ANALISIS KUALITAS BUTIR SOAL AKHIR SEMESTER (SAS) IPAS KELAS V BERDASARKAN PENDEKATAN *CLASSICAL TEST THEORY* (CTT)

Sitti Hajar[1] Siti Zahra Mulianti Natsir[2] Aedil Akmal[3] Suci Rahmadani Syam[4]
[124]PGSD FIP Universitas Negeri Makassar
[3]Pendidikan Teknik Elektronika FT Universitas Negeri Jakarta
Alamat e-mail: [1]sittihajar@unm.ac.id,
Alamat e-mail: [2]siti.zahra.mulianti@unm.ac.id,
Alamat e-mail: [3]aedilakmal@unj.ac.id,
Alamat e-mail: [4]suciramadhanisyam05@gmail.com

ABSTRACT

This study aims to analyze the quality of Final Semester Assessment items in the Science and Social Studies subject for Grade V at SDN Minasa Upa, Makassar City, in the 2025/2026 academic year using the Classical Test Theory (CTT) approach. This research employed a descriptive quantitative design involving 30 students. The primary data were collected through documentation methods, including test blueprints, test items, answer keys, and students' answer sheets, and were supported by questionnaires and interviews with teachers. The quality of test items was analyzed based on five aspects: validity, reliability, difficulty level, discriminating power, and distractor effectiveness. The results showed that 64% of the items were valid and 36% were invalid, with high reliability (KR-20 = 0.79). The distribution of difficulty levels consisted of 12% difficult, 52% moderate, and 36% easy items. A total of 56% of the items had very good discriminating power, although some items showed low and negative indices. All items had at least two functioning distractors, although several options were not optimally functioning. The findings indicate that the overall quality of the test items was categorized as adequate, although several items still require revision, particularly those with low validity and negative discrimination indices.

Keywords: Item analysis, validity, reliability, difficulty level, discriminating power, distractor effectiveness.

ABSTRAK

Penelitian ini bertujuan untuk menganalisis kualitas butir Soal Akhir Semester (SAS) mata pelajaran Ilmu Pengetahuan Alam dan Sosial (IPAS) kelas V SDN Minasa Upa Kota Makassar Tahun Pelajaran 2025/2026 menggunakan pendekatan *Classical Test Theory* (CTT). Penelitian ini merupakan penelitian deskriptif kuantitatif dengan pendekatan evaluatif. Jumlah subjek penelitian sebanyak 30 orang siswa. Data dikumpulkan melalui metode dokumentasi berupa kisi-kisi soal, naskah soal, kunci jawaban, dan lembar jawaban siswa lalu dianalisis menggunakan pendekatan CTT. Kualitas butir soal dianalisis berdasarkan lima aspek, yaitu validitas, reliabilitas, tingkat kesukaran, daya beda, dan efektivitas pengecoh. Hasil penelitian

menunjukkan bahwa 64% butir soal valid dan 36% tidak valid, dengan reliabilitas tinggi (KR-20 = 0,79). Distribusi tingkat kesukaran terdiri atas 12% soal sulit, 52% sedang, dan 36% mudah. Sebanyak 56% butir soal memiliki daya beda sangat baik, meskipun masih ditemukan butir dengan daya beda rendah dan negatif. Seluruh butir soal memiliki minimal dua pengecoh yang berfungsi, namun beberapa opsi masih tidak berfungsi secara optimal. Hasil penelitian menunjukkan bahwa secara umum kualitas butir soal berada pada kategori cukup baik, meskipun masih ditemukan beberapa butir dengan validitas rendah dan daya beda negatif yang memerlukan revisi. Hasil ini juga dapat menjadi dasar perbaikan instrumen evaluasi pembelajaran dan pengembangan kompetensi pedagogik guru sekolah dasar.

Kata Kunci: Analisis butir soal, Classical Test Theory, validitas, reliabilitas, daya beda.

A. Pendahuluan

Mutu evaluasi pembelajaran erat kaitannya dengan mutu alat ukur yang digunakan dalam proses evaluasi. Pada jenjang sekolah dasar, tes pilihan ganda masih menjadi bentuk instrumen utama dalam pelaksanaan Soal Akhir Semester (SAS). Agar hasil penilaian dapat dijadikan dasar pengambilan keputusan pedagogis yang akurat dan dapat dipertanggungjawabkan, setiap butir soal perlu memenuhi karakteristik psikometrik, mencakup aspek validitas, reliabilitas, indeks kesukaran, daya pembeda, serta fungsi distraktor (Arikunto, 2018; Rumahlewang, 2025). Apabila instrumen tidak memenuhi kriteria psikometrik tersebut, informasi hasil penilaian yang diperoleh dapat menjadi kurang akurat. Kondisi ini pada akhirnya dapat melemahkan validitas evaluasi serta ketepatan keputusan pendidikan yang diambil secara ilmiah.

