

PENGEMBANGAN INSTRUMEN TES UNTUK MENGUKUR *HIGHER ORDER THINKING SKILLS (HOTS)* PESERTA DIDIK PADA MATERI SUHU DAN KALOR SMA

Oleh :

Dayang Ayu Wandira¹⁾, Hamdani²⁾, Firdaus³⁾

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

¹email: ayuwndraa@gmail.com

²email: hamdani@fkip.untan.ac.id

³email: firdaus@fkip.untan.ac.id

Informasi Artikel

Riwayat Artikel :

Submit, 19 September 2024

Revisi, 22 Desember 2024

Diterima, 27 Desember 2024

Publish, 15 Januari 2025

Kata Kunci :

Pengembangan,

HOTS,

Two-Tier Multiple Choice,

Model Rasch,

Suhu dan Kalor.

ABSTRAK

Penelitian ini bertujuan untuk menghasilkan instrumen tes *two-tier multiple choice* yang layak untuk mengukur HOTS peserta didik pada materi suhu dan kalor. Metode penelitian yang digunakan merupakan penelitian dan pengembangan dengan model 4D. Data penelitian diambil dari 200 peserta didik kelas XI di SMA Negeri 10 Pontianak dan SMA Negeri 3 Pontianak. Analisis karakteristik butir tes dilakukan dengan Model Rasch menggunakan Winstep dan SPSS. Hasil validasi ahli sangat tinggi dengan nilai 0,97, skor keterbacaan rata-rata 6,86, menunjukkan teks soal mudah dipahami siswa SMA, dan hasil validasi konstruk menunjukkan tes dapat mengukur 76,96% dari konstruk yang ingin diukur Model Rasch menunjukkan unidimensionalitas 26,9% (cukup), *person reliability* 0,73 (cukup), *item reliability* 0,90 (bagus), terdapat 13 butir soal fit dengan model Rasch, dan instrumen tes dibagi menjadi 2 soal sangat mudah, 7 soal mudah, 3 soal sangat sulit, dan 3 soal sulit. Kemudian 1 butir soal memiliki daya pembeda jelek, 3 butir soal cukup, dan 11 butir soal baik. Berdasarkan hasil tersebut, terdapat 13 butir soal layak digunakan untuk mengukur kemampuan berpikir tingkat tinggi peserta didik pada materi suhu dan kalor.

This is an open access article under the [CC BY-SA](#) license



Corresponding Author:

Nama: Dayang Ayu Wandira

Afiliasi: Universitas Tanjungpura Pontianak

Email: ayuwndraa@gmail.com

1. PENDAHULUAN

Fisika mempelajari sifat dan fenomena alam serta interaksi di dalamnya. Pembelajaran fisika di SMA bertujuan untuk melatih siswa menguasai pengetahuan, konsep, prinsip fisika, keterampilan sains, kecakapan ilmiah, serta keterampilan berpikir kritis dan kreatif melalui Higher Order Thinking Skills (HOTS). Namun, pembelajaran di sekolah sering tidak menerapkan HOTS, terutama dalam fisika, dan guru lebih sering mengukur kemampuan kognitif rendah atau Low Order Thinking Skills (LOTS). Akibatnya, siswa tidak terbiasa berpikir mandiri dan malas berpikir secara mendalam. Hal ini terlihat dari hasil tes Programme for International Students Assessment and Development (PISA), di

mana kemampuan siswa Indonesia masih rendah. Pada PISA 2018, Indonesia mendapat skor di bawah rata-rata OECD, menunjukkan bahwa siswa hanya bisa menyelesaikan soal dengan kompleksitas rendah. Meskipun peringkat PISA Indonesia naik pada 2022, skor pada kemampuan membaca, matematika, dan sains menurun dibandingkan 2018. Hasil ini menunjukkan bahwa kemampuan berpikir tingkat tinggi siswa Indonesia masih rendah (Tim Penysun Penilaian Pendidikan, 2022).

Hal serupa terjadi di SMA Negeri 10 Pontianak, guru fisika jarang menggunakan soal HOTS dalam penilaian, baik formatif maupun sumatif, sehingga instrumen tes yang digunakan terbatas pada jenjang kognitif C1-C3 (mengukur

kemampuan mengingat, memahami, dan menerapkan). Akibatnya, peserta didik tidak terbiasa menyelesaikan soal HOTS, terlihat dari nilai rata-rata ulangan harian yang hanya mencapai 62, di bawah nilai KKM 75. Pada Ujian Nasional 2018/2019, soal HOTS hanya 22,5% dan tidak merata, dominan pada analisis (C4). Kurangnya latihan dan pengalaman memecahkan soal HOTS menyebabkan rendahnya kemampuan berpikir tingkat tinggi siswa. Memberikan soal pemecahan masalah, berpikir kreatif, dan berpikir kritis dapat meningkatkan HOTS dan motivasi belajar siswa, serta membantu mereka dalam memahami masalah secara mendalam (Afifah D, 2020)

Taksonomi Bloom menjadi dasar dalam keterampilan berpikir tingkat tinggi, meliputi tiga aspek kognitif: menganalisis (C4), mengevaluasi (C5), dan mencipta (C6). (Anderson & Krathwol, 2001). Penilaian kemampuan berpikir tingkat tinggi peserta didik dilakukan melalui instrumen yang tepat dan teruji. Tes HOTS mengukur kemampuan analisis, pemrosesan informasi, dan pemecahan masalah, bukan sekadar mengingat atau menyatakan kembali informasi. Direktorat Pembinaan SMA (2017) menyatakan bahwa soal HOTS menghubungkan konsep, memproses informasi, dan menganalisis ide secara kritis.

