



Analisis Kualitas Instrumen Kemampuan Pemecahan Masalah Siswa SMA pada Materi Suhu dan Kalor Menggunakan Model *Rasch*

Silvi Lutfiah Padilah¹, Lina Aviyanti^{1*}, Nurdini Nurdini¹

¹ Pendidikan Fisika, FPMIPA, Universitas Pendidikan Indonesia, Bandung, Indonesia.

DOI: <https://doi.org/10.29303/goescienceed.v7i3.2895>

Article Info:

Received : 21 Juni 2026
Revised : 27 Juni 2026
Accepted : 26 Juli 2026
Published : 01 Agustus 2026

Correspondence:

Lina Aviyanti

Phone: +628112139600

Abstract: Problem-solving ability is an essential competency in physics learning; therefore, a high-quality assessment instrument is required to measure students' abilities accurately. This study aimed to analyze the quality of a physics problem-solving ability instrument on the topic of temperature and heat using the *Rasch* Model. This research employed a quantitative approach with a descriptive research design involving 54 senior high school students (26 males and 28 females). The instrument consisted of five essay questions comprising 25 scored sub-items based on Heller's problem-solving indicators. Data were analyzed using Ministep version 5.10.3 software by examining item validity, reliability, item discrimination, item difficulty, Differential Item Functioning (DIF), and the Wright Map. The results showed that 21 sub-items met all item fit criteria, while four sub-items met two of the three criteria. The instrument demonstrated good psychometric quality with a Cronbach's Alpha of 0.89, Person Reliability of 0.88, and Item Reliability of 0.96. Item discrimination was generally good, only four sub-items were identified with DIF, and the Wright Map indicated good targeting between students' abilities and item difficulty. Therefore, the instrument is valid, reliable, fair, and suitable for measuring students' problem-solving ability. It can also serve as an alternative assessment tool for teachers and researchers in physics education.

Keywords: Physics Assessment; Problem-Solving Ability; *Rasch* Model; Temperature And Heat; Test Instrument.

Citation: Padilah, S. L., Aviyanti, L., & Nurdini, N. (2026). Analisis Kualitas Instrumen Kemampuan Pemecahan Masalah Siswa SMA pada Materi Suhu dan Kalor Menggunakan Model *Rasch*. *Jurnal Pendidikan, Sains, Geologi, Dan Geofisika (GeoScienceEd Journal)*, 7(3), 4715–4725. <https://doi.org/10.29303/goescienceed.v7i3.2895>

Pendahuluan

Pendidikan merupakan proses yang berperan penting dalam mempersiapkan sumber daya manusia yang mampu menghadapi perkembangan ilmu pengetahuan, teknologi, serta berbagai tantangan kehidupan. Seiring dengan perubahan paradigma pembelajaran pada abad ke-21, proses pembelajaran tidak lagi hanya berorientasi pada penguasaan pengetahuan, tetapi juga pada pengembangan berbagai keterampilan yang diperlukan peserta didik, salah satunya kemampuan pemecahan masalah (Fryda dkk., 2021). Dalam pembelajaran fisika, kemampuan pemecahan masalah memiliki peranan penting karena peserta didik tidak hanya dituntut memahami konsep,

prinsip, dan hukum fisika, tetapi juga mampu menerapkan konsep tersebut untuk menganalisis serta menyelesaikan berbagai permasalahan yang berkaitan dengan fenomena alam maupun kehidupan sehari-hari (Jayadi dkk. 2020). Oleh karena itu, pengembangan kemampuan pemecahan masalah menjadi salah satu tujuan penting dalam pembelajaran fisika.

Kemampuan pemecahan masalah merupakan kemampuan yang diperlukan peserta didik untuk menyelesaikan permasalahan secara sistematis melalui penerapan konsep dan prinsip yang telah dipelajari. Dalam pembelajaran fisika, keterampilan ini tidak hanya berorientasi pada hasil akhir penyelesaian masalah, tetapi juga pada proses berpikir peserta didik dalam

memahami permasalahan, mengidentifikasi konsep fisika yang relevan, menentukan strategi penyelesaian, hingga mengevaluasi solusi yang diperoleh. Heller dkk. (1992), mengemukakan bahwa proses pemecahan masalah terdiri atas lima tahapan, yaitu memvisualisasikan masalah (*visualize the problem*), mendeskripsikan masalah ke dalam konsep fisika (*describe the problem in physics term*), merencanakan solusi (*plan the solution*), menerapkan solusi (*execute the plan*), dan mengevaluasi solusi (*evaluate the answer*). Kelima tahapan tersebut saling berkaitan dan menggambarkan proses berpikir ilmiah yang diperlukan peserta didik dalam menyelesaikan permasalahan fisika. Oleh karena itu, kelima indikator tersebut dijadikan acuan dalam mengembangkan instrumen yang mampu mengukur kemampuan pemecahan masalah peserta didik secara komprehensif (Ratnaningdyah, 2017).

Berbagai hasil penelitian menunjukkan bahwa kemampuan pemecahan masalah peserta didik dalam pembelajaran fisika masih tergolong rendah. Sari dkk. (2023), menunjukan bahwa kemampuan pemecahan masalah siswa SMA berada pada kategori sedang hingga rendah karena peserta didik belum terbiasa berpikir secara reflektif dan sistematis dalam menyusun langkah penyelesaian. Temuan tersebut diperkuat oleh Walidain dkk. (2023), yang menunjukan bahwa sebagian besar peserta didik masih belum mampu menyelesaikan permasalahan fisika dengan baik, terutama pada tahapan penyusunan dan penerapan strategi penyelesaian masalah. Kondisi tersebut mengindikasikan bahwa kemampuan pemecahan masalah peserta didik masih perlu mendapat perhatian sehingga diperlukan pengukuran yang mampu memberikan gambaran kemampuan peserta didik secara akurat sebagai dasar dalam menentukan tindak lanjut pembelajaran.

Untuk mengukur kemampuan pemecahan masalah peserta didik secara akurat, diperlukan instrumen penilaian yang memiliki kualitas pengukuran yang baik. Instrumen yang berkualitas diharapkan mampu mengukur konstruk yang dituju secara tepat sehingga menghasilkan data yang mencerminkan kemampuan peserta didik secara objektif. Oleh karena itu, analisis kualitas instrumen menjadi tahapan penting untuk memastikan bahwa setiap butir soal berfungsi sesuai dengan tujuan pengukuran sehingga hasil yang diperoleh dapat dipercaya.

Salah satu pendekatan yang banyak digunakan untuk menganalisis kualitas instrumen adalah *Model Rasch*, yang merupakan bagian dari *Item Response Theory* (IRT). *Model Rasch* mampu memberikan informasi mengenai karakteristik butir soal dan kemampuan peserta didik secara lebih komprehensif dibandingkan pendekatan teori tes klasik. Selain menganalisis validitas dan reliabilitas instrumen, *Model Rasch* juga mampu

memberikan informasi mengenai tingkat kesukaran butir, kesesuaian butir terhadap model (*item fit*), serta kemampuan instrumen dalam membedakan tingkat kemampuan peserta didik. Analisis tersebut disajikan pada skala *logit* sehingga karakteristik butir dan kemampuan peserta didik dapat dibandingkan secara langsung (Sumintono & Widhiarso, 2015). Berbagai penelitian menunjukkan bahwa *Model Rasch* efektif digunakan untuk mengevaluasi kualitas instrumen karena mampu menghasilkan informasi psikometrik yang lebih objektif dan komprehensif (Qirom & Nurlaelah, 2020; Susdelina dkk., 2018).

Analisis kualitas instrumen menggunakan *Model Rasch* telah banyak diterapkan pada berbagai instrumen Pendidikan. Namun, penerapannya pada instrumen kemampuan pemecahan masalah fisika yang disusun berdasarkan indikator Heller, khususnya pada materi Suhu dan Kalor masih terbatas. Dengan demikian, masih diperlukan penelitian yang mengevaluasi kualitas instrumen kemampuan pemecahan masalah fisika yang dikembangkan berdasarkan indikator Heller pada materi tertentu. Oleh karena itu, penelitian ini bertujuan untuk menganalisis kualitas instrumen kemampuan pemecahan masalah fisika pada materi Suhu dan Kalor menggunakan *Model Rasch* yang ditinjau dari aspek validitas, reliabilitas, daya pembeda, tingkat kesukaran butir soal, *Differential Item Functioning* (DIF), dan *Wright Map* yang digunakan untuk membandingkan distribusi kemampuan peserta didik dan tingkat kesukaran butir.

