

DIDAKTIKA

Jurnal Pendidikan Sekolah Dasar

Volume 3, Nomor 2, 23–34, 2020

Journal homepage: <https://journal.uny.ac.id/index.php/didaktika>



Item Analysis of Force Material Problem in Elementary School

Gibta Paskalin^{1,✉}, Maria Melani Ika Susanti¹

¹Universitas Sanata Dharma, Indonesia

Abstract

Item analysis was conducted to assess the quality of items in the style material in grade IV elementary school. To achieve this goal, four points were examined in this study, including (1) item validity, (2) reliability, (3) difficulty level, and (4) item discrimination power. This research is a quantitative descriptive research. The subjects of this study were 27 grade IV students at Kentungan Elementary School in February 2020. Quantitative data analysis techniques were carried out to determine the validity, reliability, difficulty level, and distinguishing features using the SPSS 17 application program. The results of this study were: (1) there are 15 valid questions in each cycle, (2) the level of reliability in the high category both in cycle I and cycle II, (3) in cycle I, there are 8 easy questions, 6 medium questions, and 1 difficult problem, whereas in cycle II, there are 8 easy questions and 7 moderate questions, (4) Related to the power of distinguishing questions, in cycle I, there are 8 questions in the excellent category, 6 good questions, and 1 sufficient problem, whereas in cycle II, there are 14 good questions and 1 question is enough.

Keywords: *item analysis, validity, reliability, item difficulty level, item discrimination power*

Analisis Butir Soal untuk Materi Gaya di Sekolah Dasar

Abstrak

Analisis butir soal dilakukan untuk mengkaji kualitas butir soal pada materi gaya di kelas IV sekolah dasar. Untuk mencapai tujuan tersebut, empat poin dikaji dalam studi ini, diantaranya (1) Validitas butir, (2) Reliabilitas, (3) Tingkat kesukaran, dan (4) Daya pembeda. Penelitian ini merupakan jenis penelitian deskriptif kuantitatif. Subjek dari penelitian ini adalah 27 peserta didik kelas IV di SD Negeri Kentungan pada bulan february 2020. Teknik analisis data kuantitatif dilakukan untuk mengetahui validitas, reliabilitas, tingkat kesukaran, dan daya pembeda menggunakan program aplikasi SPSS 17. Hasil dari penelitian ini adalah: (1) terdapat 15 soal valid pada masing-masing siklus, (2) tingkat reliabilitas pada kategori tinggi baik pada siklus I maupun siklus II, (3) Pada siklus I, terdapat 8 soal mudah, 6 soal sedang, dan 1 soal sukar, sedangkan pada siklus II, terdapat 8 soal mudah dan 7 soal sedang, (4) Terkait daya pembeda soal, pada siklus I, terdapat 8 soal dalam kategori baik sekali, 6 soal baik, dan 1 soal cukup, sedangkan pada siklus II, terdapat 14 soal baik dan 1 soal cukup.

Kata Kunci: Analisis butir soal, validitas, reliabilitas, tingkat kesukaran, daya pembeda

✉ *Corresponding Author:*

Affiliation Address: Caturtunggal, Kec. Depok, Kabupaten Sleman, Daerah Istimewa Yogyakarta

E-mail: gibtapaskalin@gmail.com

PENDAHULUAN

Evaluasi merupakan salah satu aspek penting yang harus ada dalam kurikulum di sekolah. [Basuki dan Hariyanto \(2014: 9\)](#) mengartikan evaluasi sebagai suatu proses pengumpulan, analisis, dan penafsiran yang sistematis untuk menetapkan sampai sejauh mana peserta didik mencapai tujuan pembelajaran sesuai yang dinyatakan dalam kurikulum. Evaluasi dilakukan untuk mengetahui ketercapaian tujuan pembelajaran oleh peserta didik. Kegiatan evaluasi dilakukan oleh guru dengan memberikan tes pada peserta didik. Tes adalah pengumpulan data untuk mengukur kemampuan peserta didik dalam aspek kognitif atau tingkat penguasaan materi ([Sanjaya, 2011: 99](#)). Tes yang diberikan kepada peserta didik harus mempunyai kualitas yang baik. Kualitas tes sangat diperlukan karena dapat membantu peneliti memilih soal yang terbaik yaitu valid dan *reliable*. Tes dikatakan valid apabila mampu memberikan informasi empirik sesuai dengan apa yang diukur dan dikatakan *reliable* apabila memberikan hasil yang sama pada berkali-kali pengulangan pengukuran. Tes dianalisis secara keseluruhan maupun setiap butir pada soal tes tersebut. Analisis pada butir soal meliputi tingkat kesukaran dan daya pembeda. Analisis butir soal dilakukan untuk mengetahui terlalu mudah atau sulit soal yang dikerjakan peserta didik dan mengetahui kemampuan butir soal membedakan peserta didik yang sudah atau belum menguasai materi.

