



Analisis Kualitas Butir Soal Konsep Sistem Gerak Manusia Kelas VI Sekolah Dasar Berdasarkan Validitas, Reliabilitas, Tingkat Kesukaran dan Daya Beda

Ika Isna Umiyati^{1*}, Fina Fakhriyah², Sumaji³

¹⁻³ Program Studi Magister Pendidikan Dasar, Universitas Muria Kudus, Indonesia

*Penulis Korespondensi: 202403103@std.umk.ac.id

Abstract. *The quality of assessment instruments is crucial in determining the accuracy of measuring student learning outcomes in elementary school science learning. A good test instrument must meet criteria such as validity, reliability, difficulty level, and discrimination power. This study aims to analyze the quality of daily science test items in grade VIc of elementary school based on these four criteria. The approach used was quantitative with 19 grade VI students as research subjects. The instrument analyzed consisted of 25 multiple-choice questions. Data processing was carried out using Microsoft Excel to calculate item validity using item-total score correlation, reliability using internal consistency, and difficulty level and discrimination power indices. The analysis results showed that 17 questions (68%) were valid, while 8 questions (32%) were invalid and needed to be improved. The reliability test showed that the test instrument had good reliability, suitable for use as a measuring tool. The level of difficulty of the questions was relatively balanced, with 80% of the questions having moderate difficulty. Based on discrimination power, most of the questions (66%) had very good discrimination power. This study concludes that the quality of the daily test questions for grade VI science is quite good, but still needs improvement on questions that are invalid and have low discrimination power to optimize the assessment.*

Keywords: *Elementary School; Instrument Quality; Item Analysis; Science Learning; Validity.*

Abstrak. Kualitas instrumen penilaian sangat penting dalam menentukan ketepatan pengukuran hasil belajar siswa pada pembelajaran IPAS di sekolah dasar. Instrumen tes yang baik harus memenuhi kriteria seperti validitas, reliabilitas, tingkat kesukaran, dan daya beda. Penelitian ini bertujuan untuk menganalisis kualitas butir soal ulangan harian IPAS di kelas VIc sekolah dasar berdasarkan empat kriteria tersebut. Pendekatan yang digunakan adalah kuantitatif dengan subjek penelitian 19 siswa kelas VI. Instrumen yang dianalisis terdiri dari 25 butir soal pilihan ganda. Pengolahan data dilakukan menggunakan Microsoft Excel untuk menghitung validitas butir soal dengan korelasi butir skor total, reliabilitas menggunakan konsistensi internal, serta indeks tingkat kesukaran dan daya beda. Hasil analisis menunjukkan bahwa 17 soal (68%) valid, sementara 8 soal (32%) tidak valid dan perlu diperbaiki. Uji reliabilitas menunjukkan bahwa instrumen tes memiliki reliabilitas yang baik, layak digunakan sebagai alat ukur. Tingkat kesukaran soal relatif seimbang, dengan 80% soal berkesulitan sedang. Berdasarkan daya beda, sebagian besar soal (66%) memiliki daya beda sangat baik. Penelitian ini menyimpulkan bahwa kualitas soal ulangan harian IPAS kelas VI tergolong baik, namun masih perlu perbaikan pada soal yang tidak valid dan memiliki daya beda rendah untuk mengoptimalkan penilaian.

Kata kunci: Analisis Butir Soal; Kualitas Instrumen; Pembelajaran IPAS; Sekolah Dasar; Validitas

1. LATAR BELAKANG

Pembelajaran Ilmu Pengetahuan Alam dan Sosial (IPAS) di sekolah dasar merupakan dasar yang penting untuk membangun pemahaman mendalam pada siswa terhadap fenomena alam dan kehidupan sosial di lingkungan (Azzahra et al., 2023). Hakikat pendidikan IPAS tidak hanya menekankan pada penguasaan kumpulan pengetahuan yang berupa fakta, konsep, atau prinsip saja, tetapi juga sebagai proses penemuan. Menurut Elisa et al. (2023), pembelajaran pengetahuan alam dan sosial harus mampu memberikan pengalaman langsung kepada siswa untuk mengembangkan kompetensi agar menjelajahi, memahami alam sekitar dan lingkungan sosial. Pengalaman pada proses pembelajaran IPAS sangat bergantung pada kemampuan guru

melalui berbagai metode dan pendekatan yang efektif untuk memberikan pengalaman yang bermakna pada siswa.

Sebagai bagian integral dari proses pembelajaran, evaluasi atau asesmen memegang peranan penting dalam menentukan sejauh mana tujuan pembelajaran telah tercapai. Evaluasi hasil belajar pada mata pelajaran IPAS kelas VI sering kali menghadapi tantangan terkait kedalaman materi, seperti sistem gerak manusia yang mencakup pemahaman tentang rangka, sendi, dan otot. Fauziah et al. (2021), menyatakan bahwa materi sistem gerak sering kali dianggap sulit oleh siswa karena melibatkan banyak istilah latin dan konsep visual yang kompleks. Oleh karena itu, diperlukan instrumen penilaian yang presisi untuk mengukur keterampilan proses sains siswa secara objektif (Hamidah & Istiyono, 2022).