Metode

Penelitian ini merupakan penelitian kuantitatif dengan desain penelitian deskriptif. Penelitian bertujuan untuk menganalisis karakteristik psikometrik instrumen kemampuan pemecahan masalah fisika menggunakan *Model Rasch*. Subjek penelitian dipilih menggunakan teknik *purposive sampling*, yaitu berdasarkan kriteria peserta didik yang berasal dari kelas yang tersedia dan telah mempelajari materi Suhu dan Kalor. Subjek penelitian terdiri atas 54 peserta didik SMA, yang terdiri atas 26 peserta didik laki-laki dan 28 peserta didik perempuan. Data yang diperoleh kemudian dianalisis menggunakan perangkat lunak Ministep versi 5.10.3 untuk mengevaluasi validitas butir, reliabilitas instrumen, daya pembeda, *Differential Item Functioning* (DIF), tingkat kesukaran butir, serta kesesuaian antara kemampuan peserta didik dan tingkat kesukaran butir melalui *Wright Map*.

Instrumen penelitian berupa lima soal uraian pada materi Suhu dan Kalor yang disusun berdasarkan lima indikator kemampuan pemecahan masalah menurut Heller, yaitu memvisualisasikan masalah, mendeskripsikan masalah ke dalam konsep fisika, merencanakan solusi, menerapkan solusi, dan mengevaluasi solusi. Setiap soal terdiri atas lima

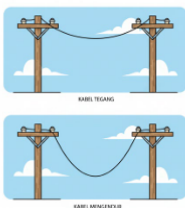
subbutir yang merepresentasikan masing-masing indikator kemampuan pemecahan masalah. Untuk memudahkan identifikasi pada proses analisis, setiap subbutir diberi kode berdasarkan indikator, yaitu A (memvisualisasikan masalah), B (mendeskripsikan masalah ke dalam konsep fisika), C (merencanakan solusi), D (menerapkan solusi), dan E (mengevaluasi solusi). Angka setelah huruf menunjukkan nomor soal, sehingga kode A1 menunjukkan indikator memvisualisasikan masalah pada soal nomor 1, sedangkan E5 menunjukkan indikator mengevaluasi solusi pada soal nomor 5. Dengan demikian, terdapat 25 subbutir penilaian. Penskoran dilakukan menggunakan rubrik analitik dengan skor maksimum 20 pada setiap soal. Dengan demikian, skor maksimum keseluruhan instrumen adalah 100. Berdasarkan karakteristik tersebut, spesifikasi instrumen kemampuan pemecahan masalah yang dikembangkan disajikan pada Tabel 1.

Tabel 1. Spesifikasi Instrumen Kemampuan Pemecahan Masalah

Kode Indikator	Indikator Kemampuan Pemecahan Masalah	Skor Maksimum
A	Memvisualisasikan masalah	3
B	Mendeskripsikan masalah ke dalam konsep fisika	4
C	Merencanakan solusi	4
D	Menerapkan solusi	6
E	Mengevaluasi solusi	3

Untuk memberikan gambaran mengenai karakteristik instrumen yang dikembangkan, salah satu contoh subbutir disajikan pada Gambar 1.

4. Perhatikan gambar di bawah ini!



Pada suatu hari yang cerah, udara sejak pagi terasa cukup sejuk. Beberapa warga yang sedang beraktivitas di luar rumah memperhatikan bahwa kabel listrik yang membentang di antara dua tiang terlihat cukup tegang seperti biasanya. Kabel tersebut sudah terpasang sejak lama dan panjangnya sekitar 20 meter ketika pertama kali dipasang pada suhu udara sekitar 25°C.

Menjelang siang hari, matahari bersinar lebih terik dari biasanya. Permukaan jalan mulai terasa panas dan udara menjadi lebih hangat dibandingkan pagi hari. Tanpa disadari, suhu udara meningkat hingga sekitar 34°C. Salah satu warga kemudian memperhatikan bahwa kabel listrik yang tadi pagi tampak tegang kini terlihat sedikit lebih kendur dan melengkung ke bawah.

Padahal tidak ada beban tambahan yang digantungkan pada kabel tersebut. Tidak ada layang-layang tersangkut, tidak ada benda lain yang menempel, dan tidak ada perubahan pada jarak antara kedua tiang. Kabel tersebut terbuat dari aluminium yang memiliki koefisien muai panjang sebesar $2,4 \times 10^{-5}/^{\circ}\text{C}$. Beberapa warga menduga kabel tersebut mengendur karena usia pemakaian, sementara yang lain berpendapat bahwa perubahan suhu mungkin memengaruhi panjang kabel. Diasumsikan kabel dapat memuai bebas dan tidak terhambat oleh gaya luar.

a. Visualisasikan permasalahan tersebut dengan sederhana dan menurutmu, apa yang sebenarnya terjadi pada kabel tersebut ketika suhu udara meningkat?.

Gambar 1. Contoh subbutir A4 pada instrument kemampuan pemecahan masalah

Contoh yang ditampilkan merupakan subbutir A4 yang mengukur indikator memvisualisasikan masalah. Melalui contoh tersebut, pembaca dapat memperoleh gambaran mengenai bentuk stimulus, konteks permasalahan, serta penyusunan subbutir pada instrumen kemampuan pemecahan masalah.

Subbutir A4 mengukur kemampuan peserta didik dalam memvisualisasikan permasalahan melalui identifikasi informasi yang diketahui, informasi yang ditanyakan, serta penyajian kondisi permasalahan ke dalam representasi yang sesuai sebagai dasar penyusunan langkah penyelesaian.

Data hasil uji coba instrumen dianalisis menggunakan *Model Rasch* dengan bantuan perangkat lunak Ministep versi 5.10.3. Analisis dilakukan untuk memperoleh informasi mengenai validitas butir, reliabilitas instrumen, daya pembeda, *Differential Item Functioning* (DIF), tingkat kesukaran butir, serta kesesuaian antara kemampuan peserta didik dan tingkat kesukaran butir melalui *Wright Map*. Validitas butir dianalisis menggunakan statistik *item fit*, sedangkan daya pembeda butir ditinjau berdasarkan nilai *Point Measure Correlation* (PTMEA Corr). Reliabilitas instrumen dianalisis berdasarkan nilai *Cronbach's Alpha*, *Person Reliability*, serta *Item Reliability*. Analisis *Differential Item Functioning* digunakan untuk mengidentifikasi potensi keberpihakan butir berdasarkan *gender*. Interpretasi hasil analisis dilakukan berdasarkan parameter *Model Rasch* sebagaimana dijelaskan oleh Sumintono & Widhiarso (2015).

Instrumen terlebih dahulu divalidasi oleh lima orang validator yang terdiri atas dua dosen ahli dan tiga guru fisika menggunakan indeks *Aiken's V*. Validasi dilakukan terhadap empat aspek, yaitu kesesuaian indikator kemampuan pemecahan masalah dengan butir soal, kesesuaian alternatif jawaban dengan butir soal, kesesuaian rubrik penilaian dengan jawaban pada butir soal, dan kesesuaian tata bahasa dalam butir soal menggunakan skala penilaian 1–4. Instrumen yang telah memenuhi kriteria validitas isi kemudian diujicobakan kepada 54 peserta didik. Data hasil uji coba selanjutnya dianalisis menggunakan *Model Rasch* melalui perangkat lunak Ministep versi 5.10.3 untuk memperoleh informasi mengenai validitas butir, reliabilitas instrumen, daya pembeda, *Differential Item Functioning* (DIF), tingkat kesukaran butir, dan *Wright Map*.

