

Analisis Soal Sumatif IPA Mengenai Gaya pada Peserta Didik Kelas IV di SD YP Nasional

Leony Eka Pratiwi¹, H. Ruffi'i²

^{1,2} Program Studi Pascasarjana Teknologi Pendidikan Fakultas Pascasarjana Teknologi Pendidikan Universitas PGRI Adi Buana, Surabaya, Indonesia

¹leonypratiwi6@gmail.com

Abstrak

Tujuan dari penelitian ini adalah untuk menganalisis butir soal dalam bentuk pilihan ganda dengan maksud mengevaluasi tingkat validitasnya, mengukur reliabilitasnya, menilai tingkat kesulitannya, serta mengidentifikasi daya pembedanya. Subjek penelitian terdiri dari 15 peserta didik kelas IV di SD YP Nasional. Pendekatan yang diterapkan adalah metode penelitian deskriptif berbasis data kuantitatif. Data dalam penelitian ini dikumpulkan menggunakan instrumen tes berupa 20 soal pilihan ganda yang terkait dengan materi Konsep Gaya. Data dianalisis menggunakan software SPSS. Hasil studi ini mengindikasikan bahwa sebanyak 55% dari soal masuk dalam kategori valid, sedangkan 45% soal termasuk dalam kategori tidak valid. Selain itu, reliabilitas soal dikatakan rendah dengan nilai R sebesar 0,647. Tingkat kesukaran soal menunjukkan bahwa satu soal termasuk dalam kategori sulit, sembilan soal berada pada tingkat kesulitan sedang, dan sepuluh soal termasuk dalam kategori mudah. Adapun dalam hal daya pembeda, terdapat 35% soal yang memiliki daya pembeda baik, 20% soal dengan daya pembeda sedang, dan 45% soal dengan daya pembeda rendah.

Kata Kunci: Soal Pilihan Ganda, SPSS, Konsep Gaya

Pendahuluan

Pembelajaran adalah sebuah proses di mana siswa berinteraksi dengan sumber belajar. Dalam proses pembelajaran, terdapat banyak faktor yang dapat mempengaruhi agar proses pembelajaran tersebut memiliki makna. Beberapa faktor tersebut meliputi peran guru, karakteristik siswa, sarana prasarana, media pembelajaran, lingkungan, dan faktor lainnya. Faktor-faktor ini perlu saling mendukung untuk menciptakan pembelajaran yang berkualitas dan bermakna bagi siswa. Dengan adanya pembelajaran yang berkualitas, diharapkan bahwa tujuan pembelajaran yang telah ditetapkan dapat dicapai oleh siswa. Ketercapaian tujuan pembelajaran ini dapat diukur menggunakan berbagai alat ukur seperti tes, lembar observasi, dan instrumen penilaian lainnya (Muluki, 2020). IPA sebagai mata pelajaran yang mencakup berbagai konsep, salah satunya adalah konsep gaya. Gaya adalah salah satu topik yang penting dalam pemahaman dasar sains, karena itu berkaitan erat dengan pemahaman dasar ilmu pengetahuan alam.

Kegiatan penilaian dalam bidang pendidikan sering disebut dengan istilah evaluasi. Evaluasi merupakan bagian yang tak terpisahkan dari seluruh proses pembelajaran. Regulasi tentang kegiatan evaluasi telah diatur dalam Undang-Undang Republik Indonesia No. 20 Tahun 2003 tentang Sistem Pendidikan Nasional Bab XVI Pasal 58 Ayat 1. Pasal tersebut menyatakan bahwa "Evaluasi hasil belajar peserta didik dilakukan oleh pendidik untuk mengawasi perkembangan, kemajuan, dan perbaikan hasil belajar peserta didik secara berkesinambungan." Oleh karena itu, tujuan dari evaluasi hasil belajar adalah untuk menilai pencapaian kompetensi siswa dan meningkatkan proses pembelajaran, serta memberikan pedoman dalam penyusunan

laporan kemajuan hasil belajar siswa (Undang-Undang Republik Indonesia No. 20 Tahun 2003 Tentang Sistem Pendidikan Nasional, 2003).

Untuk menjalankan kegiatan evaluasi dengan lebih terstruktur, digunakan alat atau teknik penilaian. Dalam konteks pendidikan, alat evaluasi yang digunakan untuk mengumpulkan data bisa berupa tes atau non-tes (Purwanto, 2011). Tes merupakan suatu instrumen atau metode yang digunakan untuk mengidentifikasi atau mengukur sesuatu dengan menggunakan prosedur atau aturan yang telah ditetapkan. Apabila tes yang disusun sesuai dan berkualitas, maka tujuan dari tes tersebut akan tercapai (Arikunto, 2016). Tes adalah suatu teknik untuk melakukan penilaian yang melibatkan serangkaian tugas yang harus dijalankan oleh siswa atau kelompok siswa, yang kemudian akan menghasilkan penilaian (Septiana, 2016).