Berdasarkan hasil wawancara dan penilaian guru pada penilaian tengah semester II peserta didik kelas IV SD Negeri Kentungan materi gaya tahun ajaran 2018/2019 didapatkan data sebagai berikut:

Tabel 1. Data Kondisi Awal Hasil Belajar Peserta Didik

KKM	Rata-rata	Persentase yang Mencapai KKM	Jumlah Peserta Didik
68	65,38	43,20%	26

Berdasarkan tabel di atas diketahui bahwa kondisi awal hasil belajar peserta

didik ditinjau dari nilai penilaian tengah semester II IPA materi gaya tahun ajaran 2018/2019 rendah. Rata-rata hasil belajar IPA materi gaya 26 peserta didik sebesar 65,38. Nilai rata-rata tersebut masih di bawah Kriteria Ketuntasan Minimal (KKM) IPA yaitu 68. Jumlah peserta didik yang mendapatkan nilai lebih dari KKM ada 11 peserta didik dan 15 lainnya mendapatkan nilai kurang dari KKM.

Rendahnya hasil belajar merupakan salah satu dampak dari kurangnya analisis guru terhadap soal yang diberikan kepada peserta didik. Kurangnya pemahaman peserta didik terhadap materi juga menjadi salah satu penyebab rendahnya hasil belajar ([Martini, 2019: 61](#)). Berkaitan dengan hal tersebut peneliti melakukan pengkajian ulang terkait analisis butir soal pilihan ganda materi gaya pada peserta didik kelas IV di SD Negeri Kentungan. Penelitian ini dilakukan untuk menganalisis validitas empiris, reliabilitas, tingkat kesukaran, dan daya pembeda pada butir soal tersebut.

Rumusan masalah dari penelitian ini adalah (1) Sejauh mana validitas butir soal pilihan ganda materi gaya pada peserta didik kelas IV di SD Negeri Kentungan?, (2) Sejauh mana reliabilitas butir soal pilihan ganda materi gaya pada peserta didik kelas IV di SD Negeri Kentungan?, (3) Sejauh mana tingkat kesukaran butir soal pilihan ganda materi gaya pada peserta didik kelas IV di SD Negeri Kentungan?, dan (4) Sejauh mana daya pembeda butir soal pilihan ganda materi gaya pada peserta didik kelas IV di SD Negeri Kentungan?. Penelitian ini bertujuan untuk mengetahui (1) Validitas butir soal pilihan ganda materi gaya pada peserta didik kelas IV di SD Negeri Kentungan, (2) Reliabilitas butir soal pilihan ganda materi gaya pada peserta didik kelas IV di SD Negeri Kentungan, (3) Tingkat kesukaran butir soal pilihan ganda materi gaya pada peserta didik kelas IV di SD Negeri Kentungan, dan (4) Daya pembeda butir soal pilihan ganda materi gaya pada peserta didik kelas IV di SD Negeri Kentungan.

KAJIAN PUSTAKA

Evaluasi

Basuki dan Hariyanto (2014: 9) mengartikan evaluasi sebagai suatu proses pengumpulan, analisis, dan penafsiran yang sistematis untuk menetapkan sampai sejauh mana peserta didik mencapai tujuan pembelajaran sesuai yang dinyatakan dalam kurikulum. Paparan lain dikemukakan oleh Tyler (dalam Arikunto, 2012: 3) bahwa evaluasi merupakan suatu proses untuk menentukan sejauh mana, dalam hal apa, dan bagian mana tujuan pendidikan telah tercapai. Senada dengan paparan tersebut, Brinkerhoff (dalam Widoyoko, 2009: 4) mengemukakan bahwa evaluasi merupakan proses yang menentukan sejauh mana tujuan pendidikan dapat dicapai. Berdasarkan paparan ketiga ahli di atas, peneliti merumuskan bahwa evaluasi merupakan suatu proses pengumpulan, analisis, dan penafsiran untuk menentukan sejauh mana tujuan pendidikan tercapai.

Keefektifan dan keakuratan informasi yang diperoleh sangat tergantung pada kualitas evaluasi yang dilaksanakan (Herwin, 2019: 42). Evaluasi perlu dilaksanakan dengan ketelitian dan pertimbangan. Alat ukur yang efektif menentukan informasi yang diperoleh akurat serta menggambarkan kondisi nyata.

Penilaian

Penilaian merupakan kegiatan yang dilakukan oleh guru untuk memperoleh informasi secara objektif, berkelanjutan, dan menyeluruh tentang proses dan hasil belajar yang dicapai peserta didik, yang hasilnya digunakan sebagai dasar untuk menentukan perlakuan selanjutnya (Depdiknas, dalam Jihad & Abdul, 2012: 54). Paparan lain dikemukakan oleh Sani (2016: 15) bahwa penilaian adalah upaya yang dilakukan melalui pengumpulan data atau informasi yang valid dan *reliable*, dan selanjutnya data atau informasi tersebut diolah sebagai upaya melakukan pertimbangan untuk pengambilan kebijakan. Kemudian, Gronlund dan Linn (dalam Kusaeri & Suprananto, 2012: 8)

mengartikan penilaian sebagai suatu proses yang sistematis dan mencakup kegiatan mengumpulkan, menganalisis, serta menginterpretasikan informasi untuk menentukan seberapa jauh peserta didik mencapai tujuan pembelajaran. Berdasarkan paparan ketiga ahli di atas, peneliti merumuskan bahwa penilaian merupakan pengumpulan data dan informasi tentang proses dan hasil belajar untuk menentukan seberapa jauh peserta didik mencapai tujuan pembelajaran.

