

Analisis Akurasi Soal Pilihan Ganda Materi Rangka Manusia pada Ulangan Harian Siswa di Sekolah Dasar

Putri Arwinda¹, Sigit Prasetyo¹

¹Universitas Islam Negeri Sunan Kalijaga Yogyakarta, Indonesia

ABSTRACT

Purpose – This study aims to determine the quality of multiple choice items on human skeletal material in terms of validity, reliability, difficulty level, and differentiating power.

Method – This research uses a quantitative descriptive approach. The subjects in this study were grade V students of State Elementary School 1 Sungai Bungin, totaling 22 students. The sample selection technique used saturated sampling technique. The data collection technique was carried out by the documentation method from the test results. The collected data were analyzed using the SPSS IBM Statistics 27 application to determine the criteria for the items.

Findings – This study provides a detailed description of the validity, reliability, differentiating power, and difficulty of the questions. It found 17 valid items (68%), with high reliability (Cronbach's Alpha: 0.798 and 0.792). A total of 4 items had sufficient differentiating power, 19 items were good, and 2 items were excellent. The level of difficulty of the questions varied: 4 difficult questions (16%), 13 medium questions (52%), and 8 easy questions (32%).

Research Implications – This research has important implications for improving the quality of education in elementary schools. The results of analyzing the accuracy of multiple choice questions on human skeletal material can be used to improve and develop questions that are more appropriate in measuring student knowledge. Teachers can adjust teaching methods based on the difficulty level of the questions, improve learning effectiveness, and provide more constructive feedback. In addition, the findings can assist in more targeted curriculum development and teacher training in the creation of high-quality questions. Thus, this study contributes to the improvement of student learning outcomes and the implementation of data-driven evaluation in the primary education system.

 OPEN ACCESS

ARTICLE HISTORY

Received: 06-06-2024

Revised: 06-07-2024

Accepted: 08-07-2024

KEYWORDS

item analysis, validity, reliability, difficulty level, discriminatory power

Corresponding Author:

Putri Arwinda

Universitas Islam Negeri Sunan Kalijaga Yogyakarta, Indonesia

Email: putriarwinda2@gmail.com

Pendahuluan

Pendidikan adalah proses sistematis untuk memperoleh pengetahuan, keterampilan, nilai, dan sikap melalui pengalaman belajar atau pengajaran yang terstruktur, baik di lingkungan formal seperti sekolah maupun informal seperti kehidupan sehari-hari (Syamsul Arifin, 2021). Pendidikan yang bermutu memiliki kaitan erat dengan proses pembelajaran di sekolah. Terdapat tiga unsur utama dalam pembelajaran, yaitu tujuan pengajaran, pengalaman belajar, dan hasil belajar (Jayanti et al., 2024).

Kegiatan penilaian dalam pendidikan dikenal sebagai evaluasi, yang merupakan bagian integral dari proses belajar mengajar. Evaluasi hasil belajar dilakukan oleh pendidik untuk memantau proses, kemajuan, pencapaian, dan perbaikan hasil belajar peserta didik secara berkelanjutan (Baharuddin, 2024). Zahra berpendapat bahwa dalam praktek evaluasi di sekolah, seringkali guru mengabaikan kualitas pertanyaan tes karena yang lebih diutamakan adalah ketersediaan perangkat tes. Guru umumnya menggunakan berbagai macam evaluasi seperti kuis, ulangan harian, penilaian tengah semester, dan penilaian akhir semester (Zahra, 2022). Evaluasi terdiri dari dua katagori yaitu tes dan non-tes.

Menurut Faiz, Tes merupakan bentuk penilaian yang menghasilkan nilai terkait tingkah laku atau prestasi anak, yang harus diselesaikan sesuai dengan standar yang telah ditetapkan (Faiz et al., 2022). Lebih lanjut Arikunto, Tes adalah suatu instrumen atau langkah yang dipakai untuk mengetahui atau mengukur sesuatu dengan mematuhi metode atau aturan yang telah ditetapkan (Suharmi, 2010). Apabila tes disusun dengan baik, maka tujuannya akan tercapai. Tes memiliki beberapa peran, dan menurut Asrul dan rekan-rekannya, terdapat dua fungsi utama tes, yakni: 1) Sebagai alat untuk mengevaluasi peserta didik, digunakan untuk mengukur kemampuan yang diperoleh siswa selama pembelajaran, dan 2) Sebagai penilaian keberhasilan dari program pengajaran (Asrul, Abdul, 2022).

Tes digunakan untuk mengevaluasi sejauh mana kesuksesan program pembelajaran yang telah dilaksanakan. Dalam rangka pentingnya tes dalam pembelajaran, guru harus merencanakan tes dengan teliti. Tes yang efektif adalah yang mampu membedakan antara siswa yang telah belajar dan yang belum belajar, serta memberikan hasil yang konsisten meskipun digunakan di tempat berbeda (Magdalena, 2020). Menurut Sopiha dalam Muluki, Tes yang efektif harus memiliki pertanyaan yang dirumuskan secara baik dan mampu menilai kemampuan serta keterampilan siswa sesuai dengan tujuan pembelajaran (Muluki, 2020). Tes yang efektif harus memiliki validitas, reliabilitas, objektivitas, kemudahan penggunaan, dan efisiensi biaya. Validitas menjamin keakuratan pengukuran, reliabilitas menunjukkan konsistensi hasil, objektivitas menjamin ketidakberpihakan, praktisitas mengacu pada kemudahan

pelaksanaan, dan ekonomisitas menekankan penggunaan sumber daya yang efisien (Faisal, 2022).