Tes HOTS biasanya berbentuk uraian atau pilihan ganda, termasuk *two-tier multiple choice* (TTMC). TTMC mengurangi kekurangan tes pilihan ganda biasa dengan meminta peserta memberikan alasan di tier kedua, mengurangi faktor lucky guess dan meningkatkan kemampuan berpikir tingkat tinggi. Penelitian menunjukkan bahwa TTMC efektif untuk mengukur HOTS. Instrumen TTMC dikembangkan dan dianalisis menggunakan Model Rasch untuk mengukur HOTS pada materi suhu dan kalor, menjadi alternatif tes HOTS dalam asesmen dan evaluasi pembelajaran fisika.

Hal ini sesuai dengan beberapa penelitian sebelumnya. Penelitian Sari dkk. (2019) menunjukkan bahwa tes pilihan ganda dengan alasan dapat mengukur keterampilan berpikir tingkat tinggi siswa. Penelitian Husnawati dkk. (2019) menghasilkan soal HOTS pada materi gerak benda di SMP yang dianalisis dengan teori tes klasik dan dibandingkan dengan soal OSN IPA menggunakan uji t-test. Sholekhah (2018) mengembangkan soal tes HOTS untuk materi Hukum Newton di SMA, sedangkan Swastika (2021) mengembangkan instrumen tes HOTS untuk materi gerak dan gaya di SMP yang dianalisis dengan model Rasch.

Penelitian ini menghasilkan instrumen tes *two-tier multiple choice* yang dapat mengukur kemampuan berpikir tingkat tinggi siswa dan dianalisis menggunakan Model Rasch. Instrumen tes ini dikembangkan untuk mengukur HOTS pada materi suhu dan kalor, sehingga dapat digunakan sebagai alternatif tes HOTS untuk asesmen maupun evaluasi pembelajaran fisika.

2. METODE PENELITIAN

Metode yang digunakan dalam penelitian ini adalah penelitian dan pengembangan (Research & Development) dengan pendekatan kuantitatif. Penelitian R&D adalah langkah-langkah terstruktur untuk menyelesaikan masalah melalui produk. Metode ini digunakan untuk menghasilkan, menguji keefektifan, dan mengembangkan produk baru. Dalam penelitian ini, dilakukan analisis kebutuhan lapangan sebelum pengembangan produk agar sesuai dengan kebutuhan. Metode R&D digunakan untuk mengembangkan instrumen tes HOTS untuk mengukur keterampilan berpikir kritis peserta didik dan diuji kelayakannya berdasarkan model Rasch.

Penelitian ini menggunakan metode pengembangan 4D oleh Thiagarajan (1974) yaitu Define, Design, Development, dan Dissemination. Langkah-langkah pengembangan 4D dalam penelitian ini meliputi:

1. Studi Pendahuluan (Define)

- Analisis front-end : Menentukan permasalahan dasar dan urgensi tes HOTS.
- Analisis peserta didik : Menilai kemampuan berpikir tingkat tinggi dan penggunaan tes HOTS.
- Analisis tugas : Mengidentifikasi tugas utama yang akan dilakukan oleh peserta didik.
- Analisis konsep : Menentukan indikator soal dan butir soal.
- Analisis tujuan pembelajaran : Menyusun indikator soal HOTS berdasarkan analisis sebelumnya.

2. Perancangan (Design)

- Menyusun kisi-kisi soal, menentukan bentuk dan jumlah soal, serta penulisan soal sesuai langkah-langkah pengembangan soal HOTS oleh Widana (2017).