Evaluasi dalam proses belajar mengajar memiliki peran yang sangat penting dan tak dapat dipisahkan. Seperti yang diungkapkan oleh (Agustiana et al., 2019), evaluasi merupakan tindakan penilaian yang dilakukan setelah melakukan pengukuran. Evaluasi sebagai proses penentuan nilai, apakah pencapaian sudah mencapai standar atau belum (Ariyana, 2019). Untuk memastikan bahwa proses evaluasi berfungsi sesuai dengan tujuan yang ditetapkan, penting untuk memperhatikan kualitas alat evaluasi. Sayangnya, seringkali para praktisi pendidikan di lapangan hanya fokus pada pelaporan hasil evaluasi tanpa memeriksa sejauh mana alat evaluasi yang mereka gunakan berkualitas baik. Alat evaluasi dalam konteks pendidikan bisa berupa tes atau non-tes (Solichin, 2017). Evaluasi merupakan aktivitas pengumpulan data dan informasi tentang kemampuan belajar siswa, yang bertujuan untuk menilai sejauh mana pencapaian target yang telah diprogramkan (Mahirah B., 2017).

Evaluasi proses belajar mengajar dapat memanfaatkan tes yang telah distandardisasi (Standardized test) maupun tes yang dibuat oleh guru sendiri (Teacher-made test). Standardized test adalah tes yang telah melewati proses standardisasi, termasuk pengujian validitas dan reliabilitas, sehingga tes tersebut dianggap valid (tepat) dan reliabel (konsisten) untuk tujuan tertentu dan bagi kelompok tertentu. Pemerintah pusat menggunakan tes standar ini dalam ujian nasional. Sementara itu, tes buatan guru sendiri adalah tes yang disusun oleh guru untuk mengevaluasi proses belajar mengajar. Tes ini sering digunakan di tingkat sekolah dan seringkali digunakan untuk mengukur kemajuan siswa dalam satu kelas atau sekolah tertentu (Harjanti, 2006).

Guru perlu memiliki instrumen penilaian dalam bentuk pertanyaan atau soal-soal untuk melakukan evaluasi, yang mencakup pengujian aspek kognitif, afektif, dan psikomotorik. Biasanya, instrumen penilaian yang digunakan guru untuk mengukur pencapaian peserta didik diperoleh dari berbagai sumber, seperti buku-buku atau koleksi soal-soal ujian. Soal-soal tersebut dapat berupa soal pilihan ganda atau soal uraian. Namun, setelah observasi di lapangan, ditemukan bahwa soal-soal cenderung lebih menekankan pengujian ingatan. Peserta didik perlu diajarkan untuk berpikir secara lebih mendalam dan kritis. Guru dapat melatih peserta didik agar memiliki kemampuan berpikir tingkat tinggi dengan memberikan soal-soal yang mendorong siswa untuk berpikir secara analitis, sintesis, dan evaluatif (Septiana, 2016). Namun, pentingnya pemahaman gaya dalam pembelajaran IPA sering kali diabaikan, dan peserta didik mungkin menghadapi kesulitan dalam memahaminya. Berbagai faktor, seperti kurikulum, metode pengajaran, atau materi ajar yang kurang memadai, dapat menjadi hambatan dalam pembelajaran konsep gaya. Oleh karena itu, perlu dilakukan analisis mendalam terhadap soal-sumatif IPA yang menguji pemahaman konsep gaya pada peserta didik kelas IV di SD YP Nasional.

Menurut (Harjanti, 2006) situasi di lapangan menunjukkan bahwa analisis butir-soal sering diabaikan. Oleh karena itu, informasi mengenai kualitas materi, konstruksi soal, bahasa yang digunakan, validitas, reliabilitas, serta analisis butir soal seperti tingkat kesulitan, daya pembeda,

dan distraktor soal sering kali tidak terukur secara jelas. Analisis butir soal merupakan kegiatan yang penting bagi seorang guru untuk meningkatkan kualitas soal yang dibuat (Sumiati et al., 2018). Kegiatan ini melibatkan pengumpulan, ringkasan, dan pemanfaatan informasi dari jawaban siswa guna membuat keputusan terkait setiap aspek penilaian. Tujuan dari analisis soal adalah untuk mengidentifikasi soal-soal yang berkualitas baik, kurang baik, dan buruk. Melalui analisis soal, kita dapat memperoleh wawasan mengenai kelemahan dalam suatu soal serta petunjuk untuk melakukan perbaikan. Soal yang berkualitas adalah soal yang mampu memberikan informasi yang sesuai dengan tujuan penilaiannya, termasuk dalam hal menilai sejauh mana peserta didik memahami materi yang diajarkan oleh guru (Hasibuan et al., 2013).