Upaya untuk mengetahui kualitas suatu instrumen evaluasi dilakukan melalui analisis butir soal. Arikunto (2018) menjelaskan bahwa analisis butir soal bertujuan mengidentifikasi butir soal yang berkualitas baik, kurang baik, maupun tidak layak digunakan sehingga dapat menjadi dasar perbaikan instrumen secara berkelanjutan. Senada dengan itu, Sudijono (2011) menyatakan bahwa analisis butir soal mampu menghasilkan informasi spesifik mengenai kualitas instrumen yang tidak dapat diperoleh hanya melalui penilaian kualitatif. Laila dkk. (2024) juga menegaskan bahwa kualitas siklus evaluasi pendidikan hanya dapat dipertahankan melalui analisis

instrumen yang dilakukan secara sistematis dan berkelanjutan.

Pentingnya analisis butir soal turut diperkuat oleh berbagai temuan empiris yang menunjukkan bahwa soal yang disusun guru masih memiliki kelemahan psikometrik yang cukup signifikan. Rezigalla dkk. (2024) menemukan bahwa efektivitas pengecoh berpengaruh terhadap tingkat kesukaran dan daya pembeda butir soal, sehingga pengecoh yang tidak berfungsi dapat menurunkan kualitas instrumen evaluasi pembelajaran. Selain itu, Konakci (2024) menegaskan bahwa kualitas soal dapat dianalisis berdasarkan aspek tingkat kesukaran, kemampuan membedakan peserta didik, dan fungsi pengecoh. Nafs dkk. (2023) juga menyatakan bahwa kualitas instrumen evaluasi dapat ditinjau melalui aspek validitas, reliabilitas, tingkat kesukaran, daya beda, dan efektivitas pengecoh. Sejalan dengan itu, Batool et al. (2023) menegaskan bahwa analisis butir soal menggunakan pendekatan Classical Test Theory (CTT) maupun Item Response Theory (IRT) merupakan langkah penting untuk memastikan validitas dan akurasi instrumen dalam mengukur kemampuan peserta didik.

Berbagai penelitian sebelumnya menunjukkan bahwa kualitas soal yang disusun guru masih belum optimal. Verawati et al. (2023), misalnya, melaporkan bahwa hanya 54% butir soal UAS Matematika yang dinyatakan valid, sedangkan 43% pengecoh berada pada kategori sangat buruk. Temuan tersebut menunjukkan bahwa kelemahan kualitas soal bukan merupakan fenomena yang bersifat insidental, melainkan permasalahan sistemik yang memerlukan penanganan melalui analisis butir soal secara berkelanjutan. Kondisi serupa juga ditemukan pada konteks yang lebih spesifik, yaitu di SDN Minasa Upa Kota Makassar.

Berdasarkan hasil wawancara awal, guru pembuat soal (G1) telah memiliki pengalaman mengajar selama sembilan tahun dan pernah mengikuti pelatihan penyusunan soal, analisis butir soal, serta *Higher Order Thinking Skills* (HOTS). Namun demikian, analisis butir soal belum dilaksanakan secara konsisten setelah evaluasi pembelajaran dilakukan. Kondisi tersebut menunjukkan adanya kesenjangan antara pemahaman teoritis guru dan implementasi praktis evaluasi

pembelajaran di sekolah. Fachri (2018) dalam Rumahlewang (2025) menegaskan bahwa tanpa analisis instrumen yang dilakukan secara periodik, hasil evaluasi tidak akan mampu memberikan umpan balik yang optimal bagi perbaikan pembelajaran.

Berbeda dengan penelitian terdahulu yang pada umumnya hanya menitikberatkan pada analisis statistik butir soal, penelitian ini tidak hanya menganalisis kualitas psikometrik instrumen berdasarkan pendekatan *Classical Test Theory* (CTT), tetapi juga mengidentifikasi faktor pedagogik dan institusional yang memengaruhi kualitas penyusunan soal oleh guru sekolah dasar. Oleh sebab itu, penelitian ini diharapkan mampu memberikan kontribusi praktis terhadap pengembangan budaya evaluasi pembelajaran yang berbasis data di sekolah dasar.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan: (1) menganalisis kualitas butir soal SAS IPAS kelas V SDN Minasa Upa berdasarkan validitas, reliabilitas, tingkat kesukaran, daya beda, dan efektivitas pengecoh; dan (2) mengidentifikasi faktor-faktor yang memengaruhi kemampuan guru menyusun butir soal berkualitas.