Tabel 1 Indikator Soal

Tujuan Pembelajaran	Indikator soal
Peserta didik mampu menganalisis perubahan suhu pada suatu benda terhadap kalor pada benda tersebut	Peserta didik dapat menganalisis grafik hubungan waktu terhadap perubahan suhu pada dua zat sejenis
	Peserta didik dapat menganalisis kenaikan suhu pada ketiga zat yang berbeda saat dipanaskan dengan waktu dan suhu yang sama
	Peserta didik dapat menganalisis pengaruh kalor terhadap perubahan suhu benda
Peserta didik mampu mengaitkan konsep perpindahan kalor dalam kehidupan sehari-hari	Peserta didik dapat menganalisis perbedaan titik didih air di dataran tinggi dan dataran rendah
Peserta didik mampu menentukan laju kalor secara konduksi, konveksi, maupun radiasi	Peserta didik dapat menganalisis pernyataan yang benar mengenai laju konduksi kalor yang melalui dinding
	Peserta didik dapat menyimpulkan hasil dari percobaan mengenai hubungan laju kalor dengan koefisien konduksi termal
Peserta didik mampu menentukan panjang benda setelah mengalami muai panjang maupun muai volume	Peserta didik dapat mengevaluasi pernyataan mengenai pemuai yang terjadi pada kaca jendela
Peserta didik mampu menentukan konduktivitas termal pada logam	Peserta didik dapat membandingkan kesetimbangan termal antara tembaga dan aluminium.
	Peserta didik dapat membandingkan hubungan panjang dengan konduktivitas pada besi dan

	kuningan
Peserta didik mampu menganalisis jumlah kalor yang diterima dan kalor yang dilepas pada suatu benda memiliki besar yang sama	Peserta didik dapat membuat hipotesis berdasarkan rumusan masalah yang telah disajikan dalam soal mengenai kalor
Peserta didik mampu melaksanakan percobaan perpindahan kalor	Peserta didik dapat mencipatkan benda yang paling cepat mengalami perpindahan kalor
	Peserta didik dapat merancang percobaan untuk menentukan konduktivitas termal beberapa bahan digunakan dalam kehidupan sehari-hari
	Peserta didik dapat merancang percobaan pengaruh kalor terhadap massa zat
Peserta didik mampu menentukan jumlah kalor yang dibutuhkan untuk menaikkan suhu	Peserta didik dapat merancang percobaan pengaruh kalor terhadap kalor jenis
	Peserta didik dapat mengevaluasi pernyataan yang diberikan mengenai jenis logam apa yang digunakan berdasarkan data pada grafik hubungan jumlah kalor dengan kenaikan suhu

- Menentukan bentuk dan jumlah soal, dalam penelitian ini tes dibuat sebanyak 15 butir soal tes pilihan ganda dua tingkat yang mengukur HOTS atau dari tingkatan berpikir C4 hingga C6. Banyaknya butir soal ditentukan dengan pertimbangan waktu peserta didik dalam menyelesaikan soal tes.
 - Penulisan Soal, butir soal dibuat berdasarkan kisi-kisi atau indikator soal yang telah dibuat.
3. Pengembangan (Development)
- Tes yang telah dirancang ditelaah tingkat keterbacaannya dengan nilai >6 untuk peserta didik tingkat SMA.
 - Divalidasi oleh ahli dan dianalisis hasil validasi isi dengan menggunakan formula aiken.

$$V = \frac{\sum s}{n(c-1)}$$

$$S = r - l0$$

Keterangan :

l0= angka penilaian validitas terendah

c = angka penilaian validitas tertinggi

r = angka yang diberikan oleh penilai

n = jumlah penilai

Tabel 2 Kriteria Validitas

Koefisien	Keterangan
0,00 – 0,20	Sangat Rendah
0,21 – 0,40	Rendah
0,41 – 0,60	Cukup
0,61 – 0,80	Tinggi
0,81 – 1,00	Sangat Tinggi

(Sumber; Arikunto, 2018)

- Uji cobakan pada kelompok kecil, ujicoba dilakukan untuk mengetahui apakah waktu yang diberikan cukup, dan bahasa yang ada dalam soal dapat dimengerti oleh peserta didik.
4. Penyebaran (Disseminate)
- Instrumen tes yang telah dikembangkan diuji coba skala besar di dua sekolah yaitu SMAN 3 Pontianak dan SMAN 10 Pontianak dengan jumlah peserta didik sebanyak 200 orang.
 - Kemudian hasil uji coba skala besar dianalisis menggunakan model rasch untuk mengetahui unidimensional, tingkat kesesuaian butir soal, reliabilitas, tingkat kesukaran dan daya pembeda.

- Unidimensional, suatu tes yang memenuhi syarat unidimensionalitas minimal memiliki nilai *Raw Variance* sebesar 20%.
- Tingkat Kesesuaian Butir Soal, suatu soal dapat dikatakan *Fit* apabila memenuhi minimal 2 kriteria berikut :
 1. Nilai Outfit Mean Square (MNSQ) yang diterima : $0,5 < MNSQ < 1,5$
 2. Nilai Outfit Z-Standart (ZSTD) yang diterima : $-2,0 < ZSTD < +2,0$
 3. Nilai Point Measure Correlation (Pt Measure Corr) yang diterima : $0,4 < Pt Measure Corr < 0,85$
- Reliabilitas, diukur menggunakan *alpha cronbach's* Nilai alpha cronbach's digunakan untuk mengukur reliabilitas yaitu interaksi antara person dan item secara keseluruhan.

