan memiliki kualitas baik. Adapun Gronlund (1998) menekankan bahwa instrumen yang reliabel menjadi syarat utama agar hasil tes dapat dipakai untuk mengambil keputusan pendidikan yang objektif. Selain itu, Kuder dan Richardson dalam KR-20 juga menegaskan bahwa reliabilitas yang tinggi menunjukkan adanya konsistensi antarbutir soal yang baik. Dengan demikian, instrumen yang digunakan pada penelitian ini dapat dikatakan cukup kuat dalam mengukur kemampuan siswa, meskipun tetap diperlukan kajian lebih lanjut pada aspek validitas dan distribusi tingkat kesukaran soal agar mutu evaluasi lebih optimal.

Hasil Analisis Tingkat Kesukaran

Analisis tingkat kesukaran bertujuan untuk mengklasifikasikan butir soal ke dalam kategori mudah, sedang, atau sulit Arikunto (2010). Sebuah soal dinilai layak digunakan apabila tingkat kesukarannya berada pada batas yang wajar, tidak terlalu mudah maupun terlalu sukar. Dengan demikian, tingkat kesulitan yang seimbang menunjukkan kualitas soal yang baik. Perhitungan tingkat kesukaran dilakukan menggunakan rumus berikut

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

Keterangan:

P : indeks kesukaran

B : jumlah siswa menjawab benar

N : jumlah siswa

Hasil perhitungan tingkat kesukaran soal PAT matematika tahun ajaran 2024/2025 di SDN 072 Sibaungbaung dengan bantuan aplikasi Microsoft Excel tersaji pada Tabel 3.

Tabel 3. Hasil analisis tingkat kesukaran butir soal

Tingkat Kesukaran	Nomor Soal	Banyaknya	Persentase
Sangat Sulit	-	0	0%
Sulit	-	0	0%
Sedang	10,11, 15, 16, 19	5	25%
Mudah	1, 2, 3, 4, 5, 6, 7, 8, 9, 12, 13, 14, 17, 18, 20	15	75%
Sangat Mudah	-	0	0%

Berdasarkan Tabel 3, diketahui bahwa sebagian besar soal UAS Matematika berada pada kategori mudah, yaitu sebanyak 15 butir soal (75%). Selanjutnya, terdapat 5 butir soal (25%) yang masuk dalam kategori sedang, sedangkan tidak ada soal yang termasuk dalam kategori sulit maupun sangat sulit. Demikian pula, tidak ditemukan soal yang berada dalam kategori sangat mudah. Hasil ini menunjukkan bahwa distribusi tingkat kesukaran soal belum seimbang, karena terlalu didominasi oleh soal yang tergolong mudah. Idealnya, proporsi tingkat kesukaran soal disusun dalam komposisi seimbang, sehingga instrumen dapat mengukur variasi kemampuan siswa secara lebih menyeluruh.

Temuan ini mengindikasikan bahwa instrumen tes UAS Matematika yang dianalisis belum memenuhi standar proporsi kesukaran soal yang dianjurkan oleh Arikunto (2010), yaitu sekitar 30% soal mudah, 40% sedang, dan 30% sulit. Ketidakseimbangan ini berpotensi menyebabkan tes kurang mampu membedakan kemampuan siswa yang beragam, karena dominasi soal mudah cenderung mempermudah sebagian besar siswa untuk memperoleh skor tinggi. Hasil ini selaras dengan penelitian Febriani (2016) yang menemukan bahwa instrumen evaluasi dengan dominasi soal mudah cenderung memiliki daya diskriminasi rendah. Adapun Halik et al. (2019) juga menegaskan bahwa distribusi tingkat kesukaran soal yang proporsional penting untuk menjaga validitas dan reliabilitas instrumen evaluasi. Oleh karena itu, dalam penyusunan soal selanjutnya, perlu ditingkatkan jumlah soal kategori

sedang dan sulit agar hasil evaluasi lebih representatif dalam mengukur kemampuan siswa secara menyeluruh.

Hasil Analisis Daya Beda

Suatu butir soal memiliki fungsi daya beda, yaitu kemampuan soal tersebut dalam membedakan peserta didik dengan kemampuan tinggi dan rendah. Butir soal dengan daya beda tinggi menandakan soal efektif dalam mengidentifikasi variasi kemampuan siswa. Analisis daya beda dilakukan melalui rumus berikut

$$DB = \frac{B_A}{J_A} - \frac{B_B}{J_B}$$

Keterangan:

DB : daya beda

B_A : jumlah benar kelompok atas

J_A : jumlah siswa kelompok atas

B_B : jumlah benar kelompok bawah

J_B : jumlah siswa kelompok bawah

Hasil analisis daya beda untuk soal PAT matematika di SDN 072 Sibaungbaung tahun ajaran 2024/2025 yang diolah menggunakan aplikasi Microsoft Excel ditunjukkan pada Tabel 4.

