

PENGEMBANGAN INSTRUMEN ASESMEN SUMATIF BERBASIS HOTS PADA MATERI WUJUD ZAT DAN PERUBAHANNYA KELAS IV SEKOLAH DASAR

Arma Thiana¹, Kartono², Dian Miranda³

^{1,2,3} PGSD FKIP Universitas Tanjungpura

1armathn08@gmail.com, 2kartono@fkip.untan.ac.id,

3dian.miranda@fkip.untan.ac.id

ABSTRACT

This research is This study aims to produce a HOTS-based summative assessment instrument product on the material of the state of matter and its changes for grade IV that is feasible in terms of content validity, question readability, empirical validity, reliability, discriminating power, level of difficulty, and effectiveness of distractors. The research method used is research and development with a qualitative and quantitative descriptive approach. Data collection techniques through interviews, observations, and documentation to obtain qualitative data, and through tests and questionnaires to obtain quantitative data. The study was conducted at SD Negeri 66, Pontianak City District. The results showed that the validation of the material expert's assessment was 99% with a very feasible category, the assessment of the evaluation expert was 94% with a very feasible category, and the assessment of the language expert was 98% with a very feasible category. The readability of the questions from student responses was 89% with a very feasible category. The empirical validity test was 100% with a valid category. The multiple choice reliability test was 0.780 with a high reliability category and the description was 0.591 with a moderate reliability category. The discriminatory power test was 8% in the excellent category, 44% in the good category, and 48% in the adequate category. The difficulty level test was 16% in the easy category, 76% in the moderate category, and 8% in the difficult category. The distractor effectiveness test was 97% effective.

Keywords: *summative assesment instruments, hots, states of matter and their changes, elementary school*

ABSTRAK

Penelitian ini bertujuan untuk menghasilkan produk instrumen asesmen sumatif berbasis *HOTS* pada materi wujud zat dan perubahannya kelas IV yang layak ditinjau dari validitas isi, keterbacaan soal, validitas empiris, reliabilitas, daya pembeda, tingkat kesukaran, dan efektivitas pengecoh. Metode penelitian yang digunakan adalah *research and development* dengan pendekatan deskriptif kualitatif dan kuantitatif. Teknik pengumpulan data melalui wawancara, observasi, dan dokumentasi untuk memperoleh data kualitatif, serta melalui tes dan angket untuk memperoleh data kuantitatif. Penelitian dilakukan di SD Negeri 66 Kecamatan Pontianak Kota. Hasil penelitian menunjukkan bahwa validasi dari penilaian ahli materi sebesar 99% dengan kategori sangat layak, penilaian dari ahli evaluasi sebesar 94% dengan kategori sangat layak, dan penilaian dari ahli bahasa sebesar 98% dengan kategori sangat layak. Keterbacaan soal dari respon peserta didik sebesar 89% dengan kategori sangat layak. Uji validitas empiris sebesar 100% dengan kategori valid. Uji reliabilitas pilihan ganda sebesar 0,780 dengan kategori

reliabilitas tinggi dan uraian sebesar 0,591 dengan kategori reliabilitas sedang. Uji daya pembeda sebesar 8% dengan kategori baik sekali, 44% dengan kategori baik, dan 48% dengan kategori cukup. Uji tingkat kesukaran sebesar 16% dengan kategori mudah, 76% dengan kategori sedang, dan 8% dengan kategori sukar. Uji efektivitas pengecoh sebesar 97% efektif.

Kata kunci: instrumen asesmen sumatif, hots, wujud zat dan perubahannya, sekolah dasar

A. Pendahuluan

Asesmen sumatif merupakan jenis asesmen yang dilakukan untuk memastikan ketercapaian tujuan pembelajaran pada akhir lingkup materi, akhir semester, dan akhir jenjang pendidikan (kemendikbudristek, 2024). Kurikulum merdeka menekankan pentingnya mengembangkan kemampuan peserta didik dalam berpikir tingkat tinggi guna mempersiapkan peserta didik untuk menghadapi tantangan dalam kehidupan sehari-hari. Penilaian yang berfokus pada kemampuan berpikir tingkat tinggi (*Higher Order Thinking Skill/ HOTS*) bertujuan untuk mengukur serta mengembangkan kemampuan peserta didik secara mendalam melalui analisis dan pemecahan masalah.

Berpikir tingkat tinggi sangat penting untuk membantu mengembangkan kemampuan menalar, mengolah dan menyajikan informasi sehingga adanya kesiapan dalam menghadapi tantangan dengan

kreatif dan solutif (Yuliani et al., 2024). Untuk menghadapi masalah kompleks dalam kehidupan sehari-hari, perlu adanya solusi yang kreatif. Dalam hal ini, kemampuan berpikir tingkat tinggi (*HOTS*) berperan dalam membuat peserta didik terlatih dan siap dalam menghadapi tantangan secara kritis dan sistematis. Selain itu, *HOTS* dapat meningkatkan kemampuan peserta didik dalam memahami materi yang kompleks melalui keterampilan dalam memecahkan masalah, menganalisis, dan mengambil keputusan. Kemampuan berpikir tingkat tinggi peserta didik dapat diukur melalui evaluasi pembelajaran. Untuk itu, instrumen yang didesain haruslah memuat butir soal yang *HOTS*.

Instrumen penilaian sumatif yang efektif harus mampu mengukur prestasi siswa dalam berpikir tingkat tinggi. Instrumen penilaian sumatif yang dirancang untuk mengukur keterampilan berpikir tingkat tinggi (*HOTS*) siswa dapat menciptakan

situasi di mana siswa harus menerapkan pengetahuan dalam konteks dunia nyata. Pengembangan pertanyaan HOTS membutuhkan penyediaan stimulus kontekstual yang relevan dengan masalah di lingkungan sekitar (Hanifah, 2019). Ada dua aspek penting dari pertanyaan HOTS: berpikir kritis dan kreativitas. Berpikir kritis berkaitan dengan kemampuan untuk menganalisis dan mengevaluasi suatu masalah (Kartono et al., 2025). Sementara itu, berpikir kreatif berkaitan dengan kemampuan untuk menciptakan, yang mencakup aktivitas seperti membuat dan menyimpulkan (Primayana, 2019). Siswa akan menjadi kreatif ketika mereka menerima dorongan untuk memecahkan suatu masalah (Miranda, 2018).. Selain mengukur keterampilan berpikir tingkat tinggi, instrumen yang dikembangkan harus memberikan informasi yang akurat untuk mencerminkan prestasi dan kemampuan siswa secara keseluruhan.