Sekolah Dasar YP Nasional merupakan salah satu institusi pendidikan di Indonesia yang bertujuan untuk memberikan pendidikan yang berkualitas kepada peserta didiknya. Kelas IV adalah salah satu tingkatan di Sekolah Dasar yang menjadi tahap penting dalam pembelajaran IPA. Pada tingkat ini, peserta didik diperkenalkan pada konsep-konsep dasar IPA yang lebih kompleks, termasuk konsep gaya. Pemahaman yang kuat tentang gaya sangat penting, karena gaya memainkan peran kunci dalam menjelaskan fenomena-fenomena fisika di dunia sekitar kita. Memahami gaya dapat membantu peserta didik dalam menjelaskan mengapa benda-benda bergerak, mengapa benda tertentu tetap diam, dan mengapa benda-benda berubah arah atau kecepatan. Selain itu, pemahaman gaya juga relevan dalam berbagai aspek kehidupan sehari-hari, seperti transportasi, olahraga, dan konstruksi. Analisis soal-sumatif ini bertujuan untuk mengidentifikasi tingkat kesulitan soal, apakah soal-soal tersebut telah sesuai dengan tingkat pemahaman peserta didik kelas IV, serta untuk memastikan bahwa materi ajar yang disajikan sudah sesuai dengan kurikulum yang berlaku. Hasil analisis ini akan memberikan pandangan yang jelas tentang efektivitas pembelajaran konsep gaya di SD YP Nasional.

Dalam rangka meningkatkan kualitas pendidikan dan memastikan bahwa peserta didik memahami konsep gaya dengan baik, analisis soal-sumatif IPA menjadi langkah awal yang penting. Dengan demikian, diharapkan hasil analisis ini dapat menjadi landasan untuk pengembangan kurikulum, metode pengajaran, serta bahan ajar yang lebih efektif dalam pembelajaran IPA kelas IV. Dengan mempertimbangkan permasalahan yang telah disebutkan sebelumnya, penelitian ini difokuskan pada evaluasi kualitas butir soal yang telah dibuat oleh guru di SD YP Nasional. Setelah itu, dilakukan revisi pada soal-soal yang dinilai kurang baik berdasarkan hasil analisis yang telah dilakukan.

Metode

Penelitian ini mengadopsi metode penelitian deskriptif kuantitatif. Pendekatan kuantitatif bertujuan untuk menggambarkan dan menginterpretasikan objek penelitian sesuai dengan keadaan yang sebenarnya (Sukardi, 2021). Di sisi lain, penelitian kuantitatif adalah jenis penelitian yang fokus pada fenomena objektif yang diselidiki dengan menggunakan pendekatan kuantitatif, seperti penggunaan angka, analisis statistik, struktur data, dan eksperimen yang terkendali (Sukmadinata, 2016). Oleh karena itu, penelitian deskriptif kuantitatif adalah jenis penelitian yang menekankan pada analisis data dengan menggunakan perhitungan angka atau statistik. Peneliti berusaha untuk mengumpulkan data, menjelaskan, menganalisis, dan akhirnya menyimpulkan temuan dari penelitian ini. Langkah-langkah dalam penelitian ini mencakup (1) penyusunan instrumen tes pilihan ganda, (2) pengumpulan data, (3) deskripsi data yang terkumpul, (4) analisis data, dan (5) penyusunan kesimpulan dari hasil analisis.

Penelitian ini dilaksanakan di SD YP Nasional pada bulan Oktober dan melibatkan subjek penelitian yang terdiri dari 15 siswa kelas IV pada tahun ajaran 2023/2024. Teknik pengumpulan data yang digunakan adalah melalui instrumen tes berupa soal pilihan ganda yang terdiri dari 20

butir soal. Data yang digunakan dalam penelitian ini adalah data kuantitatif yang berupa jawaban siswa terhadap soal ujian harian yang menguji pemahaman konsep gaya. Proses analisis data menggunakan bantuan perangkat lunak SPSS. Data yang diperoleh kemudian diolah dengan bantuan SPSS untuk mengevaluasi validitas, reliabilitas, tingkat kesulitan, dan daya pembeda dari instrumen tes yang digunakan.

Hasil

Penelitian ini dilakukan untuk menganalisis kualitas butir soal pilihan ganda yang digunakan dalam tes sumatif pada materi konsep gaya di SD YP Nasional Surabaya. Oleh karena itu, penelitian ini melibatkan uji validitas, uji reliabilitas, uji tingkat kesukaran, dan uji daya pembeda untuk setiap butir soal yang dievaluasi.

Uji Validitas Soal

Uji validitas ini dilakukan dengan menggunakan soal pilihan ganda yang terdiri atas 15 Siswa Kelas 4 di SD YP Nasional. Uji validitas ini menggunakan korelasi. Soal dapat dikatakan valid jika nilai korelasi $r > r_{\text{tabel}}$ dengan hasil yang didapatkan dari 20 butir soal, ada 11 soal yang valid dan 9 soal yang tidak valid. Penjabaran uji validitas soal dengan SPSS sebagai berikut:

Tabel 1. Hasil Uji Validitas Instrumen Tes

Nomor Soal	r-tabel	r-hitung	Status Butir Soal
1.	0.534	0.040	Tidak Valid
2.	0.533	0.041	Tidak Valid
3.	0.791	0.000	Tidak Valid
4.	0.694	0.004	Tidak Valid
5.	0.109	0.700	Valid
6.	- 0.282	0.308	Valid
7.	0.301	0.275	Tidak Valid
8.	0.599	0.018	Tidak Valid
9.	0.000	1.000	Valid
10.	0.266	0.337	Valid
11.	0.000	1.000	Valid
12.	0.272	0.327	Valid
13.	0.067	0.814	Valid
14.	0.422	0.118	Tidak Valid
15.	0.067	0.814	Valid
16.	0.313	0.225	Tidak Valid
17.	0.266	0.337	Valid
18.	0.508	0.053	Tidak Valid
19.	0.107	0.705	Valid
20.	0.214	0.445	Valid