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	-	0	0%
0,21-0,40	Sedang	9, 16	2	10%
0-0,20	Lemah	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 17, 18, 19, 20	18	90%
< 0	Sangat Lemah	-	0	0%

Berdasarkan Tabel 4, terlihat bahwa sebagian besar butir soal berada pada kategori daya beda lemah, yaitu sebanyak 18 soal (90%). Sementara itu, hanya terdapat 2 soal (10%) yang tergolong dalam kategori sedang. Tidak ada butir soal yang termasuk kategori kuat, sangat kuat, maupun sangat lemah. Hasil ini menunjukkan bahwa sebagian besar soal belum mampu membedakan siswa berkemampuan tinggi dengan siswa berkemampuan rendah secara efektif. Dengan dominasi soal dalam kategori lemah, maka instrumen tes ini belum sepenuhnya memenuhi fungsi evaluasi yang ideal.

Dominasi butir soal dengan daya beda lemah menunjukkan bahwa instrumen yang digunakan kurang optimal dalam mengukur variasi kemampuan siswa. Kondisi ini dapat disebabkan oleh komposisi soal yang terlalu mudah atau kurang variatif, sehingga tidak mampu menguji kemampuan analitis siswa berkemampuan tinggi. Hasil ini selaras dengan penelitian Magdalena et al. (2021) yang

menyebutkan bahwa soal dengan daya beda lemah sering kali menunjukkan ketidaksesuaian antara kompetensi yang diukur dengan tingkat kesulitan soal. Sementara itu, Febriani (2016) menegaskan bahwa butir soal dengan daya beda rendah tidak layak digunakan dalam evaluasi pembelajaran karena tidak mampu memberikan informasi yang representatif. Secara internasional, penelitian oleh Sim & Rasiah (2006) juga menekankan pentingnya soal dengan daya beda tinggi untuk memastikan instrumen tes benar-benar mengukur perbedaan kemampuan siswa. Oleh karena itu, diperlukan revisi pada sebagian besar soal agar instrumen dapat menghasilkan evaluasi yang lebih objektif dan akurat.

Hasil Analisis Distraktor (Pengecoh)

Analisis distraktor bertujuan untuk mengevaluasi sejauh mana setiap opsi jawaban selain kunci mampu membedakan siswa dengan tingkat kemampuan yang berbeda. Distraktor dikatakan efektif apabila lebih banyak dipilih oleh siswa dengan kemampuan rendah dibandingkan dengan siswa yang memiliki kemampuan tinggi. Sebaliknya, distraktor dinilai tidak efektif jika tidak dipilih sama sekali atau justru lebih sering dipilih oleh siswa berkemampuan tinggi. Efektivitas pengecoh juga dapat ditinjau dari persentase siswa yang memilih opsi tersebut, dengan kriteria minimal 5% agar distraktor dianggap berfungsi baik. Penilaian efektivitas pengecoh dilakukan berdasarkan perbandingan jumlah siswa dari kelompok atas dan bawah, sebagaimana kriteria berikut

$$P_i = \frac{n_i}{N} \times 100\%$$

Keterangan:

n_i : Jumlah siswa memilih opsi ke- i

N : Jumlah siswa

Selain itu, untuk mengetahui berfungsi atau tidaknya distraktor, digunakan pertidaksamaan 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 2024/2025, yang dianalisis menggunakan Microsoft Excel.

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

Kategori	Banyaknya	Persentase opsi jawaban	Persentase pengecoh
Kunci Jawaban	20	25%	0%
Pengecoh Berfungsi	43	53,75%	71,67%
Pengecoh tidak berfungsi	17	21,25%	28,33%
Pengecoh Menyesatkan	0	0%	0%

Berdasarkan Tabel 5, sebagian besar pengecoh berfungsi dengan baik, yaitu sebanyak 43 opsi (71,67%). Hal ini menunjukkan bahwa sebagian besar pilihan jawaban selain kunci dapat menarik perhatian siswa berkemampuan rendah sesuai fungsinya. Namun demikian, terdapat 17 pengecoh (28,33%) yang tidak berfungsi dengan baik karena tidak mampu mengelabui siswa kelompok bawah. Tidak ditemukan pengecoh yang menyesatkan, sehingga tidak ada opsi salah yang justru lebih banyak dipilih oleh siswa berkemampuan tinggi. Secara keseluruhan, hasil ini menunjukkan bahwa efektivitas pengecoh dalam soal sudah cukup baik, meskipun masih ada sebagian kecil opsi yang perlu diperbaiki.