Rahmaini menjelaskan bahwa evaluasi tes melibatkan beberapa langkah penting. Pertama, tes dievaluasi untuk memastikan keaslian soal, termasuk tingkat kesulitan dan kejelasan instruksi. Kedua, dilakukan analisis soal secara sistematis untuk mendapatkan informasi khusus tentang setiap butir soal. Selanjutnya, validitas kurikuler dievaluasi untuk memastikan bahwa setiap soal sesuai dengan tujuan pembelajaran. Terakhir, reliabilitas dinilai dengan menghitung daya pembeda pada setiap soal (Rahmaini & Taufiq, 2018). Agar mempermudah guru dalam mengevaluasi hasil tes, analisis butir soal dapat dilakukan. Analisis butir soal ini merupakan proses pengujian terhadap kualitas setiap soal guna memperoleh informasi tentang karakteristiknya (Putri, 2020). Soal yang baik adalah yang memberikan informasi yang tepat sesuai dengan tujuannya, termasuk dalam menilai apakah peserta didik telah memahami materi yang diajarkan atau belum (Kaka et al., 2024).

Menurut penelitian Eyrene M., ditemukan bahwa beberapa soal yang dibuat oleh guru matematika kelas X di SMA Negeri 1 Airmadidi untuk tahun pelajaran 2019/2020 tidak valid dan memiliki tingkat kesulitan yang terlalu tinggi. Dalam penelitian Astuti tentang analisis soal matematika di SD Negeri 005 Binuang tahun ajaran 2018/2019, dari 30 soal, 26 soal dinyatakan valid dan 4 soal tidak valid, dengan reliabilitas soal yang sangat tinggi yaitu 0,823. Indeks kesulitan soal menunjukkan 15 soal mudah, 13 soal sedang, dan 2 soal sulit. Daya pembeda menunjukkan 2 soal jelek, 7 soal cukup, 19 soal baik, dan 2 soal sangat baik. Efektivitas pengecoh menunjukkan 2 pengecoh sangat baik, 3 pengecoh baik, 7 pengecoh cukup, 9 pengecoh kurang baik, dan 9 pengecoh jelek. Penelitian Tilaar, Anetha L.F. menyatakan bahwa hasil penelitian di kelas VIII SMPN Pineleng tahun ajaran 2018/2019 menunjukkan validitas soal lebih dari 50% dengan 21 soal yang valid. Reliabilitas tes termasuk kategori cukup dengan indeks reliabilitas sebesar 0,5. Daya pembeda soal baik dengan 3 soal cukup dan 6 soal baik. Hasil penelitian ini menunjukkan bahwa masih terdapat butir-butir soal yang dibuat oleh guru yang tidak berkualitas, sejalan dengan masalah yang terjadi di Sekolah Dasar Negeri 1 Sungai Bungin.

Hasil pengamatan di Sekolah Dasar Negeri 1 Sungai Bungin menunjukkan bahwa evaluasi pembelajaran sering menggunakan tes tertulis, namun analisis soal jarang dilakukan. Hal ini berdampak pada ketidakpastian terhadap kualitas soal. Sehingga, kualitas soal tersebut belum dapat dipastikan apakah memenuhi persyaratan sebagai alat ukur. Untuk menghasilkan soal yang bermutu, diperlukan analisis untuk mengetahui kualitas soal sebelum diuji coba pada siswa. Kualitas soal sangat mempengaruhi hasil belajar siswa, sehingga analisis butir soal sangat diperlukan untuk mengidentifikasi masalah yang mungkin ada pada soal tersebut. Penelitian ini bertujuan untuk mengetahui kualitas butir soal pilihan ganda pada materi rangka manusia pada ulangan harian kelas

V Sekolah Dasar Negeri 1 Sungai Bungin ditinjau dari segi validitas reliabilitas, tingkat kesukaran, dan daya pembeda.

Metode

Penelitian ini merupakan penelitian deskriptif dengan pendekatan kuantitatif. Penelitian kuantitatif adalah penelitian yang dimulai dengan pengumpulan data, dilanjutkan dengan analisis data, dan kemudian menyajikan data tersebut dalam bentuk gambar, tabel, grafik, atau format lain dengan menerapkan prinsip-prinsip objektivitas (Regeta et al., 2022). Populasi penelitian berjumlah 22 siswa. Pengambilan sampel menggunakan teknik sampling jenuh, yaitu teknik penentuan sampel bila semua anggota populasi digunakan sebagai sampel. Hal ini sering dilakukan bila jumlah populasi relatif kecil, kurang dari 30 orang (Munawwaroh et al., 2024). Variabel dalam penelitian ini meliputi validitas, reliabilitas, tingkat kesukaran, dan daya beda sebagai kriteria butir soal pada materi rangka manusia, penelitian ini melibatkan 25 butir soal pilihan ganda. Teknik pengumpulan data dilakukan dengan metode dokumentasi dari hasil tes. Data dianalisis mencakup validitas, reliabilitas, tingkat kesulitan, dan daya beda dengan menggunakan software IBM SPSS Statistics 27. Analisis validitas dilakukan dengan menggunakan korelasi produk moment, yaitu:

$$r_{xy} = \frac{\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}}{\sqrt{[\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}] [\sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n}]}}$$

Nilai r_{xy} akan dibandingkan dengan koefisien tabel nilai "r" product moment pada taraf signifikansi 5%. Jika nilai r_{xy} dari hasil koefisien korelasi lebih besar (>) dari nilai r tabel, maka hasil tersebut dinyatakan signifikan, yang berarti butir soal tes dinyatakan valid. Nilai r_{xy} adalah nilai koefisien korelasi dari setiap butir/item soal sebelum dikoreksi. Selanjutnya, dilakukan pencarian koefisien korelasi item-total yang telah dikoreksi dengan menggunakan rumus berikut:

$$R_{x(y-1)} = \frac{r_{x,s} - s_x}{\sqrt{s^2 + s^2 - 2r_{x,(s)} (s_x)}}$$

Nilai $r_{x(y-1)}$ akan dibandingkan dengan koefisien korelasi dalam tabel rtabel $df = (n - 2)$. Jika $r_{x(y-1)} \geq r_{tabel}$, maka instrumen dianggap valid. Pada output SPSS, jika koefisien korelasi item-total yang telah dikoreksi $t \geq r_{tabel}$, maka instrumen juga dinyatakan valid.

