

Analisis Kualitas Butir Soal Asesmen Sumatif Akhir Semester Gasal Matematika Kelas IV Sekolah Dasar

Priska Amanda^{1*}, Bistari², Hery Kresnadi³

¹⁻³Pendidikan Guru Sekolah Dasar, Universitas Tanjungpura, Jl. Profesor Dokter H. Hadari Nawawi

E-mail: priska.amanda@student.untan.ac.id

* Corresponding Author

 <https://doi.org/10.31004/jerkin.v4i2.2973>

ARTICLE INFO

Article history

Received: 20 August 2025

Revised: 15 September 2025

Accepted: 04 October 2025

Kata Kunci:

Analisis Butir Soal,
Matematika, Asesmen
Sumatif Akhir

Keywords:

Item Analysis, Mathematics,
Summative Assessment



ABSTRACT

Penelitian ini bertujuan untuk mendeskripsikan kualitas butir soal asesmen sumatif akhir semester gasal mata pelajaran matematika tahun ajaran 2024/2025 kelas IV di SD Negeri 39 Pontianak Kota dari segi validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Metode penelitian yang digunakan adalah deskriptif dengan bentuk penelitian kuantitatif berbantuan Microsoft Excel. Subjek penelitian ini adalah butir soal ASAS Matematika kelas IV. Teknik pengumpulan data menggunakan teknik dokumentasi untuk memperoleh soal ASAS Matematika dan data rekapitulasi jawaban peserta didik kelas IV-A. Hasil analisis menunjukkan bahwa 11 soal (44%) valid, reliabilitas soal tinggi (0,70), 12 soal mudah, 10 sedang, dan 3 sukar. Daya pembeda menunjukkan 1 soal sangat baik, 6 baik, 5 cukup, 8 jelek, dan 5 sangat jelek. Secara keseluruhan, 8 soal (32%) dikategorikan baik dan layak digunakan kembali, sedangkan 17 soal (68%) perlu direvisi atau diganti karena belum memenuhi kriteria soal yang baik.

This study aims to describe the quality of the multiple-choice items in the odd-semester summative assessment of mathematics for Grade IV students at SD Negeri 39 Pontianak Kota in the 2024/2025 academic year, in terms of validity, reliability, difficulty level, and discriminating power. The research employed a descriptive method with a quantitative approach, supported by Microsoft Excel. The subjects of this study were the Grade IV mathematics summative assessment items. Data were collected through documentation techniques to obtain the test items and the answer recapitulation of Grade IV-A students. The analysis results showed that 11 items (44%) were valid, with a high reliability coefficient (0.70). Regarding the level of difficulty, 12 items were easy, 10 were moderate, and 3 were difficult. In terms of discriminating power, 1 item was categorized as very good, 6 as good, 5 as fair, 8 as poor, and 5 as very poor. Overall, 8 items (32%) were considered good and suitable for reuse, while 17 items (68%) needed revision or replacement as they did not yet meet the criteria for high-quality test items.



This is an open access article under the CC-BY-SA license.

How to Cite: Priska Amanda, et al (2025). Analisis Kualitas Butir Soal Asesmen Sumatif Akhir Semester Gasal Matematika Kelas IV Sekolah Dasar, 4 (2) 7203-7210. <https://doi.org/10.31004/jerkin.v4i2.2973>

PENDAHULUAN

Evaluasi merupakan bagian penting dalam proses pembelajaran dan memiliki peran strategis dalam dunia pendidikan. Melalui evaluasi, pendidik dapat mengetahui sejauh mana ketercapaian hasil belajar peserta didik dan memperoleh informasi yang relevan untuk meningkatkan kualitas pembelajaran. Evaluasi hasil belajar bertujuan untuk mengetahui tingkat pemahaman peserta didik secara individu maupun kelompok, serta memberikan umpan balik yang bermanfaat dalam proses pembelajaran selanjutnya.

Dalam praktiknya, evaluasi melibatkan penggunaan instrumen penilaian untuk mengumpulkan data yang objektif. Instrumen ini membantu pendidik dalam menilai kemampuan peserta didik dengan

lebih terarah. Di antara jenis instrumen penilaian, tes merupakan salah satu yang sering digunakan, karena mampu memberikan gambaran yang jelas mengenai pemahaman peserta didik terhadap materi pelajaran. Arikunto (2013) menjelaskan bahwa "Tes merupakan alat atau prosedur yang digunakan untuk mengetahui atau mengukur sesuatu dalam suasana tertentu, dengan cara dan aturan yang telah ditetapkan" (p.66).

