

ANALISIS KUALITAS INSTRUMEN TES HASIL BELAJAR IPAS MATERI WUJUD ZAT DAN PERUBAHANNYA

Adi Darma Surya¹, Sumarno², Muhtarom³
Universitas PGRI Semarang
adidarmasurya1999@gmail.com

Abstract

The instruments used to measure student learning outcomes, especially in science learning at the elementary school level, are required to be of good quality and have been tested. This study aims to determine the quality of the science achievement test instrument for class IV SD, especially in the matter of matter and its changes based on content validity, reliability, discriminating power and level of difficulty of the test. The subjects of this study were lecturers, class teachers and fourth grade elementary school students. The approach used in this research is descriptive quantitative. Quantitative data analysis techniques were used to determine the quality level of the instrument based on empirical data while qualitative data were used to determine the category level of validity, reliability, discriminatory power and level of test difficulty. The results showed that the quality of the 30-item 100% validity of the test instrument for elementary science class IV material form of matter and its changes with a reliability of 0.901 (very good category). Based on the differentiating power test, there are 1 item in the Very Good category, 16 items in the moderate category, 13 items in the good category. Meanwhile, based on the difficulty index of the test, there are 2 items in the easy category, 27 items in the moderate category, and only 1 item in the difficult category. Thus, overall the test instrument has good quality and is feasible and can be used to measure students' learning outcomes in science learning on material form of matter and its changes

Keywords: *Instruments for Science Learning Outcomes, Validity of Test Instruments, Tests for Science Learning Outcomes*

Abstrak : Instrumen yang digunakan untuk mengukur hasil belajar peserta didik khususnya dalam pembelajaran IPAS di jenjang sekolah dasar diharuskan memiliki kualitas yang baik dan telah teruji. Penelitian ini bertujuan untuk mengetahui kualitas instrumen tes hasil belajar IPAS kelas IV SD khususnya pada materi wujud zat dan perubahannya berdasarkan validitas isi, reliabilitas, daya pembeda dan tingkat kesukaran tes. Subjek penelitian ini adalah dosen, guru kelas dan peserta didik kelas IV SD. Pendekatan yang digunakan dalam penelitian ini yaitu deskriptif kuantitatif. Teknik analisis data kuantitatif digunakan untuk menentukan tingkat kualitas instrumen secara data empiris sedangkan kualitatif untuk menentukan kategori tingkat validitas, reliabilitas, daya beda dan tingkat kesukaran tes. Hasil penelitian menunjukkan bahwa kualitas instrumen tes hasil belajar IPAS SD kelas IV materi wujud zat dan perubahannya yang berjumlah 30 butir soal 100% valid dengan

reliabilitas sebesar 0,901 (kategori sangat baik). Berdasarkan uji daya pembeda terdapat 1 butir kategori Baik Sekali, 16 butir kategori cukup, 13 butir kategori baik. Sedangkan berdasarkan indeks kesukaran tes terdapat 2 butir soal kategori mudah, 27 butir kategori sedang, dan hanya 1 butir soal yang termasuk kategori sukar. Dengan demikian, secara keseluruhan instrumen tes memiliki kualitas yang baik dan layak serta dapat dipergunakan untuk mengukur hasil belajar peserta didik pada pembelajaran IPAS materi wujud zat dan perubahannya.

Kata Kunci: Instrumen Hasil Belajar IPAS; Validitas Instrumen Tes; Tes Hasil Belajar

PENDAHULUAN

Di dalam proses pembelajaran di sekolah dasar terdapat satu komponen yang harus senantiasa dilakukan oleh seorang pendidik yaitu evaluasi pembelajaran. Evaluasi pembelajaran merupakan suatu cara untuk memperoleh dan menyajikan suatu informasi yang digunakan sebagai bahan pertimbangan dalam pengambilan keputusan pembelajaran selanjutnya (Setiyawan & Wijayanti, 2020). Undang-Undang Nomor 20 Tahun 2003 Tentang Sistem Pendidikan Nasional juga menjelaskan bahwa dalam rangka pengendalian mutu pendidikan nasional maka diperlukan evaluasi di setiap jenjang pendidikan sebagai bentuk akuntabilitas penyelenggara pendidikan kepada pihak-pihak yang berkepentingan. Salah satu bentuk evaluasi pembelajaran yang ada di sekolah adalah tes. Tes ini dapat dipergunakan untuk mengukur hasil belajar peserta didik baik sebelum maupun sesudah pembelajaran.