Kualitas sebuah tes ditentukan oleh seberapa akurat, konsisten, sulit, dan soal tersebut membedakan kemampuan siswa seperti uji validitas, reliabilitas, tingkat kesukaran, dan daya beda. Instrumen yang valid akan menjamin bahwa hasil pengukuran sesuai dengan kompetensi yang diuji, sedangkan reliabilitas memastikan konsistensi hasil tersebut (Ramadhan et al., 2024). Namun, pada materi sistem gerak manusia sering ditemukan soal yang hanya menguji aspek kognitif rendah (C1-C2). Rini & Budijastuti (2022) menyatakan bahwa pengembangan instrumen dapat digunakan sebagai pengukur keterampilan pemecahan masalah pada materi sistem gerak. Hal tersebut digunakan agar evaluasi tidak sekadar menjadi formalitas administratif saja (Mutiah, 2023).

Masalah lain yang muncul dalam asesmen IPAS adalah potensi terjadinya miskonsepsi pada siswa. Miskonsepsi sering kali tidak terdeteksi jika butir soal yang digunakan tidak memiliki kualitas daya beda dan validitas yang memadai. Menurut Sari et al. (2025), miskonsepsi pada siswa sekolah dasar sering ditemukan pada konsep-konsep abstrak, termasuk fungsi organ dalam sistem gerak. Hal ini dipertegas oleh Rahmawati & Kurniawati (2024) yang menyatakan bahwa identifikasi miskonsepsi memerlukan instrumen yang mampu menggali pemahaman mendalam, bukan sekadar hafalan, sehingga analisis kualitas butir soal menjadi prasyarat mutlak sebelum tes diberikan secara luas.

Analisis butir soal secara berkala merupakan salah satu strategi untuk meningkatkan kualitas bank soal di sekolah. Penggunaan teknologi informasi, seperti *Microsoft Excel*, memberikan kemudahan bagi guru untuk melakukan pengolahan data statistik secara praktis dan akurat. Syaifudin (2023) menyebutkan bahwa analisis butir soal secara mandiri oleh guru dapat membantu dalam memetakan kualitas instrumen secara cepat sehingga perbaikan terhadap soal yang tidak layak dapat segera dilakukan. Dengan melakukan analisis tingkat

kesukaran dan daya beda, guru dapat memastikan bahwa soal yang diberikan tetap mampu mengukur ketercapaian kompetensi secara adil (Saputri et al., 2023).

Berdasarkan uraian tersebut, penelitian ini difokuskan pada analisis empiris terhadap 25 butir soal ulangan harian IPAS materi sistem gerak manusia kelas VIc SD yang diikuti oleh 19 siswa. Tujuan penelitian ini adalah mengevaluasi kualitas instrumen ditinjau dari aspek validitas, reliabilitas, tingkat kesukaran, dan daya beda. Melalui penelitian ini, diharapkan guru mendapatkan gambaran nyata mengenai kualitas instrumen yang disusunnya serta dapat melakukan perbaikan yang diperlukan pada butir-butir soal yang belum memenuhi kriteria fungsional (Lapera et al., 2022).

2. KAJIAN TEORITIS

Evaluasi pendidikan merupakan proses sistematis untuk menentukan nilai atau makna dari suatu objek, yang dalam konteks ini adalah capaian belajar siswa pada materi IPAS. Pengukuran yang akurat memerlukan instrumen tes yang memenuhi syarat metodologis agar data yang dihasilkan bersifat objektif. Analisis instrumen asesmen melalui pendekatan teori tes klasik memberikan informasi penting bagi guru mengenai kelayakan tiap butir soal dalam mengukur kompetensi dasar (Saputri et al., 2023). Kualitas instrumen ini secara teoretis dibangun oleh empat pilar utama, yakni validitas, reliabilitas, tingkat kesukaran, dan daya beda.

Validitas merupakan parameter utama yang menunjukkan sejauh mana sebuah tes mampu mengukur apa yang seharusnya diukur. Dalam pembelajaran IPAS, validitas memastikan bahwa soal benar-benar menguji pemahaman konsep saintifik, bukan kemampuan membaca atau aspek kognitif luar. Ramadhan et al. (2024) menjelaskan bahwa validitas empiris yang dihitung melalui korelasi skor butir dengan skor total (*Product Moment*) mencerminkan ketepatan butir soal tersebut dalam mendukung fungsi tes secara keseluruhan. Tanpa validitas yang memadai, hasil evaluasi tidak dapat dianggap sebagai representasi kemampuan siswa yang sebenarnya.

Reliabilitas merujuk pada derajat konsistensi atau kestabilan hasil pengukuran. Sebuah instrumen dikatakan reliabel jika memberikan hasil yang relatif sama ketika diujikan pada kelompok yang sama dalam waktu atau kondisi yang berbeda. Arlindianie et al. (2024), menyatakan bahwa reliabilitas instrumen dipengaruhi oleh panjang tes, homogenitas butir, dan objektivitas penskoran. Dalam penelitian ini, reliabilitas diukur menggunakan rumus *Kuder-Richardson* (KR-20) yang sesuai untuk tes pilihan ganda dengan skor benar atau salah, guna memastikan perangkat soal ulangan harian IPAS memiliki tingkat keterpercayaan yang tinggi.