Reliabilitas adalah indeks yang menunjukkan sejauh mana suatu alat pengukuran dapat dipercaya atau diandalkan. Sebuah instrumen pengukuran dianggap reliabel jika hasil pengukurannya konsisten, tepat, dan akurat. Analisis reliabilitas menggunakan rumus Alfa Cronbach, yaitu:

$$r_i = \frac{k}{(k-1)} \left\{ 1 - \frac{Zst^2}{st^2} \right\}$$

Nilai koefisien (r_i) akan dibandingkan dengan koefisien korelasi tabel $r_{tabel} = r(n - 2)$. Jika $r_{11} > r_{tabel}$, maka instrumen dianggap reliabel. Pada output SPSS, jika nilai Cronbach's Alpha $> r_{tabel}$, maka instrumen juga dinyatakan reliabel. Analisis tingkat kesulitan soal bertujuan untuk mengkaji soal-soal tes berdasarkan tingkat kesulitannya, sehingga dapat diketahui mana soal yang termasuk mudah, sedang, dan sulit.

Keseimbangan soal berarti adanya distribusi proporsional antara soal-soal yang mudah, sedang, dan sulit. Perbandingan antara soal mudah-sedang-sukar bisa dibuat 3-4-3, yang berarti 30% soal kategori mudah, 40% soal kategori sedang, dan 30% soal kategori sulit. Alternatif lain yang mirip adalah 3-5-2, yang berarti 30% soal kategori mudah, 50% soal kategori sedang, dan 20% soal kategori sulit. Rumus yang digunakan untuk mengetahui tingkat kesukaran dari tiap butir soal yaitu:

$$I = \frac{B}{J}$$

Penafsiran tingkat kesukaran butir tes menggunakan kriteria yang dikemukakan oleh Witherington dalam Anas Sudijono sebagai berikut:

Tabel 1: Interpretasi Tingkat Kesukaran Butir Soal

Interval	Interprestasi
0,00 - 0,30	Soal Sukar
0,31 - 0,70	Soal Sedang
0,71 - 1,00	Soal Mudah

Menurut Novalia Menganalisis daya pembeda berarti mengkaji soal-soal tes berdasarkan kemampuan tes tersebut dalam membedakan butir soal yang termasuk dalam kategori rendah dan kategori tinggi dalam presentasinya. Penentuan daya beda butir soal pada Anates dapat diketahui dalam tabel daya pembeda pada kolom DP persen. Butir soal yang memiliki indeks daya beda $\geq 0,30$ dinyatakan baik dan butir soal yang indeks daya beda $< 0,30$ dinyatakan tidak baik. Daya pembeda butir soal memiliki manfaat yaitu untuk meningkatkan mutu setiap-butir soal melalui data empiriknya dan untuk mengetahui seberapa jauh masing-masing butir soal dapat membedakan kemampuan siswa, yaitu siswa yang telah memahami atau belum memahami materi yang diajarkan pendidik. Rumus yang digunakan yaitu:

$$D = P_A - P_B$$

Sebagai catatan, dalam keadaan di mana jumlah teste adalah cukup besar. Daya pembeda item cukup dihitung berdasarkan 27% teste kelompok atas dan 27% dari testee

kelompok bawah, sedangkan testee yang terletak diantara dua ujung ekstrem itu tidak perlu diikuti sertakan dalam perhitungan analisis (Sudijono, 2017).

Tabel 2. Interpretasi Indeks Daya Pembeda Butir Soal

Daya Pembeda	Intreprestasi
0,70 – 1,00	Baik Sekali
0,40 – 0,69	Baik
0,20 – 0,39	Cukup
0,00 – 0,19	Kurang Baik

Hasil

1. Uji Validitas

Setelah melakukan perhitungan pada taraf signifikansi 5%. Dengan jumlah subjek penelitian sebanyak 22 siswa, pada tingkat signifikansi 5% dengan $n=20$, nilai tabel adalah 0.422. Jika nilai r_{tabel} lebih kecil daripada r_{hitung} , maka butir soal dianggap valid, sebaliknya jika nilai r_{tabel} lebih besar daripada r_{hitung} , maka butir soal dianggap tidak valid. Untuk menilai validitas masing-masing pertanyaan tes, peneliti menggunakan perangkat lunak SPSS IBM Statistic 27 dan menghitung korelasi menggunakan rumus Korelasi Product Moment, dengan hasil sebagai berikut:

Tabel 3. Korelasi Antara Skor Tiap Butir Soal Dengan Skor Total Berdasarkan Validitas Item

No. Soal	R_{tabel} dengan R_{hitung}	Interpretasi
Soal 1	$0,422 < 0,489$	Valid
Soal 2	$0,422 < 0,657$	Valid
Soal 3	$0,422 < 0,500$	Valid
Soal 4	$0,422 < 0,510$	Valid
Soal 5	$0,422 > 0,422$	Tidak Valid
Soal 6	$0,422 > 0,452$	Tidak Valid
Soal 7	$0,422 < 0,619$	Valid
Soal 8	$0,422 > 0,453$	Tidak Valid
Soal 9	$0,422 < 0,553$	Valid
Soal 10	$0,422 < 0,636$	Valid
Soal 11	$0,422 < 0,532$	Valid
Soal 12	$0,422 > 0,407$	Tidak Valid
Soal 13	$0,422 < 0,539$	Valid
Soal 14	$0,422 < 0,472$	Valid
Soal 15	$0,422 > 0,420$	Tidak Valid
Soal 16	$0,422 < 0,580$	Valid
Soal 17	$0,422 < 0,659$	Valid
Soal 18	$0,422 > 0,455$	Tidak Valid
Soal 19	$0,422 > 0,418$	Tidak Valid
Soal 20	$0,422 > 0,402$	Tidak Valid

Soal 21	0,422 < 0,576	Valid
Soal 22	0,422 < 0,513	Valid
Soal 23	0,422 < 0,506	Valid
Soal 24	0,422 < 0,642	Valid
Soal 25	0,422 < 0,641	Valid

Hasil perhitungan validitas item tersebut kemudian dibandingkan dengan nilai tabel pada tingkat signifikansi 5%. Dengan jumlah sampel $n=20$, nilai tabel (r_{tabel}) adalah 0,422. Jika nilai tabel (r_{tabel}) kurang dari nilai perhitungan (r_{hitung}), maka butir soal tersebut dianggap valid. Sebaliknya, jika nilai r_{tabel} lebih besar dari nilai r_{hitung} , maka butir soal dianggap tidak valid.