Dengan hasil analisis validitas soal:

Tabel 2. Hasil Analisis Validitas Instrumen Tes

Kategori	Jumlah	Presentase	Nomor soal
Valid	11	55%	5, 6, 9, 10, 11, 12, 13, 15, 17, 19, 20
Tidak Valid	9	45%	1, 2, 3, 4, 7, 8, 14, 16, 18
Jumlah	20	100%	20

Dari tabel di atas, data menunjukkan bahwa 55% dari pertanyaan masuk ke dalam kategori yang valid, sementara 45% sisanya dikategorikan sebagai tidak valid. Karena sebagian besar pertanyaan diklasifikasikan sebagai valid, dapat disimpulkan bahwa tingkat validitas keseluruhan pertanyaan dapat dianggap baik.

Uji Reabilitas

Hasil analisis reliabilitas untuk butir soal pilihan ganda pada materi Konsep Gaya, yang dilakukan dengan menggunakan perangkat lunak SPSS, tercantum dalam tabel berikut.

Tabel 3. Reliability Statistics

Cronbach's Alpha	N of Items
.467	20

Berdasarkan tabel diatas, bahwa uji reabilitas pada siswa yang terdiri dari 15 orang diperoleh nilai 0,467 yang merupakan lebih kecil. Seperti yang diketahui bahwa $R < 0,70$ maka soal tidak reliabel (masih dipertanyakan).

Tingkat Kesukaran

Hasil analisis tingkat kesulitan dari instrumen tes pilihan ganda pada materi Konsep Gaya yang dianalisis menggunakan perangkat lunak SPSS tersedia dalam tabel berikut:

Tabel 4. Hasil Tingkat Kesukaran Instrumen Tes

Nomor Soal	Tingkat Kesukaran	Status Butir Soal
1.	0.47	Sedang
2.	0.80	Mudah
3.	0.67	Sedang
4.	0.47	Sedang
5.	0.60	Sedang
6.	0.67	Sedang
7.	0.73	Mudah
8.	0.80	Mudah
9.	0.47	Sedang
10.	0.20	Sukar
11.	1.00	Mudah
12.	0.60	Sedang
13.	0.80	Mudah
14.	0.73	Sedang
15.	0.80	Mudah
16.	0.87	Mudah
17.	0.80	Mudah
18.	0.67	Sedang
19.	0.93	Mudah
20.	0.93	Mudah

Berdasarkan hasil uji tingkat kesukaran bahwa 1 soal berkategori baik, 9 soal berkategori sedang, dan 10 soal berkategori mudah.

Daya Pembeda

Dengan menggunakan SPSS, hasil daya pembeda instrumen tes pilihan ganda materi Konsep Gaya dapat dilihat pada tabel berikut:

Tabel 5. Hasil Daya Pembeda Instrumen Tes

Nomor Soal	Daya Pembeda	Status Butir Soal
1.	0.534	Baik
2.	0.533	Baik
3.	0.791	Baik
4.	0.694	Baik
5.	0.109	Buruk
6.	-0.282	Buruk
7.	0.301	Sedang
8.	0.599	Baik
9.	0.000	Buruk
10.	0.226	Sedang
11.	0	Buruk
12.	0.272	Buruk
13.	0.067	Buruk
14.	0.422	Baik
15.	0.067	Buruk
16.	0.313	Sedang
17.	0.266	Sedang
18.	0.508	Baik
19.	0.107	Buruk
20.	0.214	Buruk

Maka, jumlah soal yang termasuk kategori daya pembeda baik, cukup, dan buruk adalah sebagai berikut:

Tabel 6. Hasil Analisis Daya Pembeda Instrumen Tes

Jumlah Soal	Kategori Daya Pembeda		
	Baik	Cukup	Buruk
20	7	4	9
Presentase	35%	20%	45%

Hasil evaluasi terhadap daya pembeda, seperti yang tergambar dalam tabel di atas, mengungkapkan bahwa dalam instrumen tes pilihan ganda yang menguji materi Konsep Gaya pada siswa kelas IV pada tahun pelajaran 2023/2024 di SD YP Nasional, 45% atau 9 soal tergolong memiliki daya pembeda yang rendah, 20% atau 4 soal memiliki tingkat daya pembeda yang sedang, dan 35% atau 7 soal dinilai memiliki daya pembeda yang baik.

Pembahasan

Uji Validitas Soal

Validitas adalah ukuran yang mengindikasikan sejauh mana suatu instrumen dapat diandalkan atau sah. Prinsip validitas mengacu pada kemampuan instrumen dalam mengukur atau mengamati hal yang seharusnya diukur. Dengan kata lain, validitas lebih menekankan pada alat pengukuran atau pengamatan itu sendiri. Validitas mencerminkan sejauh mana data yang dicatat oleh peneliti sesuai dengan apa yang sebenarnya terjadi dalam objek penelitian. Dalam hal ini, jika objek penelitian memiliki warna merah, peneliti melaporkan bahwa warna tersebut adalah merah; jika pegawai dalam objek penelitian bekerja keras, peneliti mencatat bahwa pegawai bekerja keras. Jika laporan peneliti tidak mencerminkan realitas objek penelitian, maka data tersebut dianggap tidak valid (Sugiyono, 2017).