Tingkat kesukaran adalah proporsi peserta tes yang menjawab butir soal dengan benar. Secara teoretis, butir soal yang baik adalah soal yang tidak terlalu mudah dan tidak terlalu sulit bagi siswa (Aldila, 2023). Distribusi tingkat kesukaran yang ideal biasanya mengikuti kurva normal atau disesuaikan dengan tujuan tes, seperti formatif atau sumatif. Soal yang terlalu mudah tidak akan memberikan tantangan intelektual, sedangkan soal yang terlalu sulit dapat menurunkan motivasi siswa. Analisis tingkat kesukaran membantu guru dalam menyeimbangkan porsi evaluasi pada materi-materi tertentu seperti sistem gerak manusia.

Daya beda adalah kemampuan suatu butir soal untuk membedakan antara siswa yang memiliki kemampuan tinggi (kelompok atas) dengan siswa yang memiliki kemampuan rendah (kelompok bawah). Rachmawati & Pradana (2025) menegaskan bahwa daya beda merupakan indikator kualitas yang krusial karena soal dengan daya beda tinggi mampu menunjukkan gradasi pemahaman siswa secara akurat. Soal yang tidak mampu membedakan kedua kelompok tersebut mengindikasikan adanya masalah pada konstruksi soal atau tingkat kecocokan soal dengan kemampuan siswa, sehingga perlu dilakukan revisi atau penggantian butir.

Sejumlah penelitian terdahulu telah memberikan landasan kuat mengenai pentingnya analisis butir soal. Penelitian yang dilakukan oleh Fiska et al. (2021) menunjukkan bahwa analisis menggunakan perangkat lunak (seperti *Anates* atau *Excel*) secara signifikan membantu guru dalam mendeteksi butir soal yang memiliki kualitas rendah dari segi daya beda dan tingkat kesukaran. Temuan mereka menekankan bahwa soal IPAS sering kali terjebak pada kategori "terlalu mudah" sehingga kurang mampu mengeksplorasi potensi kognitif siswa secara mendalam. Hal ini menjadi titik acuan bagi penelitian ini untuk melihat apakah pola yang sama terjadi pada materi sistem gerak manusia.

Penelitian lain oleh Lopera et al. (2022) berfokus pada analisis kualitas soal ulangan harian IPAS yang menemukan bahwa sebagian besar soal yang disusun mandiri oleh guru belum sepenuhnya memenuhi kriteria validitas empiris. Masalah teknis seperti redundansi kalimat dan distraktor yang tidak berfungsi menjadi penyebab utama rendahnya kualitas soal. Sejalan dengan itu, dalam studi Syaifudin (2023) mengenai efektivitas analisis bank soal menyarankan agar guru tidak hanya mengandalkan instrumen dari buku teks, melainkan melakukan validasi berkala untuk memastikan relevansi soal dengan karakteristik siswa di sekolah masing-masing.

Integrasi penelitian-penelitian tersebut menunjukkan bahwa analisis butir soal merupakan kegiatan manajerial dan pedagogis yang tidak bisa dipisahkan dari tanggung jawab seorang guru. Anggraini et al. (2022) berpendapat bahwa penguasaan teknik evaluasi statistik,

meskipun sederhana, sangat mendukung profesionalisme dalam menjalankan peran sebagai evaluator. Dengan merujuk pada temuan-temuan terdahulu, kajian teoritis ini menegaskan bahwa setiap soal yang diberikan kepada siswa harus melalui proses kurasi ilmiah agar dapat memetakan capaian pembelajaran IPAS di sekolah dasar secara akurat dan adil.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif, yang bertujuan untuk menggambarkan Validitas, Reliabilitas, Tingkat Kesukaran dan Daya Beda pada muatan pembelajaran IPAS konsep sistem gerak manusia. Penelitian dilaksanakan di SD Negeri Growong Kidul 02 pada semester ganjil tahun ajaran 2025/2026. Subjek penelitian adalah siswa kelas VIC SD Negeri Growong Kidul 02 yang berjumlah 19 siswa, yang dipilih secara keseluruhan. Pemilihan subjek didasarkan pada pertimbangan bahwa kelas tersebut telah mempelajari materi sistem gerak manusia.

Instrumen yang dianalisis adalah tes pilihan ganda materi Sistem Gerak Manusia yang terdiri dari 25 butir soal. Setiap soal memiliki empat pilihan jawaban (A, B, C, dan D) dengan hanya satu jawaban benar. Teknik skoring, yaitu skor 1 diberikan untuk jawaban yang benar dan skor 0 untuk jawaban yang salah (Syahari et al., 2025). Instrumen ini digunakan untuk mengidentifikasi kualitas soal melalui Validitas, Reliabilitas, Tingkat Kesukaran dan Daya Beda.

Tabel 1. Kisi-Kisi Soal

Capaian Pembelajaran: Siswa dapat merefleksikan sistem organ tubuh manusia yang dikaitkan dengan cara menjaga kesehatan tubuhnya.