Melalui wawancara langsung yang dilakukan dengan guru kelas IV B menunjukkan bahwa dalam implementasi soal *HOTS* , peserta didik masih mengalami kesulitan

dalam mengerjakan soal yang berarti bahwa peserta didik mengalami kesulitan dalam mengembangkan kemampuan berpikir tingkat tinggi. Berdasarkan hasil analisis soal yang digunakan guru kelas IVA dan IV B menunjukkan bahwa terbatasnya pertanyaan yang mendorong peserta didik untuk berpikir kritis, kurangnya variasi soal yang mendorong peserta didik untuk mengeluarkan ide, dan soal seringkali tidak memberikan stimulus yang merangsang peserta didik untuk berpikir kritis dan memecahkan masalah. Selain itu, pengukuran *HOTS* kurang optimal, dimana terdapat soal *HOTS* yang masih berfokus pada penguasaan fakta dan konsep, bukan pada kemampuan analisis, evaluasi, dan mencipta yang merupakan indikator dalam soal *HOTS* .

Hasil penelitian yang dilakukan oleh Suhartono et al., (2020) menunjukkan bahwa implementasi penilaian yang mengukur kemampuan dalam berpikir tingkat tinggi belum maksimal, sebagian besar guru mengukur kemampuan peserta didik hanya pada tingkat pemahaman yaitu pada tingkat C1, C2, dan C3. Selain itu, penelitian yang dilakukan oleh

Ningsi & Shaleh (2024) menunjukkan bahwa terdapat berbagai tantangan dalam mengimplementasikan asesmen berbasis *HOTS*, dimana peserta didik mengalami kesulitan dalam mengembangkan kemampuan berpikir tingkat tinggi (*HOTS*) dan kemampuan guru yang kurang dalam membedakan antara soal *HOTS* dan bukan *HOTS* sehingga kesulitan dalam merancang instrumen penilaian berbasis *HOTS*. Tantangan-tantangan yang dihadapi dalam implementasi asesmen berbasis *HOTS* mempengaruhi kualitas dan kuantitas instrumen yang dikembangkan. Kualitas dan kuantitas suatu instrumen asesmen berbasis *HOTS* dapat diketahui melalui berbagai tahap uji validitas. Selain itu, kualitas dan kuantitas instrumen tersebut juga dapat dilihat secara langsung melalui proporsi soal yang berada pada level kognitif berpikir tingkat tinggi (*HOTS*). Kualitas dan kuantitas suatu instrumen yang didominasi oleh soal *LOTS* cenderung lebih rendah dibandingkan dengan instrumen yang didominasi oleh soal *HOTS*. Hal ini tercermin dalam penelitian yang dilakukan oleh Asti et al., (2024) yang menunjukkan bahwa proporsi soal *HOTS* pada instrumen

asesmen sebanyak 33.3% jauh lebih rendah dibandingkan dengan proporsi soal *LOTS* sebanyak 66.6% dari sekolah yang diteliti.

Berdasarkan kondisi tersebut, perlu adanya solusi dengan mengembangkan instrumen asesmen sumatif yang dapat mengukur kemampuan berpikir tingkat tinggi (*HOTS*) dan teruji keakuratannya. Instrumen asesmen berbasis *HOTS* mendukung pengukuran mendalam terhadap kemampuan peserta didik dalam berpikir kritis dan memecahkan masalah pada berbagai bidang, termasuk pada konsep sains yang kompleks. Salah satu bentuk instrumen asesmen sumatif yang efektif untuk mencapai tujuan dari pengukuran tersebut adalah tes tertulis. Instrumen asesmen sumatif dalam bentuk tes tertulis memungkinkan penilaian yang lebih efektif dengan adanya kriteria yang lebih jelas dan berfokus pada hasil, mengukur berbagai kemampuan kognitif, serta pelaksanaan yang lebih efisien. Untuk mengembangkan soal *HOTS*, perlu adanya pemahaman terhadap konsep *HOTS* seperti karakteristik keterampilan berpikir tingkat tinggi. Memolo, (2019) menyatakan bahwa karakteristik soal

HOTS meliputi menganalisis, mengevaluasi dan mencipta, soal tidak rutin, relevan dengan masalah pada konteks kehidupan sehari-hari, jawaban terdiri dari berbagai solusi, soal memiliki kompleksitas, memungkinkan perspektif yang beragam, dan terdapat stimulus yang berbasis masalah. Sani & Prayitno, (2020, h.99) menyatakan bahwa soal *HOTS* memiliki kriteria utama yaitu kontekstual, mencakup aspek berpikir kritis, dan menyajikan stimulus.

Berdasarkan masalah yang telah dipaparkan, maka dianggap perlu dilakukan penelitian *Research and Development (R&D)* dengan judul “Pengembangan Instrumen Asesmen Sumatif Berbasis *HOTS* Pada Materi Wujud Zat dan Perubahannya Kelas IV SD Negeri 66 Kecamatan Pontianak Kota”. Tujuan dari penelitian ini adalah untuk menghasilkan instrumen penilaian sumatif berbasis *HOTS* yang layak berdasarkan pengujian validitas isi, menghasilkan instrumen penilaian sumatif berbasis *HOTS* yang efektif dan mudah dipahami berdasarkan respons siswa terhadap keterbacaan pertanyaan, dan mengevaluasi hasil

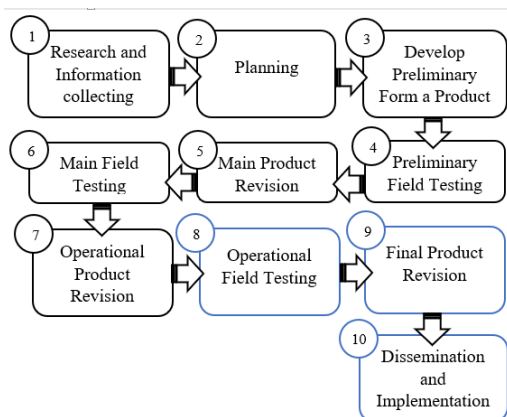
pengujian empiris instrumen penilaian sumatif berbasis *HOTS*.

B. Metode Penelitian

Penelitian ini menggunakan metode penelitian dan pengembangan (*R&D*). Penelitian Pengembangan (*Research and Development*) merupakan jembatan penghubung antara penelitian dasar dengan penelitian terapan yang bertujuan untuk menghasilkan produk melalui proses menemukan, mengembangkan dan memvalidasi produk secara sistematis (Sugiyono, 2023; h.14). Penelitian ini menggunakan pendekatan deskriptif kualitatif dan deskriptif kuantitatif (*mixed method*) untuk memperoleh hasil secara kuantitatif dan kualitatif melalui proses pengumpulan data dan analisis data, mengintegrasikan temuan, serta menarik kesimpulan.

Penelitian ini mengadopsi penelitian yang dikembangkan oleh Borg and Gall. Model yang dikemukakan Borg and Gall terdiri dari 10 langkah, yaitu *research and information collecting*, *planning*, *develop preliminary form a product*, *preliminary field testing*, *main product revision*, *main field testing*, *operational product revision*, *operational field testing*, *final product*

revision, dissemination and implementation.