Sebagai alat ukur, suatu tes dapat dikatakan baik jika telah memenuhi kriteria atau prinsip-prinsip utama dalam penyusunan tes. Tujuannya adalah agar hasil pengetesan dengan alat ukur tersebut dapat menggambarkan keadaan yang sebenarnya dari suatu objek yang hendak diukur (Yulianto, 2021). Instrumen tes yang akan digunakan tersebut paling tidak diharuskan telah memenuhi syarat valid dan reliabel serta turut juga memperhatikan daya pembeda dan tingkat kesukarannya (Iskandar & Rizal, 2018).

Serangkaian kegiatan uji validitas, reliabilitas, daya pembeda dan tingkat kesukaran termasuk dalam kegiatan analisis butir soal yang berguna untuk mengidentifikasi mana soal-soal yang termasuk dalam kategori baik, kurang baik dan jelek. Dengan melakukan kegiatan analisis butir soal maka dapat diketahui kelemahan atau kekurangan dari tes yang telah disusun dan dapat ditindaklanjuti dengan mengadakan perbaikan (Farida & Musyarofah, 2021).

Suatu instrumen tes harus diuji validitasnya. Validitas sendiri terbagi menjadi tiga diantaranya validitas isi, validitas kriteria dan validitas konstruk. Sebelum instrumen tes digunakan untuk mengukur hasil belajar maka terlebih dahulu dilakukan validitas isi. Validitas isi dilakukan untuk memastikan bahwa suatu pengukuran telah memasukkan sekumpulan item yang memadai dan mewakili konsep-konsep (Hendryadi, 2017). Validitas isi dilakukan melalui pengujian terhadap kelayakan atau relevansi isi tes oleh *expert judgement* (penilaian ahli).

Instrumen yang baik juga harus dilakukan pengujian reliabilitas. Reliabilitas merupakan koefisien yang menunjukkan suatu instrumen tes dapat dipercaya. Suatu instrumen tes dapat dikatakan reliabel apabila ketika instrumen tersebut digunakan secara berulang-ulang untuk mengukur suatu hal maka hasilnya akan relatif konsisten dan stabil (Amanda et al., 2019). Uji reliabilitas ini dapat dilakukan apabila sebelumnya suatu instrumen tes sudah diuji validitasnya. Dengan demikian apabila suatu instrumen tes banyak terdapat butir soal yang teridentifikasi tidak valid sebaiknya tidak dilakukan uji reliabilitasnya.

Selain dari kedua pengujian di atas, suatu instrumen tes yang sudah disusun juga perlu diidentifikasi daya pembeda dan tingkat kesukarannya. Ketika suatu soal dapat membedakan peserta didik yang mampu (benar-benar menguasai materi yang ditanyakan) dengan peserta didik yang tidak mampu (belum menguasai materi) disebut dengan daya beda (diskriminasi) (Loka Son, 2019). Sedangkan untuk menentukan suatu soal termasuk dalam kategori mudah, sedang dan sulit dikerjakan perlu diuji tingkat kesukarannya. Dengan dilakukan analisis tingkat kesukaran maka akan diperoleh soal-soal yang layak diberikan kepada peserta didik (Magdalena et al., 2021).

Dalam pembelajaran IPAS kelas IV sekolah dasar sangat diperlukan suatu instrumen tes yang telah teruji. Hal ini dikarenakan materi-materi dalam pembelajaran IPAS sangat banyak memuat konsep-konsep penting yang harus dikuasai oleh peserta didik. Untuk itu sebelum soal-soal tersebut digunakan untuk mengukur hasil belajar peserta didik, maka diperlukan beberapa pengujian yang meliputi uji validitas isi (*content validity*), uji reliabilitas, uji daya beda dan tingkat kesukaran. Analisis ini bertujuan untuk memastikan apakah soal-soal tersebut dapat dikatakan layak dan memiliki kualitas yang baik.