No.	Tujuan Pembelajaran (TP)	Tingkat Kognitif	Indikator Soal	No. Soal
1	Siswa dapat menganalisis struktur dan fungsi dasar sistem gerak manusia.	C1	Mengidentifikasi nama bagian sistem gerak.	1
		C2	Menjelaskan fungsi utama otot.	2, 3
		C4	Menganalisis hubungan antara tulang dan saraf.	4, 5
2	Siswa dapat merancang desain (sketsa) awal replika tangan mekanik.	C3	Memilih bahan yang tepat untuk replika tangan.	6
		C4	Menganalisis prinsip gerak dasar replika.	7, 8
		C6	Menentukan komponen utama dalam sketsa.	9
3	Siswa dapat membuat prototipe replika tangan	C3	Menyusun langkah kerja mekanik replika.	10

No.	Tujuan Pembelajaran (TP)	Tingkat Kognitif	Indikator Soal	No. Soal
4	mekanik dari bahan bekas.	C4	Menganalisis akibat dari kesalahan prosedur pembuatan replika.	11, 12
		C5	Menilai kualitas bahan yang digunakan replika.	13
	Siswa dapat menguji coba fungsionalitas replika tangan.	C3	Mengidentifikasi gerak fleksi pada replika.	14, 15
		C4	Menganalisis penyebab kegagalan gerak.	16
5	Siswa dapat menganalisis analogi antara replika (sedotan, benang) dengan sistem gerak asli (tulang, otot).	C2	Menjelaskan kegunaan komponen replika.	17
		C4	Menganalisis peran benang (otot) dalam replika.	18, 19
		C5	Menilai peran bahan dalam replika.	20
		C5	Membandingkan kelebihan/kekurangan replika.	21
6	Siswa dapat merefleksikan cara kerja sistem gerak untuk menyimpulkan cara menjaga kesehatan organ gerak.	C3	Menentukan contoh perilaku menjaga kesehatan.	22
		C5	Menilai risiko kebiasaan buruk terhadap tulang.	23, 24
		C6	Menyusun kesimpulan/saran tips kesehatan.	25

Data diolah menggunakan perangkat lunak *Microsoft Excel* untuk meminimalisir kesalahan hitung manual (Anggraini et al., 2022). Adapun prosedur statistik yang digunakan adalah:

(1) Uji Validitas yang dihitung menggunakan rumus korelasi *Product Moment* dari untuk melihat hubungan antara skor butir (x) dengan skor total (y):

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

Keterangan:

r_{xy} = Koefisien korelasi antara variabel X dan Y

N = Jumlah responden (19 siswa)

X = Skor butir soal

Y = Skor total siswa

Kriteria pengujian: Butir soal dinyatakan valid jika $r_{hitung} > r_{tabel}$ (0,456).

(2) Uji Reliabilitas digunakan untuk mengukur konsistensi instrumen, digunakan rumus *Kuder-Richardson 20* (KR-20):

$$KR\ 20 = \left(\frac{n}{n-1}\right)\left(\frac{St^2 - \sum pq}{St^2}\right)$$

Keterangan:

n = banyaknya item soal

St^2 = varians

P = proporsi siswa yang menjawab benar

q = 1 - p

KR 20 > 0,9 Reliabel

(3) Analisis Tingkat Kesukaran dihitung untuk mengetahui seberapa sulit atau mudah suatu soal bagi siswa: $\frac{\text{Jumlah skor siswa tiap soal}}{\text{Jumlah siswa}}$

Kriteria interpretasi merujuk pada ketentuan: 0,00 - 0,30 (Sukar); 0,31 - 0,70 (Sedang); 0,71 - 1,00 (Mudah).

(4) Analisis Daya Pembeda dihitung untuk mengetahui kemampuan soal dalam membedakan kelompok siswa pandai (atas) dan kurang pandai (bawah):

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

Keterangan:

D = Indeks daya pembeda

B_A = Jumlah kelompok atas yang menjawab benar

B_B = Jumlah kelompok bawah yang menjawab benar

J_A / J_B = Jumlah siswa pada kelompok atas/bawah (27% dari total)

Kriteria interpretasi merujuk pada ketentuan: $\geq 0,40$ (Sangat Baik); 0,30 - 0,39 (Baik); 0,20 - 0,29 (Cukup Baik); $< 0,20$ (Kurang Baik/Dibuang).

Hasil akhir dari seluruh analisis ini dikategorikan menjadi soal yang diterima (layak), diperbaiki (revisi), atau dibuang (tidak layak) untuk digunakan kembali dalam evaluasi mendatang.

4. HASIL DAN PEMBAHASAN

Berdasarkan pengolahan data menggunakan *Microsoft Excel* terhadap 19 siswa dan 25 butir soal IPAS, diperoleh rangkuman hasil analisis sebagai berikut:

Tabel 2. Rangkuman Hasil Analisis Butir Soal.

No. Soal	Validitas	Reliabilitas	Tingkat Kesukaran	Daya Beda
1	Valid	Reliabel	Mudah	Sangat Baik
2	Valid	Reliabel	Sedang	Cukup
3	Valid	Reliabel	Mudah	Cukup
4	Valid	Reliabel	Sedang	Sangat Baik
5	Tidak Valid	Reliabel	Mudah	Cukup
6	Valid	Reliabel	Mudah	Sangat Baik
7	Tidak Valid	Reliabel	Sedang	Sangat Baik
8	Valid	Reliabel	Sedang	Baik
9	Tidak Valid	Reliabel	Sedang	Sangat Baik
10	Tidak Valid	Reliabel	Sedang	Sangat Baik
11	Valid	Reliabel	Sedang	Sangat Baik
12	Valid	Reliabel	Sedang	Sangat Baik
13	Valid	Reliabel	Sedang	Sangat Baik
14	Valid	Reliabel	Sedang	Sangat Baik
15	Valid	Reliabel	Sedang	Sangat Baik
16	Tidak Valid	Reliabel	Sedang	Cukup
17	Valid	Reliabel	Mudah	Baik
18	Valid	Reliabel	Sedang	Sangat Baik
19	Tidak Valid	Reliabel	Sedang	Baik
20	Tidak Valid	Reliabel	Sedang	Kurang Baik
21	Valid	Reliabel	Sedang	Sangat Baik
22	Valid	Reliabel	Sedang	Sangat Baik
23	Tidak Valid	Reliabel	Sedang	Baik
24	Valid	Reliabel	Sedang	Sangat Baik
25	Valid	Reliabel	Sedang	Sangat Baik