Hasil dan Diskusi

Analisis Validitas Butir Instrumen

Validitas butir merupakan tingkat ketepatan suatu instrumen dalam mengukur konstruk yang seharusnya diukur (Sumintono & Widhiarso, 2015). Pada penelitian ini, validitas butir dianalisis menggunakan *Model Rasch* untuk mengetahui tingkat kesesuaian setiap butir soal dalam mengukur konstruk

kemampuan pemecahan masalah fisika pada materi Suhu dan Kalor. Analisis dilakukan berdasarkan tiga parameter *item fit*, yaitu *Outfit Mean Square* (MNSQ), *Outfit Z-Standard* (ZSTD), dan *Point Measure Correlation* (PTMEA Corr). Suatu butir dinyatakan sesuai apabila memenuhi ketiga parameter tersebut, sedangkan butir yang hanya memenuhi dua parameter masih dapat dipertimbangkan untuk dipertahankan apabila penyimpangannya relatif kecil. Kriteria kesesuaian butir yang digunakan dalam penelitian ini disajikan pada Tabel 2.

Tabel 2. Kriteria Kesesuaian Butir (Item Fit) Berdasarkan Model *Rasch*.

Parameter	Kriteria
<i>Outfit Mean Square</i> (MNSQ)	$0,50 \leq \text{MNSQ} \leq 1,50$
<i>Outfit Z-Standard</i> (ZSTD)	$-2,00 \leq \text{ZSTD} \leq +2,00$
<i>Point Measure Correlation</i> (PTMEA Corr)	$0,40 \leq \text{PTMEA Corr} \leq 0,85$

(Sumintono & Widhiarso, 2015)

Kualitas butir soal ditentukan melalui hasil analisis ketiga parameter yang kemudian diklasifikasikan berdasarkan acuan penilaian yang disajikan pada Tabel 3.

Tabel 3. Acuan Penilaian Kualitas Butir Soal Berdasarkan *Fit-Statistic*.

Pemenuhan Kriteria Nilai	Kategori
Seluruh kriteria terpenuhi	Sangat sesuai
Dua dari tiga kriteria terpenuhi	Sesuai
Hanya satu dari kriteria terpenuhi	Kurang sesuai
Tidak ada kriteria terpenuhi	Tidak sesuai

Tabel 4. Hasil Uji Validitas Butir Berdasarkan Model *Rasch*

Subbutir (3 Parameter <i>Item Fit</i> terpenuhi)	Kategori	Subbutir (2 Parameter <i>Item Fit</i> Terpenuhi)	Kategori
A1, A3	Sangat sesuai	A2, A4, A5	Sesuai
B1, B2, B3, B4	Sangat sesuai	B5	Sesuai
C1, C2, C3, C4, C5	Sangat sesuai		
D1, D2, D3, D4, D5	Sangat sesuai		
E1, E2, E3, E4, E5			

Sumber: Output *Item Fit Order* Ministep.

Berdasarkan Tabel 4, sebanyak 21 dari 25 subbutir memenuhi ketiga parameter *item fit* sehingga termasuk dalam kategori sangat sesuai. Subbutir tersebut tersebar pada seluruh indikator kemampuan pemecahan masalah, dengan seluruh subbutir pada indikator

(Sumintono & Widhiarso, 2015)

Hasil analisis validitas butir diperoleh melalui output table 10: *Item Fit Order Dimensionality* pada perangkat lunak Ministep. Output tersebut menampilkan nilai *Outfit* MNSQ, *Outfit* ZSTD, dan PTMEA Corr untuk setiap butir soal sebagai dasar dalam menentukan tingkat kesesuaian butir terhadap Model *Rasch*. Output analisis tersebut disajikan pada Gambar 2.

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEAS CORR.	EXP.	EXACT OBS%	MATCH EXP%	Item
22	147	54	-.02	.18	1.42	1.85	1.42	1.85	.31	.50	44.4	50.9	B5
9	211	54	-1.80	.15	1.35	1.80	1.40	1.99	.61	.54	46.3	44.0	D2
4	210	54	-1.77	.15	1.27	1.41	1.35	1.77	.62	.54	42.6	44.0	D1
2	155	54	-.27	.18	1.16	.79	1.17	.84	.47	.50	40.7	50.2	B1
13	143	54	.10	.18	1.16	.83	1.17	.84	.58	.50	38.9	50.7	C3
23	142	54	.13	.18	1.16	.79	1.16	.80	.55	.51	31.5	50.6	C5
24	202	54	-1.58	.16	1.07	.41	1.08	.45	.71	.53	57.4	44.6	D5
8	158	54	-.12	.18	1.06	.34	1.05	.30	.60	.50	42.6	50.8	C2
7	149	54	-.09	.18	1.03	.19	1.03	.21	.57	.50	40.7	50.9	B2
18	145	54	.04	.18	1.02	.18	1.03	.20	.50	.50	38.9	50.8	C4
21	117	54	.90	.17	1.03	.19	.99	.02	.37	.53	51.9	47.0	A5
1	110	54	1.11	.17	.98	-.02	.99	.01	.50	.54	42.6	46.8	A1
19	195	54	-1.41	.16	.99	.03	.96	-.13	.58	.52	42.6	45.9	D4
12	137	54	.29	.18	.96	-.14	.95	-.16	.60	.51	42.6	49.9	B3
25	115	54	.96	.17	.96	-.13	.96	-.13	.47	.53	42.6	47.1	E5
3	165	54	-.58	.17	.89	-.48	.92	-.30	.53	.50	51.9	48.5	C1
16	125	54	.66	.17	.91	-.42	.86	-.64	.32	.52	50.0	48.4	A4
14	190	54	-1.28	.16	.86	-.67	.88	-.57	.70	.52	42.6	46.2	D3
17	146	54	.01	.18	.88	-.55	.88	-.54	.50	.50	46.3	50.8	B4
20	116	54	.93	.17	.80	-1.02	.77	-1.22	.51	.53	64.8	47.0	E4
11	116	54	.93	.17	.72	-1.53	.68	-1.74	.53	.53	63.0	47.0	A3
6	131	54	.47	.18	.71	-1.51	.69	-1.65	.39	.51	53.7	49.0	A2
5	120	54	.81	.17	.70	-1.66	.68	-1.75	.44	.52	46.3	47.3	E1
15	117	54	.90	.17	.69	-1.69	.68	-1.72	.45	.53	46.3	47.0	E3
10	124	54	.69	.17	.68	-1.74	.66	-1.88	.47	.52	46.3	48.1	E2
MEAN	147.1	54.0	.00	.17	.98	-.11	.98	-.13			46.3	48.1	
P.SD	30.9	.0	.90	.01	.20	1.02	.22	1.11			7.4	2.2	

Gambar 2. Output *Item Fit Order Dimensionality*

Berdasarkan hasil analisis pada output *Item Fit Order*, ringkasan kualitas setiap subbutir berdasarkan indikator kemampuan pemecahan masalah disajikan pada Tabel 4.

merencanakan solusi (C), menerapkan solusi (D), dan mengevaluasi solusi (D) memenuhi seluruh parameter *item fit*. Sementara itu, terdapat empat subbutir, yaitu A2, A4, A5, dan B5 yang memenuhi dua dari tiga parameter *item fit* sehingga termasuk dalam kategori

sesuai. Keempat subbutir tersebut memiliki nilai *Point Measure Correlation* (PTMEA Corr) di bawah batas kriteria, sedangkan nilai *Outfit Mean Square* (MNSQ) dan *Outfit Z-Standard* (ZSTD) masih berada pada rentang yang diterima. Kondisi tersebut menunjukkan bahwa hubungan antara skor pada subbutir dengan kemampuan keseluruhan peserta didik relatif lebih rendah dibandingkan subbutir lainnya. Dengan kata lain, kemampuan keempat subbutir dalam membedakan tingkat kemampuan peserta didik belum seoptimal subbutir lain. Namun, karena masih memenuhi minimal dua parameter *item fit*, keempat subbutir tetap dinyatakan sesuai dengan *Model Rasch* dan layak dipertahankan sebagai bagian dari instrumen.

Secara keseluruhan, hasil analisis menunjukkan bahwa instrumen memiliki validitas yang baik. Dominasi subbutir yang memenuhi seluruh parameter *item fit* menunjukkan bahwa instrumen telah mampu merepresentasikan konstruk kemampuan pemecahan masalah sesuai dengan indikator yang dikembangkan. Oleh karena itu, instrumen dinilai layak digunakan untuk mengukur kemampuan pemecahan masalah peserta didik pada materi Suhu dan Kalor.