Tes

Tes adalah pengumpulan data untuk mengukur kemampuan peserta didik dalam aspek kognitif atau tingkat penguasaan materi (Sanjaya, 2011: 99). Selanjutnya, Purwanto (2009: 63) mengartikan tes sebagai instrumen alat ukur untuk mengumpulkan data dimana dalam memberikan respon atas pertanyaan dalam instrumen. Kemudian, Arikunto (2012: 67) memaparkan bahwa tes merupakan alat atau prosedur yang digunakan untuk mengetahui sesuatu dengan cara dan aturan tertentu. Berdasarkan paparan ketiga ahli di atas, peneliti merumuskan bahwa tes merupakan alat untuk mengukur kemampuan peserta didik dalam aspek kognitif dengan memberikan respon atas pertanyaan dengan cara dan aturan tertentu.

Tes Pilihan Ganda

Tes pilihan ganda terdiri dari keterangan tentang suatu pengertian yang belum lengkap. Untuk melengkapinya harus memilih satu dari beberapa kemungkinan jawaban yang telah disediakan. Tes pilihan ganda juga terdiri atas bagian keterangan dan bagian kemungkinan jawaban. Widoyoko (2009: 59) memaparkan bahwa tes pilihan ganda adalah tes dimana setiap butir soalnya memiliki lebih dari satu alternatif jawaban. Kemungkinan jawaban terdiri atas satu jawaban yang benar, yaitu kunci jawaban dan beberapa pengecoh. Peneliti membagi tes pilihan ganda menjadi dua bagian yang diujikan pada akhir siklus I dan siklus II.

Analisis Butir Soal

Analisis butir soal dirancang dengan tujuan untuk mengetahui cacat dalam butir tes sehingga dapat diperbaiki sebelum digunakan pada tes berikutnya dan untuk menemukan soal yang terlalu mudah atau sukar untuk dijawab sehingga soal-soal itu dapat diganti dengan butir soal yang lain (Basuki & Hariyanto, 2014; Wu, King, Witkiewitz, Racz, & McMahon, 2012). Senada dengan paparan tersebut, Daryanto (2007: 179) menjelaskan bahwa analisis butir soal digunakan untuk mengetahui kekurangan dalam butir tes sehingga dapat dilakukan upaya untuk memperbaiki sebelum digunakan pada tes berikutnya.

Tujuan analisis soal menurut Aiken (dalam Kusaeri & Suprananto, 2012: 163) yaitu: (1) mengkaji dan menelaah setiap butir soal agar diperoleh soal yang bermutu sebelum digunakan, (2) meningkatkan kualitas butir tes melalui revisi atau membuang soal yang tidak efektif, (3) mengetahui informasi diagnostik pada peserta didik apakah mereka telah memahami materi yang diajarkan. Berdasarkan paparan ahli di atas, peneliti merumuskan bahwa analisis butir soal dirancang untuk mengetahui kekurangan dalam butir tes sehingga dapat diperbaiki sebelum digunakan pada tes berikutnya.

Membuktikan Validitas

Subali (2012: 107) mengemukakan bahwa suatu alat ukur dikatakan sah (valid) jika alat ukur tersebut benar-benar mampu memberikan informasi empirik sesuai dengan apa yang diukur. Validitas memiliki nilai dari yang tinggi sampai rendah, semakin tinggi nilai validitas maka semakin baik tingkat validitas (Susetyo, 2011: 88). Tes sebagai salah satu alat ukur hasil belajar dapat dikatakan valid apabila tes itu dapat mengukur hasil belajar yang hendak diukur. Dengan tes yang valid akan menghasilkan data hasil belajar yang valid pula (Widoyoko, 2009; Waddington, 2001). Berdasarkan paparan ahli di atas, peneliti merumuskan bahwa validitas soal berkaitan sejauh mana butir soal dapat mengukur hasil belajar yang hendak diukur.

Peneliti menggunakan format dikotomi untuk menilai jawaban dari setiap soal. Jika jawaban peserta didik benar maka mendapat skor 1 dan jika salah mendapat skor 0. Prosedur analisis validitas ini ada empat cara (Susetyo, 2011: 91) yaitu: (1) prosentase butir yang cocok dengan indikator/tujuan, (2) indeks kecocokan butir tujuan, (3) rasio validitas isi, dan (4) indeks V dari Aiken.

Peneliti tidak menggunakan indeks V dari Aiken untuk menghitung validitas isi karena penelitian ini tidak berskala politomi. Peneliti menghitung validitas isi menggunakan analisis korelasi bivariate dalam program aplikasi *Statistical Package for the Sosial Sciences* (SPSS) 17. Menurut Pidekso (2009: 228) analisis korelasi bivariate digunakan untuk menghitung nilai hubungan antara dua variabel dengan menggunakan metode Pearson (*product moment correlation*) dalam program aplikasi SPSS. Apabila hasil perhitungan $r_{hitung} \geq r_{tabel}$ maka instrumen dinyatakan valid. Sebaliknya, jika $r_{hitung} < r_{tabel}$ maka instrumen dinyatakan tidak valid.

Reliabilitas Butir Soal

Subali (2012: 113) mengemukakan bahwa suatu alat ukur dinyatakan andal (*reliable*) apabila memberikan hasil yang sama pada berkali-kali pengulangan pengukuran. Sejalan dengan Subali, Masidjo (1995: 209) terlebih dahulu memaparkan bahwa suatu tes yang reliabel akan menunjukkan ketepatan dan ketelitian hasil dalam satu atau berbagai pengukuran. Senada dengan kedua pendapat di atas (Kusaeri & Suprananto, 2012; LeBreton, 2008) menjelaskan bahwa reliabilitas skor tes adalah tingkat ketepatan dan keajegan skor tes. Berdasarkan paparan ahli di atas, peneliti merumuskan bahwa reliabilitas tes merupakan tingkat ketepatan dan hasil yang sama pada berkali-kali pengulangan pengukuran.