Tabel 4. Distribusi Butir Soal Berdasarkan Indeks Validitas

Kategori	Jumlah	Persentase	Nomor Soal
Valid	17 Butir Soal	68%	1, 2, 3, 4, 7, 9, 10, 11, 13, 14, 16, 17, 21, 22, 23, 24, 25.
Tidak Valid	8 Butir Soal	32%	5, 6, 8, 12, 15, 18, 19, 20.
Jumlah	25 Butir Soal		

Hasil analisis butir soal pilihan ganda materi rangka manusia menunjukkan bahwa dari total 25 soal, 17 soal atau sekitar 68% dianggap valid, sedangkan 8 soal atau sekitar 32% dianggap tidak valid. Temuan ini sejalan dengan penelitian Aditya Melia Nugrahanti pada tahun 2013, yang juga menunjukkan bahwa sekitar 70% soal pilihan ganda valid dan 30% tidak valid, dan dapat mengukur sesuai dengan yang seharusnya diukur.

2. Uji Reliabilitas

Dasar penentuan reliabilitas dalam pengujian adalah jika nilai Alpha lebih besar daripada nilai R Tabel, maka item dalam kuesioner dianggap reliabel atau konsisten. Sebaliknya, jika nilai Alpha lebih kecil daripada nilai R Tabel, maka item dalam kuesioner dianggap tidak reliabel (Hermawaty, 2019). Berdasarkan dari hasil perhitungan IBM SPSS Statistics 27 Informasi tersebut tersedia dalam tabel berikut:

Tabel 5. Output Hasil Uji Reliabilitas Butir Soal (*Summary*)

Case Processing Summary			
		N	%
Cases	Valid	17	100.0
	Excluded ^a	0	.0
	Total	17	100.0

a. Listwise deletion based on all variables in the procedure.

Berdasarkan ringkasan pengolahan butir soal pada tabel, terdapat 17 data yang dapat dianggap valid. Ini menunjukkan bahwa dari seluruh data yang dikumpulkan, 17 di antaranya memenuhi kriteria validitas yang telah ditetapkan untuk analisis lebih lanjut.

Validitas data ini penting untuk memastikan bahwa hasil analisis dan kesimpulan yang diambil berdasarkan data tersebut adalah akurat dan dapat dipercaya. Data yang valid ini kemudian digunakan dalam uji reliabilitas, yang hasilnya ditampilkan dalam tabel Reliability Statistics berikut:

Tabel 6. Output Hasil Uji Reliabilitas Butir Soal (*Statistic*)

Reliability Statistics			
Cronbach's Alpha	Part	Value	.798
	1	N of Items	8 ^a
	Part	Value	.792
	2	N of Items	9 ^b
Total N of Items			17
Correlation Between Forms			.862
Spearman-Brown Coefficient	Equal Length		.926
	Unequal Length		.926
Guttman Split-Half Coefficient			.919
a. The items are: SOAL1, SOAL2, SOAL3, SOAL4, SOAL7, SOAL9, SOAL10, SOAL11.			
b. The items are: SOAL13, SOAL14, SOAL16, SOAL17, SOAL21, SOAL22, SOAL23, SOAL24, SOAL25.			

Perhatikan nilai *Cronbach's Alpha* dalam tabel Statistik Reliabilitas di bagian value bahwa nilai tersebut merupakan indikator reliabilitas keseluruhan dari tes. Semakin tinggi nilainya, semakin reliabel tes tersebut. Untuk menentukan apakah nilai-nilai tersebut reliabel, bandingkan dengan nilai R Tabel pada $DF = N - 2$ dan probabilitas 0,05. Dalam hal ini, $DF = 22 - 2 = 20$. Nilai R Tabel untuk DF 20 dengan probabilitas 0,05 adalah 0,422.

Kemudian, hasil uji reliabilitas ditampilkan pada tabel *Reliability Statistics* di atas. Dalam uji ini, nilai *Cronbach's Alpha* yang tercantum pada bagian part 1 value adalah sebesar 0,798, dengan jumlah item yang diuji sebanyak 8 butir soal dan part 2 value sebesar 0,792, dengan jumlah item yang diuji sebanyak 9 butir soal. Nilai *Cronbach's Alpha* ini menunjukkan tingkat reliabilitas dari instrumen yang digunakan.

Untuk menentukan apakah instrumen tersebut reliabel, dilakukan perbandingan antara nilai *Cronbach's Alpha* dengan nilai rtabel. Pada uji ini, nilai rtabel adalah 0,422. Karena nilai *Cronbach's Alpha* (0,798 dan 0,792) melebihi nilai rtabel (0,422), dapat disimpulkan bahwa butir soal yang digunakan reliabel. Ini mengindikasikan bahwa item-item dalam butir soal tersebut konsisten dalam mengukur aspek yang patut diukur dan dapat diandalkan untuk menghasilkan hasil yang konsisten.