Tabel 3. Hasil Analisis Uji Validitas.

Uji Butir Soal	Keterangan	Jumlah
Uji Validitas	Valid	17
	Tidak Valid	8

Berdasarkan hasil uji validitas menggunakan korelasi *Product Moment* dengan $N=19$ dan taraf signifikansi 5% $r_{tabel} = 0,456$, ditemukan bahwa dari 25 butir soal, sebanyak 17 butir soal (68%) dinyatakan valid dan 8 butir soal (32%) dinyatakan tidak valid. Soal yang tidak valid seperti nomor 5, 20, dan 23 menunjukkan korelasi yang sangat rendah dengan skor total. Hal ini mengindikasikan bahwa butir-butir tersebut tidak mampu mengukur penguasaan konsep sistem gerak secara konsisten dengan butir soal lainnya. Secara teoritis, ketidakvalidan ini dapat disebabkan oleh konstruksi kalimat yang ambigu atau materi yang ditanyakan keluar

dari batasan indikator pembelajaran IPAS kelas VI di SD Negeri Growong Kidul 02. Tingkat validitas sebuah butir soal tidak semata-mata ditentukan oleh muatan materi dan aspek kebahasaan yang digunakan, tetapi juga sangat dipengaruhi oleh karakteristik siswa sebagai subjek yang mengerjakan soal tersebut (Fatayah & Yuliana, 2022).

Tabel 4. Hasil Analisis Uji Reliabilitas.

KR-20	Jumlah Butir Soal	Keterangan
0,950	25	Reliabel

Uji reliabilitas yang dilakukan dengan rumus KR-20 menghasilkan koefisien reliabilitas sebesar 0,950. Berdasarkan kriteria *Guilford*, nilai ini berada pada rentang KR 20 > 0,9 dikategorikan sebagai "Reliabilitas". Hasil ini menunjukkan bahwa meskipun terdapat beberapa butir soal yang tidak valid, instrumen ulangan harian secara keseluruhan memiliki tingkat keajekan yang baik. Artinya, jika tes ini diberikan kembali pada kelompok yang sama dalam kondisi yang serupa, maka hasil pengukurannya cenderung tetap konsisten. Skor reliabilitas yang tinggi ini memberikan keyakinan bagi guru untuk menggunakan hasil tes sebagai dasar pengambilan keputusan akademik (Ramadhan et al., 2024).

Tabel 5. Hasil Analisis Tingkat Kesukaran.

Uji Butir Soal	Keterangan	Jumlah
Uji Tingkat Kesukaran	Mudah	5
	Sedang	20
	Sukar	0

Ditinjau dari tingkat kesukaran (P), instrumen ini didominasi oleh soal kategori "Sedang" sebanyak 20 butir (80%) dan soal kategori "Mudah" sebanyak 5 butir (20%). Tidak ditemukan soal dengan kategori "Sukar". Distribusi ini menguntungkan untuk ulangan harian yang bersifat formatif karena tidak membebani mental siswa secara berlebihan, namun dari sudut pandang bank soal, ketiadaan soal sukar membuat instrumen kurang mampu menantang siswa dengan kemampuan akademik tinggi (Aldila, 2023). Ketidaan soal sukar kemungkinan disebabkan oleh siswa dapat memahami soal yang diberikan dan tidak ada pernyataan pengecoh yang membutuhkan pemikiran tingkat tinggi. Soal sukar bukan semata-mata untuk menjatuhkan nilai siswa, melainkan sebagai alat untuk melatih ketajaman berpikir dan menjamin bahwa soal tersebut memang berkualitas secara ilmiah (Waworuntu et al., 2024).

Tabel 6. Hasil Analisis Daya Beda.

Uji Butir Soal	Keterangan	Jumlah
Uji Daya Beda	Kurang Baik	1
	Cukup Baik	4
	Baik	4
	Sangat Baik	16

Analisis daya beda menunjukkan hasil yang sangat memuaskan, di mana 16 butir soal masuk dalam kategori "Sangat Baik", 4 butir soal "Baik", 4 butir soal "Cukup Baik", dan 1 butir soal "Kurang Baik". Daya beda yang kuat ini sangat cocok dalam pembelajaran IPAS untuk mengidentifikasi siswa yang benar-benar memahami mekanisme kerja rangka dan otot dibandingkan mereka yang hanya menghafal nama tulang. Namun, satu soal memiliki daya beda "Kurang Baik", yang berarti soal tersebut dijawab benar oleh jumlah siswa yang hampir sama baik dari kelompok atas maupun kelompok bawah. Soal seperti ini harus dibuang karena gagal membedakan tingkat penguasaan kompetensi antar siswa. Dengan adanya daya beda yang tepat, pemahaman dan tingkat pengetahuan siswa dapat diukur secara lebih akurat untuk mencegah terjadinya kesalahan penilaian di mana siswa yang sebenarnya tidak menguasai materi mendapatkan nilai baik hanya karena faktor kebetulan atau soal yang kurang berkualitas (Rajagukguk, 2023).