Temuan ini sejalan dengan penelitian Qirom & Nurlaelah (2020) serta Andriani dkk. (2026) yang menunjukkan bahwa butir yang masih memenuhi kriteria *item fit* dapat dipertahankan karena tetap mampu mengukur konstruk yang dimaksud, meskipun terdapat salah satu parameter yang belum optimal. Beberapa butir yang memiliki nilai *fit statistics* relatif kurang baik tidak serta-merta dieliminasi, tetapi perlu ditinjau kembali dari aspek redaksi, rubrik penskoran, dan kesesuaian indikator agar kualitas instrumen semakin baik.

Analisis Reliabilitas

Reliabilitas instrumen menunjukkan tingkat konsistensi hasil pengukuran ketika instrumen digunakan untuk mengukur kemampuan peserta didik. Pada penelitian ini, reliabilitas dianalisis menggunakan pendekatan *Model Rasch* melalui output *Summary Statistics* pada perangkat lunak Ministep. Analisis reliabilitas meliputi nilai *Cronbach's Alpha* (KR-20), *Person Reliability*, dan *Item Reliability*. *Cronbach's Alpha* digunakan untuk menunjukkan konsistensi internal instrumen secara keseluruhan, *Person Reliability* menunjukkan konsistensi kemampuan instrumen dalam membedakan kemampuan peserta didik, sedangkan *Item Reliability* menunjukkan konsistensi tingkat kesukaran antarbutir soal (Sumintono & Widhiarso, 2015).

Panduan interpretasi nilai *Cronbach's Alpha* yang digunakan pada penelitian ini mengacu pada kriteria yang dikemukakan oleh Sumintono dan Widhiarso (2015) sebagaimana disajikan pada Tabel 5.

Tabel 5. Acuan Interpretasi Nilai Cronbach's Alpha

Rentang Nilai	Interpretasi
KR-20 < 0,50	Buruk
0,50 ≤ KR-20 < 0,60	Jelek
0,60 ≤ KR-20 < 0,70	Cukup
0,70 ≤ KR-20 < 0,80	Bagus
KR-20 ≥ 0,80	Sangat bagus

(Sumintono & Widhiarso, 2015)

Selain *Cronbach's Alpha*, analisis reliabilitas pada *Rasch Model* juga mempertimbangkan nilai *Person Reliability* dan *Item Reliability* untuk mengevaluasi konsistensi respons peserta didik dan kualitas butir soal. Acuan interpretasi kedua parameter tersebut disajikan pada Tabel 6.

Tabel 6. Acuan Interpretasi Nilai Person Reliability dan Item Reliability

Skala	Interpretasi
< 0,67	Lemah
0,67 < r ≤ 0,80	Cukup
0,80 < r ≤ 0,90	Baik
0,90 < r ≤ 0,94	Sangat baik
> 0,94	Istimewa

(Sumintono & Widhiarso, 2015)

Selanjutnya, hasil analisis reliabilitas diperoleh melalui output *Summary Statistics* pada perangkat lunak Ministep versi 5.10.3 sebagaimana ditampilkan pada Gambar 3.

SUMMARY OF 54 MEASURED Person									
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		
					MNSQ	ZSTD	MNSQ	ZSTD	
MEAN	68.1	25.0	-.31	.25	.98	-.10	.98	-.12	
SEM	1.7	.0	.11	.00	.05	.17	.05	.17	
P.SD	12.1	.0	.77	.01	.40	1.24	.39	1.22	
S.SD	12.2	.0	.78	.01	.40	1.25	.40	1.23	
MAX.	87.0	25.0	.87	.26	2.44	3.47	2.34	3.54	
MIN.	40.0	25.0	-2.06	.24	.55	-1.78	.54	-1.78	
REAL RMSE	.27	TRUE SD	.72	SEPARATION	2.69	Person RELIABILITY .88			
MODEL RMSE	.25	TRUE SD	.73	SEPARATION	2.90	Person RELIABILITY .89			
S.E. OF Person MEAN = .11									
Person RAW SCORE-TO-MEASURE CORRELATION = 1.00									
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .89 SEM = 4.06									
STANDARDIZED (50 ITEM) RELIABILITY = .94									
SUMMARY OF 25 MEASURED Item									
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		
					MNSQ	ZSTD	MNSQ	ZSTD	
MEAN	147.1	54.0	.00	.17	.98	-.11	.98	-.13	
SEM	6.3	.0	.18	.00	.04	.21	.04	.23	
P.SD	30.9	.0	.90	.01	.20	1.02	.22	1.11	
S.SD	31.6	.0	.92	.01	.20	1.04	.22	1.13	
MAX.	211.0	54.0	1.11	.18	1.42	1.85	1.42	1.99	
MIN.	110.0	54.0	-1.80	.15	.68	-1.74	.66	-1.88	
REAL RMSE	.18	TRUE SD	.88	SEPARATION	4.99	Item RELIABILITY .96			
MODEL RMSE	.17	TRUE SD	.89	SEPARATION	5.16	Item RELIABILITY .96			
S.E. OF Item MEAN = .18									
Item RAW SCORE-TO-MEASURE CORRELATION = -1.00									
Global statistics: please see Table 44.									
UMEAN=.0000 USCALE=1.0000									

Gambar 3. Output *Summary Statistics* Analisis Reliabilitas Instrumen

Berdasarkan output tersebut, ringkasan hasil analisis reliabilitas instrumen disajikan pada Tabel 7.

Tabel 7. Hasil Analisis Reliabilitas Instrumen

Parameter	Nilai	Interpretasi
<i>Cronbach's Alpha</i>	0,89	Sangat bagus
<i>Person Reliability</i>	0,88	Baik
<i>Item Reliability</i>	0,96	Istimewa

Sumber: Output *Summery Statistic* Ministep

Berdasarkan hasil analisis pada Tabel 7, instrumen memperoleh nilai *Cronbach's Alpha* sebesar 0,89 yang termasuk dalam kategori sangat bagus. Nilai tersebut menunjukkan bahwa instrumen memiliki konsistensi internal yang tinggi, sehingga setiap subbutir secara keseluruhan mampu mengukur konstruk kemampuan pemecahan masalah secara konsisten. Dengan demikian, hasil pengukuran yang diperoleh memiliki tingkat kepercayaan yang baik dan relatif stabil.

Nilai *Person Reliability* sebesar 0,88 juga berada pada kategori baik. Hasil ini menunjukkan bahwa instrumen mampu membedakan kemampuan peserta didik secara konsisten. Dengan kata lain, perbedaan skor yang diperoleh peserta didik lebih mencerminkan perbedaan kemampuan yang sebenarnya dibandingkan dengan kesalahan pengukuran. Hal ini mengindikasikan bahwa instrumen memiliki kemampuan yang baik dalam mengelompokkan peserta didik berdasarkan tingkat kemampuan pemecahan masalah.

Sementara itu, nilai *Item Reliability* sebesar 0,96 termasuk dalam kategori istimewa. Nilai tersebut menunjukkan bahwa tingkat kesukaran antarbutir memiliki konsistensi yang sangat tinggi, sehingga urutan tingkat kesukaran setiap subbutir diperkirakan akan tetap stabil apabila instrumen digunakan pada kelompok peserta didik lain dengan karakteristik yang serupa. Tingginya nilai *Item Reliability* juga mengindikasikan bahwa penyebaran tingkat kesukaran antarbutir telah tersusun dengan baik, sehingga instrumen mampu mengukur kemampuan peserta didik pada berbagai tingkat kemampuan.

Secara keseluruhan, hasil analisis reliabilitas menunjukkan bahwa instrumen memiliki tingkat konsistensi yang baik hingga istimewa. Hal ini memperkuat bahwa instrumen layak digunakan sebagai alat ukur kemampuan pemecahan masalah pada materi Suhu dan Kalor karena mampu menghasilkan data yang konsisten dan dapat dipercaya.

Daya Pembeda Butir

Daya pembeda butir menunjukkan kemampuan setiap subbutir dalam membedakan peserta didik yang memiliki kemampuan tinggi dan kemampuan rendah.