3. Uji Tingkat Kesukaran

Dari hasil perhitungan yang dilakukan menggunakan IBM SPSS Statistic versi 27, tingkat kesulitan dari setiap butir soal materi struktur dan fungsi bagian rangka manusia mata pelajaran Tematik telah tercatat dan dapat dijelaskan sebagai berikut:

Tabel 7. Hasil Evaluasi Kesulitan Butir Soal

No.	Butir Soal	Indeks	Kriteria
1	Soal 1	0,64	Soal Sedang
2	Soal 2	0,50	Soal Sedang
3	Soal 3	0,59	Soal Sedang
4	Soal 4	0,64	Soal Sedang
5	Soal 5	0,45	Soal Sedang
6	Soal 6	0,45	Soal Sedang
7	Soal 7	0,91	Soal Mudah
8	Soal 8	0,50	Soal Sedang
9	Soal 9	0,68	Soal Sedang
10	Soal 10	0,68	Soal Sedang
11	Soal 11	0,50	Soal Sedang
12	Soal 12	0,82	Soal Mudah
13	Soal 13	0,68	Soal Sedang
14	Soal 14	0,77	Soal Mudah
15	Soal 15	0,86	Soal Mudah
16	Soal 16	0,23	Soal Sukar
17	Soal 17	0,14	Soal Sukar
18	Soal 18	0,59	Soal Sedang
19	Soal 19	0,73	Soal Mudah
20	Soal 20	0,32	Soal Mudah
21	Soal 21	0,09	Soal Sukar
22	Soal 22	0,64	Soal Sedang
23	Soal 23	0,73	Soal Mudah
24	Soal 24	0,18	Soal Sukar
25	Soal 25	0,73	Soal Mudah

Hasil evaluasi yang telah dijalankan secara rinci membuktikan bahwa dari total jumlah butir soal yang diujikan, terdapat 4 soal (sekitar 16% dari keseluruhan) yang termasuk dalam kategori sukar. Selanjutnya, sebanyak 13 soal (atau sekitar 52%) masuk dalam kategori sedang. Di sisi lain, ada 8 soal (sekitar 32%) yang dinyatakan sebagai soal mudah. Berikut adalah distribusi tingkat kesulitan butir soal pilihan ganda pada materi rangka manusia siswa kelas V (lima) Sekolah Dasar:

Tabel 8. Distribusi Butir Soal Berdasarkan Tingkat Kesukaran Butir Soal

Kategori	Jumlah	Persentase	Nomor Butir Soal
Sukar (0,00-0,30)	4 butir soal	16%	16,17,21,24

Sedang (0,31-0,70)	13 butir soal	52%	1,2,3,4,5,6,8,9,10,11,13,18,22
Mudah (0,71-1,00)	8 butir soal	32%	7,12,14,15,19,20,23,25

Dari tabel yang telah dipresentasikan Informasi mengenai proporsi tingkat kesulitan butir soal, dapat disimpulkan bahwa hasil analisis atas tingkat kesulitan butir soal pilihan ganda materi struktur dan fungsi bagian rangka manusia telah memenuhi standar yang diharapkan. Standar tersebut merujuk pada proporsi yang ditetapkan untuk tingkat kesulitan butir soal, yaitu 25% untuk kategori sukar, 50% untuk kategori sedang, dan 25% untuk kategori mudah.

4. Uji Daya Pembeda

Dalam penelitian ini, daya pembeda diukur dengan melibatkan 22 siswa. Cara untuk menentukan daya pembeda adalah dengan mengurangi nilai proporsi siswa yang menjawab benar di kelompok atas dibagi dengan jumlah siswa di kelompok atas, kemudian dibagi dengan proporsi siswa yang menjawab benar di kelompok bawah dibagi dengan jumlah siswa di kelompok bawah.

Tabel 9. Hasil Analisis Daya Beda Butir Soal

No.	Butir Soal	Indeks Correlation	Interpretasi
1	Soal 1	0,431	Baik
2	Soal 2	0,527	Baik
3	Soal 3	0,268	Cukup
4	Soal 4	0,624	Baik
5	Soal 5	0,553	Baik
6	Soal 6	0,503	Baik
7	Soal 7	0,753	Sangat Baik
8	Soal 8	0,421	Baik
9	Soal 9	0,523	Baik
10	Soal 10	0,405	Baik
11	Soal 11	0,646	Baik
12	Soal 12	0,439	Baik
13	Soal 13	0,460	Baik
14	Soal 14	0,493	Baik
15	Soal 15	0,369	Cukup
16	Soal 16	0,549	Baik
17	Soal 17	0,602	Baik
18	Soal 18	0,440	Baik
19	Soal 19	0,392	Cukup
20	Soal 20	0,363	Cukup
21	Soal 21	0,552	Baik
22	Soal 22	0,431	Baik
23	Soal 23	0,543	Baik
24	Soal 24	0,706	Sangat baik
25	Soal 25	0,472	Baik

Berdasarkan tabel 7 diketahui bahwa kemampuan membedakan butir soal didominasi oleh kategori baik, yang mencakup 19 butir soal atau 76% dari total. Selain itu,

terdapat 2 butir soal atau 16% yang termasuk dalam kategori sangat baik, menunjukkan bahwa sebagian besar soal memiliki kualitas yang sangat tinggi. Kategori cukup mencakup 4 butir soal atau 16%, menunjukkan bahwa masih ada beberapa soal yang memerlukan perbaikan. Nihil butir soal atau 0% berada dalam kategori jelek.

Tabel 8. Distribusi Daya Beda Butir Soal

Kategori	Jumlah	Persentase	Nomor Butir Soal
Sangat Baik	2 butir soal	16%	7,24
Baik	19 butir soal	76%	1,2,4,5,6,8,9,10,11,12,13,14,16,17,18,21,22,23,25
Cukup	4 butir soal	16%	3,15,19,20
Jelek	-	0%	-

Pembahasan

1. Validitas

Validitas merupakan parameter yang mencerminkan keabsahan atau keandalan suatu alat ukur. Prinsip validitas berkaitan dengan keandalan alat ukur dalam mengumpulkan data yang tepat dan relevan. Validitas memverifikasi bahwa instrumen pengukur secara tepat mengukur apa yang seharusnya diukur (Fadli et al., 2023). Hasil penelitian menunjukkan bahwa validitas butir soal pilihan ganda materi rangka manusia pada ulangan harian kelas V di SDN 1 Sungai Bungin menunjukkan butir soal memiliki indeks validitas dengan kategori "valid" sebanyak 68% (17 butir) dan "tidak valid" sebanyak 32% (8 butir).