Di sekolah, tes seperti penilaian harian, asesmen sumatif tengah semester (ASTS), dan asesmen sumatif akhir semester (ASAS) merupakan bagian dari evaluasi pembelajaran yang rutin dilaksanakan. Sebagai instrumen evaluasi, tes memerlukan perencanaan yang matang agar dapat mencerminkan kemampuan peserta didik secara akurat. Tes yang baik perlu memenuhi beberapa kriteria, seperti validitas, reliabilitas, objektivitas, dan kepraktisan (Sudijono, 2017, p.93-97).

Salah satu langkah penting untuk meningkatkan kualitas tes adalah melakukan analisis butir soal. Analisis ini bertujuan untuk memastikan bahwa soal-soal yang disusun sesuai dengan kurikulum, kompetensi dasar, dan indikator pembelajaran. Analisis butir soal juga membantu dalam mengidentifikasi aspek-aspek yang dapat ditingkatkan, sehingga hasil tes dapat memberikan informasi yang lebih akurat dan bermanfaat.

Kendati demikian, guru jarang menganalisis kualitas dari setiap butir soal yang diujikan sehingga kebanyakan belum bisa mengidentifikasi soal-soal yang baik, kurang baik, dan soal yang tidak baik. Soal yang dikatakan baik memenuhi semua syarat yang telah ditentukan yang berhubungan dengan analisis soal yaitu validitas, reliabilitas, daya pembeda, dan tingkat kesukaran. Soal yang dikatakan baik adalah soal yang memiliki tingkat kesukaran yang tidak terlalu mudah dan tidak terlalu sukar.

Berdasarkan hasil wawancara dengan guru kelas IV-A SD Negeri 39 Pontianak Kota yaitu Ibu Nurhidayani, S.Pd., ditemukan bahwa analisis soal selama ini dilakukan secara sederhana, yaitu dengan memeriksa distribusi skor peserta didik per soal. Guru belum melakukan analisis mendalam terkait validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Kondisi ini menunjukkan bahwa kualitas instrumen tes belum sepenuhnya terukur secara komprehensif, sehingga hasil evaluasi yang diperoleh berpotensi tidak mencerminkan kemampuan peserta didik secara akurat. Temuan unik dari kondisi tersebut adalah adanya kesenjangan antara praktik evaluasi yang dilakukan guru dengan standar analisis butir soal yang seharusnya diterapkan, khususnya pada mata pelajaran matematika. Hal ini menegaskan bahwa variasi capaian peserta didik tidak semata-mata dipengaruhi oleh kemampuan individu, melainkan juga dapat disebabkan oleh kualitas instrumen evaluasi yang digunakan.

Hasil asesmen sumatif akhir semester gasal menunjukkan bahwa capaian peserta didik dalam mata pelajaran matematika masih bervariasi. Berdasarkan kriteria ketercapaian tujuan pembelajaran (KKTP), terdapat 5 peserta didik (18%) yang masih berada dalam kategori "Perlu Bimbingan" (0-60), yang menunjukkan bahwa hasil evaluasi belum sepenuhnya optimal. Distribusi nilai peserta didik dapat dilihat pada tabel berikut.

Tabel 1. Distribusi Skor ASAS Gasal Matematika Peserta Didik Kelas IV-A

Kategori	Rentang Skor	Jumlah Peserta Didik	Persentase (%)
Sangat Baik	90-100	6	21
Baik	79-89	6	21
Cukup	68-78	11	40
Perlu Bimbingan	0-60	5	18
Total	-	28	100

Sumber: (Data Nilai ASAS Gasal Matematika Kelas IV-A SD Negeri 39 Pontianak Kota Tahun Ajaran 2024/2025, 2024).

Kondisi ini memperlihatkan adanya kesenjangan pemahaman antar peserta didik, di mana sebagian besar sudah mencapai kategori "Cukup" hingga "Sangat Baik", namun masih ada kelompok kecil yang belum memenuhi standar minimal. Situasi tersebut semakin memperkuat pentingnya analisis butir soal untuk mengetahui apakah perbedaan capaian tersebut murni disebabkan oleh kemampuan peserta didik atau juga dipengaruhi oleh kualitas instrumen tes yang digunakan.

Sebagai bagian dari upaya untuk meningkatkan kualitas pembelajaran, analisis butir soal dapat memberikan gambaran yang lebih komprehensif mengenai kualitas soal yang digunakan. Dalam penelitian ini, analisis dilakukan menggunakan pendekatan kuantitatif berdasarkan teori pengukuran klasik, yang mencakup validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Hasil analisis ini

diharapkan dapat memberikan masukan yang bermanfaat bagi pengembangan instrumen penilaian yang lebih baik di masa mendatang.