Pada penelitian ini, daya pembeda dianalisis menggunakan pendekatan *Model Rasch* melalui nilai *Point Measure Correlation* (PTMEA Corr) pada output *Item Fit Order* perangkat lunak *Ministep*. Nilai *Point Measure Correlation* menggambarkan hubungan antara skor pada setiap subbutir dengan kemampuan keseluruhan peserta didik. Semakin tinggi nilai PTMEA Corr, semakin baik kemampuan subbutir dalam membedakan kemampuan peserta didik.

Berdasarkan Smiley (2015), subbutir dengan nilai *Point Measure Correlation* (PTMEA Corr) > 40 dikategorikan memiliki daya pembeda yang baik, kriteria tersebut disajikan pada Tabel 8.

Tabel 8. Acuan Interpretasi Daya Pembeda

Nilai Pt. Measure Corr.	Interpretasi
$Pt. Measure Corr. \geq 0,40$	Sangat Baik
$0,30 \leq Pt. Measure Corr. < 0,40$	Baik
$0,20 \leq Pt. Measure Corr. < 0,30$	Kurang Baik
$Pt. Measure Corr. < 0,20$	Buruk

(Smiley, 2015)

Hasil analisis daya pembeda diperoleh melalui output *Item Fit Order* pada perangkat lunak *Ministep* sebagaimana disajikan pada Gambar 4.

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEASUR- CORR.	AL- EXP.	EXACT OBS%	MATCH EXP%	Item
22	147	54	-.02	.18	1.42	1.85	1.42	1.85	A .31	.50	44.4	50.9	B5
9	211	54	-1.80	.15	1.35	1.80	1.40	1.99	B .61	.54	46.3	44.0	D2
4	210	54	-1.77	.15	1.27	1.41	1.35	1.77	C .62	.54	42.6	44.0	D1
2	155	54	-.27	.18	1.16	.79	1.17	.84	D .47	.50	40.7	50.2	B1
13	143	54	.10	.18	1.16	.83	1.17	.84	E .58	.50	38.9	50.7	C3
23	142	54	.13	.18	1.16	.79	1.16	.80	F .55	.51	31.5	50.6	C5
24	202	54	-1.58	.16	1.07	.41	1.08	.45	G .71	.53	57.4	44.6	D5
8	150	54	-.12	.18	1.06	.34	1.05	.30	H .60	.50	42.6	50.8	C2
7	149	54	-.09	.18	1.03	.19	1.03	.21	I .57	.50	40.7	50.9	B2
18	145	54	.04	.18	1.02	.18	1.03	.20	J .50	.50	38.9	50.8	C4
21	117	54	.90	.17	1.03	.19	.99	.02	K .37	.53	51.9	47.0	A5
1	110	54	1.11	.17	.98	-.02	.99	.01	L .50	.54	42.6	46.8	A1
19	195	54	-1.41	.16	.99	.03	.96	-.13	M .58	.52	42.6	45.9	D4
12	137	54	.29	.18	.96	-.14	.95	-.16	N .60	.51	42.6	49.9	B3
25	115	54	.96	.17	.96	-.13	.96	-.13	K .47	.53	42.6	47.1	E5
3	165	54	-.58	.17	.89	-.48	.92	-.30	J .53	.50	51.9	48.5	C1
16	125	54	.66	.17	.91	-.42	.86	-.64	I .32	.52	50.0	48.4	A4
14	190	54	-1.28	.16	.86	-.67	.88	-.57	H .70	.52	42.6	46.2	D3
17	146	54	.01	.18	.88	-.55	.88	-.54	G .50	.50	46.3	50.8	B4
20	116	54	.93	.17	.80	-.02	.77	-1.22	F .51	.53	64.8	47.0	E4
11	116	54	.93	.17	.72	-1.53	.68	-1.74	E .53	.53	63.0	47.0	A3
6	131	54	.47	.18	.71	-1.51	.69	-1.65	D .39	.51	53.7	49.0	A2
5	120	54	.81	.17	.70	-1.66	.68	-1.75	C .44	.52	46.3	47.3	E1
15	117	54	.90	.17	.69	-1.69	.68	-1.72	B .45	.53	46.3	47.0	E3
10	124	54	.69	.17	.68	-1.74	.66	-1.88	A .47	.52	46.3	48.1	E2
MEAN	147.1	54.0	.00	.17	.98	-.11	.98	-.13			46.3	48.1	
P.SD	30.9	.0	.90	.01	.20	1.02	.22	1.11			7.4	2.2	

Gambar 4. Output *Pt. Measure Corr*

Ringkasan hasil analisis daya pembeda berdasarkan indikator kemampuan pemecahan masalah disajikan pada Tabel 9. Berdasarkan Tabel 9, sebanyak 21 dari 25 subbutir memiliki nilai *Point Measure Correlation* (PTMEA Corr) > 0,40 sehingga termasuk kategori sangat baik. Subbutir tersebut tersebar pada seluruh indikator kemampuan pemecahan masalah. Sementara itu, terdapat empat subbutir, yaitu A2, A4, A5, dan B5, yang memiliki nilai *Pt. Measure Corr* di bawah 0,40, tetapi masih termasuk dalam kategori baik.

Nilai *Pt. Measure Corr* yang kurang dari 0,40 menunjukkan bahwa hubungan antara skor pada keempat subbutir dengan kemampuan keseluruhan peserta didik belum sekuat subbutir lainnya. Subbutir A2, A4, dan A5 berada pada indikator memvisualisasikan masalah, sehingga mengindikasikan bahwa kemampuan peserta didik dalam mengidentifikasi informasi dan merepresentasikan permasalahan masih belum mampu membedakan secara optimal antara peserta didik berkemampuan tinggi dan rendah. Sementara itu, subbutir B5 berada

pada indikator mendeskripsikan masalah ke dalam konsep fisika, yang menunjukkan bahwa kemampuan menghubungkan permasalahan dengan konsep fisika masih memberikan pola respon yang relatif serupa pada sebagian peserta didik. Meskipun demikian, seluruh nilai *Pt. Measure Corr* masih bernilai positif serta keempat subbutir tetap memenuhi parameter *Outfit Mean Square* (MNSQ) dan *Outfit Z-Standard* (ZSTD), sehingga masih dinyatakan layak digunakan dalam instrumen.

Tabel 9. Hasil Analisis Daya Pembeda Butir Instrumen

<i>Pt. Measure Corr</i> >0,40	Kategori	<i>Pt. Measure Corr</i> < 0,40	Kategori
A1, A3	Sangat Baik	A2, A4, A5	Baik
B1, B2, B3, B4	Sangat Baik	B5	Baik
C1, C2, C3, C4, C5	Sangat Baik		
D1, D2, D3, D4, D5			
E1, E2, E3, E4, E5			

Keterangan: Daya pembeda ditentukan berdasarkan nilai *Point Measure Correlation* (Pt. Measure Corr).

Secara keseluruhan, hasil analisis menunjukkan bahwa sebagian besar subbutir memiliki daya pembeda yang sangat baik. Hal ini mengindikasikan bahwa instrumen mampu membedakan kemampuan pemecahan masalah peserta didik secara efektif sehingga hasil pengukuran yang diperoleh lebih representatif terhadap kemampuan peserta didik.

Analisis Tingkat Kesukaran Butir

Tingkat kesukaran butir menunjukkan tingkat kesulitan setiap subbutir dalam instrumen berdasarkan nilai *logit* (JMLE Measure). Dalam pendekatan *Model Rasch*, semakin tinggi nilai *logit* suatu subbutir, semakin tinggi tingkat kesukarannya sehingga subbutir tersebut relatif lebih sulit dijawab oleh peserta didik. Sebaliknya, semakin rendah nilai *logit* menunjukkan bahwa subbutir relatif lebih mudah dijawab. Selain itu, nilai *Model S.E.* digunakan untuk menunjukkan ketelitian estimasi tingkat kesukaran setiap subbutir. Analisis tingkat kesukaran dilakukan untuk mengetahui sebaran tingkat kesukaran butir sehingga dapat dievaluasi apakah instrumen memiliki variasi tingkat kesukaran yang memadai.

Tingkat kesukaran setiap butir diklasifikasikan berdasarkan nilai *Measure (logit)* dengan menggunakan nilai rata-rata (*Mean*) dan simpangan baku (*Standard Deviation/SD*). Kriteria klasifikasi tingkat kesukaran disajikan pada Tabel 10.