Pada penelitian ini, peneliti menggunakan *cronbach's Alpha* atau koefisien Alpha untuk mengetahui reliabilitas soal evaluasi siklus I dan siklus II. Kualifikasi reliabilitas menurut Masidjo (1995: 209) tersaji pada tabel 2. berikut ini:

Tabel 2. Kualifikasi Reliabilitas

Koefisien Korelasi	Kualifikasi
0,91 – 1,00	Sangat Tinggi
0,71 – 0,90	Tinggi
0,41 – 0,70	Cukup
0,21 – 0,40	Rendah
Negatif – 0,20	Sangat Rendah

Reliabilitas suatu tes juga dipengaruhi oleh beberapa hal, faktor yang mempengaruhi reliabilitas tes menurut [Uno dan Satria \(2012: 155-156\)](#) yaitu: (1) secara umum jika tes semakin panjang, maka semakin tinggi reliabilitasnya, (2) penyebaran skor yakni semakin besar penyebaran skor, maka akan semakin tinggi perkiraan reliabilitasnya, (3) kesulitan tes yakni umumnya tes yang terlalu mudah atau terlalu sulit akan menyebabkan reliabilitas tes semakin rendah. Hal ini disebabkan terbatasnya penyebaran skor, (4) objektivitas tes yakni tes yang objektivitasnya tinggi memiliki reliabilitas yang lebih tinggi, karena hasil tesnya tidak dipengaruhi oleh prosedur penskoran, dan (5) interval waktu tes yakni tes dengan interval waktu yang pendek menyebabkan koefisien reliabilitas tes besar.

Tingkat Kesukaran Butir Soal

Tingkat kesukaran soal adalah peluang menjawab benar suatu soal pada tingkat kemampuan tertentu yang biasanya dinyatakan dalam bentuk indeks. Semakin besar indeks tingkat kesukaran yang diperoleh dan hasil hitungan, berarti semakin mudah soal tersebut ([Kusaeri & Suprananto, 2012: 175](#)). Selanjutnya, [Arikunto \(2012: 222\)](#) memaparkan bahwa soal yang berkualitas baik adalah soal yang tidak terlalu mudah atau tidak terlalu sulit. Senada dengan pendapat tersebut, [Purwanto \(2009: 10\)](#) memaparkan bahwa butir soal yang baik adalah yang memiliki tingkat kesulitan sedang yaitu tidak terlalu mudah atau tidak terlalu sulit. Berdasarkan paparan beberapa ahli di atas, peneliti merumuskan tingkat kesukaran butir soal merupakan peluang menjawab benar suatu soal yang tidak terlalu mudah atau tidak terlalu sulit. Berikut adalah klasifikasi tingkat kesukaran soal.

Tabel 3. Klasifikasi Tingkat Kesukaran Soal

Indeks Kesukaran	Kategori	Keputusan
0,00 – 0,30	Sukar	Ditolak/direvisi
0,31 – 0,70	Sedang	Diterima
0,71 – 1,00	Mudah	Ditolak/direvisi

Daya Pembeda Butir Soal

[Arikunto \(2018: 235\)](#) mengartikan daya pembeda soal sebagai kemampuan suatu soal untuk membedakan antara peserta didik yang pandai (berkemampuan tinggi) dengan peserta didik yang bodoh (berkemampuan rendah). Senada dengan pendapat tersebut, [Kusaeri dan Suprananto \(2012: 175\)](#) mengungkapkan bahwa daya pembeda soal adalah kemampuan suatu butir soal untuk dapat membedakan antara siswa yang telah menguasai materi dan siswa yang belum menguasai materi yang diujikan. Pemaparan lain dikemukakan oleh [Daryanto \(2007: 183\)](#) bahwa daya pembeda adalah kemampuan soal untuk membedakan peserta didik memiliki kemampuan tinggi dan yang memiliki kemampuan rendah. Berdasarkan paparan para ahli di atas, peneliti merumuskan daya pembeda butir soal sebagai kemampuan soal untuk membedakan peserta didik memiliki kemampuan tinggi dan yang memiliki kemampuan rendah.

Rentang proporsi daya pembeda butir soal ini adalah 0,00 sampai dengan 1,00. Semakin tinggi indeks proporsi daya pembeda butir soal, maka soal semakin mampu membedakan peserta tes yang telah memahami materi dengan peserta tes yang belum. Berikut adalah kriteria daya pembeda butir soal menurut [Arikunto \(2018: 242\)](#) dan [Kusaeri & Suprananto \(2012: 177\)](#):

Tabel 4. Kriteria Daya Pembeda Butir Soal

Daya Pembeda	Keterangan	Keputusan
0,00 – 0,20	Jelek	Direvisi total
0,21 – 0,40	Cukup	Ditolak/direvisi
0,41 – 0,70	Baik	Diterima
0,71 – 1,00	Baik Sekali	Diterima

Manfaat daya pembeda butir soal menurut [Rusdiana & Elis \(2015: 167\)](#) yaitu: (1) untuk meningkatkan mutu setiap butir soal melalui data empirisnya. Berdasarkan

indeks daya pembeda, setiap butir soal dapat diketahui bahwa soal itu baik, direvisi, atau ditolak dan (2) untuk mengetahui seberapa jauh setiap butir soal dapat mendeteksi/membedakan kemampuan siswa, yaitu siswa yang telah memahami atau belum materi yang diajarkan guru.