Gambar 1. Tahap Pengembangan Borg & Gall

Dari sepuluh langkah yang dikemukakan Borg and Gall, penelitian ini hanya melibatkan tujuh langkah. Langkah ke tujuh sebagai langkah terakhir yang digunakan dalam penelitian ini yaitu *Operational Product Revision*. Setelah produk dikembangkan dan diuji coba, evaluasi setelah pengujian sudah cukup untuk memenuhi tujuan dalam penelitian (Sugiyono, 2019). Untuk itu langkah ketujuh sudah memenuhi beberapa kriteria yang diperlukan dalam penelitian ini yang meliputi uji validitas isi, keterbacaan soal, dan uji empiris. Langkah-langkah pengembangan dalam penelitian ini yaitu: (1) *Research and information collecting*: dilakukan penelitian dan pengumpulan data melalui wawancara dan analisis soal guru. (2) *Planning*: dilakukan desain produk. (3) *Develop preliminary form a*

product: dilakukan pengembangan produk awal dan uji validitas isi produk. (4) *Preliminary field testing*: dilakukan uji coba skala kecil, (5) *Main product revision*: dilakukan revisi berdasarkan hasil uji coba skala kecil (6) *Main field testing*: dilakukan uji coba skala terbatas. (7) *Operational product revision*: dilakukan revisi berdasarkan hasil uji coba skala terbatas.

Sumber data dalam penelitian ini meliputi sumber data primer dan sekunder. Sumber data primer merujuk pada data yang dikumpulkan secara langsung dari sumber pertama, yaitu guru dan peserta didik kelas IV di SD Negeri 66 Kecamatan Pontianak Kota, serta para ahli yang menjadi validator produk yang dikembangkan. Peserta didik kelas IV yang dipilih mencakup kelas IVA dan kelas IVB yang memiliki kemampuan akademik yang relatif setara.

Pengumpulan data dalam penelitian ini dilakukan melalui wawancara, dokumentasi, dan angket. Wawancara dalam penelitian ini dilakukan dengan guru kelas IVB sebagai narasumber. Wawancara dilakukan secara terstruktur menggunakan pedoman wawancara. Dokumentasi yang dikumpulkan dalam penelitian ini berupa instrumen

asesmen sumatif pada lingkup materi wujud zat dan perubahannya yang mencakup kisi-kisi dan soal telah digunakan oleh guru kelas IVA dan IVB di SD Negeri 66 Kecamatan Pontianak Kota. Dan angket dalam penelitian ini diberikan kepada validator ahli materi, evaluasi, dan bahasa untuk mendapatkan informasi atau data mengenai tingkat validitas isi produk berupa instrumen asesmen sumatif berbasis *HOTS* . Selain itu, angket juga diberikan kepada peserta didik untuk memperoleh respon peserta didik terkait keterbacaan instrumen asesmen sumatif berbasis *HOTS* yang digunakan.

Instrumen yang digunakan untuk mengumpulkan data dalam penelitian ini yaitu pedoman wawancara, angket validasi produk, dan angket respon peserta didik. Pedoman wawancara dijadikan sebagai panduan yang memuat pertanyaan-pertanyaan terkait topik penelitian. Angket validasi produk digunakan untuk mengumpulkan informasi mengenai validitas isi produk dari aspek materi, evaluasi, dan bahasa. Angket respon peserta didik digunakan untuk memperoleh respon peserta didik terhadap keterbacaan soal.

Teknik analisis data yang digunakan dalam penelitian ini menggunakan pendekatan kualitatif dan kuantitatif. Analisis data kualitatif dilakukan setelah memperoleh saran atau masukan dari validator. Sedangkan analisis data kuantitatif dilakukan dengan metode statistik untuk mengetahui kualitas produk yang dikembangkan, meliputi validitas isi, respon peserta didik, uji empiris yang mencakup validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan efektivitas pengecoh.

1 Analisis Validitas Isi

Analisis validitas isi menggunakan skala likert dengan skor 1 sampai 5 yang menggambarkan gradasi dari sangat positif sampai sangat negatif. Rumus yang digunakan untuk memperoleh rata-rata menurut Arikunto (2016, h.276) sebagai berikut.

$$\pi = \frac{\sum x}{n \times a}$$

Keterangan :

π = Skor rata-rata setiap aspek

$\sum x$ = Jumlah total nilai jawaban responden

n = Jumlah responden

a = Jumlah butir instrumen

Persentase hasil dihitung dengan rumus:

$$P = \frac{\sum x}{\sum xi} \times 100\%$$

Keterangan :

P = Besar persentase

$\sum x$ = Jumlah total skor soal yang diperoleh

$\sum xi$ = Jumlah skor maksimum

Selain itu, dilakukan *percentage agreement* untuk menentukan kesepakatan antar validator. Instrumen asesmen sumatif berbasis *HOTS* akan disetujui jika nilainya $\geq 75\%$ (Borich, 2016). Rumus yang digunakan sebagai berikut.

$$\text{Percentage of agreement} = \left(1 - \frac{A - B}{A + B}\right) \times 100\%$$

Keterangan :

A : Frekuensi terbesar

B : Frekuensi terkecil

2 Analisis Respon Peserta Didik

Pada angket respon peserta didik menggunakan skala likert dengan 5 kategori penilaian dalam bentuk *checklist*. Analisis data hasil respon peserta didik menggunakan rumus rata-rata sebagai berikut.

$$\pi = \frac{\sum x}{n \times a}$$

Persentase produk diperoleh dengan menggunakan rumus sebagai berikut.

$$P = \frac{\sum x}{\sum xi} \times 100\%$$

Tabel 1. Kategori Tingkat Respon Peserta Didik

Skor Nilai (%)	Kelayakan
< 21 %	Sangat Tidak Layak
21 – 40 %	Tidak Layak
41 – 60 %	Cukup Layak
61 – 80 %	Layak
81 – 100 %	Sangat Layak

3 Analisis Uji Empiris

Uji validitas dianalisis dengan berbantuan SPSS versi 26.

Tabel 2. Interpretasi Validitas Butir Soal

Range Validitas	Kategori
$r_{pbi} > r_t$	Valid
$r_{pbi} = r_t$	Tidak Valid
$r_{pbi} < r_t$	Tidak Valid

Uji daya pembeda dianalisis dengan berbantuan SPSS versi 26.

Tabel 3. Interpretasi nilai reliabilitas

Range Reliabilitas	Kategori
$0,80 < r \leq 1,00$	Reliabilitas sangat tinggi
$0,60 < r \leq 0,80$	Reliabilitas tinggi
$0,40 < r \leq 0,60$	Reliabilitas sedang
$0,20 < r \leq 0,40$	Reliabilitas rendah
$-1,00 \leq r \leq 0,20$	Reliabilitas sangat rendah (tidak reliabel)

Uji daya pembeda dianalisis dengan berbantuan SPSS versi 26.