B. Metode Penelitian

Penelitian ini menggunakan desain deskriptif-kuantitatif dengan pendekatan evaluatif. Lokasi penelitian ini yakni di SDN Minasa Upa, Kecamatan Rappocini, Kota Makassar dengan akreditasi A. Subjek penelitian adalah seluruh 30 siswa kelas V Tahun Pelajaran 2025/2026.

Jumlah subjek sebanyak 30 siswa digunakan karena penelitian ini dilakukan pada satu kelas secara utuh dalam konteks evaluasi pembelajaran di tingkat sekolah dasar. Meskipun jumlah subjek relatif terbatas, analisis butir soal tetap dapat dilakukan untuk kepentingan evaluasi kelas dan perbaikan instrumen pembelajaran sebagaimana dikemukakan oleh Arikunto (2018), terutama pada penelitian evaluatif dengan pendekatan deskriptif-kuantitatif.

Data dikumpulkan melalui metode dokumentasi, meliputi: (1) kisi-kisi soal SAS Ganjil IPAS kelas V, (2) naskah soal SAS Ganjil IPAS kelas V yang terdiri atas 25 butir pilihan ganda dengan empat opsi jawaban (A–D), (3) kunci jawaban resmi, dan (4) seluruh lembar jawaban siswa. Selain itu, wawancara dilakukan

kepada guru pembuat soal untuk memperoleh informasi pendukung mengenai proses penyusunan instrumen evaluasi.

Pengolahan data dilakukan melalui tiga tahap. Pertama, setiap jawaban siswa dicocokkan dengan kunci jawaban untuk menghasilkan matriks skor dikotomus (benar = 1, salah = 0) berukuran 30 × 25. Kedua, skor total per siswa dihitung sebagai penjumlahan skor seluruh butir. Ketiga, matriks skor beserta skor total menjadi basis perhitungan analisis butir soal dengan pendekatan *Classical Test Theory* (CTT), yang mencakup lima aspek: validitas, reliabilitas, tingkat kesukaran, daya beda, dan efektivitas pengecoh.

Analisis data dilakukan menggunakan bantuan Microsoft Excel. Validitas butir soal diukur menggunakan korelasi point biserial (Arikunto, 2018):

$$r_{pbi} = \frac{M_p - M_t}{S_t} \times \sqrt{\frac{\overline{p}}{q}}$$

Keterangan:

r_{pbi} = koefisien korelasi biserial
M_p = rata-rata skor jawaban benar
M_t = rata-rata skor total
S_t = standar deviasi skor total
p =proporsi siswa yang menjawab benar
q =proporsi siswa yang menjawab salah (1 – p)

(Arikunto, 2018)

Butir soal dinyatakan valid apabila koefisien korelasi r ≥ 0,30 (Arikunto, 2018).

Reliabilitas tes dihitung menggunakan formula Kuder-Richardson 20 (KR-20) sebagai berikut (Arikunto, 2018) :

$$r_{11} = (\frac{n}{n-1}) \times (\frac{S^2 - \sum pq}{S^2})$$

Keterangan:

r44 = reliabilitas tes
n = jumlah butir soal
S^2 = varians skor total
p = proporsi jawaban benar
q = proporsi jawaban salah (1–p)

(Arikunto, 2018)

Tes dinyatakan reliabel apabila nilai KR-20 ≥ 0,70 (Arikunto, 2018).

Daya beda dihitung menggunakan rumus berikut (Arikunto,2018):

$$D = P_A - P_B$$

Keterangan:

D = indeks daya beda
P_A = proporsi kelompok atas yang menjawab benar ($\frac{BA}{JA}$)
P_B = proporsi kelompok bawah yang menjawab benar ($\frac{BB}{J_B}$)

B_A = jumlah siswa kelompok atas yang menjawab benar

B_B = jumlah siswa kelompok bawah yang menjawab benar

J_A = banyaknya siswa kelompok atas

J_B = banyaknya siswa kelompok bawah

(Arikunto, 2018)

Kategorisasi daya beda mengacu pada kriteria Arikunto (2018). Daya beda dikategorikan jelek apabila D < 0,20; cukup apabila D = 0,20–0,29; baik apabila D = 0,30–0,39; dan sangat baik apabila D ≥ 0,40.