Tes yang dilaksanakan di SD Negeri 39 Pontianak Kota, seperti halnya di sekolah lainnya, dilakukan secara rutin baik untuk evaluasi formatif maupun sumatif. Namun, hasil tes pada mata pelajaran matematika sering kali menunjukkan bahwa masih terdapat ruang untuk perbaikan. Untuk memastikan bahwa soal yang digunakan mampu mencerminkan kemampuan peserta didik secara optimal, diperlukan analisis butir soal yang lebih mendalam dan komprehensif.

Dengan demikian, peneliti tertarik untuk melakukan penelitian yang berjudul: "Analisis Kualitas Butir Soal Asesmen Sumatif Akhir Semester Gasal Matematika Kelas IV Sekolah Dasar".

METODE

Penelitian ini menggunakan desain penelitian kuantitatif dengan metode deskriptif kuantitatif. Metode ini sesuai dengan penelitian yang akan dilakukan, yaitu menggunakan fakta apa adanya untuk menggambarkan kualitas butir soal Asesmen Sumatif Akhir Semester (ASAS) mata pelajaran matematika kelas IV Sekolah Dasar Negeri 39 Pontianak Kota. Bentuk penelitian yang digunakan dalam penelitian ini adalah survei dengan pendekatan analisis isi. Pendekatan ini digunakan untuk menganalisis kualitas butir soal ASAS mata pelajaran matematika berdasarkan indikator validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Pendekatan ini bertujuan untuk menggambarkan kualitas soal-soal tersebut secara mendalam dan objektif berdasarkan data hasil tes peserta didik.

Populasi dalam penelitian ini mencakup seluruh peserta didik kelas IV di SD Negeri 39 Pontianak Kota pada tahun ajaran 2024/2025. Peserta didik kelas IV terdiri dari dua kelas, yaitu kelas IV-A dan IV-B, dengan total jumlah peserta didik sebanyak 55 orang, yang masing-masing berjumlah 28 peserta didik di kelas IV-A dan 27 peserta didik di kelas IV-B. Sampel dalam penelitian ini diambil menggunakan teknik *purposive sampling*. Sugiyono (2013) menyatakan bahwa *purposive sampling* adalah "teknik penentuan sampel dengan pertimbangan tertentu." (p.85). Pada penelitian ini, kelas IV-A dipilih sebagai sampel penelitian dengan jumlah peserta didik sebanyak 28 orang.

Objek penelitian dalam penelitian ini adalah soal asesmen sumatif akhir semester gasal mata pelajaran matematika yang digunakan di SD Negeri 39 Pontianak Kota tahun ajaran 2024/2025. Selain itu, data yang dianalisis mencakup kunci jawaban soal dan data hasil jawaban peserta didik kelas IV-A sebanyak 28 orang. Teknik pengumpulan data dalam penelitian ini menggunakan teknik dokumentasi. Teknik dokumentasi dalam penelitian ini digunakan untuk memperoleh soal asesmen sumatif akhir semester gasal mata pelajaran Matematika kelas IV SD Negeri 39 Pontianak Kota tahun ajaran 2024/2025 dan data rekapitulasi hasil jawaban peserta didik kelas IV-A yang menunjukkan pilihan jawaban setiap peserta didik untuk tiap butir soal.

Instrumen penelitian dalam penelitian ini berupa dokumen soal asesmen sumatif akhir semester gasal mata pelajaran Matematika beserta kunci jawaban, serta data rekapitulasi hasil jawaban peserta didik kelas IV-A SD Negeri 39 Pontianak Kota yang diperoleh dari guru kelas. Analisis data secara kuantitatif dalam penelitian ini dilakukan dengan menggunakan *software* (perangkat lunak) *Microsoft Excel* beserta rumus yang peneliti aplikasikan guna menganalisisnya. Untuk tes hasil belajar umumnya dipertimbangkan klasifikasinya yaitu:

Validitas

Dalam penelitian ini, pengujian validitas dilakukan secara empiris, yaitu dengan menghitung korelasi antara skor butir soal dengan skor total menggunakan teknik korelasi *point biserial* dengan rumus sebagai berikut (Arikunto, 2013):

$$y_{pbi} = \frac{M_p - M_t}{S_t} \sqrt{\frac{p}{q}}$$

Keterangan:

y_{pbi} = Koefisien korelasi biserial

M_p = Rerata skor dari subyek yang menjawab benar bagi item yang dicari validitasnya

M_t = Rerata skor total

S_t = Standar deviasi dari skor total proporsi

p = Proporsi siswa yang menjawab benar

q = Proporsi siswa yang menjawab salah ($q = 1 - p$) (p.196-197).