Iskandar & Rizal (2018) juga mengungkapkan hal yang serupa bahwa suatu tes yang telah dirancang harus dianalisis kualitasnya untuk mengetahui derajat kualitas tes, baik per

butir soal maupun secara keseluruhan. Dalam penilaian hasil belajar IPAS, suatu tes diharapkan dapat memberikan gambaran perilaku sampel dan dapat menghasilkan nilai yang objektif dan akurat. Apabila suatu tes yang digunakan kurang baik, maka hasil yang didapatkan kurang baik juga. Mengingat pentingnya kualitas suatu instrumen tes hasil belajar, maka tes tersebut harus dianalisis terlebih dahulu kelayakannya sehingga dapat dipergunakan untuk mengukur hasil belajar peserta didik kelas IV SD khususnya pada pembelajaran IPAS materi wujud zat dan perubahannya.

Adapun tujuan penelitian ini adalah untuk mendeskripsikan kualitas dan mengetahui kelayakan instrumen tes hasil belajar IPAS materi wujud zat dan perubahannya yang dilihat berdasarkan validitas isi (content validity), reliabilitas, daya beda dan tingkat kesukaran soal. Butir soal yang dinyatakan layak dan memiliki kualitas yang baik, maka dapat dipergunakan dalam pengukuran dan penilaian hasil belajar peserta didik pada pembelajaran IPAS kelas IV sekolah dasar. Melalui penelitian ini, diharapkan dapat menghasilkan instrumen tes yang dapat mengukur hasil belajar peserta didik dengan akurat pada pembelajaran IPAS kelas IV SD khususnya pada materi wujud zat dan perubahannya.

METODE

Penelitian ini menggunakan pendekatan deskriptif kuantitatif. Metode ini digunakan untuk mengambil data kemudian diolah dan dianalisis agar dapat mengambil kesimpulan (Sugiyono, 2018). Variabel dalam penelitian ini adalah variabel tunggal yaitu analisis kualitas instrumen untuk mengukur hasil belajar peserta didik kelas IV pada pembelajaran IPAS materi wujud zat dan perubahannya.

Subjek dalam penelitian ini adalah dosen, guru kelas dan peserta didik kelas IV SD yang berjumlah 30 orang. Penelitian ini dibagi menjadi 2 tahap. Tahap pertama adalah uji validitas isi dengan melibatkan jumlah rater sebanyak 4 orang yang terdiri dari dua orang dosen PGSD dan dua orang guru kelas sekolah dasar. Tahap kedua adalah uji coba instrumen tes dengan menyebarkan tes di dua sekolah. Data hasil penelitian kemudian dianalisis uji reliabilitas, daya beda dan tingkat kesukarannya. Bentuk tes yang dianalisis berupa tes pilihan ganda (*multiple choice*) yang berjumlah 30 butir soal.

Instrumen tes hasil belajar dalam penelitian ini ditentukan validitas isinya menggunakan formula/indeks Aiken. Instrumen tes hasil belajar yang telah disusun perlu dianalisis validitas isinya. Hal ini dilakukan untuk penyempurnaan item tes dan sekaligus

memastikan apakah tes tersebut dapat mengukur apa yang seharusnya hendak diukur (Ikhsanudin & Subali, 2018). Validitas isi pada penelitian ini dilakukan dengan memberikan skor 1-5 pada setiap butir soal. Hasil rata-rata penskoran oleh 4 orang rater kemudian dianalisis menggunakan formula Aiken (1985) dengan rumus sebagai berikut:

$$V = \sum S / [n (C-1)]$$

$$S = R - Lo$$

Keterangan:

V = indeks Aiken

S = skor yang diberikan oleh penilai dikurangi skor terendah dalam kategori

R = skor yang diberikan oleh penilai

Lo = skor penilaian terendah (1)

C = skor penilaian tertinggi (5)

n = jumlah validator (penilai)

Setelah butir soal dinyatakan telah valid, maka selanjutnya diuji reliabilitasnya. Reliabilitas adalah ketetapan alat evaluasi dalam melaksanakan tugas. Suatu tes dapat dikatakan mempunyai taraf kepercayaan yang tinggi jika tes tersebut dapat memberikan hasil yang tetap. Menurut Arikunto (2013), “reabilitas menunjuk pada suatu pengertian bahwa sesuatu instrumen cukup dapat dipercaya untuk digunakan sebagai alat pengumpulan data karena instrumen tersebut sudah baik”.