Hasil analisis menunjukkan bahwa mayoritas pengecoh sudah berfungsi sesuai perannya, namun masih terdapat lebih dari seperempat pengecoh yang tidak berfungsi dengan baik. Pengecoh yang tidak berfungsi biasanya disebabkan karena opsi jawaban terlalu jelas salah atau tidak relevan dengan materi (Arikunto, 2010). Kondisi ini serupa dengan hasil penelitian Febriani (2016), yang menemukan bahwa distraktor yang tidak efektif sering muncul akibat penyusunan opsi jawaban yang kurang cermat. Di sisi lain, studi internasional oleh Tarrant et al. (2009) menegaskan bahwa kualitas distraktor sangat berpengaruh terhadap validitas instrumen tes. Semakin efektif distraktor, semakin tinggi kemampuan soal dalam mengukur kompetensi siswa. Oleh karena itu, revisi perlu difokuskan pada opsi yang tidak berfungsi, agar instrumen tes lebih representatif dan sesuai dengan tujuan evaluasi pembelajaran.

KESIMPULAN

Berdasarkan hasil analisis, dapat disimpulkan bahwa instrumen soal PAT matematika di SDN 072 Sibaungbaung tahun ajaran 2024/2025 memiliki kualitas yang beragam. Dari sisi validitas, 65% butir soal tergolong valid, sementara 35% lainnya tidak valid sehingga memerlukan revisi. Reliabilitas soal menunjukkan hasil yang baik dengan koefisien 0,760 lebih besar dari rtabel 0,700, sehingga soal dapat dinyatakan reliabel. Analisis tingkat kesukaran memperlihatkan dominasi soal kategori mudah (75%) dengan hanya 25% berada pada kategori sedang, serta tidak ditemukan soal kategori sulit maupun ekstrem, sehingga distribusinya belum sesuai proporsi ideal. Analisis daya beda menunjukkan sebagian besar soal (90%) berada pada kategori lemah, sehingga kurang mampu membedakan siswa berkemampuan tinggi dan rendah. Analisis distraktor memperlihatkan bahwa 71,67% pengecoh berfungsi dengan baik, namun terdapat 28,33% pengecoh tidak berfungsi dan perlu diperbaiki. Instrumen soal ini memerlukan revisi terutama pada aspek validitas, daya beda, dan distraktor agar lebih representatif dan adil dalam mengukur kemampuan siswa. Rekomendasi untuk penelitian selanjutnya adalah memperluas kajian dengan membandingkan kualitas soal dari beberapa sekolah dasar yang berbeda, dan mengintegrasikan analisis kualitatif terhadap konstruksi butir soal untuk memperoleh gambaran yang lebih komprehensif mengenai mutu instrumen evaluasi pembelajaran matematika.

REFERENSI

Arifin, Z. (2015). Evaluasi Pembelajaran Penulis. In *Direktorat Jenderal Pendidikan Islam Kementerian Agama RI*.

- Arikunto, S. (2010). *Dasar-dasar Evaluasi Pendidikan*. Bumi Aksara.
- Basri, K., Baidowi, B., Junaidi, J., & Turmuzi, M. (2021). Analisis Butir Soal Ulangan Semester Ganjil Mata Pelajaran Matematika Kelas VIII SMP. *Griya Journal of Mathematics Education and Application*, 1(4), 682–694. <https://doi.org/10.29303/griya.v1i4.107>
- 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*, 2(1), 43–52.
- Ciftci, S. K., & Karadag, E. (2016). Developing a mathematics education quality scale. *Africa Education Review*, 13(3–4), 211–234. <https://doi.org/10.1080/18146627.2016.1224590>
- DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications*. Sage publications.
- Fatimah, L. U., & Alfath, K. (2019). Analisis kesukaran soal, daya pembeda dan fungsi distraktor. *AL-MANAR: Jurnal Komunikasi Dan Pendidikan Islam*, 8(2), 37–64.
- 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.
- Haladyna, T. M., Rodriguez, M. C., & Stevens, C. (2019). Are multiple-choice items too fat? *Applied Measurement in Education*, 32(4), 350–364. <https://doi.org/10.1080/08957347.2019.1660348>
- 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>
- Hesamamiri, R., Mahdavi Mazdeh, M., & Jafari, M. (2013). Measuring the reliability of knowledge management: instrument development and validation. *Aslib Proceedings: New Information Perspectives*, 65(5), 484–502. <https://doi.org/10.1108/AP-04-2013-0029>
- 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>
- 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.
- 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>
- Sim, S., & Rasiah, R. I. (2006). Relationship between item difficulty and discrimination indices in true/false-type multiple choice questions of a para-clinical multidisciplinary paper. *Annals-Academy of Medicine Singapore*, 35(2), 67.
- Sudjana, N. (1995). *Penilaian hasil proses belajar mengajar*.
- Tarrant, M., Ware, J., & Mohammed, A. M. (2009). An assessment of functioning and non-functioning distractors in multiple-choice questions: a descriptive analysis. *BMC Medical Education*, 9(1), 40. <https://doi.org/10.1186/1472-6920-9-40>
- 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>