Nilai y_{pbi} kemudian dibandingkan dengan nilai r_{tabel} pada taraf signifikansi 5% dengan $N = 28$, yaitu sebesar 0.374. Setelah didapatkan koefisien korelasi, untuk mengambil keputusan dengan ketentuan sebagai berikut. Jika nilai $y_{pbi} > r_{tabel}$, maka soal dinyatakan valid secara empiris (Kurniawan et al., 2022, p.154).

Reliabilitas

Dalam penelitian ini, reliabilitas butir soal dianalisis menggunakan rumus KR-20 (Kuder-Richardson Formula 20) (Kurniawan et al., 2022, p.166):

$$r_i = \frac{k}{(k-1)} \left[1 - \frac{\sum p_i q_i}{s_t^2} \right]$$

Keterangan:

r_i = Reliabilitas internal instrumen

k = Jumlah *item* soal

p_i = Proporsi banyaknya subjek yang menjawab benar

$q_i = 1 - p_i$ (proporsi subjek yang menjawab salah pada tiap soal)

s_t^2 = Variasi total skor, dihitung dengan:

$$s_t^2 = \frac{\sum (X_i - \bar{X})^2}{N}$$

di mana:

X_i = Skor total tiap peserta didik

$\bar{X}$ = Rata-rata skor total

N = Jumlah subjek

Tingkat Kesukaran

Tingkat kesukaran digunakan untuk mengetahui apakah suatu butir soal tergolong mudah, sedang, atau sukar bagi peserta didik. Tingkat kesukaran dihitung menggunakan rumus (Rohmad, 2017):

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

Keterangan:

P = Indeks kesukaran

B = Banyaknya peserta didik yang menjawab soal dengan benar

JS = Jumlah seluruh peserta tes (p.196)

Daya Pembeda

Daya pembeda mengacu pada kemampuan sebuah soal untuk membedakan antara peserta didik dengan kemampuan tinggi dan kemampuan rendah. Adapun daya pembeda dapat dihitung dengan rumus sebagai berikut (Arikunto, 2013):

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

Keterangan:

D = Daya pembeda

J = Jumlah peserta tes

B_A = Banyaknya kelompok atas yang menjawab benar

B_B = Banyaknya kelompok bawah yang menjawab benar

J_A = Jumlah kelompok atas

J_B = Jumlah kelompok bawah (p.165)

Kualitas Butir Soal Pilihan Ganda

Penentuan kualitas butir soal dalam penelitian ini dilakukan setelah melalui analisis terhadap empat aspek utama, yaitu validitas empiris, daya pembeda, tingkat kesukaran, serta didukung oleh aspek reliabilitas secara keseluruhan. Klasifikasi kualitas soal dibagi menjadi dua kategori, yaitu soal baik dan soal tidak baik, sebagaimana dijelaskan oleh Arikunto (2013), Ebel & Frisbie (1991), dan Arifin (2014). Kriteria yang digunakan sebagai acuan dalam klasifikasi kualitas soal adalah sebagai berikut:

1. Soal dikategorikan baik apabila memenuhi seluruh kriteria, yaitu valid secara empiris (dilihat dari korelasi skor butir dengan skor total), memiliki daya pembeda minimal cukup, serta tingkat kesukaran sedang. Soal dalam kategori ini dinyatakan layak untuk digunakan kembali atau disimpan dalam bank soal.
2. Soal dikategorikan tidak baik apabila tidak valid secara empiris dan/atau memiliki daya pembeda jelek atau sangat jelek. Soal dalam kategori ini tidak disarankan untuk digunakan kembali tanpa revisi menyeluruh.

Untuk memperjelas klasifikasi tersebut, berikut disajikan tabel pedoman klasifikasi kualitas butir soal:

Tabel 2. *Klasifikasi Kualitas Butir Soal dan Tindak Lanjut*

Kriteria yang Dipenuhi	Kategori Soal	Tindak Lanjut
Valid, daya pembeda $\geq$ cukup, kesukaran sedang	Baik	Disimpan/Digunakan kembali
Tidak valid dan/atau daya pembeda jelek/sangat jelek	Tidak Baik	Direvisi menyeluruh atau dibuang

HASIL DAN PEMBAHASAN

Validitas

Analisis validitas butir soal ASAS gasal matematika dengan menggunakan rumus korelasi *point biserial*. Dimana dengan mengkorelasikan skor *item* setiap butir soal dengan skor total. Setelah didapatkan nilai atau koefisien korelasi kemudian membandingkan dengan koefisien r_{tabel} . Apabila koefisien $y_{\text{pbi}} > r_{\text{tabel}}$ maka soal dinyatakan valid secara empiris. Jika r_{tabel} dengan melibatkan sebanyak 28 peserta didik diperoleh $r_{\text{tabel}} = 0,374$. Distribusi soal berdasarkan validitasnya bisa dilihat pada tabel berikut.