Tabel 3 Kriteria Reliabilitas

Nilai Alpha Cronbach	Korelasi	Ket
$0,90 \leq r \leq 1,00$	Sangat tinggi	Sangat baik
$0,70 \leq r \leq 0,90$	Tinggi	Baik
$0,40 \leq r \leq 0,70$	Sedang	Cukup baik
$0,20 \leq r \leq 0,40$	Rendah	Buruk
$r < 0,20$	Sangat rendah	Sangat buruk

(Sumber ; Susdelina, Perdana & Febrian, 2018)

- Tingkat Kesukaran Soal, untuk mengetahui tingkat kesukaran atau kesulitan butir soal dengan melihat nilai measure.

Tabel 4 Kategori Tingkat Kesukaran Soal

Nilai Measure Logit	Kategori
Measure logit < - SD logit	Sangat mudah
-SD logit ≤ measure logit ≤ 0	Mudah
$0 \leq \text{measure logit} \leq \text{SD logit}$	Sulit
Measure logit > SD logit	Sangat sulit

(Sumber ; Erfan et al, 2020)

- Daya Pembeda, kemampuan soal untuk membedakan antara siswa yang berkemampuan tinggi dengan siswa yang berkemampuan rendah (Sumintono & Widhiarso, 2015).

Tabel 5 Kriteria Daya Pembeda

Koefisien	Kriteria
0,00 – 0,19	Jelek
0,20 – 0,39	Cukup
0,40 – 0,69	Baik
0,70 – 1,00	Baik sekali
Negatif	Tidak baik, harus dibuang

(Sumber ; Erfan et al, 2020)

3. HASIL DAN PEMBAHASAN

Hasil yang diperoleh dalam penelitian ini dijabarkan sesuai urutan tahapan pengembangan 4D, yaitu :

1) Tahap Studi Pendahuluan (*Define*)

Tahap ini dilakukan analisis kebutuhan dengan penelitian dan studi literatur atau tahap awal yang dilakukan untuk meninjau perlu pengembangan instrumen tes. Adapun langkah-langkah dalam tahap ini meliputi :

a. Analisis *Front-End* :

Bertujuan untuk mengidentifikasi permasalahan pembelajaran, khususnya kemampuan berpikir tingkat tinggi peserta didik, yang ditemukan masih rendah berdasarkan data PISA 2022 dan Ujian Nasional. Peserta didik di Indonesia kesulitan dalam

mengintegrasikan informasi, memformulasikan masalah dunia nyata ke dalam konsep pelajaran, serta melakukan investigasi. Rendahnya kemampuan ini juga disebabkan oleh kurangnya penggunaan soal HOTS dalam evaluasi pembelajaran.

b. Analisis Peserta Didik :

Menunjukkan bahwa peserta didik jarang diberikan soal HOTS dalam asesmen. Ujian Nasional hanya sedikit menggunakan soal HOTS, yang menurun di tahun 2017/2018 hingga 2018/2019. Penelitian pendahuluan di beberapa SMA Pontianak menunjukkan bahwa soal yang diberikan umumnya hanya mengukur pemahaman, bukan analisis mendalam.

c. Analisis Tugas :

Analisis tugas dilakukan untuk mengetahui tugas-tugas utama yang harus dicapai oleh peserta didik. Analisis ini dilakukan dengan menelaah Capaian Pembelajaran (CP) mengenai materi suhu dan kalor dan menguraikannya menjadi Tujuan Pembelajaran

d. Analisis Konsep :

Mengidentifikasi konsep penting yang harus dipahami siswa, seperti perubahan suhu, perpindahan kalor, pemuain zat, dan konduktivitas termal.

e. Analisis Tujuan Pembelajaran :

Dilakukan untuk merumuskan tujuan pembelajaran yang dapat dikembangkan menjadi soal HOTS berdasarkan capaian pembelajaran yang telah dianalisis.

2) Tahap Perancangan (*Design*)

Tahap perancangan merupakan tahap kedua dalam penelitian ini yang dilakukan untuk menyusun spesifikasi tes. Tahapan ini terdiri dari empat langkah, yaitu menentukan tujuan tes, menentukan bentuk dan panjang tes, menentukan kisi-kisi tes, dan menulis tes. Hasil akhir dalam tahap ini adalah tes sebanyak 15 butir soal pilihan ganda tingkat dua (*two-tier*) dengan 4 pilihan jawaban dengan alasan tertutup, jenjang kognitif yang digunakan mulai dari C4 – C6, dan waktu pengerjaan tes dilaksanakan selama 60 menit.

3) Tahap Pengembangan (*Development*)

a. Hasil Validasi prototipe Produk

Setelah dikembangkan tes berdasarkan spesifikasi tes pada tahap design. Butir soal divalidasi oleh 2 dosen pendidikan fisika FKIP Untan sebagai validator ahli dan 1 orang guru SMA sebagai validator panel. Berdasarkan hasil validasi oleh ahli didapatkan perbaikan kata pada tiap butir soal pada tabel 6.