Dari pemaparan tersebut, dapat disimpulkan bahwa butir soal pilihan ganda materi rangka manusia pada ulangan harian kelas V di SDN 1 Sungai Bungin memiliki kualitas baik dari segi validitasnya. Hal ini karena 17 dari 25 dinyatakan valid. Oleh karena itu, butir soal pilihan ganda materi rangka manusia pada ulangan harian kelas V di SDN 1 Sungai Bungin dapat dikatakan mampu mengukur apa yang seharusnya diukur (Sudijono, 2017). Hal ini sejalan dengan pendapat Himawan dan Nurgiyanto (2022), yang menyatakan bahwa "alat ukur dinyatakan valid jika mampu menyampaikan informasi empirik sesuai dengan yang diukur dengan tepat."

2. Reliabilitas

Reliabilitas mengacu pada kestabilan dalam instrumen pengukuran. Ini dapat diamati ketika instrumen pengukuran yang sama memberikan hasil yang serupa, atau ketika ada kesamaan dalam penilaian subjektif, bagaimana kesamaan skor yang diberikan oleh dua penilai. Meskipun pengukuran yang dapat diandalkan akan menghasilkan hasil yang konsisten, namun belum tentu mampu mengukur dengan akurat aspek yang seharusnya diukur (Sanaky, 2021). Dalam konteks penelitian, reliabilitas merujuk pada tingkat konsistensi pengukuran dari sebuah tes ketika diulang-ulang terhadap subjek

yang sama dalam kondisi yang serupa. Suatu penelitian dianggap reliabel jika hasilnya konsisten dalam pengukuran yang sama. Sebaliknya, jika hasil pengukuran yang berulang menunjukkan variasi, penelitian dianggap tidak reliabel (Jaya, 2020).

Hasil penelitian pada nilai *Cronbach's Alpha* dalam tabel 4 Statistik Reliabilitas di bagian *value* bahwa nilai tersebut merupakan indikator reliabilitas keseluruhan dari tes. Semakin tinggi nilainya, semakin reliabel tes tersebut. Untuk menentukan apakah nilai-nilai tersebut reliabel, bandingkan dengan nilai R Tabel pada $DF = N - 2$ dan probabilitas 0,05. Dalam hal ini, $DF = 22 - 2 = 20$. Nilai R Tabel untuk DF 20 dengan probabilitas 0,05 adalah 0,422. Dalam uji ini, nilai *Cronbach's Alpha* yang tercantum pada bagian part 1 *value* adalah sebesar 0,798, dengan jumlah item yang diuji sebanyak 8 butir soal dan part 2 *value* sebesar 0,792, dengan jumlah item yang diuji sebanyak 9 butir soal. Nilai *Cronbach's Alpha* ini menunjukkan tingkat reliabilitas dari instrumen yang digunakan. Karena nilai Cronbach's Alpha (0,798 dan 0,792) melebihi nilai rtabel (0,422), dapat disimpulkan bahwa butir soal yang digunakan reliabel. Ini mengindikasikan bahwa item-item dalam butir soal tersebut konsisten dalam mengukur aspek yang patut diukur dan dapat diandalkan untuk menghasilkan hasil yang konsisten.

3. Tingkat Kesukaran

Kesulitan suatu soal adalah persentase yang menunjukkan jumlah siswa yang dapat menjawabnya dengan benar atau tidak. Persentase ini memberikan gambaran tentang seberapa mudah atau sulit sebuah soal bagi kelompok siswa tertentu. Semakin tinggi persentase siswa yang menjawab dengan benar, maka soal tersebut dianggap lebih mudah, sedangkan semakin rendah persentasenya, soal tersebut dianggap lebih sulit (Kaka et al., 2024). Kualitas butir-butir item tes pembelajaran bisa diidentifikasi berdasarkan tingkat kesulitan atau kesukaran masing-masing butir tersebut. Butir-butir tes pembelajaran dianggap baik jika tidak terlalu sulit maupun terlalu mudah, artinya, tingkat kesulitan item tersebut harus seimbang atau memadai (Mudanta et al., 2020)

Hasil evaluasi yang telah dijalankan secara rinci membuktikan bahwa dari total jumlah butir soal yang diujikan, terdapat 4 soal (sekitar 16% dari keseluruhan) yang termasuk dalam kategori sukar. Selanjutnya, sebanyak 13 soal (atau sekitar 52%) masuk dalam kategori sedang. Di sisi lain, ada 8 soal (sekitar 32%) yang dinyatakan sebagai soal mudah. Hasil analisis atas tingkat kesulitan butir soal pilihan ganda materi struktur dan fungsi bagian rangka manusia telah memenuhi standar yang diharapkan. Standar tersebut merujuk pada proporsi yang ditetapkan untuk tingkat kesulitan butir soal, yaitu 25% untuk kategori sukar, 50% untuk kategori sedang, dan 25% untuk kategori mudah.

Berdasarkan pembahasan di atas, proporsi tingkat kesulitan butir soal harus mencerminkan distribusi yang seimbang, di mana setiap kategori kesulitan memiliki bobot yang sesuai dengan kebutuhan siswa dalam menguasai materi. Dengan demikian, untuk memenuhi standar yang baik, proporsi tingkat kesulitan butir soal harus sesuai

dengan standar yang telah ditetapkan, sehingga dapat memberikan gambaran yang tepat tentang tingkat pemahaman siswa terhadap materi Pelajaran (Ida & Musyarofah, 2021). Hal ini sependapat dengan Arikunto yang menyatakan soal yang baik merupakan soal yang tidak sangat mudah dan tidak sangat sukar karena soal yang sangat mudah tidak bisa membangkitkan kemampuan peserta didik untuk memecahkan soal yang ada. Sedangkan soal yang sangat sukar mematahkan semangat peserta didik untuk mencoba kembali karena menganggap dirinya tidak mampu (Suharsimi, 2019).