Tabel 10. Kriteria Tingkat Kesukaran Butir Berdasarkan Nilai *Measure (Logit)*

Rentang Nilai Measure	Kategori
$Measure > Mean + SD$	Sangat sulit
$Mean < Measure \leq Mean + SD$	Sulit
$Mean - SD \leq Measure \leq Mean$	Sedang
$Measure < Mean - SD$	Sangat mudah

(Sumintono & Widhiarso, 2015)

Hasil analisis tingkat kesukaran diperoleh melalui Output *Measure Order* pada perangkat lunak Ministep sebagaimana disajikan pada Gambar 5.

Item STATISTICS: MEASURE ORDER													
ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIIT	OUTFIIT	PTMEASURE	AL	EXACT	HATCH			Item
1	118	54	1.11	.17	.98	-.02	.99	.01	.50	.54	42.6	46.8	A1
25	115	54	-.96	.17	.96	-.13	.96	-.13	.47	.53	42.6	47.1	E5
11	116	54	-.93	.17	.72	-1.53	.68	-1.74	.53	.53	63.0	47.0	A3
20	116	54	-.93	.17	.80	-1.02	.77	-1.22	.51	.53	64.8	47.0	E4
15	117	54	-.90	.17	.69	-1.69	.68	-1.72	.45	.53	46.3	47.0	E3
21	117	54	-.90	.17	1.03	.19	.99	.02	.37	.53	51.9	47.0	A5
5	120	54	-.81	.17	.70	-1.66	.68	-1.75	.44	.52	46.3	47.3	E1
10	124	54	-.69	.17	.68	-1.74	.66	-1.88	.47	.52	46.3	48.1	E2
16	125	54	-.66	.17	.91	-.42	.86	-.64	.32	.52	50.0	48.4	A4
6	131	54	-.47	.18	.71	-1.51	.69	-1.65	.39	.51	53.7	49.0	A2
12	137	54	-.29	.18	.96	-.14	.95	-.16	.60	.51	42.6	49.9	B3
23	142	54	-.13	.18	1.16	.79	1.16	.80	.55	.51	31.5	50.6	C5
13	143	54	-.10	.18	1.16	.83	1.17	.84	.58	.50	30.9	50.7	C3
18	145	54	-.04	.18	1.02	.18	1.03	.20	.50	.50	30.9	50.8	C4
17	146	54	-.01	.18	.88	-.55	.88	-.54	.50	.50	46.3	50.8	B4
22	147	54	-.02	.18	1.42	1.85	1.42	1.85	.31	.50	44.4	50.9	B5
7	149	54	-.09	.18	1.03	.19	1.03	.21	.57	.50	40.7	50.9	B2
8	150	54	-.12	.18	1.06	.34	1.05	.30	.60	.50	42.6	50.8	C2
2	155	54	-.27	.18	1.16	.79	1.17	.84	.47	.50	40.7	50.2	B1
3	165	54	-.58	.17	.89	-.48	.92	-.30	.53	.50	51.9	48.5	C1
14	190	54	-1.28	.16	.86	-.67	.88	-.57	.70	.52	42.6	46.2	D3
19	195	54	-1.41	.16	.99	.03	.96	-.13	.58	.52	42.6	45.9	D4
24	202	54	-1.58	.16	1.07	.41	1.08	.45	.71	.53	57.4	44.6	D5
4	210	54	-1.77	.15	1.27	1.41	1.35	1.77	.62	.54	42.6	44.0	D1
9	211	54	-1.80	.15	1.35	1.80	1.40	1.99	.61	.54	46.3	44.0	D2
MEAN	147.1	54.0	.00	.17	.98	-.11	.98	-.13			46.3	48.1	
P.SD	30.9	.0	.90	.01	.20	1.02	.22	1.11			7.2	2.1	

Gambar 5. Output *Measure Order* Analisis Tingkat Kesukaran Butir.

Selanjutnya, hasil klasifikasi tingkat kesukaran butir soal berdasarkan nilai *Measure* disajikan pada Tabel 11.

Tabel 11. Hasil Klasifikasi Tingkat Kesukaran Butir Soal

Kategori	Rentang <i>Measure</i>	Kode Butir	Jumlah
Sangat sulit	$> 0,90$	A1, E5, A3, E4	4
Sulit	$0,00 - 0,89$	E3, A5, E1, E2, A4, A2, B3, C5, C3, C4	10
Sedang	$-0,89 - 0,00$	B4, B5, B2, C2, B1, C1	6
Sangat mudah	$< -0,89$	D3, D4, D5, D1, D2	5

Berdasarkan Tabel 11, diperoleh 4 subbutir berkategori sangat sulit, 10 subbutir berkategori sulit, 6 subbutir berkategori sedang, dan 5 subbutir berkategori sangat mudah. Variasi tingkat kesukaran tersebut menunjukkan bahwa instrumen memiliki sebaran tingkat kesulitan yang baik sehingga mampu mengukur kemampuan peserta didik pada berbagai tingkat kemampuan.

Indikator memvisualisasikan masalah didominasi oleh subbutir berkategori sulit hingga sangat sulit, yaitu A1, A2, A3, A4, dan A5, yang menunjukkan bahwa peserta didik masih mengalami kesulitan dalam mengidentifikasi informasi penting serta merepresentasikan permasalahan ke dalam bentuk visual sebagai tahap awal pemecahan masalah. Demikian pula pada indikator mengevaluasi solusi, seluruh subbutir (E1–E5) berada pada kategori sulit hingga sangat sulit, yang mengindikasikan bahwa peserta didik masih mengalami kesulitan dalam meninjau kembali ketepatan prosedur maupun hasil penyelesaian yang telah diperoleh. Sebaliknya, seluruh subbutir pada indikator menerapkan solusi (D1–D5) berada pada kategori sangat mudah, yang menunjukkan bahwa peserta didik relatif lebih mampu menerapkan strategi penyelesaian dan melakukan perhitungan setelah langkah penyelesaian berhasil ditentukan.

Temuan bahwa indikator memvisualisasikan masalah menjadi salah satu indikator yang paling sulit sejalan dengan penelitian Ningsih dkk. (2024), yang dimana penelitian tersebut menunjukkan bahwa kemampuan peserta didik pada tahap *describing the problem into physical aspects* masih relatif rendah sebelum diberikan pembelajaran berbasis *problem solving*, yaitu hanya sebesar 51% pada saat *pretest*. Rendahnya kemampuan tersebut disebabkan peserta didik cenderung langsung menggunakan persamaan matematika tanpa terlebih dahulu menganalisis informasi yang terdapat pada soal serta belum terbiasa

merepresentasikan permasalahan ke dalam bentuk fisika. Berbeda dengan penelitian Tanjung dkk. (2024), yang menunjukkan bahwa indikator menerapkan solusi masih berada pada kategori rendah, penelitian ini memperoleh bahwa seluruh subbutir pada indikator menerapkan solusi berada pada kategori sangat mudah. Perbedaan tersebut diduga dipengaruhi oleh karakteristik instrumen, materi suhu dan kalor yang digunakan, serta kemampuan responden dalam melakukan perhitungan setelah strategi penyelesaian berhasil ditentukan.

Sementara itu, indikator mendeskripsikan masalah ke dalam konsep fisika (B1–B5) dan merencanakan solusi (C1–C5) memiliki variasi tingkat kesukaran, mulai dari kategori sedang hingga sulit. Hal ini menunjukkan bahwa kedua indikator tersebut mampu mengakomodasi peserta didik dengan tingkat kemampuan yang berbeda, sehingga tidak terlalu mudah maupun terlalu sulit bagi sebagian besar peserta didik.

Selain nilai *Measure*, hasil analisis juga menunjukkan bahwa nilai *Model S.E.* setiap subbutir berada pada rentang 0,15–0,18, yang mengindikasikan bahwa estimasi tingkat kesukaran memiliki tingkat ketelitian yang baik dengan kesalahan pengukuran yang relatif kecil. Secara keseluruhan, distribusi tingkat kesukaran yang bervariasi dari kategori sangat mudah hingga sangat sulit menunjukkan bahwa instrumen memiliki karakteristik yang baik untuk mengukur kemampuan pemecahan masalah peserta didik secara komprehensif.