Tingkat kesukaran dihitung menggunakan rumus berikut (Arikunto,2018):

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

Keterangan:

P = indeks tingkat kesukaran s

B = jumlah siswa yang menjawab benar

N = jumlah seluruh siswa

(Arikunto, 2018)

Kategorisasi tingkat kesukaran mengacu pada kriteria Arikunto (2018). Butir soal dikategorikan sulit apabila P ≤ 0,30; sedang apabila P= 0,31–0,70; dan mudah apabila P > 0,70.

Efektivitas pengecoh dianalisis menggunakan indeks pengecoh (IP) berikut (Arifin, 2012 dalam Verawati et al., 2023) :

$$IP = \frac{P}{(N-B)/(n-1)} \times 100$$

Keterangan:

IP = indeks pengecoh

P = banyaknya pemilih pengecoh

N = jumlah seluruh peserta tes

B = banyaknya siswa yang menjawab benar

n = jumlah alternatif jawaban (pilihan)

1 = bilangan tetap

(Arifin, 2012 dalam Verawati et al., 2023)

Pengecoh dinyatakan berfungsi apabila dipilih oleh minimal 5% dari seluruh peserta tes (Arifin, 2012 dalam Verawati et al., 2023).

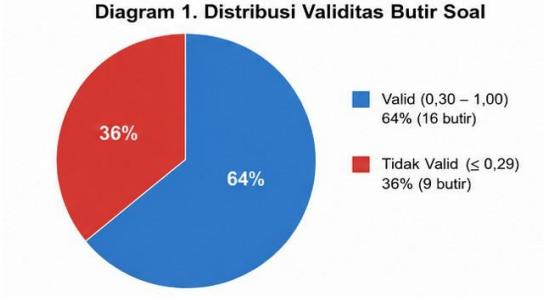
Data hasil wawancara digunakan sebagai informasi pendukung dalam menginterpretasikan hasil analisis butir soal.

C. Hasil Penelitian dan Pembahasan

1. Hasil Penelitian

a. Validitas

Validitas butir soal dianalisis menggunakan korelasi point biserial dengan kriteria r lebih dari atau sama dengan 0,30. Hasil analisis disajikan pada Diagram 1.



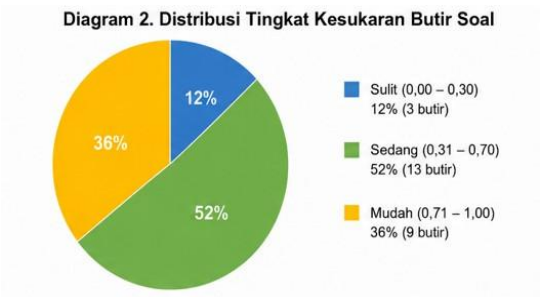
Sebanyak 16 butir soal (64%) dinyatakan valid dan 9 butir soal (36%) tidak valid. Koefisien validitas tertinggi terdapat pada soal nomor 13 (r = 0,78) dan terendah pada soal nomor 21 (r = -0,26). Dua butir soal, yaitu nomor 18 dan 21, menunjukkan indeks validitas negatif yang mengindikasikan adanya kelemahan dalam konstruksi soal sehingga perlu dilakukan tindak lanjut.

b. Reliabilitas

Reliabilitas butir soal dihitung menggunakan formula KR-20 dan diperoleh koefisien r_{44} = 0,79. Nilai tersebut menunjukkan bahwa perangkat soal memiliki reliabilitas tinggi.

c. Tingkat Kesukaran

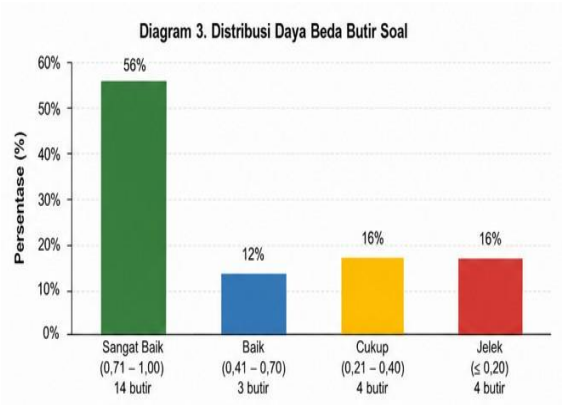
Distribusi tingkat kesukaran soal disajikan pada Diagram 2.