4. Daya Pembeda

Hasil penelitian memperlihatkan bahwa daya pembeda butir soal mengenai materi rangka manusia dalam mata pelajaran tematik kelas V diketahui bahwa kemampuan membedakan butir soal didominasi oleh kategori baik, yang mencakup 19 butir soal atau 76% dari total. Selain itu, terdapat 2 butir soal atau 16% yang termasuk dalam kategori sangat baik, menunjukkan bahwa sebagian besar soal memiliki kualitas yang sangat tinggi. Kategori cukup mencakup 4 butir soal atau 16%, menunjukkan bahwa masih ada beberapa soal yang memerlukan perbaikan. Nihil butir soal atau 0% berada dalam kategori jelek.

Dari tabel 8, butir nomor 3,15,19,20 termasuk dalam kategori cukup, berarti butir soal tersebut hampir bisa membedakan kemampuan peserta didik yang pandai dan kurang pandai, butir nomor 1, 2, 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 16, 17, 18, 21, 22, 23, 25 termasuk dalam kategori baik, berarti butir soal tersebut sedikit bisa membedakan peserta didik yang pandai dan kurang pandai. Dan butir nomor 7, 24 termasuk kategori sangat baik, berarti butir soal tersebut bisa membedakan peserta didik yang pandai dan kurang pandai.

Menurut Nurhayati, Pembeda soal adalah kemampuan soal dalam menyeleksi tingkat kemampuan peserta didik (pandai dan kurang pandai). Analisis daya pembeda dilakukan untuk memahami sejauh mana soal dapat membedakan antara peserta didik dengan kemampuan akademik yang tinggi dan rendah. Indeks diskriminasi soal digunakan untuk menghitung daya pembeda ini. Dari hasil analisis, soal dengan daya pembeda yang buruk dapat diperbaiki atau dibuang. Namun, soal dengan indeks daya pembeda negatif harus dibuang. Soal-soal dengan kategori cukup, baik, dan sangat baik dapat dimasukkan ke dalam bank soal dan digunakan kembali untuk tes mendatang (Nurhayati, 2024). Pendapat ini didukung oleh Supriani dkk., Menurut Supriani dan rekan-rekannya, perhitungan daya pembeda merupakan evaluasi seberapa baik suatu soal dapat memisahkan peserta didik yang telah memahami materi dari mereka yang belum atau kurang memahaminya, berdasarkan pada kriteria tertentu (Supriani et al., 2022).

Simpulan

Dari hasil analisis pada bagian metode dan hasil penelitian, dapat disarikan beberapa hal sebagai berikut: 1). Taraf butir soal dari validitas, dengan 17 soal atau 68%

yang dianggap valid berdasarkan kriteria validitas yang ditetapkan oleh Yudono dalam Arikunto dengan tolak ukur pengujian apabila r hitung $>$ r tabel dengan $\alpha=0,05$ maka alat ukur dinyatakan valid dan sebaliknya. Kriteria validitas butir soal, 0,800 – 1,00 = Sangat Tinggi, 0,600 – 0,799 = Tinggi, 0,400 – 0,599 = Cukup, 0,200 – 0,399 = Rendah, 0,000 – 0,199 = Sangat Rendah. 2). Taraf butir soal dari reliabilitas, di mana butir soal memiliki koefisien reliabilitas tinggi, yakni 0,798 dan 0,792. 3). Taraf butir soal dari daya pembeda, dengan tidak ada butir soal yang memiliki ketidakkampuan dalam membedakan (buruk), 4 butir atau 16% memiliki daya pembeda cukup (0.2-0.3), 19 butir atau 76% memiliki daya pembeda baik (0.3-0.4), dan 2 butir atau 8% memiliki daya pembeda sangat baik (>0.4). 4) Kualitas butir soal berdasarkan tingkat kesulitan, dengan 4 butir atau 16% yang sulit, 13 butir atau 52% yang sedang, dan 8 butir atau 32% yang mudah.

Penelitian ini menunjukkan bahwa sebagian besar soal pilihan ganda pada materi rangka manusia di ulangan harian sekolah dasar memiliki validitas dan reliabilitas yang baik. Namun, beberapa soal memerlukan revisi karena daya pembeda yang kurang memadai dan tingkat kesulitan yang tidak seimbang. Untuk meningkatkan kualitas soal, guru perlu pelatihan dalam penyusunan soal yang valid dan reliabel. Butir soal yang kurang memadai harus direvisi atau diganti. Pengembangan bank soal terstandarisasi dianjurkan agar soal-soal berkualitas dapat digunakan kembali. Evaluasi berkala terhadap soal-soal diperlukan, dan guru sebaiknya melakukan diskusi rutin dengan rekan sejawat untuk berbagi pengalaman dalam membuat soal yang berkualitas. Implementasi rekomendasi ini diharapkan dapat meningkatkan kualitas evaluasi pembelajaran dan hasil belajar siswa.

Referensi

- Asrul, Abdul, M. (2022). *Evaluasi Pembelajaran*. Medan: Perda Publishing.
- Baharuddin. (2024). The effectiveness of distance Arabic learning for Indonesia speakers using YouTube channels. *Journal of Education and Learning*, 18(3), 1029–1037. <https://doi.org/https://doi.org/10.11591/edulearn.v18i3.21034>
- Fadli, R., Hidayati, S., Cholifah, M., Siroj, R. A., & Afgani, M. W. (2023). Validitas dan Reliabilitas pada Penelitian Motivasi Belajar Pendidikan Agama Islam Menggunakan Product Moment. *JlIP - Jurnal Ilmiah Ilmu Pendidikan*, 6(3), 1734–1739. <https://doi.org/10.54371/jiip.v6i3.1419>
- Faisal, A. (2022). Administrasi Tes Dalam Evaluasi Pembelajaran. *Darussalam*, 23(1), 1–12. <https://doi.org/http://dx.doi.org/10.58791/drs.v23i1.175>
- Faiz, A., Permana Putra, N., & Nugraha, F. (2022). Memahami Makna Tes, Pengukuran (Measurement), Penilaian (Assessment), Dan Evaluasi (Evaluation) Dalam Pendidikan. *Jurnal Education and Development*, 10(3), 492–495.
- Himawan, R., & Nurgiyantoro, B. (2022). Analisis Butir Soal Latihan Penilaian Akhir Semester Ganjil Mata Pelajaran Bahasa Indonesia Kelas VIII SMPN 1 Bambanglipuro