METODE PENELITIAN

Penelitian ini merupakan jenis penelitian deskriptif kuantitatif. Penelitian deskriptif kuantitatif menekankan pada analisis data dengan menggunakan perhitungan angka atau statistik. Darmawan (2013: 49) memaparkan bahwa tujuan penelitian deskriptif kuantitatif yaitu untuk mendeskripsikan objek/kegiatan yang menjadi perhatian peneliti. Penelitian dilakukan pada bulan Februari 2020. Subjek dari penelitian ini adalah peserta didik kelas IV di SD Negeri Kentungan yang terdiri dari 27 peserta didik. Pengumpulan data dalam penelitian ini menggunakan wawancara dan dokumentasi. Peneliti melakukan wawancara langsung kepada wali kelas IV, sedangkan dokumentasi peneliti gunakan untuk mengumpulkan data kondisi awal hasil belajar peserta didik kelas IV SD Negeri Kentungan. Teknik analisis data kuantitatif dilakukan untuk mengetahui validitas, reliabilitas, tingkat kesukaran, dan daya pembeda menggunakan program aplikasi SPSS 17.

HASIL PENELITIAN DAN PEMBAHASAN

Penelitian ini membahas tentang analisis butir soal pilihan ganda materi gaya pada peserta didik kelas IV di SD Negeri Kentungan. Soal dibagi menjadi dua bagian sesuai dengan pembagian materi gaya, yaitu: (1) siklus 1 terdapat 25 soal mengenai materi pengaruh gaya, gaya otot, dan gaya listrik; (2) siklus 2 terdapat 25 soal mengenai materi gaya gravitasi, gaya magnet, dan gaya gesek. Hasil penelitian disajikan data mengenai validitas, reliabilitas, tingkat kesukaran, dan daya pembeda.

Hasil Analisis Validitas Butir Soal

Pembuktian validitas soal evaluasi peneliti lakukan dengan cara mengujikan di SD Negeri Kentungan pada tanggal 14 Februari 2020. Adapun jumlah responden sebanyak 26 peserta didik kelas VB SD Negeri Kentungan. Peserta didik mengerjakan soal evaluasi siklus I dan siklus II dengan waktu 2x35 menit. Langkah selanjutnya, peneliti menghitung validitas menggunakan korelasi Pearson dalam SPSS 17. Dalam penghitungan pembuktian validitas soal evaluasi siklus I dan siklus II peneliti menggunakan taraf signifikansi 5% dengan r_{tabel} sebesar 0,388. Berikut tabel hasil pembuktian validitas soal evaluasi siklus I.

Tabel 5. Hasil Pembuktian Validitas Soal pada Siklus I

No. Butir	Nilai Korelasi (r_{hitung})	Nilai r_{tabel}	Hasil	Keputusan
1	0,468	0,388	Valid	Dipakai
2	0,544	0,388	Valid	Dipakai
3	0,628	0,388	Valid	Dipakai
4	0,389	0,388	Valid	Dipakai
5	-0,043	0,388	Tidak valid	Tidak dipakai
6	0,293	0,388	Tidak valid	Tidak dipakai
7	0,544	0,388	Valid	Dipakai
8	0,449	0,388	Valid	Dipakai
9	0,227	0,388	Tidak valid	Tidak dipakai
10	-0,278	0,388	Tidak valid	Tidak dipakai
11	0,463	0,388	Valid	Dipakai
12	0,544	0,388	Valid	Dipakai
13	0,234	0,388	Tidak valid	Tidak dipakai
14	0,162	0,388	Tidak valid	Tidak dipakai
15	0,668	0,388	Valid	Dipakai
16	0,093	0,388	Tidak valid	Tidak dipakai
17	0,300	0,388	Tidak valid	Tidak dipakai
18	0,299	0,388	Tidak valid	Tidak dipakai
19	0,427	0,388	Valid	Dipakai
20	0,449	0,388	Valid	Dipakai
21	0,526	0,388	Valid	Dipakai
22	0,637	0,388	Valid	Dipakai
23	0,668	0,388	Valid	Dipakai
24	-0,081	0,388	Tidak valid	Tidak dipakai
25	0,637	0,388	Valid	Dipakai

r_{tabel} untuk $n = 26$

Berdasarkan Tabel 5, terdapat 15 soal valid dan 10 soal tidak valid. Peneliti menggunakan 15 soal valid untuk digunakan sebagai instrumen tes,

sedangkan 10 soal yang tidak valid tidak digunakan.

Berbeda dengan hasil pembuktian validitas soal evaluasi siklus I, hasil pembuktian validitas soal evaluasi siklus II tersaji pada Tabel 6.