Tabel 4. Kriteria Daya Pembeda

Range Daya Pembeda	Kategori
0,71 – 1,00	Baik Sekali (<i>Excellent</i>)
0,41 – 0,70	Baik (<i>Good</i>)
0,21 – 0,40	Cukup (<i>Satisfactory</i>)
0,00 – 0,20	Jelek (<i>Poor</i>)

(Sumber : Arikunto, 2018: h.242)

Uji tingkat kesukaran dianalisis dengan berbantuan SPSS versi 26.

Tabel 5. Kriteria Tingkat Kesukaran Butir Soal

Nilai Interval	Kategori
0,71 – 1,00	Mudah
0,31 – 0,70	Sedang
0,00 – 0,30	Sukar

(Sumber : Arikunto, 2018: h.235)

Uji Efektivitas pengecoh dianalisis dengan berbantuan *microsoft excel*.

Tabel 6. Kriteria Pengecoh

Kategori Pengecoh	Nilai Butir Soal
$> 0,05$	Efektif
$< 0,05$	Tidak Efektif

C. Hasil Penelitian dan Pembahasan

Produk awal yang dikembangkan berupa instrumen asesmen sumatif berbasis *HOTS* pada materi wujud zat dan perubahannya untuk kelas IV sekolah dasar yang mencakup kisi-kisi soal, lembar soal, kunci jawaban dan rubrik penilaian. Soal yang dikembangkan berjumlah 25 butir soal yang terdiri dari 20 butir soal pilihan ganda dan 5 butir soal uraian.

Tabel 7. Produk Instrumen Asesmen Sumatif Berbasis *HOTS*

Kisi-Kisi Soal

[illegible]

Soal

PEMERINTAH KOTA PONTIANAK
SD NEGERI 66 KECAMATAN PONTIANAK KOTA
Jalan Ujung Pandang Gg. Hamura Pontianak Kota
Kota Pontianak, Kalimantan Barat 78118
Telp. (0561) 586874 NPSN 30105121 NSS 101136005066
Website : - Email : sdn66pontianakkota@gmail.com

Mata Pelajaran : IPAS
 Bab : 2. Wujud Zat dan Perubahannya
 Hari / Tanggal :

Nama :
 Kelas : 4
 Waktu : 90 Menit

Petunjuk :

1. Berdoalah sebelum mengerjakan soal!
2. Tulislah identitas dengan lengkap!
3. Soal terdiri dari 20 butir dengan tipe pilihan ganda dan 5 butir soal uraian.
4. Skor untuk pilihan ganda dengan jawaban benar adalah 1 (satu) dan untuk jawaban salah atau tidak menjawab adalah 0 (nol)
5. Kerjakan soal dengan jujur dan teliti
6. Periksa kembali lembar jawabannmu sebelum dikumpulkan.

Selamat Mengerjakan

I. Berilah tanda (X) pada salah satu huruf "a", "b", "c", dan "d" pada jawaban yang paling tepat!

1. Air sebagai salah satu wujud zat cair memiliki volume yang tetap namun bentuknya berubah mengikuti wadahnya. Kegiatan yang menunjukkan sifat zat cair tersebut adalah....
- membuka keran air hingga air keluar dengan cepat tanpa henti
 - menuang satu gelas air penuh ke gelas yang kecil sampai tumpah
 - menuang satu gelas air ke teko yang ukurannya lebih besar dari gelas
 - menuang setengah dari satu botol berisi air penuh ke dalam gelas kecil



Berdasarkan gambar, pernyataan yang tepat mengenai massa adalah....

- a. balon B memiliki massa yang sama dengan balon A
b. balon A memiliki massa, sedangkan balon B tidak ada massa
c. balon B memiliki massa yang lebih besar dibandingkan balon A
d. balon A memiliki massa yang lebih besar dibandingkan balon B

Kunci Jawaban dan Rubrik Penilaian

Pedoman Penskoran Asesmen Sumatif Berbasis *HOTS*

A. Kunci Jawaban Soal Pilihan Ganda

No. Soal	Jawaban
1	c. memotong satu gelas air ke toko yang ukurannya lebih besar dari gelas
2	d. Isheq. A. mengukir, musa yang lebih besar dibandingkan balon B
3	b. (7, 4, 5), dan 83
4	a. memiliki bentuk dan volume tetap meskipun di tempat tertutup
5	d. ketika balon diletakkan terus menerus, udara akan pecah
6	c. saat Dina menyempatkan parfum ke udaranya, udaranya pecah menyebarkan
7	b. 11 dan 31
8	a. memotong puding dan memasukkan air hingga memendih
9	c. mengkilatkan
10	a. air mengisap ketika berada pada suhu tinggi, sehingga pakaian menjadi kering
11	a. letifan! lili beradu pada suhu dingin
12	d. uap air sedang sekitar ketebal Nenehingga dengan permukaan bolu yang dingin
13	c. koin memiliki suhu yang sama dengan suhu udara
14	d. pengaliran ruangan mengisap langsung menjadi gas tanpa melalui proses memisalkan
15	a. berubah menjadi air
16	d. semakin tinggi suhu, semakin cepat air akan paman pendinginan
17	a. es batu mencair tanpa proses memisalkan
18	d. suhu yang rendah akan mempercepat proses pembekuan
19	b. jalan raya akan cepat kering saat matahari bersinar setelah hujan
20	a. proses ini merupakan proses pendinginan

$$\text{Nilai} = \frac{\text{jumlah benar}}{\text{jumlah soal}} \times 100$$

B. Rubrik Penilaian Soal Uraian

Mata Pelajaran :	IPAS
Kelas :	4 (Empat)
Fase B	
Indikator	1. Memahami proses pembentukan wujud dan dalam kehidupan sehari-hari a. Diketahui soal, peserta didik mampu <i>mengidentifikasi</i> asal usul materi yang ada di alam b. Diketahui soal, peserta didik mampu <i>mengidentifikasi</i> proses perubahan wujud dan berakibatkan fenomena yang terjadi dalam kehidupan sehari-hari c. Diketahui gambar, peserta didik mampu <i>mengidentifikasi</i> perubahan wujud yang terjadi d. Diketahui gambar, peserta didik mampu <i>mengidentifikasi</i> proses perubahan wujud dan dalam kehidupan sehari-hari e. Diketahui teks tentang fenomena yang terjadi dalam kehidupan sehari-hari, peserta didik mampu <i>mengidentifikasi</i> perubahan wujud yang akan terjadi dari proses perubahan wujud