Tabel 6 Saran dan Perbaikan Oleh Validator

Saran No. 9 :
Cek kembali pernyataan soal agar tepat dan mudah dipahami
Sebelum :
Beberapa peserta didik di kelas XI A diminta oleh guru fisika untuk membuat benda yang paling cepat mengalami perpindahan kalor, benda tersebut dibuat menggunakan barang bekas yang ada disekitar berdasarkan kreativitas mereka masing-masing. Maka benda yang paling tepat dibuat adalah....
Sesudah :
Beberapa siswa di kelas XI A diminta oleh guru fisika untuk menciptakan benda yang mampu menghantarkan panas dengan cepat. Benda tersebut harus dibuat dari bahan bekas yang tersedia di sekitar

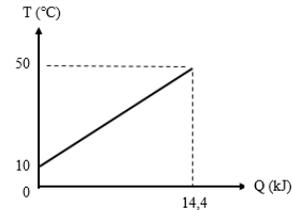
mereka, dengan memanfaatkan kreativitas mereka masing-masing. Oleh karena itu, benda yang paling tepat dibuat adalah....

Saran No. 10 :

Alasan dalam soal bersifat hafalan bukan evaluasi, sehingga bentuk soal dapat diganti, dengan indikator soal yang sama

Sebelum :

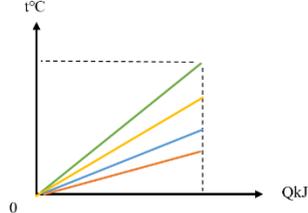
Retno, Farel, Ratih, dan Sela mengamati grafik yang menunjukkan hubungan antara jumlah kalor yang diperlukan dengan kenaikan suhu dari 800 gram logam.



Setelah mengamati grafik diatas, mereka memiliki perbedaan pendapat mengenai jenis logam yang sesuai dengan grafik tersebut. Retno menyatakan bahwa jenis logam tersebut adalah perak, Farel menyatakan bahwa jenis logam tersebut adalah tembaga, Ratih menyatakan bahwa jenis logam tersebut besi, dan Sela menyatakan bahwa jenis logam tersebut baja. Berdasarkan pendapat mereka berempat, pernyataan yang paling sesuai adalah ...

Sesudah :

Retno, Farel, Ratih, dan Sela mengamati grafik yang menunjukkan hubungan antara jumlah kalor yang diperlukan dengan kenaikan suhu dari 800 gram logam.



Berdasarkan grafik diatas, mereka memiliki perbedaan pendapat mengenai jenis logam yang sesuai dengan grafik yang berwarna hijau. Retno menyatakan bahwa jenis logam tersebut adalah perak dengan kapasitas kalor 235 J/kg°C, Farel menyatakan bahwa jenis logam tersebut adalah tembaga dengan kapasitas kalor 385 J/kg°C, Ratih menyatakan bahwa jenis logam tersebut besi dengan kapasitas kalor 450 J/kg°C, dan Sela menyatakan bahwa jenis logam tersebut baja dengan kapasitas kalor 490 J/kg°C.

Berdasarkan pendapat mereka berempat, pernyataan yang paling sesuai adalah ...

Instrumen tes divalidasi oleh ahli dengan rata-rata koefisien validitas Aiken V' sebesar 0,97 hingga 0,98, menunjukkan validitas yang sangat baik dalam segi materi, konstruksi, dan bahasa. Setiap butir soal dinilai valid dan dapat mengukur HOTS sesuai dengan dimensi kognitifnya. Selain itu, semua butir soal juga memenuhi tingkat keterbacaan untuk peserta didik jenjang SMA dengan nilai rata-rata tingkat keterbacaan atau *readability index* 6,86.

Tabel 7 Hasil Validitas Isi dan *Readability Index*

No	Validitas Isi	RI	Ket
1	0,95	6,12	Valid dan Tepenuhi
2	0,98	6,33	Valid dan Tepenuhi
3	0,98	6,77	Valid dan Tepenuhi
4	0,99	6,32	Valid dan Tepenuhi
5	1	6,08	Valid dan Tepenuhi
6	0,98	7,13	Valid dan Tepenuhi
7	1	7,31	Valid dan Tepenuhi
8	0,98	7,30	Valid dan Tepenuhi
9	0,98	7,26	Valid dan Tepenuhi
10	0,88	7,98	Valid dan Tepenuhi
11	0,98	8,21	Valid dan Tepenuhi
12	0,96	6,52	Valid dan Tepenuhi
13	1	6,04	Valid dan Tepenuhi
14	1	6,95	Valid dan Tepenuhi
15	1	6,68	Valid dan Tepenuhi

b. Hasil Ujicoba Skala Kecil

Setelah soal dinyatakan valid dan terbaca dengan baik, dilakukan uji coba skala kecil pada kelas XI G SMA Negeri 10 Pontianak pada 3 Mei 2024 untuk mengevaluasi kecukupan waktu dan pemahaman bahasa soal. Hasil uji coba menunjukkan bahwa petunjuk, bahasa, grafik, gambar, serta istilah dalam soal jelas dan dapat dimengerti, serta waktu pengerjaan cukup. Oleh karena itu, soal siap digunakan untuk uji coba skala besar.