Tabel 3. Distribusi Soal ASAS Matematika Berdasarkan Validitas

Klasifikasi	Jumlah Soal	Nomor Soal
Valid	14	2, 3, 4, 5, 6, 7, 11, 13, 15, 16, 17, 18
Tidak Valid	11	1, 4, 8, 9, 10, 12, 14, 19, 20, 21, 22, 23, 24, 25

Berdasarkan hasil analisis tersebut, ditemukan bahwa sebanyak 8 soal (32%) termasuk dalam kategori validitas tinggi hingga sangat tinggi (soal nomor 3, 6, 7, 11, 13, 15, 16, dan 17). Soal-soal ini menunjukkan hubungan yang kuat antara skor butir dan skor total, menandakan bahwa butir soal tersebut relevan terhadap konstruk yang diukur; sebanyak 6 soal (24%) berada pada kategori cukup (soal nomor 5, 10, 12, 18, 20, dan 21), yang masih dapat diterima tetapi memerlukan perhatian lebih lanjut untuk penyempurnaan; sebanyak 7 soal (28%) memiliki validitas rendah hingga sangat rendah (soal nomor 2, 8, 9, 22, 23, 24, dan 25), sehingga kurang mencerminkan kompetensi yang diukur secara akurat. 4 soal (16%) berada dalam kategori tidak valid atau invalid (soal nomor 1, 4, 14, dan 19), karena koefisien korelasi negatif atau sangat rendah, sehingga tidak layak digunakan tanpa revisi signifikan.

Walaupun pengujian validitas bersifat statistik, terdapat beberapa kemungkinan penyebab rendahnya korelasi antara skor butir dan skor total. Beberapa faktor yang dapat memengaruhi validitas antara lain, redaksi soal yang kurang jelas atau membingungkan, tingkat kesukaran yang terlalu tinggi atau terlalu rendah, serta ketidaksesuaian antara isi soal dengan indikator pembelajaran. Selain itu, faktor dari peserta didik seperti ketidaktelitian dalam membaca soal atau menjawab secara asal juga dapat memengaruhi validitas butir soal (Ebel & Frisbie, 1991; (Kurniawan et al., 2022).

Dengan demikian, soal-soal yang telah memenuhi kriteria validitas dapat disimpan dalam bank soal dan digunakan kembali dalam asesmen berikutnya. Sedangkan soal dengan validitas rendah dan tidak valid perlu direvisi, baik dari segi substansi dan redaksi, agar instrumen dapat memberikan hasil pengukuran yang lebih akurat dan representatif.

Reliabilitas

Analisis reliabilitas soal pilihan ganda ASAS matematika dihitung dengan menggunakan rumus *Kuder Richardson* atau KR-20. Reliabilitas soal ditentukan dengan membandingkan nilai reliabilitas *Kuder Richardson* terhadap koefisien tetap sebesar 0,60. Jika hasil perhitungan *Kuder Richardson* $\geq 0,60$ maka soal dinyatakan reliabel. Namun sebaliknya jika hasil perhitungan reliabilitas $\leq 0,60$ maka

soal dinyatakan tidak reliabel. Dari hasil perhitungan diperoleh nilai koefisien KR-20 sebesar 0,70. Berdasarkan hasil tersebut, dapat disimpulkan bahwa soal ASAS matematika reliabel dengan kategori reliabilitas tinggi.

Menurut (Febriana, 2019), faktor yang mempengaruhi reliabilitas ditentukan oleh banyaknya *sampling*, perbedaan kemampuan dan peserta didik yang diuji, suasana dan kondisi peserta didik (p.127). Hal ini semakin banyak sampel maka nilai reliabilitas semakin tinggi, kemudian jika dalam pelaksanaan dalam suasana yang bising, tenang, atau banyak gangguan akan mempengaruhi nilai dan keandalan suatu tes.

Tingkat Kesukaran

Berdasarkan hasil perhitungan analisis tingkat kesukaran terhadap soal pilihan ganda ASAS matematika, terdapat 12% soal termasuk sukar, 40% soal sedang, dan 48% soal mudah. Tingkat kesukaran dihitung dengan membandingkan jumlah peserta didik yang menjawab benar tiap butir soal dengan seluruh jumlah peserta didik. Butir soal dianggap sukar jika $0 \leq P \leq 0,30$. Butir soal dianggap sedang jika memiliki indeks $0,30 \leq P \leq 0,70$. Kemudian, butir soal dianggap mudah jika indeks diantara $0,70 \leq P \leq 1,00$. Berikut ini distribusi soal berdasarkan tingkat kesukarannya.