Tabel 6. Hasil Pembuktian Validitas Soal pada Siklus II

No. Butir	Nilai Korelasi (r _{hitung})	Nilai r _{tabel}	Hasil	Keputusan
1	0,549	0,388	Valid	Dipakai
2	0,048	0,388	Tidak valid	Tidak dipakai
3	0,436	0,388	Valid	Dipakai
4	-0,035	0,388	Tidak valid	Tidak dipakai
5	0,574	0,388	Valid	Dipakai
6	0,475	0,388	Valid	Dipakai
7	0,279	0,388	Tidak valid	Tidak dipakai
8	0,559	0,388	Valid	Tidak dipakai
9	0,422	0,388	Valid	Dipakai
10	-0,144	0,388	Tidak valid	Tidak dipakai
11	0,305	0,388	Tidak valid	Tidak dipakai
12	0,315	0,388	Tidak valid	Tidak dipakai
13	0,549	0,388	Valid	Dipakai
14	0,549	0,388	Valid	Dipakai
15	0,010	0,388	Tidak valid	Tidak dipakai
16	0,446	0,388	Valid	Dipakai
17	0,475	0,388	Valid	Dipakai
18	0,659	0,388	Valid	Dipakai
19	0,098	0,388	Tidak valid	Tidak dipakai
20	0,539	0,388	Valid	Dipakai
21	0,408	0,388	Valid	Dipakai
22	0,549	0,388	Valid	Dipakai
23	0,539	0,388	Valid	Dipakai
24	0,297	0,388	Tidak valid	Tidak dipakai
25	0,659	0,388	Valid	Dipakai

r_{tabel} untuk n = 26

Berdasarkan Tabel 6, hasil pembuktian validitas soal evaluasi siklus II terdapat 16 soal valid dan 9 soal tidak valid. Agar jumlah soal sama dengan soal evaluasi siklus I, peneliti menggunakan 15 soal valid untuk digunakan sebagai instrumen tes. Peneliti tidak menggunakan soal valid

nomor 8 (1 soal) dan 9 soal tidak valid lainnya.

Hasil Estimasi Reliabilitas

Peneliti menggunakan koefisien Alpha masih dalam SPSS 17 untuk mengetahui reliabilitas soal evaluasi siklus I dan siklus II. Hasil perhitungan reliabilitas soal pilihan ganda siklus I terdapat pada Tabel 7.

Tabel 7. Hasil Reliabilitas Soal Evaluasi Siklus I

Cronbach's Alpha	N	Kualifikasi
0,836	15	Tinggi

Tabel 7 menunjukkan bahwa reliabilitas soal evaluasi siklus I adalah 0,836. Tingkat reliabilitas soal diketahui dengan membandingkan hasil perhitungan *cronbach's Alpha* dengan koefisien korelasi. Maka, 0,836 masuk dalam kualifikasi tinggi. Untuk mengetahui reliabilitas soal evaluasi siklus II, berikut hasil perhitungan yang terdapat pada tabel 8.

Tabel 8. Hasil Reliabilitas Soal Evaluasi Siklus II

Cronbach's Alpha	N	Kualifikasi
0,816	15	Tinggi

Tabel 8 menunjukkan bahwa reliabilitas soal evaluasi siklus II adalah 0,816. Tingkat reliabilitas soal diketahui dengan membandingkan hasil perhitungan *cronbach's Alpha* dengan koefisien korelasi. Maka, 0,816 masuk dalam kualifikasi tinggi.

Hasil Analisis Tingkat Kesukaran Butir Soal

Berikut merupakan hasil analisis tingkat kesukaran butir soal pilihan ganda materi gaya pada siklus:

Tabel 9. Tingkat Kesukaran Soal Evaluasi Siklus I

No	No. Butir	Mean	Kategori	Keputusan
1	1	0,769	Mudah	Ditolak/direvisi
2	2	0,730	Mudah	Ditolak/direvisi
3	3	0,846	Mudah	Ditolak/direvisi
4	4	0,730	Mudah	Ditolak/direvisi
5	7	0,730	Mudah	Ditolak/direvisi
6	8	0,576	Sedang	Diterima
7	11	0,500	Sedang	Diterima
8	12	0,730	Mudah	Ditolak/direvisi
9	15	0,576	Sedang	Diterima

No	No. Butir	Mean	Kategori	Keputusan
10	19	0,307	Sukar	Ditolak/direvisi
11	20	0,423	Sedang	Diterima
12	21	0,615	Sedang	Diterima
13	22	0,807	Mudah	Ditolak/direvisi
14	23	0,576	Sedang	Diterima
15	25	0,807	Mudah	Ditolak/direvisi

Berdasarkan Tabel 9, dari 15 butir soal pilihan ganda siklus I terdapat 8 soal kategori mudah ditolak atau direvisi. Terdapat 6 soal kategori sedang diterima, dan 1 soal sukar ditolak atau direvisi. Untuk mengetahui tingkat kesukaran soal evaluasi siklus II, berikut hasil perhitungan yang terdapat pada Tabel 10.

Tabel 10. Tingkat Kesukaran Soal Evaluasi Siklus II

No	No. Butir	Mean	Kategori	Keputusan
1	1	0,961	Mudah	Ditolak/direvisi
2	3	0,346	Sedang	Diterima
3	5	0,692	Sedang	Diterima
4	6	0,730	Mudah	Ditolak/direvisi
5	9	0,769	Mudah	Ditolak/direvisi
6	13	0,961	Mudah	Ditolak/direvisi
7	14	0,961	Mudah	Ditolak/direvisi
8	16	0,807	Mudah	Ditolak/direvisi
9	17	0,730	Mudah	Ditolak/direvisi
10	18	0,346	Sedang	Diterima
11	20	0,653	Sedang	Diterima
12	21	0,653	Sedang	Diterima
13	22	0,961	Mudah	Ditolak/direvisi
14	23	0,653	Sedang	Diterima
15	25	0,346	Sedang	Diterima

Berdasarkan Tabel 10, dari 15 butir soal pilihan ganda siklus II, terdapat 8 soal kategori mudah ditolak atau direvisi dan 7 soal kategori sedang diterima.