4) Tahap Penyebaran (*Disseminate*)

Uji coba skala besar dilakukan di 2 sekolah, yaitu SMA Negeri 10 Pontianak dan SMA Negeri 3 Pontianak. Uji coba di SMA Negeri 10 Pontianak dengan peserta didik sebanyak 100 orang pada tanggal 6-8 Mei 2024 dan di SMA Negeri 3 Pontianak dengan peserta didik sebanyak 100 orang pada tanggal 13-16 Mei 2024 berdasarkan uji coba skala besar kemudian didapatkan hasil unidimensional, reliabilitas, tingkat kesukaran butir soal, tingkat kesesuaian butir soal berdasarkan Model Rasch.

a. Nilai Unidimensional

Nilai unidimensional berada pada kategori cukup dengan nilai *raw variance* 26,7%. Unidimensional menunjukkan ketepatan sebuah tes untuk mengukur kemampuan yang ingin diukur. Dalam penelitian ini, instrumen tes yang dikembangkan bertujuan untuk mengukur kemampuan berpikir tingkat tinggi peserta didik pada materi Suhu dan Kalor. Standar nilai minimum unidimensionalitas adalah 20%.

TABLE 23.8 C:\Users\ASUS\Desktop\JAHABAH.prm ZOU912MS.TXT May 27 2024 16: 8
INPUT: 200 Person 15 Item REPORTED: 200 Person 15 Item 2 CATS WINSTEPS 5.1.4.0

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information

	Eigenvalue	Observed	Expected
Total raw variance in observations =	20.5273	100.0%	100.0%
Raw variance explained by measures =	5.5273	26.9%	26.7%
Raw variance explained by persons =	2.6537	12.9%	12.8%
Raw Variance explained by items =	2.8736	14.0%	13.9%
Raw unexplained variance (total) =	15.0000	73.1%	73.3%
Unexplained variance in 1st contrast =	1.5287	7.4%	10.1%
Unexplained variance in 2nd contrast =	1.5090	7.4%	10.1%
Unexplained variance in 3rd contrast =	1.3971	6.8%	9.3%
Unexplained variance in 4th contrast =	1.2366	6.0%	8.2%
Unexplained variance in 5th contrast =	1.1802	5.7%	7.9%

Gambar 1 Hasil Unidimensional

b. Reliabilitas

Dalam penelitian ini terdapat 3 reliabilitas yaitu, *alpha Cronbach*, *item reliability*, dan *person reliability*. Nilai alpha Cronbach sebesar 0,80 menunjukkan interaksi yang baik antara responden dan item tes. Dalam model Rasch, nilai item reliability sebesar 0,90 (kategori bagus) menunjukkan kualitas butir soal yang konsisten, sedangkan person reliability sebesar 0,74 (kategori cukup) menunjukkan konsistensi jawaban peserta tes.

c. Item Fit

Dalam model Rasch, validitas butir soal dinilai berdasarkan tingkat kesesuaian butir soal atau *item fit*, menggunakan tiga kriteria: Outfit MNSQ, Outfit ZSTD, dan Point Measure Correlation. Sebuah butir soal dianggap valid jika memenuhi setidaknya dua dari tiga kriteria tersebut. Hasil analisis menunjukkan bahwa butir soal nomor 1 dan 3 tidak memenuhi kriteria Outfit MNSQ dan Outfit ZSTD. Selain itu, butir soal nomor 1 dan 3 juga tidak

memenuhi kriteria Point Measure Correlation. Oleh karena itu, butir soal nomor 1 dan 3 perlu dibuang atau diganti. Sebanyak 13 butir soal lainnya dinyatakan valid dan dapat digunakan.

Tabel 8 Hasil Analisis Tingkat Kesesuaian Butir Soal

No.	Tingkat Kesesuaian Butir Soal		
	Mnsq	Zstd	M Corr
1	1,73	4,89	0,29
2	1,05	0,45	0,49
3	1,23	2,03	0,36
4	0,85	-1,40	0,53
5	0,92	-0,73	0,53
6	0,92	-0,57	0,56
7	0,88	-1,06	0,55
8	0,89	-1,06	0,54
9	0,80	-1,26	0,61
10	1,12	1,05	0,50
11	0,90	-0,96	0,54
12	1,00	0,05	0,52
13	0,85	-1,45	0,56
14	0,96	-0,29	0,52
15	1,14	0,79	0,42

d. Tingkat Kesukaran

Tingkat kesukaran menyatakan seberapa sukar atau sulit butir soal yang dikerjakan oleh peserta didik Tingkat kesukaran tes bervariasi dengan 2 butir soal sangat mudah, 7 butir soal mudah, 3 butir soal sulit, dan 3 butir soal sangat sulit, menunjukkan distribusi kesukaran yang baik. Berdasarkan analisis tingkat kesukaran, semua butir soal berada dalam jangkauan kemampuan peserta didik. Soal nomor 15 paling sulit dengan nilai logit 1,20, dan soal nomor 1 paling mudah dengan nilai logit -0,70.