Hasil uji validitas menunjukkan bahwa dari seluruh pertanyaan yang dievaluasi, sebanyak 55% dianggap valid, sedangkan 45% sisanya dianggap tidak valid. Proporsi yang lebih besar dari pertanyaan yang dikategorikan sebagai valid mengindikasikan bahwa sebagian besar pertanyaan dalam instrumen penelitian ini memiliki tingkat kevalidan yang baik. Hal tersebut sesuai dengan teori (Sudijono, 2017) "validitas item dari suatu tes adalah ketepatan mengukur yang dimiliki dalam mengukur apa yang seharusnya diukur lewat butir item tersebut". Soal yang valid (70%) berarti butir soal tersebut sudah dapat menjalankan fungsinya yaitu dapat mengukur apa yang seharusnya diukur. Sementara soal yang tidak valid dapat dipengaruhi oleh tiga faktor yang memengaruhi validitas hasil tes. Ketiga faktor tersebut adalah instrumen yang digunakan untuk tes, administrasi dan proses penilaian, serta jawaban yang diberikan oleh siswa (Arifin, 2017). Dengan kata lain, instrumen ini mampu mengukur dengan baik apa yang seharusnya diukur dalam konteks penelitian. Oleh karena itu, keseluruhan tingkat validitas dari pertanyaan-pertanyaan ini dapat dianggap memadai dan sesuai untuk digunakan dalam penelitian ini. Ini berarti bahwa instrumen penelitian tersebut cukup sah dan dapat diandalkan untuk mengumpulkan data yang diperlukan dalam penelitian ini.

Uji Reabilitas

Reliabilitas berasal dari kata "*rely*" yang berarti kepercayaan, dan "*reliable*" yang berarti dapat diandalkan. Kepercayaan terkait dengan keteguhan dan konsistensi. Reliabilitas menunjukkan sejauh mana instrumen mampu mengukur dengan akurat apa yang seharusnya diukur, tingkat ketepatan hasil pengukuran, dan seberapa akurat hasilnya jika diukur ulang. Reliabilitas menggambarkan kemampuan instrumen untuk melakukan pengukuran dengan cermat. Ini merujuk pada akurasi dan ketepatan yang diberikan oleh instrumen dalam proses pengukuran. Instrumen yang dapat diandalkan akan memberikan hasil pengukuran yang sesuai dengan keadaan yang sebenarnya.

Hasil perhitungan koefisien reliabilitas tes dapat diinterpretasikan dengan menggunakan patokan. Jika koefisien reliabilitas (r_{11}) sama dengan atau lebih besar dari 0,70, maka tes hasil belajar yang sedang diuji reliabilitasnya dianggap memiliki reliabilitas yang tinggi (dapat diandalkan). Sebaliknya, jika koefisien reliabilitas (r_{11}) lebih kecil dari 0,70, maka tes hasil belajar yang sedang diuji reliabilitasnya dianggap belum memiliki reliabilitas yang tinggi (tidak dapat diandalkan). Dengan demikian, patokan ini membantu untuk menilai sejauh mana tes dapat diandalkan dalam mengukur hasil belajar. Jika koefisien reliabilitas memenuhi atau melebihi 0,70, maka tes dianggap memiliki reliabilitas yang baik, sementara jika kurang dari 0,70, tes dianggap memiliki reliabilitas yang kurang baik.

Hasil uji validitas menunjukkan bahwa untuk sampel siswa yang terdiri dari 15 orang, koefisien reliabilitas yang diperoleh adalah 0,467. Nilai ini lebih kecil daripada batas standar

reliabilitas yang biasanya ditetapkan pada 0,70. Sehingga, berdasarkan nilai koefisien reliabilitas (R) yang kurang dari 0,70, dapat disimpulkan bahwa soal yang diujikan belum dapat dianggap sebagai reliabel (masih dipertanyakan). Artinya, hasil uji reliabilitas ini menimbulkan keraguan terhadap keandalan instrumen tes yang digunakan. Instrumen tes yang tidak dapat diandalkan mungkin mengakibatkan hasil yang bervariasi atau tidak konsisten saat digunakan pada sampel yang berbeda. Oleh karena itu, perlu dilakukan peninjauan dan perbaikan pada instrumen tes ini untuk meningkatkan reliabilitasnya sebelum digunakan dalam pengukuran atau evaluasi lebih lanjut.