Tabel 4. Distribusi Soal ASAS Matematika Berdasarkan Tingkat Kesukaran

Klasifikasi	Jumlah Soal	Nomor Soal
Sukar	3	9, 12, 8
Sedang	10	3, 4, 6, 7, 10, 13, 16, 17, 18, 24
Mudah	12	1, 2, 5, 8, 11, 14, 15, 19, 21, 22, 23, 25

Rata-rata indeks kesukaran keseluruhan (mean P) adalah sebesar 63,8%, yang termasuk dalam kategori sedang. Hal ini menunjukkan bahwa secara umum, tingkat kesukaran soal telah disusun cukup proporsional dan berada dalam rentang ideal untuk mengukur kemampuan peserta didik secara umum. Namun demikian, terdapat beberapa hal yang perlu diperhatikan. Salah satunya adalah soal nomor 9, yang tergolong sukar bukan karena tingkat kesulitan materi, melainkan karena diduga terdapat kesalahan pada kunci jawaban. Banyak peserta menjawab salah bukan karena tidak memahami konsep, tetapi karena tidak tersedia jawaban yang benar dalam pilihan opsi. Kesalahan teknis ini berdampak langsung pada penurunan indeks kesukaran dan dapat menurunkan validitas serta daya pembeda soal. Hal ini sejalan dengan pernyataan Ebel & Frisbie (1991), bahwa kualitas butir soal tidak hanya dipengaruhi oleh isi materi, tetapi juga oleh konstruk teknis dan ketepatan penyusunan opsi jawaban.

Di sisi lain, soal-soal dalam kategori mudah juga perlu menjadi perhatian. Soal yang terlalu mudah, menunjukkan bahwa sebagian besar peserta dapat menjawabnya dengan benar. Meskipun dapat mencerminkan keberhasilan pembelajaran, soal yang terlalu mudah berisiko tidak menantang peserta didik berkemampuan tinggi, dan menurunkan daya pembeda soal secara keseluruhan. Menurut Heaton (1995), soal dengan tingkat kesukaran tinggi atau rendah yang berlebihan sebaiknya dikurangi agar evaluasi dapat mencerminkan perbedaan kemampuan peserta secara lebih akurat. Oleh karena itu, penyusunan soal di masa mendatang perlu memperhatikan komposisi proporsional antara soal sukar, sedang, dan mudah, dengan dominasi soal pada tingkat kesukaran kategori sedang yang dinilai paling efektif dalam membedakan kemampuan peserta didik dan menggambarkan penguasaan materi secara representatif.

Daya Pembeda

Daya pembeda soal pilihan ganda ASAS matematika dianalisis dengan menggunakan rumus indeks diskriminasi (D). Daya pembeda soal dihitung dengan menghitung selisih proporsi antara peserta didik yang menjawab benar dari kelompok atas dan kelompok bawah. Suatu soal memiliki daya pembeda yang sangat jelek jika $D < 0$; jelek jika $0,00 \leq D \leq 0,20$; cukup jika $0,20 \leq D \leq 0,40$; baik jika $0,40 \leq D \leq 0,70$; baik sekali jika $0,70 \leq D \leq 1,00$. Berikut ini distribusi soal berdasarkan daya pembedanya.

Tabel 5. Distribusi Soal ASAS Matematika Berdasarkan Daya Pembeda

Klasifikasi	Jumlah Soal	Nomor Soal
Sangat Jelek	5	1, 14, 19, 20, 23
Jelek	8	2, 5, 8, 9, 12, 22, 24, 25
Cukup	5	4, 10, 11, 15, 21

Baik	6	3, 6, 7, 13, 15, 18,
Sangat Baik	1	17

Berdasarkan hasil penelitian yang telah dianalisis menggunakan program *Microsoft Excel*, diperoleh bahwa, sebanyak 5 butir soal termasuk dalam kategori sangat jelek. Butir soal dalam kategori ini tidak layak digunakan karena gagal membedakan antara peserta didik yang mampu dan yang kurang mampu; sebanyak 8 butir soal tergolong jelek. Soal dalam kategori ini juga tidak mampu membedakan secara efektif, karena peserta didik dengan kemampuan tinggi dan rendah memiliki peluang yang hampir sama dalam menjawab soal dengan benar. Sebanyak 5 butir soal termasuk dalam kategori soal cukup. Meskipun daya pembeda pada kategori ini tergolong sedang, (Elviana, 2020) berpendapat bahwa terdapat kemungkinan peserta didik yang kurang memahami materi menjawab soal dengan benar secara kebetulan (p.280); sebanyak 6 butir soal tergolong dalam kategori baik, yaitu soal nomor 3, 6, 7, 13, 15, dan 18. Butir-butir soal ini telah menunjukkan kemampuan membedakan secara signifikan antara peserta didik yang memahami materi dengan baik dan yang kurang memahami; hanya 1 butir soal yang tergolong sangat baik. Soal ini memiliki daya pembeda yang tinggi dan akurat dalam mengukur tingkat penguasaan materi. Soal ini dinilai sangat efektif dalam memisahkan peserta didik berdasarkan tingkat kemampuan.