Berdasarkan analisis empat parameter di atas, peneliti melakukan klasifikasi akhir terhadap 25 butir soal materi Sistem Gerak Manusia. Analisis mendalam terhadap data menunjukkan adanya korelasi:

Tabel 7. Keputusan dari Hasil Analisi Butir Soal.

Keputusan	Jumlah	Nomor Soal	Alasan
Diterima (Gunakan lagi)	15	1, 2, 4, 6, 8, 11, 12, 13, 14, 15, 18, 21, 22, 24, 25	Sudah valid, reliabel, dan mampu membedakan kemampuan siswa.
Revisi (perlu perbaikan)	9	3, 5, 7, 9, 10, 16, 17, 19, 23	Validitasnya rendah atau terlalu mudah, perlu diganti kalimatnya.
Dibuang	1	20	Tidak valid dan tidak bisa membedakan kemampuan siswa sama sekali.

Hal ini memperkuat teori Fajariani & Hanum (2025) bahwa evaluasi kualitas instrumen secara berkala adalah wajib bagi guru untuk menjamin keadilan dalam penilaian. Sesuai dengan

anjuan Rachmawati & Pradana (2025), soal-soal dalam kategori ini memerlukan revisi pada tingkat redaksi kalimat atau penyesuaian kembali dengan indikator pembelajaran sebelum digunakan kembali dalam penilaian sumatif. Secara keseluruhan, kelayakan instrumen ulangan harian IPAS materi sistem gerak manusia ini dapat dikategorikan "Layak dengan Revisi". Kelebihan utama instrumen terletak pada nilai reliabilitas yang tinggi dan daya beda yang tajam pada sebagian besar butir soal. Hal ini sejalan dengan penelitian Hamidah & Istiyono (2022) yang menyatakan bahwa instrumen IPAS yang baik harus mampu mendiskriminasi kemampuan saintifik siswa secara akurat. Penggunaan *Microsoft Excel* dalam proses analisis terbukti memberikan presisi data yang tinggi, memungkinkan guru di SD Negeri Growong Kidul 02 untuk melihat kelemahan spesifik pada butir soal nomor 5, 16, dan 20 yang memiliki kualitas teknis rendah.

20. Jika sedotan yang mereplikasi tulang jari dibuat dari bahan yang sangat lunak (mudah bengkok), hal yang terjadi ketika benang ditarik adalah
- jari replika akan bergerak sangat cepat
 - gerakan fleksi akan sempurna
 - replika menjadi lebih ringan
 - gerakan jari tidak stabil

Gambar 1. Butir soal nomor 20.

Evaluasi soal yang kurang baik pada soal nomor 20 yaitu karena memiliki kelemahan mendasar pada aspek kontekstual dan konstruksi pilihan jawaban. Dari sisi konteks, soal tidak menyertakan stimulus berupa gambar atau deskripsi teknis mengenai bagaimana benang dan sedotan tersebut dirangkai, sehingga siswa dipaksa berimajinasi tanpa dasar mekanis yang jelas. Selain itu, terdapat masalah pada homogenitas pilihan jawaban; pilihan A, B, C, dan D tidak berada dalam satu kategori logika yang setara, di mana pilihan "menjadi lebih ringan" (C) sama sekali tidak relevan dengan sifat elastisitas bahan dalam konteks mekanika gerak. Soal ini terjebak dalam ambiguitas istilah. Penggunaan kata "tidak stabil" pada pilihan D bersifat terlalu umum dan multitafsir, karena ketidakstabilan tersebut bisa berarti bergetar, melengkung di tempat yang salah, atau gagal bergerak sama sekali. Hal ini menyebabkan tidak adanya satu jawaban yang paling mutlak benar secara ilmiah, karena pilihan B (gerakan tidak sempurna) dan pilihan D (tidak stabil) sebenarnya saling berkaitan dan bisa dianggap benar secara bersamaan (tumpang tindih). Akibatnya, soal ini tidak mampu mengukur pemahaman konsep siswa secara akurat dan lebih cenderung mengandalkan tebakan logika sederhana. Soal dalam kategori ini hampir tidak memberikan perbedaan hasil antara kelompok siswa pandai dan siswa kurang pandai. Padahal, soal yang baik seharusnya dijawab benar oleh kelompok atas dan dijawab salah oleh kelompok bawah (Fatayah & Yuliana, 2022).