Tiga butir soal (12%) berada pada kategori sulit, 13 butir soal (52%) pada kategori sedang, dan 9 butir soal (36%) pada kategori mudah. ITK tertinggi diperoleh soal nomor 2 dan 23 (p = 0,87), sedangkan terendah pada soal nomor 25 (p = 0,27).

d. Daya Beda

Hasil analisis daya beda disajikan pada Diagram 3.



Sebanyak 14 butir soal (56%) berkategori sangat baik (D lebih dari atau sama dengan 0,40), 3 butir (12%) baik, 4 butir (16%) cukup, dan 4 butir (16%) jelek. Dua butir soal memiliki indeks daya beda negatif: soal nomor 18 (D=-0,125) dan nomor 21 (D=-0,25). Daya beda tertinggi diperoleh soal nomor 20 (D = 0,875).

e. Efektivitas Pengecoh

Efektivitas pengecoh dianalisis berdasarkan frekuensi pemilihan opsi non-kunci oleh seluruh peserta tes. Hasil analisis disajikan pada Tabel 4.

Tabel 4. Distribusi Butir Soal Berdasarkan Efektivitas Pengecoh

No.	Kategori Efektivitas Pengecoh	Jumlah	Persentase
1	Sangat Baik (3 pengecoh aktif)	17	68%
2	Baik (2 pengecoh aktif)	8	32%
3	Kurang Berfungsi (<=1 pengecoh aktif)	0	0%

Sebanyak 17 butir soal (68%) memiliki seluruh pengecoh yang berfungsi (3 opsi non-kunci dipilih) yaitu soal nomor 1, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 16, 17, 18, 22, 24, 25, sedangkan 8 butir soal (32%) memiliki 2 pengecoh yang berfungsi yakni soal nomor 2, 9, 14, 15, 19, 20, 21, 23. Tidak ditemukan butir soal dengan pengecoh yang seluruhnya tidak berfungsi. Namun terdapat sejumlah opsi tertentu yang tidak dipilih satu pun siswa, khususnya pada soal nomor 9 dan 14.

Berdasarkan wawancara kepada guru pembuat soal, tiga faktor struktural teridentifikasi sebagai penentu kualitas soal. Pertama, distribusi pelatihan yang tidak merata meskipun guru telah mengikuti pelatihan analisis soal. Kedua, dukungan institusi berupa alokasi waktu dan program pengembangan profesional masih belum konsisten. Ketiga, literasi teknologi asesmen masih rendah sehingga analisis berbasis perangkat lunak belum terlaksana secara sistematis.

2. Pembahasan

Secara umum, kualitas butir soal SAS IPAS kelas V SDN Minasa Upa berada pada kategori cukup baik. Dari aspek validitas, menunjukkan bahwa sebanyak 64% butir soal telah memenuhi kriteria valid. Temuan ini menunjukkan bahwa sebagian besar soal mampu mengukur kompetensi yang seharusnya diukur sesuai dengan tujuan evaluasi pembelajaran. Menurut Magdalena et al. (2021), validitas merupakan aspek utama dalam menentukan kualitas instrumen evaluasi karena berkaitan langsung dengan ketepatan pengukuran hasil belajar siswa. Instrumen yang valid akan menghasilkan data yang lebih akurat sehingga dapat dijadikan dasar dalam pengambilan keputusan pembelajaran.

Selain itu, Tarmizi et al. (2020) menjelaskan bahwa analisis validitas butir soal amat pentinh untuk dilakukan guna mengidentifikasi soal yang layak dipertahankan maupun yang perlu direvisi. Namun demikian, masih ditemukannya 36% butir soal yang tidak valid pada penelitian ini menunjukkan bahwa sebagian soal belum mampu merepresentasikan

konstruk yang diukur secara optimal. Kondisi tersebut kemungkinan disebabkan oleh redaksi soal yang ambigu, ketidaksesuaian indikator dengan materi pembelajaran, atau kurang tepatnya penyusunan opsi jawaban. Dua butir dengan validitas negatif (soal 18: r = -0,16; soal 21: r = -0,26) bersifat kontraproduktif karena siswa berkemampuan tinggi justru lebih sering menjawab salah. Menurut Rumahlewang (2025), kondisi ini lazim disebabkan oleh ambiguitas stem soal, ketidaktepatan kunci jawaban, atau cakupan materi yang tidak diajarkan secara eksplisit. Temuan ini selaras dengan Saputri et al. (2023) yang menegaskan bahwa validitas rendah pada suatu butir soal mengindikasikan ketidaktepatan instrumen dalam mengukur konstruk yang dimaksud.