Analisis Differential Item Functioning (DIF)

Differential Item Functioning (DIF) merupakan analisis yang digunakan untuk mengidentifikasi adanya perbedaan karakteristik suatu subbutir ketika dikerjakan oleh dua kelompok peserta didik yang memiliki tingkat kemampuan yang setara. Pada

penelitian ini, analisis DIF dilakukan berdasarkan jenis kelamin menggunakan pendekatan *Model Rasch* melalui perangkat lunak *Ministep* versi 5.10.3. Parameter yang digunakan untuk mendeteksi DIF adalah nilai *Rasch-Welch Probability* dan *DIF Contrast*. Subbutir dinyatakan terindikasi mengalami DIF apabila memiliki nilai *Rasch-Welch Probability* < 0,05. Sementara itu, nilai *DIF Contrast* digunakan untuk menunjukkan arah perbedaan tingkat kesukaran subbutir antar kelompok, sehingga dapat diketahui kelompok peserta didik yang relatif lebih diuntungkan dalam menjawab subbutir tersebut.

Berdasarkan Sumintono dan Widhiarso (2015), subbutir yang memiliki nilai *Rasch-Welch Probability* < 0,05 dinyatakan terindikasi mengalami *Differential Item Functioning* (DIF), sedangkan subbutir dengan nilai *Rasch-Welch Probability* ≥ 0,05 dinyatakan tidak terindikasi DIF. Hasil analisis DIF diperoleh melalui output *DIF Class/Group Specification* pada perangkat lunak *Ministep*, sebagaimana disajikan pada Gambar 6.

DIF class/group specification is: DIF=4

Person CLASS/	Obs-Exp Average	DIF MEASURE	DIF S.E.	Person CLASS/	Obs-Exp Average	DIF MEASURE	DIF S.E.	DIF CONTRAST	JOINT RASCH-CHEN	Mantel	Size Active Item					
									S.E.	Prob.	CUMULOR Slices Number	Name				
L	-.34	1.63	.24	P	.31	.60	.24	1.03	.34	3.00	51	.0041	3.9705 .0463	1.66	12	1 A1
L	-.13	-.05	.25	P	.12	-.48	.24	.43	.35	1.22	51	.2268	2.2097 .1371	1.36	12	2 B1
L	-.07	-.69	.25	P	-.06	-.47	.24	-.22	.35	-.62	51	.5364	1.3402 .2456	-.77	12	3 C1
L	-.21	-1.50	.23	P	.20	-2.02	.21	.52	.31	1.68	51	.0992	1.8750 .1709	.87	12	4 O1
L	-.06	.91	.25	P	.06	.71	.24	-.20	.35	.57	51	.5714	.0022 .2625	-.04	12	5 A1
L	.00	.47	.25	P	.00	.47	.24	.00	.35	.00	51	1.000	.7388 .3900	.69	12	6 A2
L	-.41	.60	.25	P	.38	-.70	.24	1.30	.35	3.77	51	.0004	5.8640 .0155	2.00	12	7 B2
L	.03	-.10	.26	P	-.03	-.06	.24	-.12	.35	-.34	51	.7368	.3635 .5466	-.44	12	8 C2
L	-.04	-1.75	.22	P	.04	-1.84	.21	.00	.31	.30	51	.7685	.0102 .3195	.07	12	9 O2
L	-.02	.73	.25	P	.02	.66	.24	.07	.35	.21	51	.8352	.0098 .2212	-.07	12	10 E2
L	-.10	1.10	.25	P	.09	.77	.24	.32	.35	.93	51	.3566	.7305 .3927	.64	12	11 A3
L	-.03	.35	.25	P	.03	.24	.25	.11	.35	.31	51	.7578	.0250 .0744	-.13	12	12 B3
L	.01	.10	.25	P	-.01	.10	.25	.00	.35	.00	51	1.000	.8883 .7653	-.20	12	13 C3
L	-.08	-1.17	.24	P	.07	-1.38	.22	.21	.32	.66	51	.5129	.5649 .4523	.46	12	14 O3
L	.03	.85	.25	P	-.03	.95	.24	-.10	.35	-.30	51	.7644	.2981 .5851	.39	12	15 E3
L	.04	.60	.25	P	-.04	.72	.24	-.12	.35	-.34	51	.7325	.0000 1.000	.00	12	16 A4
L	.11	-.18	.26	P	-.10	.18	.25	-.36	.35	-1.02	51	.3144	.7893 .3743	-.65	12	17 B4
L	.09	-.11	.26	P	-.08	.18	.25	-.29	.35	-.83	51	.4897	1.2327 .2669	-.76	12	18 C4
L	.01	-1.39	.23	P	.02	-1.43	.22	.04	.32	.13	51	.8964	.0160 .3094	.07	12	19 O4
L	-.20	.60	.25	P	-.19	1.24	.24	-.64	.35	-1.04	51	.0712	.5650 .4522	-.56	12	20 E4
L	.22	.54	.25	P	-.21	1.24	.24	-.70	.35	-2.02	51	.0484	1.2777 .2583	-.79	12	21 A5
L	.21	-.37	.25	P	-.19	.30	.25	-.67	.35	-1.90	51	.0626	.4955 .4815	-.43	12	22 B5
L	.18	-.18	.26	P	-.17	.42	.24	-.60	.35	-1.70	51	.0959	.9423 .3317	-.71	12	23 C5
L	.28	-1.94	.22	P	-.25	-1.23	.22	-.71	.31	-2.27	51	.0275	5.9798 .0145	-2.93	12	24 O5
L	-.05	1.04	.25	P	.04	.89	.24	-.14	.35	-.42	51	.6786	.0203 .3067	-.80	12	25 E5

Gambar 6. Output DIF Class/Group Specification

Berdasarkan output yang dihasilkan oleh *Ministep*, berikut ringkasan butir yang terdeteksi adanya *Differential Item Functioning* (DIF) disajikan pada Tabel 12.

Tabel 12. Hasil Analisis DIF Butir Soal

Kode Butir	DIF Contrast	Rasch-Welch Prob.	Interpretasi
A1	+1,03	0,0041	Lebih menguntungkan perempuan Lebih menguntungkan laki-laki
A5	-0,70	0,0464	Lebih menguntungkan perempuan Lebih menguntungkan laki-laki
B2	+1,30	0,0004	
D5	-0,71	0,0275	

Berdasarkan Tabel 12, terdapat empat subbutir, yaitu A1, A5, B2, dan D5, yang terindikasi mengalami *Differential Item Functioning* (DIF), sedangkan 21 subbutir lainnya tidak menunjukkan indikasi DIF. Hasil tersebut menunjukkan bahwa sebagian besar subbutir memiliki karakteristik yang relatif sama ketika digunakan pada peserta didik laki-laki maupun perempuan.

Berdasarkan nilai *DIF Contrast*, subbutir A1 dan B2 memiliki nilai positif, yang menunjukkan bahwa kedua subbutir tersebut relatif lebih sulit dijawab oleh peserta didik laki-laki sehingga cenderung lebih menguntungkan peserta didik perempuan. Subbutir A1 berada pada indikator memvisualisasikan masalah, sedangkan B2 berada pada indikator mendeskripsikan masalah ke dalam konsep fisika, sehingga perbedaan fungsi butir diduga berkaitan dengan perbedaan cara peserta didik dalam merepresentasikan informasi dan menghubungkan permasalahan dengan konsep fisika. Sebaliknya, subbutir A5 dan D5 memiliki nilai *DIF Contrast* negatif, yang menunjukkan bahwa kedua subbutir tersebut relatif lebih sulit dijawab oleh peserta didik perempuan sehingga cenderung lebih

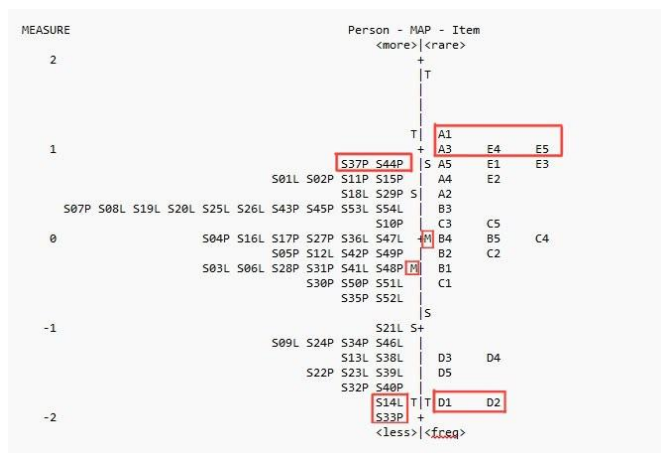
menguntungkan peserta didik laki-laki. Subbutir D5 berada pada indikator menerapkan solusi, sedangkan A5 berada pada indikator memvisualisasikan masalah, yang mengindikasikan bahwa perbedaan respons hanya muncul pada subbutir tertentu, bukan pada keseluruhan indikator.