Pada prinsipnya, suatu instrumen tes yang memiliki validitas yang baik pada setiap butir soalnya juga cenderung memiliki tingkat reliabilitas yang tinggi. Ini sejalan dengan teori (Arikunto, 2016) yang menyatakan bahwa tes yang terdiri dari banyak butir soal biasanya lebih valid daripada tes yang hanya terdiri dari beberapa butir soal. Tingkat validitas yang tinggi cenderung menunjukkan tingkat reliabilitas yang juga tinggi, dan semakin panjang tes (dengan lebih banyak butir soal), reliabilitasnya cenderung semakin tinggi. Dengan demikian, dapat disimpulkan bahwa soal sumatif IPA mengenal gaya pada peserta didik kelas IV di SD YP Nasional masih belum dianggap baik. Tes yang valid dan terdiri dari banyak butir soal umumnya memiliki tingkat reliabilitas yang tinggi, sehingga tes ini dianggap baik dari segi reliabilitasnya.

Tingkat Kesukaran

Tingkat kesulitan sebuah soal merujuk pada kemungkinan seorang peserta dapat menjawab soal tersebut dengan benar pada tingkat kemampuan tertentu, yang sering kali diukur dengan indeks tertentu. Indeks kesukaran, juga dikenal sebagai difficulty index, adalah ukuran yang menggambarkan seberapa banyak peserta tes berhasil menjawab soal tersebut dengan benar (Yusdiana et al., 2019). Kualitas butir-butir soal ujian hasil belajar dapat dievaluasi pertama kali berdasarkan tingkat kesulitan atau tingkat kesulitan yang dimiliki oleh masing-masing butir soal tersebut. Butir-butir soal ujian hasil belajar dianggap baik jika tingkat kesulitan mereka tidak terlalu tinggi atau terlalu rendah, dengan kata lain, tingkat kesulitan butir soal tersebut adalah sedang atau cukup. Untuk menghitung indeks kesukaran ini, digunakan program Anates versi 4, dan hasilnya diinterpretasikan dalam tiga kriteria: soal dengan P 0,00 sampai 0,30 dianggap sulit; soal dengan P 0,31 sampai 0,70 dianggap sedang; dan soal dengan P 0,71 sampai 1,00 dianggap mudah. Berdasarkan hasil uji tingkat kesukaran, terdapat 1 soal yang masuk dalam kategori baik, 9 soal dalam kategori sedang, dan 10 soal dalam kategori mudah.

Setelah menganalisis tingkat kesukaran butir soal, tindak lanjut yang dapat diambil sesuai dengan panduan yang disajikan oleh (Sudijono, 2017) adalah sebagai berikut: Pertama, untuk butir soal yang memiliki tingkat kesukaran dalam kategori baik, disarankan untuk menyimpannya dalam bank soal agar dapat digunakan kembali di masa depan. Kedua, untuk butir soal yang tergolong sulit, ada tiga kemungkinan tindakan: pertama, butir soal tersebut dapat dibuang dan tidak digunakan lagi dalam tes hasil belajar berikutnya; kedua, dapat dilakukan penelitian ulang untuk memahami faktor-faktor yang membuat butir soal sulit dijawab oleh peserta tes, dan perbaikan dapat dilakukan dengan menyederhanakan kalimat soal sehingga tidak menimbulkan tafsiran ganda, dan setelah itu, butir soal dapat digunakan kembali dalam tes berikutnya; ketiga, butir soal dapat tetap dipertahankan untuk digunakan dalam tes seleksi yang sangat ketat, di mana sebagian besar peserta tidak akan lulus. Ketiga, untuk butir soal yang termasuk dalam kategori mudah, juga ada tiga kemungkinan tindakan: pertama, butir soal dapat dibuang dan tidak digunakan lagi dalam tes hasil belajar berikutnya; kedua, faktor-faktor yang membuat butir soal tersebut mudah dijawab oleh hampir semua peserta dapat diteliti ulang, dan perbaikan dapat dilakukan dengan memperbaiki opsi jawaban dan membuat kalimat soal menjadi lebih kompleks, dan setelah itu, soal dapat digunakan kembali dalam tes berikutnya; ketiga, butir soal dapat

dipertahankan untuk digunakan dalam tes yang sifatnya lebih longgar, di mana tes tersebut mungkin hanya sebagai formalitas.

Daya Pembeda

Daya pembeda merujuk pada kemampuan suatu butir soal untuk membedakan siswa yang memiliki pemahaman materi tinggi dengan siswa yang memiliki pemahaman materi rendah. Daya pembeda dari suatu soal adalah kemampuan soal tersebut dengan skornya untuk membedakan antara peserta tes yang termasuk dalam kelompok tinggi dan kelompok rendah (Hanifah, 2014). Klasifikasi yang digunakan untuk menafsirkan hasil perhitungan daya pembeda adalah sebagai berikut: Butir soal dengan indeks daya pembeda antara 0,71 hingga 1,00 dianggap sangat baik (*excellent*), antara 0,41 hingga 0,70 dianggap baik (*good*), antara 0,21 hingga 0,40 dianggap cukup (*satisfactory*), dan antara 0,00 hingga 0,20 dianggap buruk (*poor*). Dengan kata lain, jika indeks daya pembeda suatu soal kurang dari 0,20, maka soal tersebut dianggap tidak memadai karena tidak mampu membedakan antara kemampuan peserta tes dalam kelompok tinggi dan kelompok rendah.