Reliabilitas perangkat soal pada penelitian ini tergolong tinggi dengan koefisien KR-20 sebesar 0,79. Hasil tersebut menunjukkan bahwa instrumen memiliki konsistensi internal yang baik dalam mengukur kemampuan peserta didik. Menurut Suharsimi Arikunto (2018), reliabilitas yang tinggi mengindikasikan bahwa instrumen evaluasi mampu menghasilkan pengukuran yang relatif konsisten ketika digunakan dalam kondisi yang serupa. Magdalena et al. (2021) sejalan dengan Arikunto. Ia menyatakan bahwa reliabilitas merupakan indikator penting dalam menjamin konsistensi hasil evaluasi pembelajaran. Selain itu, penelitian Tarmizi et al. (2020) menemukan bahwa instrumen dengan reliabilitas tinggi cenderung lebih efektif digunakan dalam proses evaluasi karena mampu meminimalkan kesalahan pengukuran. Dengan demikian, nilai reliabilitas pada penelitian ini menunjukkan bahwa perangkat soal SAS IPAS telah memenuhi syarat sebagai instrumen evaluasi yang cukup konsisten meskipun masih terdapat beberapa butir soal yang perlu direvisi.

Distribusi tingkat kesukaran pada penelitian ini didominasi oleh kategori sedang, yaitu sebesar 52%. Temuan tersebut sesuai prinsip Arikunto (2018) bahwa soal yang baik tidak boleh terlalu mudah maupun terlalu sukar. Imam (2023) juga menyatakan bahwa soal dengan tingkat kesukaran sedang lebih efektif digunakan dalam evaluasi pembelajaran karena mampu mengukur kemampuan siswa secara lebih proporsional. Selain itu,

penelitian oleh Rezigalla et al. (2024) menjelaskan bahwa distribusi tingkat kesukaran yang seimbang dapat meningkatkan kualitas instrumen evaluasi karena mampu mengakomodasi berbagai tingkat kemampuan peserta didik. Distribusi tingkat kesukaran yang belum proporsional menunjukkan bahwa jumlah soal sulit masih relatif sedikit. Kondisi ini berpotensi mengurangi fungsi diagnostik SAS, khususnya dalam konteks Kurikulum Merdeka yang menekankan pengembangan Higher Order Thinking Skills (HOTS). Kondisi ini berpotensi memunculkan efek atap (*ceiling effect*) bagi siswa berkemampuan tinggi, sehingga distribusi yang lebih proporsional sekitar 20% sampai 25% pada kategori sulit direkomendasikan (Arikunto, 2018).

Dari aspek daya beda, mayoritas soal memiliki kemampuan pembeda yang sangat baik. 56% butir berkategori sangat baik (D lebih dari atau sama dengan 0,40) merupakan kekuatan psikometrik utama butir soal ini. Butir dengan daya beda tinggi berkontribusi langsung pada akurasi pengukuran dengan memisahkan siswa yang menguasai materi dari yang belum (Rumahlewang, 2025). Hal ini juga memperlebar variansi skor total yang menopang tingginya reliabilitas. Sebaliknya, empat butir berkategori jelek dan dua butir berindeks negatif (soal 18: D = -0,125; soal 21: D = -0,25) menunjukkan disfungsi pengukuran serius. Perlu digarisbawahi bahwa kedua butir berindeks negatif tersebut juga tidak valid, yang mengonfirmasi bahwa permasalahannya bersifat konstruktif mendasar. Vincent dan Shanmugam (2020) menegaskan bahwa butir soal dengan indeks daya beda negatif atau mendekati nol harus segera direvisi atau dieliminasi karena tidak memenuhi fungsi pengukuran yang valid dalam sebuah tes standar.

Menurut Apriliani (2026), daya beda merupakan indikator penting dalam menentukan kualitas instrumen evaluasi karena berkaitan dengan kemampuan soal dalam mengidentifikasi tingkat penguasaan materi peserta didik. Selain itu, Rezigalla et al. (2024) menyatakan bahwa butir soal dengan daya beda tinggi memberikan kontribusi positif terhadap kualitas pengukuran hasil belajar siswa. Akan tetapi, masih ditemukan beberapa butir soal dengan daya beda negatif pada penelitian ini. Kondisi tersebut menunjukkan adanya

kemungkinan kesalahan dalam konstruksi soal, ketidaktepatan kunci jawaban, atau penggunaan redaksi yang ambigu sehingga perlu dilakukan revisi secara menyeluruh.