Pembahasan mengenai butir soal yang tidak valid menunjukkan adanya pola miskonsepsi pada materi sendi dan gangguan tulang. Sebagai contoh, soal yang menanyakan perbedaan sendi engsel dan sendi peluru sering kali memiliki distraktor (pilihan pengecoh) yang kurang berfungsi, sehingga siswa dapat menebak jawaban tanpa memahami konsep secara utuh. Hal ini memperkuat temuan Sari et al. (2025) bahwa kelemahan kualitas soal sering kali berakar pada kurangnya perhatian guru terhadap fungsionalitas pilihan jawaban. Oleh karena itu, 8 butir soal yang tidak valid disarankan untuk diperbaiki narasinya atau diganti dengan soal baru yang memiliki tingkat kognitif lebih tinggi (C4-C6) guna meningkatkan kualitas evaluasi pembelajaran di SD Negeri Growong Kidul 02. Jika soal tersebut sudah tidak bisa diperbaiki maknanya perlu dibuang, sedangkan soal diperbaiki ganti bahasanya supaya pas dengan indikator dan coba gunakan lagi di evaluasi berikutnya (Ayu & Realita, 2024).

5. KESIMPULAN DAN SARAN

Kesimpulan dari penelitian ini adalah kualitas butir soal ulangan harian IPAS kelas VI berada pada kategori cukup baik dengan tingkat reliabilitas yang memadai. Dari 25 butir soal, sebanyak 17 soal (68%) layak digunakan kembali tanpa revisi. Terdapat 8 soal yang tidak valid, di mana soal nomor 20 disarankan untuk dibuang karena tidak valid dan memiliki daya beda yang kurang baik, sedangkan soal tidak valid lainnya dapat diperbaiki. Komposisi tingkat kesukaran menunjukkan dominasi soal sedang dan mudah tanpa adanya soal sukar. Dominasi soal sedang dan mudah memberikan informasi bahwa instrumen ini efektif untuk mengukur kompetensi dasar, namun perlu pengayaan soal bertipe HOTS (*Higher Order Thinking Skills*) di masa depan. Mulai mengonstruksi soal yang tidak hanya menuntut kemampuan menghafal (*Lower Order Thinking Skills*), melainkan soal yang memicu kemampuan analisis, evaluasi, dan kreasi. Hal ini berguna untuk mengidentifikasi dengan jelas seberapa tinggi tingkat kecerdasan siswa.

Saran bagi guru adalah melakukan evaluasi terhadap soal-soal tidak valid untuk mengidentifikasi apakah terdapat kesalahan konsep pada materi sistem gerak. Penggunaan media visual dalam soal juga disarankan untuk mempermudah siswa memahami instruksi pada butir soal yang bersifat anatomis. Peneliti selanjutnya disarankan untuk menganalisis efektivitas distraktor (pilihan jawaban yang salah) secara lebih mendalam untuk menyempurnakan kualitas soal.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Kepala SD Negeri Growong Kidul 02 dan para guru atas kerja sama dan bantuannya selama pengambilan data. Apresiasi khusus juga disampaikan kepada siswa kelas VIc SD Negeri Growong Kidul 02 atas antusiasmenya sebagai responden penelitian. Terima kasih juga kepada rekan sejawat yang telah membantu dalam pengolahan data sehingga artikel ini dapat terselesaikan dengan baik.

DAFTAR REFERENSI

- Aldila, O. S. (2023). Analisis Tingkat Kesulitan Butir Soal Pilihan Ganda pada Penilaian Tengah Semester Genap di MI Ma'arif Beji. In *In Jurnal Ilmu Pendidikan*.
- Anggraini, F. D. P., Aprianti², Setyawati, V. A. V., & Hartanto, A. A. (2022). Pembelajaran Statistika Menggunakan Software SPSS untuk Uji Validitas dan Reliabilitas. *Jurnal Basicedu*, 6(4), 6491–6504.
- Arlindianie, N. K., A'isyah, E. T., Rahmah, E. A., Hakim, L., & Pratiwi, V. (2024). Analisis Kualitas Butir Soal Akuntansi Perusahaan. *Jurnal Nirta: Studi Inovasi*, 4(1), 1–18.
- Ayu, D., & Realita, R. (2024). Analisis Butir Soal pada Mata Pelajaran Pendidikan Agama Islam di Kelas V SDN 3 Tanjung Pura. *ALACRITY: Journal of Education*, 4(3), 140–151. <https://doi.org/10.52121/alacrity.v4i3.418>
- Azzahra, I. E., Nurhasanah, A., & Hermawati, E. (2023). Implementation of the Independent Curriculum in Learning Ips At Sdn 4 Purwawinangun. *Jurnal Ilmiah PGSD FKIP Universitas Mandiri*, 9(2).
- Elisa, D. T., Juliana, J., Bundel, B., Bumbun, M., Silvester, S., & Purnasari, P. D. (2023). Analisis Karakteristik Hakikat Pembelajaran IPA di Sekolah Dasar. *Jurnal Pedagogik Pendidikan Dasar*, 10(1), 37–44. <https://doi.org/10.17509/jppd.v10i1.54868>
- Fajariani, K., & Hanum, E. (2025). Evaluasi Kualitas Instrumen Tes Pilihan Ganda pada Mata Pelajaran IPA Kelas V Sekolah Dasar Berdasarkan Analisis Butir Soal. *Jurnal Pengabdian Masyarakat Dan Riset Pendidikan*, 4(2), 9224–9229. <https://doi.org/10.31004/jerkin.v4i2.2984>
- Fatayah, F., & Yuliana, I. F. (2022). Analisis Validitas Dan Reliabilitas Dalam Mendukung Ketuntasan Belajar Model STEM. 18(1), 49–60.
- Fauziah, A., Hasibuan, M. F., & Yendrawaty. (2021). MENINGKATKAN HASIL BELAJAR SISWA PADA PELAJARAN IPA DENGAN MENGGUNAKAN MODEL PEMBELAJARAN VISUAL, AUDITORI, KINESTHETI(VAK) PADA MATERI SISTEM GERAK MANUSIA DI KELAS VI SD NEGERI 067241 MEDAN DENAI. *Jurnal Ilmiah PGSD FKIP Universitas Mandiri*, 32(3), 167–186.
- Fiska, J. M., Hidayati, Y., Qomaria, N., & Hadi, W. P. (2021). Analisis Butir Soal Ulangan Harian Ipa Menggunakan Software Anates Pada Pendekatan Teori Tes Klasik. *Natural Science Education Research*, 4(1), 65–76. <https://doi.org/10.21107/nser.v4i1.8133>
- Hamidah, N., & Istiyono, E. (2022). The quality of test on National Examination of Natural science in the level of elementary school. *International Journal of Evaluation and Research in Education*, 11(2), 604–616. <https://doi.org/10.11591/ijere.v11i2.22233>