Untuk menentukan reliabilitas tes digunakan rumus *Spearman Brown* (Asrul, 2014):

$$r_{11} = \frac{2 \times r^{1/2} \ 1/2}{(1 + r^{1/2} \ 1/2)}$$

Keterangan:

r_{11} = koefisien reliabilitas yang sudah disesuaikan

$r^{1/2} \ 1/2$ = korelasi antara skor-skor setiap belahan tes

Tabel 1. Klasifikasi Reliabilitas

No	Indeks Reliabilitas	Klasifikasi
1.	$0,80 \leq r_{11} < 1,00$	Tinggi
2.	$0,60 \leq r_{11} < 0,80$	Cukup
3.	$0,40 \leq r_{11} < 0,60$	Agak Rendah
4.	$0,20 \leq r_{11} < 0,40$	Rendah
5.	$0,00 \leq r_{11} < 0,20$	Sangat rendah

Sumber: Arikunto (2013)

Analisis tingkat kesukaran soal artinya mengkaji soal-soal tes dari segi kesulitannya sehingga dapat diperoleh soal-soal mana yang termasuk mudah, sedang dan sukar. Rumus yang digunakan untuk mengetahui tingkat/indek kesukaran dari tiap butir soal yaitu:

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

Dengan:

P = Indeks kesukaran

B = Banyaknya siswa yang menjawab soal itu dengan betul

Js = Jumlah seluruh siswa peserta tes

Tabel 2. Interpretasi Tingkat Kesukaran Tes

No	Interval	Klasifikasi
1.	0,00 – 0,30	Soal Sukar
2.	0,31 – 0,70	Soal Sedang
3.	0,71 – 1,00	Soal Mudah

Sumber: Arikunto (2013)

Selanjutnya instrumen tes juga perlu dianalisis daya pembedanya. Analisis daya pembeda yaitu mengkaji butir soal dari segi kesanggupan suatu tes untuk membedakan peserta didik yang termasuk ke dalam kategori lemah dan kategori yang kuat. Untuk menghitung daya beda maka digunakan rumus berikut:

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

Keterangan:

D = besarnya daya beda

J_A = banyaknya peserta kelompok atas

J_B = banyaknya peserta kelompok bawah

B_A = jumlah peserta kelompok atas yang menjawab benar

B_B = jumlah peserta kelompok bawah yang menjawab benar

Tabel 3. Indeks Daya Pembeda Soal

No	Indeks daya beda	Klasifikasi
1.	0,71 - 1,00	baik sekali
2.	0,41 - 0,70	baik
3.	0,21 - 0,40	cukup
4.	0,00 - 0,20	jelek

Sumber: Arikunto (2013)

HASIL

Hasil perhitungan validitas isi (*content validity*) butir-butir instrumen tes hasil belajar IPAS materi wujud zat dan perubahannya menggunakan formula Aiken menggunakan empat orang rater pada setiap aspek dapat dilihat pada tabel berikut ini:

Tabel 4. Hasil Analisis Indeks Aiken Instrumen Tes Hasil Belajar IPAS

Butir Soal	Indeks Aiken	Hasil	Keterangan
1	0,88	0,89	Valid
2	0,88	0,91	Valid
3	0,88	0,9	Valid
4	0,88	0,91	Valid
5	0,88	0,93	Valid
6	0,88	0,93	Valid
7	0,88	0,9	Valid
8	0,88	0,93	Valid
9	0,88	0,9	Valid
10	0,88	0,93	Valid
11	0,88	0,9	Valid
12	0,88	0,9	Valid
13	0,88	0,91	Valid
14	0,88	0,9	Valid

15	0,88	0,89	Valid
16	0,88	0,91	Valid
17	0,88	0,91	Valid
18	0,88	0,89	Valid
19	0,88	0,9	Valid
20	0,88	0,88	Valid
21	0,88	0,89	Valid
22	0,88	0,91	Valid
23	0,88	0,89	Valid
24	0,88	0,93	Valid
25	0,88	0,91	Valid
26	0,88	0,91	Valid
27	0,88	0,88	Valid
28	0,88	0,89	Valid
29	0,88	0,9	Valid
30	0,88	0,91	Valid

Berdasarkan formula Aiken (1985), koefisien validitas isi dari empat orang penilai (rater) dengan lima kategori jawaban dapat dikatakan baik jika memperoleh koefisien validitas minimal 0,88. Hasil validitas isi menunjukkan bahwa setiap butir soal memperoleh hasil lebih dari 0,88. Dengan demikian, semua butir soal telah memenuhi kriteria validitas isi. Artinya, semua butir soal dapat dikatakan layak secara validitas isi dan kemudian bisa diujikan reliabilitasnya.