Hasil Analisis Daya Pembeda Butir Soal

Berikut merupakan hasil analisis daya pembeda butir soal pilihan ganda materi gaya pada siklus I.

Tabel 11. Daya Pembeda Soal Evaluasi Siklus I

No	No. Butir	Nilai Korelasi (r_{hitung})	Kategori	Keputusan
1	1	0,468	Baik sekali	Diterima
2	2	0,544	Baik sekali	Diterima
3	3	0,628	Baik sekali	Diterima

No	No. Butir	Nilai Korelasi (r_{hitung})	Kategori	Keputusan
4	4	0,389	Baik sekali	Diterima
5	7	0,544	Baik sekali	Diterima
6	8	0,449	Baik	Diterima
7	11	0,463	Baik	Diterima
8	12	0,544	Baik sekali	Diterima
9	15	0,668	Baik	Diterima
10	19	0,427	Cukup	Ditolak/direvisi
11	20	0,449	Baik	Diterima
12	21	0,526	Baik	Diterima
13	22	0,637	Baik sekali	Diterima
14	23	0,668	Baik	Diterima
15	25	0,637	Baik sekali	Diterima

Berdasarkan tabel 11. Dari 15 butir soal pilihan ganda siklus I terdapat 8 soal baik sekali dengan keputusan diterima, 6 soal baik dengan keputusan diterima, dan 1 soal cukup dengan keputusan ditolak atau direvisi. Untuk mengetahui daya pembeda soal evaluasi siklus II, berikut hasil perhitungan yang terdapat pada tabel 12.

Tabel 12. Daya Pembeda Soal Evaluasi Siklus II

No	No. Butir	Nilai Korelasi (r_{hitung})	Kategori	Keputusan
1	1	0,549	Baik	Diterima
2	3	0,436	Baik	Diterima
3	5	0,574	Baik	Diterima
4	6	0,475	Baik	Diterima
5	9	0,422	Baik	Diterima
6	13	0,549	Baik	Diterima
7	14	0,549	Baik	Diterima
8	16	0,446	Baik	Diterima
9	17	0,475	Baik	Diterima
10	18	0,659	Baik	Diterima
11	20	0,539	Baik	Diterima
12	21	0,408	Cukup	Ditolak/direvisi
13	22	0,549	Baik	Diterima
14	23	0,539	Baik	Diterima
15	25	0,659	Baik	Diterima

Berdasarkan Tabel 12, dari 15 butir soal pilihan ganda siklus II, terdapat 14 soal baik dengan keputusan diterima dan 1 soal cukup dengan keputusan ditolak atau direvisi.

PEMBAHASAN

Penelitian ini dilakukan untuk menganalisis butir soal dengan mengkaji

kualitas butir soal penelitian pada materi gaya kelas IV di SD Negeri Kentungan. Analisis butir soal dirancang untuk mengetahui kekurangan dalam butir tes sehingga dapat diperbaiki sebelum digunakan pada tes berikutnya. Analisis butir soal meliputi: validitas, reliabilitas, tingkat kesukaran, dan daya pembeda.

Validitas memiliki nilai dari yang tinggi sampai rendah, semakin tinggi nilai validitas maka semakin baik tingkat validitas (Susetyo, 2011: 88). Validitas soal berkaitan sejauh mana butir soal dapat mengukur hasil belajar yang hendak diukur. Hasil pengujian pertama menunjukkan bahwa soal evaluasi siklus I dari 25 soal pilihan ganda terdapat 15 soal valid. Sementara itu, soal evaluasi siklus II dari 25 soal pilihan ganda terdapat 16 soal valid. Perbedaan jumlah soal valid pada soal evaluasi siklus I dan siklus II, membuat peneliti tidak menggunakan salah satu soal valid pada siklus II agar keduanya berjumlah masing-masing 15 soal. Soal yang tidak valid tidak digunakan dalam penelitian.

Kusaeri dan Suprananto (2012: 177) menjelaskan bahwa reliabilitas skor tes adalah tingkat ketepatan dan keajegan skor tes. Reliabilitas tes merupakan tingkat ketepatan dan hasil yang sama pada berkali-kali pengulangan pengukuran. Reliabilitas soal evaluasi siklus I adalah 0,836 masuk dalam kualifikasi tinggi, sedangkan reliabilitas soal evaluasi siklus II adalah 0,816 masuk dalam kualifikasi tinggi. Jadi soal evaluasi siklus I dan II yang peneliti gunakan dapat digunakan untuk penelitian selanjutnya.

Tingkat kesukaran butir soal merupakan peluang menjawab benar suatu soal yang tidak terlalu mudah atau tidak terlalu sulit. Semakin besar indeks tingkat kesukaran yang diperoleh dan hasil hitungan, berarti semakin mudah soal tersebut (Kusaeri & Suprananto, 2012: 175). Pada soal evaluasi siklus I terdapat 8 soal kategori mudah, 6 soal kategori sedang, dan 1 soal sukar. Pada soal evaluasi siklus II terdapat 8 soal kategori mudah dan 7 soal kategori sedang.