Tabel 9 Hasil Analisis Tingkat Kesukaran Butir Soal

No.	Tingkat Kesukaran	Keterangan
1.	-0,70	Sangat Mudah
2.	-0,59	Sangat Mudah
3.	-0,22	Mudah
4.	-0,33	Mudah
5.	-0,19	Mudah
6.	0,55	Sangat Sulit
7.	-0,30	Mudah
8.	-0,22	Mudah
9.	1,01	Sangat Sulit
10.	0,22	Sulit
11.	-0,30	Mudah
12.	-0,19	Mudah
13.	0,06	Sulit
14.	0,11	Sulit
15.	1,20	Sangat Sulit

e. Daya Pembeda

Daya pembeda soal merupakan kemampuan sesuatu soal untuk membedakan antara peserta didik yang berkemampuan tinggi dengan peserta didik yang berkemampuan rendah (Arikunto, 2018). Uji daya pembeda menunjukkan soal nomor 1 memiliki daya pembeda yang buruk (0,159), sementara soal nomor 2, 3, dan 15 memiliki daya pembeda yang cukup. Sebanyak 11 butir soal lainnya memiliki daya pembeda yang baik.

Tabel 9 Hasil Analisis Daya Pembeda

No.	Daya Pembeda	Keterangan
1.	0,159	Jelek
2.	0,374	Cukup
3.	0,228	Cukup
4.	0,448	Baik
5.	0,447	Baik
6.	0,481	Baik
7.	0,453	Baik
8.	0,449	Baik

9.	0,539	Baik
10.	0,415	Baik
11.	0,470	Baik
12.	0,453	Baik
13.	0,482	Baik
14.	0,426	Baik
15.	0,326	Cukup

Selain dianalisis menggunakan model rasch, tes diuji validitas konstruksinya menggunakan *Exploratory Factor Analysis* (EFA) dengan SPSS. EFA berfungsi untuk mengelompokkan indikator yang berkorelasi. Analisis faktor dilakukan jika nilai KMO lebih dari 0,5 dan signifikansi uji Bartlett kurang dari 0,05, yang menunjukkan kelayakan instrumen. Didapatkan nilai hasil analisis uji KMO dan Bartlett's ditunjukkan pada gambar 2.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,601
Bartlett's Test of Sphericity	Approx. Chi-Square	549,111
	df	45
	Sig.	,000

Gambar 2. Hasil Nilai KMO dan Bartlett

Hasil uji Bartlett menunjukkan signifikansi 0,000 dan KMO 0,601, sehingga analisis faktor dapat dilakukan. Analisis Measure Sampling Adequacy (MSA) dilakukan melalui tabel Anti-image Matrices untuk menentukan butir soal yang layak. Setelah reduksi, butir soal nomor 1, 3, 7, 11, dan 12 dinyatakan tidak layak pada tabel 3.6.

Tabel 10 Analisis Nilai MSA

No	Nilai MSA	Keterangan
2	0,633	Terpenuhi
4	0,709	Terpenuhi
5	0,547	Terpenuhi
6	0,532	Terpenuhi
8	0,626	Terpenuhi
9	0,629	Terpenuhi
10	0,594	Terpenuhi
13	0,592	Terpenuhi
14	0,645	Terpenuhi
15	0,546	Terpenuhi

Hasil uji Bartlett (signifikansi 0,000) dan KMO 0,601 menunjukkan analisis faktor dapat dilakukan. Berdasarkan analisis MSA pada tabel Anti-image Matrices, butir soal 1, 3, 7, 11, dan 12 dinyatakan tidak layak. Hasil analisis *communalities* ditunjukkan pada gambar 3.

Communalities

	Initial	Extraction
S2	1,000	,767
S4	1,000	,533
S5	1,000	,758
S6	1,000	,784
S8	1,000	,713
S9	1,000	,694
S10	1,000	,705
S13	1,000	,842
S14	1,000	,781
S15	1,000	,777

Extraction Method: Principal Component Analysis.

Gambar 3 Nilai Communalities

Berdasarkan analisis communalities, 10 butir soal memenuhi syarat dengan nilai lebih dari 0,50, yaitu soal 2, 4, 5, 6, 8, 9, 10, 13, 14, dan 15. Analisis Total Variance Explained menunjukkan 5 faktor dengan total varians 76,969%. Hasil EFA menunjukkan validitas konstruk tes menjelaskan sebagian besar varians data.

Selanjutnya butir soal dikelompokkan berdasarkan 5 faktor yang ditemukan pada tabel *Total Variance Explained*. Butir soal dikelompokkan pada faktor yang memiliki nilai korelasi terbesar diantara kelima faktor tersebut. Sehingga didapatkan persebaran variabel berdasarkan faktor pada Tabel 4.10.