Butir soal yang telah dinyatakan valid, maka selanjutnya diuji reliabilitas, daya pembeda dan tingkat kesukaran butir soal. Berdasarkan hasil uji coba lapangan dengan jumlah sampel sebanyak 30 orang peserta didik didapatkan hasil sebagai berikut:

Tabel 5. Rangkuman Hasil Uji Coba Instrumen

Reliabilitas	Daya Pembeda				Tingkat Kesukaran		
	Baik Sekali	Baik	Cukup	Jelek	Sukar	Sedang	Mudah
0,901	1	13	16	0	1	27	2

Sumber: Pengolahan Data Penelitian 2023

Berdasarkan tabel di atas, diketahui hasil analisis instrumen tes hasil belajar IPAS dengan jumlah butir soal sebanyak 30 butir memperoleh nilai reliabilitas sebesar 0,901 dengan kategori sangat baik atau tinggi. Jika dilihat dari daya pembeda soal diperoleh butir soal dengan kategori baik sekali sebanyak satu butir, kategori baik 13 butir dan kategori cukup sebanyak 16 butir dan sama sekali tidak terdapat butir soal yang termasuk kategori jelek. Hasil analisis juga menunjukkan bahwa terdapat satu butir soal yang termasuk kategori sukar, 27 butir kategori sedang dan 2 butir kategori mudah

PEMBAHASAN

Penelitian ini menghasilkan instrumen tes hasil belajar yang layak dan memiliki kualitas yang baik sehingga dapat dipergunakan untuk mengukur hasil belajar peserta didik pada pembelajaran IPAS kelas IV SD khususnya pada materi wujud zat dan perubahannya. Instrumen yang telah melalui serangkaian pengujian sekiranya mampu untuk mengatasi permasalahan kelayakan dan kualitas instrumen yang rendah (Setiyawan & Wijayanti, 2020). Apabila suatu instrumen tes memiliki kualitas yang rendah dan tidak layak maka hasil tes tersebut akan menjadi bias dan tidak dapat mengukur kemampuan peserta didik secara akurat.

Terdapat beberapa aspek yang dapat menentukan kualitas suatu instrumen tes hasil belajar dikatakan layak digunakan. Pertama, jika dilihat dari validitas isinya, suatu tes dapat dikatakan memiliki validitas isi yang baik jika nilai indeks Aiken v melebihi tabel nilai V yang sudah ditentukan. Berdasarkan hasil analisis validitas isi dengan jumlah soal sebanyak 30 butir yang dinilai oleh expert judgement sebanyak 4 orang (rater) didapatkan bahwa semua butir soal memiliki nilai indeks Aiken yang melebihi tabel nilai v (0,88). Menurut Aiken (1985), indeks validitas Aiken dinyatakan berada dalam kategori rendah apabila penilaian ahli menghasilkan nilai dibawah 0,4. Suatu instrumen memiliki validitas sedang apabila berada pada rentang 0,4-0,8 dan dapat dikatakan memiliki validitas yang tinggi apabila menghasilkan nilai lebih dari 0,8. Sejalan dengan itu, jika suatu nilai indeks Aiken semakin mendekati 1 maka dapat dinyatakan item soal tersebut semakin baik juga. Hal ini juga mengindikasikan bahwa item tersebut lebih relevan dengan indikator yang sudah ditetapkan sebelumnya (Retnawati, 2016). Dengan demikian, berdasarkan hasil analisis pada tabel 4 diketahui bahwa semua butir soal memiliki validitas isi yang tinggi dan sangat memadai.

Instrumen tes yang telah memenuhi kriteria dalam validitas isi (*Content Validity*) juga diperlukan pengujian lapangan untuk memastikan apakah instrumen tersebut benar-benar layak untuk digunakan. Instrumen tes hasil belajar perlu diuji reliabilitas, daya beda dan tingkat kesukarannya. Hal ini dilakukan mengingat soal-soal yang akan dipergunakan untuk mengukur hasil belajar IPAS harus benar-benar akurat dan memiliki kualitas yang baik.