Efektivitas pengecoh pada penelitian ini tergolong baik karena seluruh butir soal memiliki minimal dua pengecoh yang berfungsi. Temuan ini menunjukkan bahwa sebagian besar opsi jawaban mampu menarik perhatian siswa yang belum memahami materi dengan baik. Rezigalla et al. (2024) menjelaskan bahwa efektivitas pengecoh berpengaruh langsung terhadap tingkat kesukaran dan daya beda soal. Distraktor (pengecoh) yang tidak berfungsi dapat meningkatkan peluang siswa menjawab benar melalui tebakan sehingga menurunkan kualitas instrumen evaluasi. Selain itu, penelitian oleh Magdalena et al. (2021) menegaskan bahwa pengecoh yang baik harus disusun berdasarkan miskonsepsi nyata peserta didik agar mampu berfungsi secara optimal. Oleh karena itu, meskipun hasil penelitian menunjukkan efektivitas pengecoh yang cukup baik, beberapa opsi jawaban yang tidak dipilih siswa tetap perlu diperbaiki agar kualitas instrumen evaluasi menjadi lebih optimal.

Temuan penelitian menunjukkan bahwa masih terdapat beberapa butir soal dengan validitas rendah dan daya beda negatif. Kondisi ini mengindikasikan bahwa proses penyusunan soal belum sepenuhnya memperhatikan prinsip psikometrik dalam pengembangan instrumen evaluasi. Berdasarkan hasil wawancara, guru belum melaksanakan analisis butir soal secara sistematis setelah pelaksanaan evaluasi sehingga revisi soal belum dilakukan secara optimal. Hal tersebut disebabkan beberapa faktor, yakni: (1) keterbatasan pelatihan teknis analisis butir soal yang tidak terdistribusi secara merata; (2) lemahnya dukungan institusional berupa tidak adanya program analisis butir soal yang terlembaga pascaujian; dan (3) rendahnya pemanfaatan teknologi asesmen. Zulkarnain et al. (2025) menemukan bahwa implementasi kompetensi pedagogik guru sekolah dasar sangat bergantung pada akses terhadap program pengembangan profesional yang berkelanjutan. Absennya kebijakan sekolah yang mewajibkan analisis butir soal pascaujian

menyebabkan kualitas soal tidak pernah dievaluasi secara formal. Daga et al. (2022) juga menegaskan bahwa ekosistem sekolah yang mendukung praktik reflektif merupakan prasyarat bagi implementasi kompetensi pedagogik yang optimal.

E. Kesimpulan

Kualitas butir soal SAS IPAS kelas V SDN Minasa Upa Tahun Pelajaran 2025/2026 secara umum berada pada kategori cukup baik berdasarkan pendekatan *Classical Test Theory* (CTT). Sebagian besar butir soal telah memenuhi kriteria validitas, reliabilitas, tingkat kesukaran, daya beda, dan efektivitas pengecoh. Reliabilitas tes tergolong tinggi sehingga instrumen cukup konsisten dalam mengukur kemampuan peserta didik. Meskipun demikian, masih ditemukan beberapa butir soal dengan validitas rendah dan daya beda negatif yang menunjukkan perlunya revisi pada aspek konstruksi soal, ketepatan kunci jawaban, dan redaksi soal. Selain itu, keterbatasan pelaksanaan analisis butir soal secara sistematis serta rendahnya pemanfaatan teknologi asesmen menjadi faktor yang memengaruhi kualitas penyusunan instrumen evaluasi. Oleh karena itu, pelaksanaan analisis butir soal secara berkelanjutan perlu dioptimalkan untuk meningkatkan kualitas evaluasi pembelajaran di sekolah dasar.

DAFTAR PUSTAKA

Apriliani, A. (2026). Analisis Kualitas Butir Soal Menggunakan ANATES pada Pembelajaran Sekolah Dasar. *Pendas: Jurnal Ilmiah Pendidikan Dasar*, *11*(1), 961–970. https://journal.unpas.ac.id/index.php/pendas/article/download/38965/23542.

Arikunto, S. (2018). *Dasar-Dasar Evaluasi Pendidikan* (Edisi 3). Jakarta: Bumi Aksara.