Kusaeri dan Suprananto (2012: 175) mengungkapkan bahwa daya pembeda soal adalah kemampuan suatu butir soal untuk dapat membedakan antara siswa yang telah menguasai materi dan siswa yang belum menguasai materi yang diujikan. Daya pembeda butir soal merupakan kemampuan soal untuk membedakan peserta didik memiliki kemampuan tinggi dan yang memiliki kemampuan rendah (Hofmeister, 2013). Pada soal evaluasi siklus I terdapat 8 soal baik sekali, 6 soal baik, dan 1 soal cukup. Pada soal siklus II terdapat 14 soal baik dan 1 soal cukup.

KESIMPULAN

Berdasarkan hasil analisis dan pembahasan kesimpulan dari penelitian ini adalah: (1) Validitas butir soal pilihan ganda materi gaya pada peserta didik kelas IV di SD Negeri Kentungan terdapat 15 soal valid pada masing-masing siklus I dan II, (2) Reliabilitas butir soal pilihan ganda materi gaya pada peserta didik kelas IV di SD Negeri Kentungan memiliki tingkat reliabilitas tinggi pada masing-masing siklus I dan II, (3) Tingkat kesukaran butir soal pilihan ganda materi gaya pada peserta didik kelas IV di SD Negeri Kentungan pada siklus I terdapat 8 soal mudah, 6 soal sedang, dan 1 soal sukar, sedangkan pada siklus II terdapat 8 soal mudah dan 7 soal sedang, (4) Daya pembeda butir soal pilihan ganda materi gaya pada peserta didik kelas IV di SD Negeri Kentungan pada siklus I terdapat 8 soal baik sekali, 6 soal baik, dan 1 soal cukup, sedangkan pada siklus II terdapat 14 soal baik dan 1 soal cukup.

DAFTAR PUSTAKA

- Arikunto, S. (2012). *Dasar-dasar evaluasi pendidikan*. Jakarta: PT Bumi Aksara.
- Arikunto, S. (2018). *Dasar-dasar evaluasi pendidikan*. Jakarta: PT Bumi Aksara.
- Basuki, I. dan Hariyanto. (2014). *Asesmen pembelajaran*. Bandung: PT Remaja Rosdakarya.
- Darmawan, R. (2013). *Metode penelitian kuantitatif*. Bandung: PT Remaja Rosdakarya.

- Daryanto, H. (2017). *Evakuasi pendidikan: komponen MKDK*. Jakarta: Rineka Cipta.
- Herwin. (2019). Evaluasi program pembelajaran IPS di Sekolah Dasar Negeri 126 Lagoe. *Didaktika Jurnal Pendidikan Sekolah Dasar*, Vol. 2. No. 2.
- Hofmeister, P., Jaeger, T. F., Arnon, I., Sag, I. A., & Snider, N. (2013). The source ambiguity problem: Distinguishing the effects of grammar and processing on acceptability judgments. *Language and Cognitive Processes*, 28(1-2), 48-87.
- Jihad, A. & Abdul H. (2012). *Evaluasi pembelajaran*. Yogyakarta: Multi Pressindo.
- Kusaeri dan Suprananto. (2012). *Pengukuran dan penilaian pendidikan*. Yogyakarta: Graha Ilmu.
- LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational research methods*, 11(4), 815-852.
- Martini, S. (2019). Peningkatan hasil belajar siswa pada materi energi gerakan melalui model Contextual Teaching and Learning (CTL) pada siswa sekolah dasar. *Didaktika Jurnal Pendidikan Sekolah Dasar*, Vol 2. No. 2.
- Masidjo. (1995). *Penilaian pencapaian hasil belajar siswa di sekolah*. Yogyakarta: Penerbit Kanisius.
- Pidekso, A. (2009). *SPSS 17 untuk pengolahan data statistik*. Yogyakarta: CV Andi Offset.
- Purwanto. (2009). *Evaluasi hasil belajar*. Yogyakarta: Pustaka Pelajar.
- Rusdiana & Elis R. (2015). *Evaluasi pembelajaran*. Bandung: CV Pustaka Setia.
- Sani, R.A. (2016). *Penilaian autentik*. Jakarta: PT Bumi Aksara.
- Sanjaya, W. (2011). *Penelitian tindakan kelas*. Jakarta: Kencana Prenada Media Gr
- Subali, B. (2012). *Prinsip asesmen & evaluasi pembelajaran*. Yogyakarta: UNY Press.
- Susetyo, B. (2011). *Menyusun tes hasil belajar*. Bandung: CV Cakra.
- Uno, H.B. & Satria, K. (2012). *Assessment pembelajaran*. Jakarta: PT Bumi Aksara.
- Waddington, C. (2001). Different methods of evaluating student translations: The question of validity. *Meta: journal des traducteurs/Meta: Translators' Journal*, 46(2), 311-325.
- Widoyoko, E.S. (2009). *Evaluasi dan program pembelajaran*. Yogyakarta: Pustaka Pelajar.
- Wu, J., King, K. M., Witkiewitz, K., Racz, S. J., & McMahon, R. J. (2012). Item analysis and differential item functioning of a brief conduct problem screen. *Psychological assessment*, 24(2), 444.