Uji reliabilitas dalam penelitian ini diperlukan untuk memastikan sejauh mana alat ukur yang sudah dikembangkan dapat dipercaya. Reliabilitas tes ini merupakan ketetapan suatu alat untuk menilai apa yang hendak dinilainya (Magdalena et al., 2021). Hasil penelitian menunjukkan bahwa instrumen tes hasil belajar IPAS dalam penelitian ini menghasilkan nilai realibilitas sebesar 0,91. Jika dilihat tabel 1 maka hasil pengujian instrumen menunjukkan reliabilitas yang sangat tinggi.

Analisis tingkat kesukaran soal diperlukan untuk mengkaji butir-butir soal dilihat dari tingkat kesulitannya. Tingkat kesulitan suatu butir soal dapat dilihat dari seberapa banyak suatu peserta didik dapat menjawabnya. Apabila semakin banyak peserta didik salah dalam menjawab soal maka soal tersebut termasuk kategori sulit atau sukar. Dari 30 butir soal yang diujicobakan didapatkan 2 butir soal berkategori mudah (terdapat pada butir nomor 23 dan 30), 1 butir soal berkategori sukar (terdapat pada butir nomor 11) dan selebihnya termasuk kategori soal dengan tingkat kesukaran sedang. Butir soal yang termasuk dalam kategori sedang dapat dipergunakan pada penilaian hasil belajar peserta didik. Kemudian untuk butir soal yang termasuk kategori sukar, soal tersebut dapat dimungkinkan untuk tidak dipakai atau bisa juga dilakukan pengkajian ulang dengan meneliti atau menelusuri penyebab peserta didik tidak mampu menjawab soal tersebut dengan benar.

Suatu instrumen tes juga perlu diuji daya bedanya. Uji daya beda ini merupakan kemampuan butir soal untuk membedakan mana peserta didik yang memiliki kemampuan tinggi dan peserta didik yang berkemampuan rendah (Nani, 2014). Berdasarkan tabel 5 diketahui bahwa dari 30 butir soal yang diujicobakan terdapat daya beda baik sekali sebanyak 1 butir, daya beda baik 13 butir dan daya beda kategori cukup sebanyak 16 butir. Dengan demikian tidak terdapat butir soal yang memiliki daya beda jelek, sehingga semua butir soal dapat dipergunakan untuk mengukur hasil belajar peserta didik.

Dilihat dari ke empat aspek pengujian yaitu uji validitas isi, uji reliabilitas, uji daya beda dan uji tingkat kesukaran. Maka secara umum seluruh soal dapat dipergunakan untuk

mengukur hasil belajar peserta didik khususnya dalam pembelajaran IPAS materi wujud zat dan perubahannya. Hal ini dikarenakan butir tes paling tidak telah memenuhi sebagian besar dari kriteria kelayakan suatu instrumen. Suatu instrumen dapat dikatakan tidak layak apabila hanya sebagian kecil saja yang memenuhi kriteria kelayakan yang sudah ditetapkan. Oleh sebab itu, instrumen dalam penelitian ini sudah memenuhi sebagian kriteria minimal sehingga dapat dipergunakan untuk mengukur hasil belajar baik yang dilaksanakan oleh pendidik maupun peneliti itu sendiri.

KESIMPULAN

Berdasarkan hasil dan pemaparan di atas, diketahui bahwa dari 30 butir soal yang dikembangkan secara umum instrumen tersebut dapat dikatakan memiliki kualitas yang baik berdasarkan hasil uji validitas isi, reliabilitas, daya beda dan tingkat kesukaran. Dari segi validitas isi, sebanyak 30 butir soal dinyatakan valid dengan tingkat validitas isi yang baik. Hasil uji validitas isi ini merupakan hasil skor dari *expert judgement* yang kemudian dianalisis menggunakan indeks Aiken. Instrumen tes juga dilakukan uji reliabilitas untuk memastikan bahwa instrumen tersebut dapat dipergunakan secara terus menerus (dapat dipercaya). Hasil penelitian juga menunjukkan instrumen tes sudah memiliki reliabilitas yang sangat baik dengan nilai 0,901. Selain itu instrumen tes hasil belajar juga dilakukan analisis daya beda dan tingkat kesukaran. Berdasarkan hasil analisis daya beda terdapat 1 butir kategori Baik Sekali, 16 butir kategori cukup, 13 butir kategori baik. Sedangkan berdasarkan indeks kesukaran tes terdapat 2 butir soal kategori mudah, 27 butir kategori sedang, dan hanya 1 butir soal yang termasuk kategori sukar. Dengan demikian, dapat diartikan bahwa instrumen tes hasil belajar ini memiliki kualitas yang baik dan dapat dikatakan layak digunakan untuk mengukur hasil belajar peserta didik kelas IV SD pada pembelajaran IPAS khususnya materi wujud zat dan perubahannya.