Temuan ini menunjukkan bahwa sebagian besar soal (52%) memiliki daya pembeda yang jelek atau sangat jelek, yang berarti lebih dari separuh soal belum mampu menjalankan fungsi selektif secara optimal. Menurut Nurhalimah et al. (2022), salah satu penyebab rendahnya daya pembeda adalah karena soal tidak valid dan/atau memiliki tingkat kesukaran yang tidak proporsional, seperti terlalu mudah atau terlalu sulit (p.249–257). Hal ini sejalan dengan pendapat Heaton (1995), yang menyatakan bahwa soal dengan daya pembeda rendah seringkali merupakan soal yang ambigu, terlalu sederhana, atau tidak sesuai dengan indikator kemampuan yang seharusnya diukur. Selain itu, struktur soal yang kurang jelas serta kesalahan dalam kunci jawaban juga dapat menurunkan kualitas diskriminatif soal (Ebel & Frisbie, 1991). Oleh karena itu, perlu dilakukan evaluasi dan revisi terhadap soal-soal dengan daya pembeda rendah untuk memastikan bahwa setiap butir soal dapat berfungsi secara optimal dalam membedakan kemampuan peserta didik secara objektif dan adil.

Kualitas Butir Soal Pilihan Ganda

Hasil analisis kuantitatif yang dilakukan terhadap 25 butir soal pilihan ganda mata pelajaran Matematika kelas IV SD Negeri 39 Pontianak Kota mencakup lima aspek, yaitu validitas empiris, reliabilitas, daya pembeda, dan tingkat kesukaran. Berdasarkan hasil analisis tersebut, kualitas setiap butir soal kemudian diklasifikasikan ke dalam dua kategori, yaitu soal baik dan soal tidak baik. Kriteria soal baik mencakup soal yang memenuhi keempat aspek utama kualitas soal, yaitu valid secara empiris, memiliki daya pembeda minimal cukup, dan tingkat kesukaran sedang. Sementara itu, soal dikategorikan tidak baik apabila tidak valid dan/atau memiliki daya pembeda yang jelek atau sangat jelek. Soal-soal yang tergolong tidak baik disarankan untuk ditindaklanjuti dengan revisi ringan apabila hanya satu aspek minor yang belum terpenuhi (seperti validitas yang rendah), atau revisi menyeluruh jika dua atau lebih aspek tidak terpenuhi. Adapun soal yang tidak valid dan memiliki daya pembeda sangat jelek sebaiknya tidak digunakan kembali dan diganti dengan soal baru. Distribusi penjabaran butir soal berdasarkan kualitas butir soal bisa dilihat pada tabel berikut.

Tabel 6. Distribusi Soal Berdasarkan Kualitas

Klasifikasi	Jumlah Soal	Nomor Soal
Baik	8	3, 6, 7, 13, 15, 16, 17, 18
Tidak Baik	17	1, 2, 4, 5, 8, 9, 10, 11, 12, 14, 19, 20, 21, 22, 23, 24, 25

Berdasarkan hasil analisis terhadap 25 butir soal pilihan ganda mata pelajaran Matematika tahun ajaran 2024/2025, diperoleh bahwa, sebanyak 8 butir soal termasuk dalam kategori soal baik, karena telah memenuhi seluruh atau sebagian besar indikator kualitas seperti validitas minimal cukup, daya pembeda baik atau sangat baik, dan tingkat kesukaran sedang. Soal-soal ini direkomendasikan untuk digunakan kembali dengan sedikit revisi bila diperlukan; sebanyak 17 butir soal dikategorikan tidak baik, karena tidak memenuhi satu atau lebih kriteria penting. Soal nomor 2, 5, dan 23 memiliki validitas rendah serta daya pembeda sangat jelek, sehingga sebaiknya dibuang. Sementara itu, soal seperti nomor 1, 8, 9, 12, 14, 19, 20, 22, 24, dan 25 memerlukan revisi menyeluruh karena kelemahan signifikan di

aspek validitas dan daya pembeda. Adapun soal nomor 4, 10, 11, dan 21 masih bisa direvisi ringan, karena hanya gagal di satu kriteria saja.