No		Kriteria		Skor	Bobot
1	Mendeskripsikan tentang wujud zat pada dasar ilmu	Jawaban sesuai dengan konsep tentang wujud, kelipatan, jalin dan bentuk partikel suatu zat		4	
		Jawaban sesuai dengan konsep, kelipatan, jalin dan bentuk partikel suatu zat		2	
2	Menggunakan simbol kimia untuk mendeskripsikan fenomena yang terjadi dalam kehidupan sehari-hari	Jawaban sesuai simbol kimia sesuai dengan nilai dan terjemah Indonesian		4	20
		Jawaban sesuai dengan konsep, kelipatan, jalin dan bentuk partikel suatu zat		2	
3	Memprediksi perubahan wujud yang terjadi	Jawaban sesuai konsep, namun tidak sesuai dengan partikel suatu zat		2	20
		Jawaban tidak sesuai dengan partikel suatu zat		1	
3	Memprediksi perubahan wujud yang terjadi	Jawaban sesuai dengan konsep, kelipatan, jalin dan bentuk partikel suatu zat		4	20
		Jawaban sesuai dengan konsep, kelipatan, jalin dan bentuk partikel suatu zat		2	

1. Uji Validitas Isi

Uji validitas isi dilakukan oleh beberapa validator, yaitu oleh ahli materi, ahli evaluasi, dan ahli bahasa. Setiap aspek pada instrumen asesmen sumatif berbasis *HOTS* yang diuji validitasnya, divalidasi oleh 2 ahli agar memperoleh perbandingan dari hasil validasi oleh kedua ahli. Hasil uji validitas isi disajikan secara berturut-turut pada Tabel 8, Tabel 9, dan Tabel 10.

Tabel 8. Hasil Validasi ahli materi

Ahli	Jumlah Skor	Nilai	%	Kategori
Ahli 1	38	4,75	95%	Sangat Layak
Ahli 2	39	4,90	98%	Sangat Layak
Persentase rata-rata			97%	Sangat Layak
<i>Percentage of Agreement (PoA)</i>			98%	Sangat Baik

Data hasil validasi oleh ahli materi pada Tabel 8 menunjukkan bahwa persentase rata-rata yang diperoleh sebesar 97% dengan kategori sangat layak, dan *Percentage of Agreement (PoA)* sebesar 98% dengan kriteria sangat baik.

Tabel 9. Hasil Validasi Ahli Evaluasi

Ahli	Jumlah Skor	Nilai	%	Kategori
Ahli 1	60	4,61	92%	Sangat Layak
Ahli 2	62	4,76	95%	Sangat Layak
Persentase rata-rata			94%	Sangat Layak
<i>Percentage of Agreement (PoA)</i>			97%	Sangat Baik

Data hasil validasi oleh ahli evaluasi pada Tabel 9 menunjukkan bahwa persentase rata-rata yang diperoleh sebesar 94% dengan kategori sangat layak, dan *Percentage of Agreement (PoA)* sebesar 97% dengan kriteria sangat baik.

Tabel 10. Hasil Validasi Ahli Bahasa

Ahli	Jumlah Skor	Nilai	%	Kategori
Ahli 1	20	5,00	100%	Sangat Layak
Ahli 2	19	4,75	95%	Sangat Layak
Persentase rata-rata			98%	Sangat Layak
<i>Percentage of Agreement (PoA)</i>			97%	Sangat Baik

Berdasarkan data hasil validasi oleh ahli bahasa pada Tabel 10 menunjukkan bahwa persentase rata-rata yang diperoleh sebesar 98% dengan kategori sangat layak, dan *Percentage of Agreement (PoA)* sebesar 97% dengan kriteria sangat baik.

Validitas isi bertujuan untuk menentukan ketepatan suatu alat tes yang akan diujikan dan relevansinya terhadap aspek yang akan dikur (Nurgiyantoro, 2016:173). Validitas isi dalam penelitian ini ditinjau dari aspek materi, evaluasi, dan bahasa.

Instrumen asesmen sumatif berbasis *HOTS* pada materi wujud zat dan perubahannya telah divalidasi oleh dua ahli materi yang memiliki

kualifikasi dalam menilai kesesuaian butir soal dengan lingkup materi dan tingkat kognitif kelas IV Sekolah Dasar. Penilaian aspek materi dilakukan terhadap kesesuaian materi dan ketercapaian karakteristik *HOTS* pada soal yang terdiri dari 20 butir soal pilihan ganda dan 5 butir soal uraian. Berdasarkan validasi yang telah dilakukan diperoleh persentase rata-rata dari ahli pertama dan kedua sebesar 97% dengan kriteria sangat layak, dan *Persentage of Agreement (POA)* 98% dengan kriteria sangat baik. Penilaian aspek materi dilakukan terhadap kesesuaian materi dan ketercapaian karakteristik *HOTS*. Instrumen dapat dikatakan valid apabila setiap butir tes mencerminkan keseluruhan materi yang diujikan dan seharusnya dikuasai (Rahayu et al., 2016). Materi dalam instrumen yang dikembangkan telah sesuai dengan capaian pembelajaran serta jenjang sekolah dan tingkat kelas. Selain itu, butir soal telah sesuai dengan indikator soal, butir soal telah sesuai dengan dimensi kognitif berpikir tingkat tinggi dan indikator *HOTS*, serta terdapat stimulus yang mendorong peserta didik untuk berpikir tingkat tinggi.

Aspek evaluasi pada instrumen asesmen sumatif berbasis *HOTS* mencakup aspek konstruk dan *HOTS* yang divalidasi oleh dua ahli evaluasi. Berdasarkan analisis hasil penilaian oleh dua ahli validasi evaluasi, diperoleh rata-rata persentase dari ahli pertama dan kedua 94% dengan kriteria sangat layak, dan *Persentage of Agreement (POA)* 97% dengan kriteria sangat baik. Konstruksi soal yang baik yang baik adalah yang pokok soal, petunjuk, butir soal, dan berbagai elemen dalam soal tersebut tersusun dengan jelas dan terperinci (Rahmawati & Trimulyono, 2022). Setiap soal disusun secara sistematis dengan konstruk yang baik yang terdiri dari kesesuaian perumusan butir soal, petunjuk pada soal jelas, kesesuaian stimulus yang digunakan, dan kesesuaian penulisan soal dengan standar penulisan soal yang tepat. Selain itu, soal telah disusun berbasis *HOTS* dengan adanya indikator berpikir tingkat tinggi dan sesuai dengan karakteristik soal *HOTS* yaitu kontekstual, terdapat stimulus, dan mendorong peserta didik untuk berpikir kritis sehingga soal dapat mengukur kemampuan berpikir tinggi peserta didik dengan optimal.