Untuk kedepannya, soal ini dapat dipergunakan sebagaimana mestinya sehingga dapat menghasilkan penilaian hasil belajar yang lebih akurat. Bagi peneliti lain sekiranya dapat mengembangkan indikator soal sehingga dapat memperluas dan menambah bank soal IPAS kelas IV SD khususnya pada materi wujud zat dan perubahannya.

DAFTAR PUSTAKA

- Aiken, L. R. (1985). *Educational and Psychological Measurement*. Pearson.
- Amanda, L., Yanuar, F., & Devianto, D. (2019). Uji Validitas dan Reliabilitas Tingkat Partisipasi Politik Masyarakat Kota Padang. *Jurnal Matematika UNAND*, 8(1), 179. <https://doi.org/10.25077/jmu.8.1.179-188.2019>
- Arikunto, S. (2013). *Prosedur Penelitian: Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta.
- Asrul. (2014). *Evaluasi Pembelajaran*. Bandung: Citapustaka Media.
- Undang-undang Nomor 20 Tahun 2003 Tentang Sistem Pendidikan Nasional, (2022).
- Farida, & Musyarofah, A. (2021). Validitas dan Reliabilitas dalam Analisis Butir Soal. *Al-Mu'arrif: Jurnal Pendidikan Bahasa Arab*, 1(1), 34–44. <https://jurnal.lp2msasbabel.ac.id/index.php/AL-MUARRIB>
- Hendryadi, H. (2017). Validitas Isi: Tahap Awal Pengembangan Kuesioner. *Jurnal Riset Manajemen Dan Bisnis (JRMB) Fakultas Ekonomi UNLAT*, 2(2), 169–178. <https://doi.org/10.36226/jrmb.v2i2.47>
- Ikhsanudin, & Subali, B. (2018). Content validity analysis of first semester formative test on biology subject for senior high school. *Journal of Physics: Conference Series*, 1097(1). <https://doi.org/10.1088/1742-6596/1097/1/012039>
- Iskandar, A., & Rizal, M. (2018). Analisis kualitas soal di perguruan tinggi berbasis aplikasi TAP. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 22(1), 12–23. <https://doi.org/10.21831/pep.v22i1.15609>
- Loka Son, A. (2019). Instrumentasi Kemampuan Pemecahan Masalah Matematis: Analisis Reliabilitas, Validitas, Tingkat Kesukaran Dan Daya Beda Butir Soal. *Gema Wiralodra*, 10(1), 41–52. <https://doi.org/10.31943/gemawiralodra.v10i1.8>
- Magdalena, I., Fauziah, S. N., Faziah, S. N., & Nupus, F. S. (2021). Analisis Validitas, Reliabilitas, Tingkat Kesulitan dan Daya Beda Butir Soal Ujian Akhir Semester Tema 7 Kelas III SDN Karet 1 Sepatan. *BINTANG: Jurnal Pendidikan Dan Sains*, 3(2), 198–214.
- Nani, H. (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), 46.
- Retnawati, H. (2016). Proving Content Validity Of Self-Regulated Learning Scale (The Comparison Of Aiken Index And Expanded Gregory Index). *Research and Evaluation in Education*, 2(2), 155–164.
- Rizky Ananda Setiyawan, & Palupi Sri Wijayanti. (2020). Analisis Kualitas Instrumen Untuk Mengukur Kemampuan Pemecahan Masalah Siswa Selama Pembelajaran Daring Di Masa Pandemi. *Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika*, 1(2), 130–139. <https://doi.org/10.46306/lb.v1i2.26>
- Sugiyono. (2018). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Yulianto, H. (2021). Analisis Indeks Aiken untuk Mengukur Validitas Isi Instrumen Komitmen Tugas Bermain Sepakbola. *Jurnal Pendidikan Jasmani Indonesia*, 17(1), 19–23.