Analisis ini menunjukkan bahwa mayoritas soal masih belum sepenuhnya memenuhi standar kualitas evaluasi. Meskipun demikian, instrumen soal memiliki reliabilitas tinggi dengan koefisien sebesar 0,70, yang menandakan konsistensi pengukuran cukup baik secara keseluruhan. Oleh karena itu, perlu dilakukan upaya peningkatan kualitas melalui peninjauan dan revisi butir soal yang belum optimal, guna menghasilkan instrumen asesmen yang benar-benar valid, reliabel, dan bermakna bagi proses evaluasi pembelajaran.

SIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan terhadap soal pilihan ganda Asesmen Sumatif Akhir Semester (ASAS) Gasal mata pelajaran matematika kelas IV SD Negeri 39 Pontianak Kota tahun ajaran 2024/2025, maka diperoleh beberapa kesimpulan sebagai berikut ini: 1) Validitas menunjukkan bahwa dari 25 butir soal yang dianalisis, sebanyak 11 butir soal (44%) dinyatakan valid karena memiliki korelasi positif yang signifikan terhadap skor total. Sementara itu, 14 butir soal (56%) dinyatakan tidak valid, karena memiliki korelasi yang rendah atau negatif, sehingga perlu diperbaiki atau diganti. 2) Reliabilitas soal mencapai indeks sebesar 0,70 yang menandakan bahwa soal ASAS matematika reliabel dengan kategori reliabilitas tinggi. 3) Tingkat kesukaran menunjukkan bahwa 12 butir soal (48%) berada dalam kategori mudah, 10 butir soal (40%) tergolong sedang, dan 3 butir soal (12%) termasuk dalam kategori sukar. 4) Daya pembeda soal bervariasi di mana terdapat 1 butir soal (4%) yang memiliki daya pembeda sangat baik, 6 butir soal (24%) baik, 5 butir soal (20%) cukup, 8 butir soal (32%) tergolong jelek, dan 5 butir soal (20%) sangat jelek. 5) Berdasarkan analisis menyeluruh terhadap 25 butir soal yang mencakup aspek validitas, reliabilitas, tingkat kesukaran, dan daya pembeda, diperoleh bahwa 8 butir soal (32%) tergolong soal baik karena memenuhi sebagian besar kriteria kualitas, dan dapat digunakan kembali setelah revisi ringan. Sementara itu, 17 butir soal (68%) tergolong tidak baik, dengan rekomendasi tindak lanjut berupa revisi menyeluruh atau penggantian. Temuan ini menunjukkan bahwa meskipun instrumen tergolong reliabel, kualitas butir soal secara keseluruhan masih perlu ditingkatkan agar dapat digunakan sebagai alat ukur yang valid dan akurat dalam mengevaluasi hasil belajar peserta didik.

UCAPAN TERIMA KASIH

Peneliti menyampaikan ucapan terima kasih kepada pihak yang sudah berkontribusi dalam pelaksanaan penelitian dan penyusunan artikel ini.

REFERENSI

- Arifin, Z. (2014). *Evaluasi Pembelajaran*. PT Remaja Rosdakarya.
- Arikunto, S. (2013). *Dasar-Dasar Evaluasi Pendidikan*. PT Bumi Aksara.
- Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of Educational Measurement*. Prentice-Hall.
- Elviana. (2020). Analisis Butir Soal Evaluasi Pembelajaran Pendidikan Agama Islam Menggunakan Program Pendahuluan Anates. *Jurnal Mudarissuna*, 10(2), 58–74.
- Febriana, R. (2019). *Evaluasi Pembelajaran*. PT Bumi Aksara.
- Heaton, J. B. (1995). *Writing English Language Tests*. Longman.
- Kurniawan, A., Febrianti, A. N., Hardianti, T., Ichsan, Desy, Risan, R., Sari, D. M. M., Sitopu, J. W., Dewi, R. S., Fitriyah, L. A., Zulkarnaini, Jalal, N. M., G, H., & Hasyim, F. (2022). *Evaluasi Pembelajaran*. PT Global Eksekutif Teknologi.
- Nurhalimah, S., Hidayati, Y., Rosidi, I., & Hadi, W. P. (2022). Hubungan antara Validitas Item Dengan Daya Pembeda dan Tingkat Kesukaran Soal Pilihan Ganda PAS. *Jurnal Natural Science Education Research*, 4(3), 249–257. <https://doi.org/10.21107/nser.v4i3.8682>
- Rohmad. (2017). *Pengembangan Instrumen Evaluasi dan Penelitian*. KALIMEDIA.
- Sudijono, A. (2017). *Pengantar Evaluasi Pendidikan*. Raja Grafindo.
- Sugiyono. (2013). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. alfabeta.