- Lapera, L., Asmara, Y., & Sofiarini, A. (2022). Analisis Kualitas Soal Ulangan Harian pada Mata Pelajaran IPA Kelas V SD Negeri 32 Lubuklinggau. *Linggau Jurnal Of Elementary School Education*, 2(1), 23–32. <https://jurnal.lp3mkil.or.id/index.php/ljese%0A>
- Mutiah, R. (2023). Analisis Kualitas Butir Soal Ujian Sekolah Mata Pelajaran IPA di SD N 1 LodanKulon. *Jurnal Kajian Islam AI Kamal*, 3(2), 127–140. <http://e-journal.hamzanwadi.ac.id/index.php/gdk/article/view/3696>
- Rachmawati, D., & Pradana, A. B. (2025). Analisis butir soal mata pelajaran ekonomi: validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. *Jurnal Pendidikan Ekonomi (JUPE)*, 13(3), 273–284. <https://doi.org/10.26740/jupe.v13n3.p273-284%0>
- Rahmawati, P., & Kurniawati, W. (2024). Miskonsepsi Pembelajaran IPAS Siswa Kelas IV Sekolah Dasar. *Jurnal Penelitian & Artikel Pendidikan*, 16(02), 383–394.
- Rajagukguk, M. J. T. (2023). MAMPU MEMILIH SOAL BERDASARKAN TINGKAT KESUKARAN. *Jurnal Pendidikan Sosial Dan Humaniora*, 2(4), 12736–12747.
- Ramadhan, M. F., Siroj, R. A., & Afgani, M. W. (2024). Validitas and Reliabilitas. *Journal on Education*, 6(2), 10967–10975. <https://doi.org/10.31004/joe.v6i2.4885>
- Rini, A. F., & Budijastuti, W. (2022). Pengembangan Instrumen Soal Hots Untuk Mengukur Keterampilan Pemecahan Masalah Pada Materi Sistem Gerak Manusia. *Berkala Ilmiah Pendidikan Biologi (Bioedu)*, 11(1), 127–137. <https://ejournal.unesa.ac.id/index.php/bioedu>
- Saputri, H. A., Zulhijrah, Larasati, N. J., & Shaleh. (2023). ANALISIS INSTRUMEN ASSESMEN : VALIDITAS, RELIABILITAS, TINGKAT KESUKARAN DAN DAYA BEDA BUTIR SOAL. *Jurnal Ilmiah PGSD FKIP Universitas Mandiri*, 09, 2986–2995.
- Sari, N. M., Lestari, I., & Nasbey, H. (2025). Studi Literatur : Miskonsepsi IPA pada Siswa Sekolah Dasar. *Seminar Nasional Pendidikan Sarjanawiyata Tamansiswa*, 2(1), 462–473.
- Syahari, Y. F., Azizah, A., Diana, S., Adhani, R., Arifin, R., Gigi, F. K., & Mangkurat, U. L. (2025). DENTIN JURNAL KEDOKTERAN GIGI Vol IX . No 1 . APRIL 2025 EFEKTIVITAS DENTAL HEALTH EDUCATION MENGGUNAKAN PERMAINAN TRADISIONAL BADAMPRAK TERHADAP PENGETAHUAN DAN SKOR OHI-S (Tinjauan Pada Siswa Umur 10-14 Tahun di Sekolah Buddhis Dhammasoka Kota Banjarma. *JURNAL KEDOKTERAN GIGI*, IX(1), 52–56.
- Syaifudin. (2023). Analisis Butir Soal Penilaian Akhir Tahun Mata Pelajaran Matematika dengan Item And Test Analysis. *Jurnal Jendela Matematika*, 1(2), 34–42.
- Waworuntu, F., Tuerah, J. M., Tengker, S. M., Saiya, A., Paat, V., & Rumengan, S. (2024). Uji Validitas Sebagai Penentuan Kualitas Butir Soal Pilihan Ganda " Pacu Kimia " FMIPAK Universitas Negeri Manado. *OXYGENIUS*, 6(1), 29–38. <https://doi.org/10.37033/ojce.v6i1.680>