Hasil evaluasi daya pembeda mengindikasikan bahwa dalam instrumen tes pilihan ganda yang menguji materi Konsep Gaya pada siswa kelas IV di SD YP Nasional pada tahun pelajaran 2023/2024, terdapat variasi dalam tingkat daya pembeda butir soal. Sebanyak 45% atau 9 soal tergolong memiliki daya pembeda yang rendah. Ini menandakan bahwa sebagian besar soal tidak efektif dalam membedakan antara siswa yang memiliki pemahaman materi tinggi dengan siswa yang memiliki pemahaman materi rendah. Sebanyak 20% atau 4 soal memiliki tingkat daya pembeda yang sedang. Ini menunjukkan bahwa sebagian kecil soal mampu membedakan siswa dengan tingkat pemahaman materi yang berbeda dengan cukup baik. Sebanyak 35% atau 7 soal dinilai memiliki daya pembeda yang baik. Ini berarti bahwa sebagian soal dalam instrumen tes mampu secara efektif membedakan antara siswa yang memiliki tingkat pemahaman materi yang berbeda. Kesimpulannya, perlu perhatian khusus terhadap perbaikan butir soal dengan daya pembeda rendah, agar instrumen tes menjadi lebih efektif dalam membedakan kemampuan siswa. Dengan demikian, evaluasi dan perbaikan terhadap instrumen tes ini dapat meningkatkan keandalan dalam mengukur pemahaman materi siswa.

Butir soal yang memiliki daya pembeda yang baik sekali, baik, dan cukup harus dipertahankan dengan memasukkannya ke dalam bank soal. Sementara itu, butir soal yang memiliki daya pembeda yang sangat jelek dan jelek perlu mengalami perbaikan menyeluruh dengan mencari penyebab kegagalan tersebut. Salah satu langkah yang dapat diambil adalah memperbaiki kejelasan maksud dari soal sehingga tidak membingungkan siswa yang memiliki pemahaman materi yang baik. Butir soal harus mampu mencerminkan adanya perbedaan dalam pemahaman materi antara siswa yang telah menguasainya dengan siswa yang belum atau kurang menguasainya. Sebagaimana dikatakan oleh (Arifin, 2017) perhitungan daya pembeda merupakan pengukuran sejauh mana suatu butir soal mampu membedakan peserta didik yang sudah menguasai materi dengan peserta didik yang belum atau kurang menguasainya berdasarkan kriteria tertentu. Hal ini didukung oleh (Sudijono, 2017) yang menjelaskan bahwa pengetahuan tentang daya pembeda pada setiap butir soal sangat penting, mengingat variasi kemampuan peserta didik yang berbeda.

Berdasarkan pembahasan di atas, dapat disimpulkan bahwa memiliki daya pembeda yang cukup. Sebagian besar butir soal, yaitu sebanyak 20% atau 4 soal memiliki tingkat daya pembeda yang sedang. Sebanyak 35% atau 7 soal dinilai memiliki daya pembeda yang baik. Butir soal dengan daya pembeda yang baik (baik dan sangat baik) sebaiknya dimasukkan ke dalam bank soal tes hasil belajar. Sementara butir soal dengan daya pembeda yang jelek dapat diperbaiki.

Namun, butir soal dengan daya pembeda yang negatif sebaiknya tidak digunakan lagi untuk tes yang akan datang, karena kualitas butir soal tersebut sangat buruk.

Kesimpulan

Berdasarkan hasil penelitian ini, beberapa kesimpulan dapat ditarik. Pertama, hasil uji validitas menunjukkan bahwa sebagian besar pertanyaan dalam instrumen tes dapat dikategorikan sebagai valid (55%), menunjukkan bahwa instrumen ini memiliki tingkat validitas yang baik. Kedua, pada uji reliabilitas, hasil menunjukkan nilai koefisien reliabilitas yang lebih rendah dari batas standar ($0,467 < 0,70$), yang menandakan bahwa instrumen tes masih perlu perbaikan untuk meningkatkan tingkat reliabilitasnya. Selanjutnya, dalam hal tingkat kesukaran soal, sebagian besar soal diklasifikasikan dalam kategori sedang (9 soal) dan mudah (10 soal), dengan hanya satu soal yang masuk dalam kategori baik. Ini menunjukkan variasi dalam tingkat kesulitan soal dalam instrumen tes. Terakhir, hasil evaluasi terhadap daya pembeda soal menunjukkan bahwa sebagian besar soal memiliki daya pembeda yang rendah (45%), dengan sebagian kecil soal yang memiliki tingkat daya pembeda yang baik (35%). Oleh karena itu, perlu perhatian khusus terhadap perbaikan butir soal dengan daya pembeda rendah untuk meningkatkan efektivitas instrumen tes dalam membedakan kemampuan siswa.