Aspek bahasa divalidasi oleh dua ahli bahasa, diperoleh persentase rata-rata dari ahli pertama dan kedua sebesar 98% dengan kriteria sangat layak, dan *Persentase of Agreement (POA)* 97% dengan kriteria sangat baik. Bahasa yang baik dilihat dari tata bahasa dan kejelasan bahasa dalam soal (Rahmawati & Trimulyono, 2022). Kaidah bahasa yang digunakan dalam instrumen telah sesuai dengan kaidah ejaan bahasa indonesia, baik itu dari segi penulisan ejaan, tanda baca, dan struktur kalimat. Bahasa yang digunakan jelas dan tidak bersifat ambigu sehingga informasi pada instrumen soal *HOTS* dapat tersampaikan dengan tepat dan efektif. Selain itu, bahasa yang digunakan sudah baku atau tidak menggunakan bahasa daerah, dan pilihan jawaban telah sesuai dengan kaidah penulisan yang berlaku, mulai dari frasa yang digunakan, huruf yang digunakan, bahkan urutannya yang dilihat berdasarkan panjang pendek kalimat yang digunakan.

Validitas isi yang baik ditandai dengan adanya kesesuaian antara butir tes dengan kriteria tertentu (Nurgiyantoro, 2016; h.174). Hasil validitas isi produk berdasarkan aspek materi, evaluasi, dan bahasa tersebut

menunjukkan bahwa produk yang dikembangkan telah sesuai dengan indikator yang ditetapkan. Ketiga aspek yang divalidasi berada pada kategori sangat layak dan kelayakan produk tidak terlepas dari masukan validator untuk perbaikan, sehingga secara keseluruhan instrumen asesmen sumatif berbasis *HOTS* sangat layak untuk diujikan.

2. Respon Peserta Didik Terhadap Keterbacaan Soal

Hasil uji respon peserta didik terhadap keterbacaan soal disajikan pada Tabel 11 sebagai berikut.

Tabel 11. Hasil Respon Peserta Didik terhadap Keterbacaan Soal

Responden	Jumlah Skor	%	Kategori
F	48	87%	Sangat Layak
FA	53	96%	Sangat Layak
NN	44	80%	Layak
R	45	82%	Sangat Layak
D	50	91%	Sangat Layak
A	52	95%	Sangat Layak
RA	51	93%	Sangat Layak
SAA	48	87%	Sangat Layak
AF	48	87%	Sangat Layak
SN	50	91%	Sangat Layak
Persentase rata-rata	489	89%	Sangat Layak

Data hasil analisis keterbacaan instrumen asesmen sumatif berbasis *HOTS* diperoleh melalui respon

peserta didik yang melibatkan sepuluh peserta didik kelas IVA SD Negeri 66 Kecamatan Pontianak Kota.

Respon peserta didik diperlukan untuk memastikan produk layak digunakan ditinjau dari keterbacaannya. Respon peserta didik diperoleh melalui angket respon peserta didik terhadap keterbacaan instrumen asesmen sumatif berbasis *HOTS*. Angket respon peserta didik memuat beberapa aspek, diantaranya yaitu kejelasan bahasa, kejelasan huruf, kejelasan angka, kejelasan gambar, dan kejelasan grafik.

Berdasarkan analisis hasil respon peserta didik diperoleh persentase rata-rata keterbacaan instrumen asesmen sumatif berbasis *HOTS* sebesar 89% dengan kategori sangat layak. Keterbacaan yang baik tercermin dari struktur kalimat yang jelas serta kata-kata yang mudah dipahami oleh pembaca sehingga informasi tersampaikan dengan baik (Arvianto, 2016). Hasil respon peserta didik menunjukkan bahwa soal-soal telah disusun dengan jelas dan dapat dipahami dari segi bahasa, huruf, angka, gambar, dan grafik.

3. Uji Empiris

Uji empiris dilakukan untuk mendapatkan hasil mengenai kualitas

butir soal pada instrumen asesmen sumatif berbasis *HOTS* yang dikembangkan. Uji coba dilakukan melalui dua tahap uji, yaitu uji coba skala kecil pada 10 peserta didik kelas IVA dan uji coba skala terbatas pada 29 peserta didik kelas IVB. Analisis uji empiris mencakup validitas (Tabel 12), reliabilitas (Tabel 13), daya pembeda (Tabel 14), tingkat kesukaran (Tabel 15), dan efektivitas pengecoh (Tabel 16).

Tabel 12. Hasil Uji Validitas

Tahap Uji	Valid		Tidak Valid	
	Jumlah Skor	%	Jumlah Skor	%
Skal Kecil	17	68%	8	32%
Skala Terbatas	25	100%	0	0%

Instrumen asesmen sumatif berbasis *HOTS* dinyatakan valid apabila telah memenuhi beberapa kriteria validitas, baik secara isi maupun konstruk. Butir soal dinyatakan valid apabila signifikansi $< 0,05$ dan nilai r hitung $> r$ tabel. Suatu instrumen dikatakan valid jika mampu mengukur aspek yang sedang diukur, dengan acuan nilai r -hitung lebih tinggi dari r -tabel (Anam, 2017). Data yang diperoleh dari uji coba skala kecil dan uji coba skala terbatas dianalisis secara kuantitatif menggunakan SPSS versi 26 sebagai alat bantu pengolahan data. Nilai r -tabel pada uji

coba skala kecil sebesar 0,632 dan pada uji skala terbatas sebesar 0,367 yang disesuaikan dengan jumlah subjek uji di masing-masing tahap uji. Berdasarkan uji coba skala kecil dan skala terbatas menunjukkan adanya peningkatan hasil validitas butir soal yang ditunjukkan oleh meningkatnya butir soal yang berkategori valid. Peningkatan hasil validitas yang signifikan dipengaruhi oleh proses revisi butir soal yang dilakukan secara sistematis yang didasarkan pada masukan oleh ahli dan analisis statistik. Revisi dilakukan berdasarkan aspek materi, evaluasi, dan bahasa, serta dilakukan penyelarasan antara butir soal dengan indikator soal. Perbaikan butir soal juga didasarkan pada hasil statistik, pengecoh pada soal pilihan ganda juga turut direvisi karena pengecoh yang tidak baik akan menurunkan kualitas soal. Pengecoh berperan penting dalam membedakan peserta didik yang kurang mampu dengan yang mampu (Alwi, 2015). Melalui perbaikan tersebut, soal mengalami penyempurnaan secara keseluruhan sehingga berdampak pada kualitas instrumen.