- Bantul Menggunakan program ITEMAN (Analysis of exercise items for odd semester end of semester Indonesian language subjects class. *Kembara: Jurnal Keilmuan Bahasa, Sastra, Dan Pengajarannya*, 8(1), 160–180. <https://doi.org/https://doi.org/10.22219/kembara.v8i1.20530>
- Ida, F. F., & Musyarofah, A. (2021). Validitas dan Reliabilitas dalam Analisis Butir Soal. *Al-Mu'Arrib: Journal of Arabic Education*, 1(1), 34–44. <https://doi.org/10.32923/al-muarrib.v1i1.2100>
- Jaya, M. L. M. (2020). *Metode Penelitian Kuantitatif dan Kualitatif (Teori, Penerapan dan Riset Nyata)*. Yogyakarta: Quadrant.
- Jayanti, T., Siregar, N., Jalal, M., Guru Madrasah Ibtidaiyah, P., & Jambi, U. (2024). Implementasi ke mampuan kerja sama siswa dalam pembelajaran tematik Madrasah Ibtidaiyah. *Trihayu: Jurnal Pendidikan Ke-SD-An*, 10(2), 203–212.
- Kaka, L., Bano, V. O., & Njoeroemana, Y. (2024). Efektivitas Analisis Butir Soal Pilihan Ganda Menggunakan Aplikasi Anates di SMPN 2 Kanatang. *Jurnal Inovasi Penelitian*, 4(9), 1441–1450.
- Magdalena, I. (2020). *Evaluasi Pembelajaran SD (Teori dan Praktik)*. Sukabumi: CV Jejak, anggota IKAPI.
- Mudanta, K. A., Astawan, I. G., & Jayanta, I. N. L. (2020). Instrumen Penilaian Motivasi Belajar dan Hasil Belajar IPA Siswa Kelas V Sekolah Dasar. *Mimbar Ilmu*, 25(2), 101. <https://doi.org/10.23887/mi.v25i2.26611>
- Muluki, A. (2020). Analisis Kualitas Butir Tes Semester Ganjil Mata Pelajaran IPA Kelas IV Mi Radhiatul Adawiyah. *Jurnal Ilmiah Sekolah Dasar*, 4(1), 86. <https://doi.org/10.23887/jisd.v4i1.23335>
- Munawwaroh, M., Amirullah, M. K., Alfani, M. A. U., Nisa', K., & Sa'idah, N. (2024). Efektivitas Penerapan Media Audio Visual terhadap Pengaruh Motivasi Belajar Pendidikan Agama Islam pada Anak Usia Sekolah Dasar. *Jurnal Basicedu*, 8(1), 286–290. <https://doi.org/10.31004/basicedu.v8i1.6969>
- Nurhayati, S. (2024). Analisis Daya Pembeda Butir Pada Materi Persamaan dan Pertidaksamaan Nilai Mutlak dengan Metode Teori Tes Klasik. *Jurnal MathEdu (Mathematic Education Journal)*, 7(1), 57–63. <https://doi.org/https://doi.org/10.37081/mathedu.v7i1.5681>
- Putri, M. (2020). *Analisis Butir Soal Mata Pelajaran Bahasa Jawa Kelas Iv Sd Negeri Se-Gugus Teuku Umar Kota Tegal Tahun Ajaran 2019/2020*. <https://lib.unnes.ac.id/38571/>
- Rahmaini, A., & Taufiq, A. N. (2018). Analisis Butir Soal Pendidikan Agama Islam Di SMK N 1 Sedayu Tahun Ajaran 2017/2018 (Analisis Tingkat Kesukaran, Daya Pembeda dan Fungsi Distraktor pada Soal Pilihan Ganda Kelas XI). *Jurnal Mudarrisuna*, 8(1), 1–24. <https://www.jurnal.ar-raniry.ac.id/index.php/mudarrisuna/article/view/2787>
- Regeta, N. R., Chintya, R., Nurilah, F., Sujadi, J. A., Khaerullah, M. I., & Karimah, S. (2022). Analisis Butir Soal Pilihan Ganda Pada Materi Plsv Dan Ptlsv Siswa Smp Negeri 2

- Wiradesa. *Prosiding Seminar Nasional Pendidikan Matematika IV (Sandika IV)*, 4(1), 473–480.
- Sanaky, M. M. (2021). Analisis Faktor-Faktor Keterlambatan Pada Proyek Pembangunan Gedung Asrama Man 1 Tulehu Maluku Tengah. *Jurnal Simetrik*, 11(1), 432–439. <https://doi.org/10.31959/js.v11i1.615>
- Sudijono, A. (2017). *Pengantar Evaluasi Pendidikan*. Jakarta: PT RajaGrafindo.
- Suharmi, A. (2010). *Prosedur Penelitian Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta.
- Suharsimi, A. (2019). *Prosedur Penelitian: Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta.
- Supriani, Y., Supriyadi, A., Arifin, B. S., & Ruswandi, U. (2022). Islamic Education in the Merdeka Belajar Curriculum. *Edumaspul - Jurnal Pendidikan*, 6(2), 2780–2787.
- Syamsul Arifin, Nurul Abidin, F. A. A. (2021). Kebijakan Merdeka Belajar dan Implikasinya terhadap Pengembangan Desain Evaluasi Pembelajaran Pendidikan Agama Islam. *Dirāsāt: Jurnal Manajemen Dan Pendidikan Islam*, 7. <https://journal.unipdu.ac.id/index.php/dirasat/article/view/2394/1248>
- Zahra, F. (2022). *Evaluasi Pembelajaran SD/MI* (S. Arif (ed.)). Kediri: CV Kreator Cerdas Indonesia.