Keseluruhan, penelitian ini mengindikasikan bahwa instrumen tes ini memiliki tingkat validitas yang baik, namun perlu perbaikan dalam hal reliabilitas, tingkat kesukaran soal, dan daya pembeda. Hasil penelitian ini memberikan panduan untuk pengembangan instrumen tes yang lebih baik dan dapat diandalkan dalam mengukur pemahaman materi siswa. Berdasarkan hasil analisis di atas, dapat disimpulkan bahwa terdapat perbedaan dalam kualitas soal pilihan ganda yang digunakan dalam tes sumatif materi Konsep Gaya. Meskipun sebagian besar soal dianggap valid, reliabilitasnya rendah. Selain itu, terdapat variasi dalam tingkat kesulitan dan daya pembeda antara soal-soal tersebut. Oleh karena itu, perlu dilakukan perbaikan dalam penyusunan soal untuk memastikan kualitas yang lebih baik dalam pengukuran pemahaman siswa terhadap materi Konsep Gaya di SD YP Nasional Surabaya.

Ucapan Terima Kasih

Saya ingin menyampaikan rasa terima kasih kepada semua individu yang telah berkontribusi dalam penyelesaian artikel ini. Saya mengucapkan terima kasih yang mendalam kepada para pembimbing saya, atas bimbingan dan dukungan tak ternilai selama proses penelitian. Keahlian dan bimbingan mereka sangat berperan dalam membentuk arah penelitian ini.

References

- Agustiana, M., Mayrita, H., & Muchti, A. (2019). Analisis Butir Soal Ulangan Akhir Semester Mata Pelajaran Bahasa Indonesia Kelas XI. *Jurnal Ilmiah Bina Edukasi*, 11(01), 26–35. <https://doi.org/10.33557/jedukasi.v11i01.203>
- Arifin, Z. (2017). *Evaluasi Pembelajaran*. PT Remaja Rosdakarya.
- Arikunto, S. (2016). *Dasar-Dasar Evaluasi Pendidikan*. Bumi aksara.
- Ariyana. (2019). Evaluasi Pembelajaran Bahasa dan Sastra Indonesia. *Prosiding Seminar Nasional Bulan Bahasa (Semiba)*, 55–63. <https://ejournal.unib.ac.id/index.php/semiba>
- Hanifah, N. (2014). Perbandingan Tingkat Kesukaran, Daya Pembeda Butir Soal Dan Reliabilitas Tes Bentuk Pilihan Ganda Biasa Dan Pilihan Ganda Asosiasi Mata Pelajaran Ekonomi. *Sosio E-Kons*, 6(1), 41–55.
- Harjanti. (2006). *Perencanaan Pengajaran*. Rineka Cipta.

- Hasibuan, E. S. B., Sukamto, & Syambasril. (2013). Analisis Soal Ulangan Tengah Semester Bahasa Indonesia Kelas Xii Mas Raudhatul Ulum Meranti. *Jurnal Pendidikan dan Pembelajaran Khatulistiwa*, 2(4).
- Mahirah B . (2017). Evaluasi Belajar Peserta Didik (Siswa). *Idaarah: Jurnal Manajemen Pendidikan*, 1(2). <https://doi.org/10.24252/idaarah.v1i2.4269>
- 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>
- Undang-Undang Republik Indonesia No. 20 Tahun 2003 tentang Sistem Pendidikan Nasional, Pub. L. No. 20 (2003).
- Purwanto. (2011). *Evaluasi Hasil Belajar*. Pustaka Pelajar.
- Septiana, N. (2016). Analisis Butir Soal Ulangan Akhir Semester (UAS) Biologi Tahun Pelajaran 2015/2016 Kelas X Dan XI Pada MAN Sampit. *Edu Sains: Jurnal Pendidikan Sains & Matematika*, 4(2).
- Solichin, M. (2017). Analisis Daya Beda Soal, Taraf Kesukaran, Validitas Butir Tes, Interpretasi Hasil Tes Dan Validitas Ramalan Dalam Evaluasi Pendidikan. *Dirasat: Jurnal Manajemen Dan Pendidikan Islam*, 2(2), 193–213. <https://doi.org/10.26594/dirasat.v2i2.879>
- Sudijono, A. (2017). *Pengantar Evaluasi Pendidikan*. PT. Raja Grafindo.
- Sugiyono. (2017). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Alfabeta.
- Sukardi. (2021). *Metode Penelitian Pendidikan Tindakan Kelas: Implementasi dan Pengembangannya*. PT Bumi Aksara.
- Sukmadinata, N. S. (2016). *Metode Penelitian Pendidikan*. PT. Remaja Rosdakarya.
- Sumiati, A., Widiastuti, U., & Suhud, U. (2018). Workshop Teknik Menganalisis Butir Soal dalam Meningkatkan Kompetensi Guru di SMK Cileungsi Bogor. *Jurnal Pemberdayaan Masyarakat Madani (JPMM)*, 2(1), 136–153. <https://doi.org/10.21009/JPMM.002.1.10>
- Yusdiana, Y., Zamsir, Z., & Kodirun, K. (2019). Kualitas Tes Sumatif Mata Pelajaran Matematika Kelas VIII SMP Negeri 5 Kendari Tahun Ajaran 2016/2017. *Jurnal Penelitian Pendidikan Matematika*, 6(3), 141. <https://doi.org/10.36709/jppm.v6i3.9146>

---Halaman ini sengaja dikosongkan---