Tabel 13. Hasil Uji Reliabilitas

Uji	Pilihan Ganda		Uraian	
	<i>Cronbach's Alpha</i>	Interpretasi	<i>Cronbach's Alpha</i>	Interpretasi

Skala Kecil	0,780	Reliabilitas Tinggi	0,506	Reliabilitas Sedang
Skala Terbatas	0,867	Reliabilitas Sangat Tinggi	0,591	Reliabilitas Sedang

Data uji reliabilitas diperoleh melalui hasil uji coba skala kecil dan uji coba skala terbatas yang kemudian dianalisis menggunakan SPSS versi 26. Hasil analisis yang disajikan pada Tabel 13 menunjukkan bahwa pada uji coba skala kecil diperoleh nilai *Cronbach's Alpha* untuk soal pilihan ganda sebesar 0,780 dengan interpretasi reliabilitas tinggi, dan untuk soal uraian sebesar 0,506 dengan interpretasi reliabilitas sedang. Sedangkan pada uji coba skala terbatas diperoleh nilai *Cronbach's Alpha* untuk soal pilihan ganda sebesar 0,867 dengan interpretasi reliabilitas sangat tinggi, dan untuk soal uraian sebesar 0,591 dengan interpretasi reliabilitas sedang. Reliabilitas soal mengalami peningkatan secara signifikan pada soal pilihan ganda sehingga interpretasi reliabilitas sangat tinggi. Instrumen dapat dikatakan reliabel jika hasil pengukuran stabil dan menunjukkan nilai koefisien reliabilitas yang lebih baik (Farida & Musyarofah, 2021). Secara keseluruhan, soal yang

telah diujikan dinyatakan reliabel dengan tingkat reliabilitas soal pilihan ganda yang tinggi dan soal uraian yang sedang atau cukup reliabel. Dengan demikian, instrumen yang telah dikembangkan dapat digunakan secara konsisten untuk mengukur kemampuan berpikir tingkat tinggi (*HOTS*) peserta didik.

Tabel 14. Uji Daya Pembeda

Tahap Uji	Kategori			
	Baik Sekali	Baik	Cukup	Jelek
Skala Kecil	8%	56%	8%	28%
Skala Terbatas	8%	44%	48%	0%

Uji daya pembeda bertujuan untuk mengetahui kemampuan butir soal dalam membedakan peserta didik yang memiliki kemampuan berpikir tingkat tinggi (*HOTS*) dan berpikir tingkat rendah (*LOTS*). Butir soal yang baik adalah butir soal yang memiliki daya untuk membedakan kemampuan peserta didik kelompok tinggi dan rendah berdasarkan besarnya indeks daya beda (Nurgiyantoro, 2016: h.221). Secara keseluruhan, hasil analisis pad uji coba skala kecil dan terbatas menunjukkan adanya peningkatan. Hal ini dapat dilihat dari hasil akhir yang menunjukkan tidak adanya butir soal berkategori jelek. Dengan demikian, butir soal dalam

instrumen asesmen sumatif berbasis *HOTS* mampu memberikan perbedaan antara peserta didik yang menguasai kemampuan berpikir tingkat rendah (*LOTS*) dan kemampuan berpikir tingkat (*HOTS*).

Tabel 15. Uji Tingkat Kesukaran

Tahap Uji	Kategori		
	Mudah	Sedang	Sukar
Skala Kecil	28%	64%	8%
Skala Terbatas	16%	76%	8%

Uji tingkat kesukaran dilakukan untuk mengetahui tingkat kesulitan setiap butir soal dalam instrumen. Soal yang berada pada tingkat kesukaran yang tepat akan menghasilkan distribusi jawaban yang optimal antara peserta didik yang memiliki kemampuan berpikir tingkat tinggi dan rendah. Tingkat kesukaran yang seimbang akan menghasilkan soal yang berkualitas (Rajagukguk & Naibaho, 2023). Hasil akhir menunjukkan bahwa sebagian besar soal berada pada tingkat kesukaran sedang yaitu sebanyak 19 butir soal dari 25 butir soal, yang berarti jawaban benar dan salah peserta didik seimbang. Dengan demikian, instrumen asesmen sumatif berbasis *HOTS* memiliki kualitas yang baik dari segi tingkat kesukaran sehingga dapat digunakan untuk

mengukur kemampuan berpikir tingkat tinggi peserta didik.

Tabel 16. Uji Efektivitas Pengecoh

Tahap Uji	Kategori	
	Efektif	Tidak Efektif
Skala Kecil	57%	43%
Skala Terbatas	97%	3%

Analisis efektivitas pengecoh dilakukan terhadap 20 butir soal pilihan ganda menggunakan microsoft excel. Data hasil analisis pada Tabel 16 menunjukkan bahwa pada uji coba skala kecil, sebesar 57% pengecoh dengan kategori efektif dan 43% pengecoh dengan kategori tidak efektif. Pada uji coba skala terbatas, sebesar 97% pengecoh dengan kategori efektif dan 3% pengecoh dengan kategori tidak efektif. Secara keseluruhan, pengecoh soal tergolong efektif dan pengecoh yang tidak efektif dijadikan dasar untuk dilakukan revisi agar kualitas soal menjadi optimal.

Uji efektivitas pengecoh dilakukan untuk menilai kualitas setiap opsi jawaban dalam soal pilihan ganda. Pengecoh yang efektif akan dipilih oleh peserta didik yang kurang memahami materi. Pengecoh yang baik hendaknya tidak terlalu dominan sehingga membuat peserta didik yang paham menjadi bingung. Pengecoh dapat dikatakan efektif apabila ada

peserta didik yang memilih opsi jawaban tersebut (Qadir et al., 2024). Hasil analisis efektivitas pengecoh pada tahap uji coba skala kecil dan uji coba skala terbatas menunjukkan terjadinya peningkatan efektivitas pengecoh melalui revisi yang dilakukan terhadap opsi pengecoh yang tidak efektif, sehingga hasil akhir pengecoh jawaban dalam soal pilihan ganda sebagian besar berada pada kategori efektif.

D. Kesimpulan

Berdasarkan hasil penelitian dan pembahasan, dapat disimpulkan sebagai berikut.

1. Validitas isi menunjukkan bahwa instrumen telah memenuhi kriteria kelayakan ditinjau dari aspek materi, evaluasi, dan bahasa dengan kategori sangat layak.
2. Respon peserta didik terhadap keterbacaan soal menunjukkan bahwa soal telah disusun dengan jelas dan dapat dipahami oleh peserta didik dengan kategori sangat layak.
3. Uji empiris menunjukkan hasil yang positif. Validitas empiris yang diperoleh yaitu seluruh butir soal memiliki nilai r -hitung diatas r -tabel, sehingga dinyatakan valid dan

memiliki validitas empiris yang sangat baik. Reliabilitas instrumen asesmen sumatif berbasis *HOTS* menunjukkan hasil yang reliabel dengan kategori reliabilitas sangat tinggi, dan untuk soal uraian dengan kategori reliabilitas sedang. Daya pembeda yang diperoleh menunjukkan bahwa butir soal memiliki daya untuk membedakan tingkat berpikir peserta didik. Tingkat kesukaran yang diperoleh menunjukkan bahwa soal yang dikembangkan memiliki kualitas yang baik yang ditandai dengan tingkat kesukaran yang seimbang. Serta efektivitas pengecoh butir soal pada soal pilihan ganda secara umum efektif.