Batool, I., Shah, A. A., & Naseer, S. (2023). Construction, Analysis and Calibration of Multiple-Choice Questions: IRT versus CTT. *Archives of Educational Studies (ARES), 3*(2), 242–257. https://ares.pk/ojs/index.php/ares/article/view/69

Daga, A. T., Wahyudin, D., & Susilana, R. (2022). Implementation of Teacher's Pedagogic Competence In The 2013 Curriculum Learning Process In Elementary School. *Eduvest: Journal of Universal Studies, 2*(5), 900–909. https://eduvest.greenvest.co.id/index.php/edv/article/view/439.

Imam, M. K. (2023). Analisis Daya Pembeda Dan Distraktor Soal Sekolah Dasar. *Jurnal Pendidikan Indonesia*, Vol. 9 No. 1 (2023). https://journal.upy.ac.id/index.php/JPI/article/view/4856

Konakci, S. (2024). Item analysis in multiple choice questions: A study on question difficulty and authors' evaluation. *Journal of Basic and Clinical Health Sciences*, *8*(2), 490–497. https://doi.org/10.30621/jbachs.1427464

Laila, L., Nabila, A., & Widyanti, E. (2024). Konsep Dasar Evaluasi Pembelajaran. *Jurnal Manajemen dan Pendidikan Agama Islam, 2*(5), 252–262. https://doi.org/10.61132/jmpai.v2i5.536

Magdalena, I., Fauzi, H. N., & Putri, R. (2021). Analisis Validitas, Reliabilitas, Tingkat Kesukaran Dan Daya Beda Butir Soal Ujian Akhir Semester Tema 7 Kelas III SDN Karet 1 Sepatan. *BINTANG: Jurnal Pendidikan dan Sains, 3*(2), 198–214. https://ejournal.stitpn.ac.id/index.php/bintang/article/download/1291/902

Nafs, H., Sridana, N. ., Hikmah, N. ., & Soeprianto, H. . (2023). Analisis Kualitas Butir Soal Ulangan Akhir Semester Genap Mata Pelajaran Matematika Kelas VII SMPN 6 Mataram Tahun Ajaran 2022/2023. *Jurnal Ilmiah Profesi Pendidikan*, *8*(4), 2324–2331. https://doi.org/10.29303/jipp.v8i4.1785

Rezigalla, A. A., et al. (2024). Item Analysis: The Impact of Distractor Efficiency on The Difficulty Index and Discrimination Power of Multiple-Choice Items. *BMC Medical Education, 24*(1),

445. https://doi.org/10.1186/s12909-024-05433-y

Rumahlewang, E. (2025). *Evaluasi Pembelajaran: Konsep dan Manajemen*. Bandung: Penerbit Widina Media Utama.

Saputri, H. A., Zulhijrah, Z., Larasati, N. J., & Shaleh, S. (2023). Analisis Instrumen Asesmen: Validitas, Reliabilitas, Tingkat Kesukaran dan Daya Beda Butir Soal. *Didaktik: Jurnal Ilmiah PGSD FKIP Universitas Mandiri, 9*(5), 2986–2995. https://doi.org/10.51574/didaktik.v9i5.1527

Sudijono, A. (2011). *Pengantar Evaluasi Pendidikan*. Jakarta: PT RajaGrafindo Persada.

Tarmizi, P., Nebula, N., & Anggraini, D. (2020). Analisis Butir Soal Tematik Sekolah Dasar. *ELSE (Elementary School Education Journal), 4*(2), 142–151. https://journal.um-surabaya.ac.id/pgsd/article/view/7090/3507

Verawati, Y., Siskawati, F. S., & Susilaningtyas, T. (2023). Analisis Butir Soal Ujian Akhir Semester (UAS) Mata Pelajaran Matematika Pada Tahun Ajaran 2020/2021 Kelas VII SMP Islam At Tanwir Kecamatan Ledokombo Kabupaten Jember. *Jurnal Jendela Pendidikan, 3*(1), 114–121. https://ejournal.jendelaedukasi.id/index.php/JJP/article/view/422.

Vincent, W., & Shanmugam, S. K. S. (2020). The Role of Classical Test Theory to Determine The Quality Of Classroom Teaching Test Items. *PEDAGOGIA: Jurnal Pendidikan, 9*(1), 5–14. https://doi.org/10.21070/pedagogia.v9i1.123

Zulkarnain, Z., Adiputra, S., & Amir, A. (2025). The Implementation of Teachers' Pedagogical Competence In Elementary Schools. *Kontribusia: Research Dissemination for Community Development, 8*(1), 61–69. https://journal.umg.ac.id/index.php/kontribusia/article/view/8985