Tabel 11 Persebaran Faktor

Faktor	Butir Soal
1	2
2	4, 5
3	6
4	8, 9
5	10, 13, 14, 15

4. KESIMPULAN

Berdasarkan data hasil penelitian dan pembahasan didapatkan bahwa secara umum instrumen tes yang dikembangkan untuk mengukur kemampuan berpikir tingkat tinggi pada materi Suhu dan Kalor yang layak untuk digunakan.

Secara khusus disimpulkan bahwa :

1. Instrumen tes yang dikembangkan sudah valid berdasarkan validasinya. Hasil validasi ahli berada dalam kategori sangat tinggi dengan nilai 0,97, bahasa yang disusun mudah dipahami oleh peserta didik jenjang SMA kelas XI dengan tingkat keterbacaan 6,86, dan hasil validasi konstruk instrumen tes yang dikembangkan memiliki persentasi varians konstruk yang dapat dijelaskan oleh tes sebesar 76,96%.
2. Instrumen tes yang dikembangkan sudah valid dan reliabel dan diketahui tingkat kesukarannya menurut Model Rasch. Unidimensionalitas dalam kategori cukup dengan nilai sebesar 26,9%; 13 butir soal memenuhi setidaknya dua dari tiga kriteria tingkat kesesuaian butir soal; reliabilitas *alpha cronbach's* dalam kategori baik dengan nilai 0,80; *person reliability* dalam kategori cukup dengan nilai 0,74%; *item reliability* dalam kategori bagus dengan nilai 0,90; instrumen tes terdiri dari 2 butir soal sangat mudah, 7 butir soal mudah, 3 butir soal sulit, dan 3 butir soal sangat sulit; dan instrumen tes memiliki daya pembeda yang terdiri dari 1 butir soal jelek, 3 butir soal cukup, dan 11 butir soal baik

5. REFERENSI

Afifah, N. (2020). Analisis Soal Ujian Nasional Mata Pelajaran IPA SMP/MTs Tahun Ajaran 2018/2019 Berbasis Higher Order Thinking Skill (HOTS) (Institut Agama Islam Negeri (IAIN) Salatiga). Institut Agama Islam Negeri (IAIN) Salatiga. Retrieved from <http://erepository.perpus.iainsalatiga.ac.id/id/e>

- print/10432.
- Anderson, L. W., & Krathwohl, D.R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing; A revision of Bloom's taxonomy of educational objectives*. Longman.
- Arikunto, S. (2018). *Dasar - Dasar Evaluasi Pendidikan* (3rd ed.). Jakarta: PT Bumi Aksara.
- Direktorat. 2017. *Modul Penyusunan Soal Higher Order Thinking Skill (HOTS)*. Jakarta : Departemen Pendidikan Dan Kebudayaan.
- Erfan, M., Maulyda, M. A., Hidayati, V. R., Astria, F. P., & Ratu, T. (2020). Analisis Kualitas Soal Kemampuan Membedakan Rangkaian Seri dan Paralel Melalui Teori Tes Klasik dan Model Rasch. *Indonesian Journal of Educational Research and Review*, 3(1), 11–19.
- Husnawati, A., Hartono, H., & Masturi, M. (2019). Pengembangan Soal Higher Order Thinking Skill (HOTS) Fisika Kelas VIII SMP Materi Gerak Pada Benda. *UPEJ Unnes Physics Education Journal*, 8(2), 133–140. <https://doi.org/10.15294/upej.v8i2.33320>
- Sari, R, A, (2019). *Pembelajaran Berbasis HOTS (Higher Order Thinking Skills)*. Tirs Smart Jakarta.
- Sholekhah, S. (2018). *Efektivitas Metode Pembelajaran Berbasis Inkuiri pada Peningkatan Pemahaman Siswa*. *Jurnal Pendidikan Indonesia*, 15(3), 45-57.
- Sumintono, B., & Widhiarsono, W. (2015). Aplikasi Permodelan Rasch pada Assessment Pendidikan. Cimahi: Trim Kominika.
- Susdelina, S., Perdana, S. A., & Febrian, F. (2018). Analisis Kualitas Instrumen Pengukuran Pemahaman Konsep Persamaan Kuadrat Melalui Teori Tes Klasik Dan Rasch Model. *Jurnal Kiprah*, 6(1), 41–48. <https://doi.org/10.31629/kiprah.v6i1.574>
- Swastika, R. Z. (2021). Pengembangan Instrumen Tes Untuk Mengukur *Higher Order Thinking Skills (HOTS)* Pada Materi Gerak dan Gaya.
- Tim Penyusun Penilaian Pendidikan. (2022). *Panduan Penilaian dalam Kurikulum 2022*. Kementerian Pendidikan dan Kebudayaan.
- Widana, I. W. (2017). Higher Order Thinking Skills Assessment (HOTS). *Jisae: Journal of Indonesian Student Assesment and Evaluation*, 3(1), 32–44. <https://doi.org/10.21009/jisae.031.04>