DAFTAR PUSTAKA

- Aditomo, A. (2024). *Panduan Pembelajaran dan Asesmen Pendidikan Anak Usia Dini, Pendidikan Dasar, dan Pendidikan Menengah Edisi Revisi Tahun 2024*. 1–72.
- Alwi, I. (2015). Pengaruh Jumlah Alternatif Jawaban Tes Obyektif Bentuk Pilihan Ganda Terhadap Reliabilitas, Tingkat Kesukaran Dan Daya Pembeda. *Jurnal Ilmiah Faktor Exacta*, 3(2), 184–193. https://journal.lppmunindra.ac.id/index.php/Faktor_Exacta/article/view/19/18
- Anam, S. (2017). Instrumen Penelitian Yang Valid Dan Terpercaya. *Jurnal Pendidikan STKIP Sebelas April*, 1(1). <https://d1wqtxts1xzle7.cloudfront.net/88937333/6-17-1-PB-libre.pdf>
- Arikunto, S. (2016). *Prosedur Penelitian: Suatu pendekatan Praktik*. Rineka Cipta.
- Arvianto, F. (2016). Analisis kualitas dan keterbacaan soal ujian nasional bahasa indonesia. *Utile Jurnal Kependidikan*, 2(2), 184–203. <https://www.jurnal.ummi.ac.id/index.php/JUT/article/view/287>
- Asti, N. R., Ngazizah, N., & Ratnaningsih, A. (2024). Analisis Soal Penilaian Akhir Semester Muatan Pembelajaran Ipa Berbasis Hots Kelas V Sd Muhammadiyah Purwodadi. *Journal Binagogik*, 11(1), 189–194. <https://ejournal.stkipbbm.ac.id/index.php/pgsd>
- Borich, G. D. (2016). Observation Skills for Effective Teaching: Research-Based Practice. In *Routledge* (Vol. 11, Issue 1). Acmillan Publishing Company.
- Farida, & Musyarofah, A. (2021). Validitas dan Reliabilitas dalam Analisis Butir Soal. *Al-Mu'arrib: Jurnal Pendidikan Bahasa Arab*, 1(1), 34–44. <https://doi.org/https://doi.org/10.32923/al-muarrib.v1i1.2100>
- Hanifah, N. (2019). Pengembangan instrumen penilaian Higher Order Thinking Skill (HOTS) di sekolah dasar. *Conference Series Journal*, 1(1), 18–23. <https://doi.org/https://doi.org/10.51612/teunuleh.v6i3.203>
- Kartono, Gasya, D. A. V., Pranata, R., Muskania, R. T., Izudin, M. I., Fatmadiwi, A., & Mu'afiyah, R. (2025). A Textbook Oriented Towards Critical Thinking Skills On The IPAS Subject (Energy Crisis) For Primary School. *Teunuleh Scientific Journal*, 6(September), 299–315. <https://doi.org/https://doi.org/10.51612/teunuleh.v6i3.203>
- Memolo, T. (2019). Karakteristik

- Keterampilan Berpikir Tingkat Tinggi dalam Soal Standar Ujian Matematika SMP. *Seminar Nasional Edusainstek*, 1(2), 465–469.
<https://prosiding.unimus.ac.id/index.php/edusaintek/article/viewFile/268/271>
- Miranda, D. (2018). Pengembangan Buku Cerita Berbasis Pendidikan Karakter Untuk Meningkatkan Kreativitas Aud. *Jurnal Visi Ilmu Pendidikan*, 10(1), 18.
<https://doi.org/10.26418/jvip.v10i1.25975>
- Ningsi, A., & Shaleh, S. (2024). Problematika Penerapan Asesmen Berbasis HOTS (Higher Order Thinking Skills) di Sekolah Dasar. *SAP (Susunan Artikel Pendidikan)*, 8(3), 447.
<https://doi.org/10.30998/sap.v8i3.20998>
- Nurgiyantoro, B. (2016). *Penilaian Pembelajaran Bahasa Berbasis Kompetensi* (2nd ed.). BPFE-YOGYAKARTA.
- Primayana, K. H. (2019). Menciptakan Pembelajaran Berbasis Pemecahan Masalah Dengan Berorientasi Pembentukan Karakter Untuk Mencapai Tujuan Higher Order Thingking Skilss (HOTS) Pada Anak Sekolah Dasar. *Purwadita: Jurnal Agama Dan Budaya*, 3(2), 85–92.
<https://jurnal.stahnmpukuturan.ac.id/index.php/Purwadita/article/view/367>
- Qadir, A., Huda, N., & Hermina, D. (2024). Analisis Butir Tes: Tingkat Kesukaran, Daya Pembeda dan Efektivitas Pengecoh. *Cv.Eureka Media Aksara*, 3(3), 105–134.
<https://publisherqu.com/index.php/AI-Furqan/article/view/1069/962>
- Rahayu, R., Djazari, M., Pengajar, S., Universitas, J. P. A., & Yogyakarta, N. (2016). Analisis Kualitas Soal Pra Ujian Nasional Mata Pelajaran Ekonomi Akuntansi Quality Analysis of Pre National Examination Questions in Economic-Accounting Subject. *Jurnal Pendidikan Akuntansi Indonesia*, XIV(1), 85–94.
<https://doi.org/https://doi.org/10.21831/jpai.v14i1.11370>
- Rahmawati, D. E., & Trimulyono, G. (2022). Validitas Instrumen Penilaian Higher Order Thinking Skills (Hots) pada Materi Keanekaragaman Hayati. *Berkala Ilmiah Pendidikan Biologi (BioEdu)*, 11(1), 138–147.
<https://doi.org/10.26740/bioedu.v11n1.p138-147>
- Rajagukguk, M. J. ., & Naibaho, D. (2023). Mampu Memilih Soal Berdasarkan Tingkat Kesukaran. *Jurnal Pendidikan Sosial Dan Humaniora*, 01(4), 1–23.
<https://publisherqu.com/index.php/pediaqu/article/view/701/645>
- Sani, R. A., & Prayitno, W. (2020). *Asesmen Kompetensi Minimum*. PT Remaja Rosdakarya.
- Sugiyono. (2019). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D* (2nd ed.). ALFABETA.
- Sugiyono. (2023). Metode Penelitian Kuantitatif, Kualitatif, dan R&D. In *ALFABETA*. ALFABETA.
- Suhartono, S., Susiani, T. S., Salimi, M., & Hidayah, R. (2020). Strategi dan Implementasi Penyusunan Soal Higher Order Thinking Skills (HOTS) Bagi Guru Sekolah Dasar di Kabupaten Kebumen. *DEDIKASI: Community Service Reports*, 2(1), 58–68.
<https://doi.org/10.20961/dedikasi.v2i1.34494>
- Yuliani, D., Supriyadi, & Sutopo, Y. (2024). Pengembangan Instrumen Asesmen Tes Higher Order Thinking Skills Pada Pembelajaran IPAS Kelas V SD. *Pendas : Jurnal Ilmiah Pendidikan Dasar*, 09(04).
<https://doi.org/https://doi.org/10.23969/jp.v9i04.18793>