Meskipun demikian, jumlah subbutir yang terindikasi DIF hanya 4 dari 25 subbutir (16%), sedangkan 21 subbutir (84%) tidak menunjukkan adanya perbedaan fungsi berdasarkan jenis kelamin. Hasil tersebut menunjukkan bahwa secara umum instrumen memiliki tingkat keadilan (*fairness*) yang baik karena sebagian besar subbutir memberikan kesempatan yang setara bagi peserta didik laki-laki maupun perempuan dalam menunjukkan kemampuan pemecahan masalah. Keberadaan empat subbutir yang terindikasi DIF tidak serta-merta menunjukkan bahwa instrumen bersifat bias secara keseluruhan, namun subbutir tersebut dapat menjadi bahan evaluasi pada pengembangan instrumen selanjutnya, terutama dalam penyempurnaan redaksi soal maupun penyajian stimulus agar potensi bias dapat diminimalkan.

Analisis Wright Map

Wright Map merupakan representasi grafis dalam *Model Rasch* yang digunakan untuk membandingkan distribusi kemampuan peserta didik (*person ability*) dengan tingkat kesukaran butir (*item difficulty*) pada satu skala logit yang sama. Melalui *Wright Map*, dapat diketahui kesesuaian (*targeting*) antara kemampuan peserta didik dan tingkat kesukaran instrumen, sehingga memberikan gambaran mengenai apakah instrumen telah sesuai untuk mengukur kemampuan responden. Pada penelitian ini, analisis *Wright Map* dilakukan menggunakan perangkat lunak *Ministep* versi 5.10.3.

Hasil analisis *Wright Map* diperoleh melalui output *Person-Item Map* sebagaimana disajikan pada Gambar 7.



Gambar 7. Output *Wright Map* Hasil Analisis Ministep

Berdasarkan Gambar 7, distribusi kemampuan peserta didik ditampilkan pada sisi kiri (*person*), sedangkan distribusi tingkat kesukaran subbutir ditampilkan pada sisi kanan (*item*). Kedua distribusi tersebut berada pada skala logit yang sama sehingga memungkinkan perbandingan secara langsung antara kemampuan peserta didik dan tingkat kesukaran subbutir. Kode SxxL menunjukkan peserta didik laki-laki, sedangkan kode SxxP menunjukkan peserta didik perempuan.

Pada sisi *person*, peserta didik S37P dan S44P berada pada posisi logit tertinggi sehingga memiliki kemampuan pemecahan masalah paling tinggi dibandingkan peserta didik lainnya. Sebaliknya, S14L dan S33P berada pada posisi logit terendah yang mengindikasikan bahwa kedua peserta didik tersebut memiliki kemampuan pemecahan masalah paling rendah.

Pada sisi *item*, subbutir A1, A3, E4, dan E5 berada pada posisi logit tertinggi sehingga merupakan subbutir dengan tingkat kesukaran paling tinggi. Sebaliknya, subbutir D1 dan D2 berada pada posisi logit terendah

sehingga menjadi subbutir yang paling mudah dijawab oleh peserta didik. Hasil tersebut konsisten dengan analisis tingkat kesukaran berdasarkan nilai *JMLE Measure*.

Selain menggambarkan distribusi kemampuan peserta didik dan tingkat kesukaran subbutir, *Wright Map* juga menunjukkan kesesuaian (*targeting*) antara instrumen dan responden. Berdasarkan Gambar 7, posisi rata-rata kemampuan peserta didik (*Mean Person*) berada sedikit di atas rata-rata tingkat kesukaran subbutir (*Mean Item*), yang mengindikasikan bahwa secara umum instrumen telah sesuai dengan kemampuan peserta didik. Sebagian besar subbutir berada pada rentang kemampuan responden sehingga instrumen mampu mengukur kemampuan pemecahan masalah secara optimal.

Meskipun demikian, masih terdapat beberapa subbutir yang berada jauh di atas sebagian besar kemampuan peserta didik, yaitu A1 dan A3 pada indikator memvisualisasikan masalah, serta E4 dan E5 pada indikator mengevaluasi solusi. Hal ini menunjukkan bahwa kedua indikator tersebut menuntut kemampuan berpikir yang lebih kompleks dibandingkan indikator lainnya. Sebaliknya, subbutir D1 dan D2 pada indikator menerapkan solusi berada pada posisi logit terendah sehingga lebih mudah dijawab oleh peserta didik. Secara keseluruhan, sebaran tingkat kesukaran subbutir yang bervariasi menunjukkan bahwa instrumen mampu mengukur kemampuan pemecahan masalah peserta didik pada berbagai tingkat kemampuan. Temuan ini sejalan dengan Boone (2016), yang menyatakan bahwa *Wright Map* pada *Model Rasch* memberikan informasi mengenai kesesuaian antara kemampuan responden dan tingkat kesukaran butir sehingga dapat digunakan untuk mengevaluasi kualitas instrumen secara menyeluruh.

Kesimpulan

Berdasarkan hasil analisis menggunakan *Model Rasch*, instrumen kemampuan pemecahan masalah pada materi Suhu dan Kalor memiliki kualitas yang baik sebagai alat ukur. Hasil analisis menunjukkan bahwa sebagian besar subbutir memenuhi kriteria *item fit*, memiliki reliabilitas yang tinggi, daya pembeda yang baik, serta tingkat kesukaran yang bervariasi dari kategori sangat mudah hingga sangat sulit. Selain itu, hanya sebagian kecil subbutir yang terindikasi mengalami *Differential Item Functioning* (DIF), sehingga secara umum instrumen memiliki tingkat keadilan (*fairness*) yang baik. Hasil *Wright Map* juga menunjukkan bahwa distribusi tingkat kesukaran subbutir telah sesuai dengan kemampuan peserta didik (*targeting*), sehingga instrumen mampu mengukur kemampuan pemecahan masalah pada berbagai tingkat kemampuan peserta didik. Dengan demikian, instrumen yang

dikembangkan dinyatakan layak digunakan untuk mengukur kemampuan pemecahan masalah peserta didik pada materi suhu dan kalor. Instrumen ini diharapkan dapat menjadi salah satu alternatif alat ukur yang valid dan reliabel bagi guru maupun peneliti dalam mengevaluasi kemampuan pemecahan masalah peserta didik pada pembelajaran fisika, khususnya materi suhu dan kalor. Meskipun demikian, penelitian ini masih memiliki keterbatasan, yaitu jumlah sampel yang relatif terbatas dan hanya melibatkan peserta didik pada materi Suhu dan Kalor. Oleh karena itu, penelitian selanjutnya disarankan melibatkan jumlah sampel yang lebih besar serta cakupan materi fisika yang lebih luas untuk menguji konsistensi karakteristik psikometrik instrumen.

Ucapan Terimakasih

Penulis mengucapkan terima kasih kepada Program Studi Pendidikan Fisika, Universitas Pendidikan Indonesia, atas dukungan selama pelaksanaan penelitian. Ucapan terima kasih juga disampaikan kepada dosen pembimbing yang telah memberikan arahan, masukan, dan bimbingan selama proses penyusunan instrumen dan penelitian. Terima kasih juga penulis sampaikan kepada kepala salah satu SMA Negeri yang ada di Kota Bandung, serta kepada seluruh peserta didik yang telah berpartisipasi dalam proses pengumpulan data sehingga penelitian ini dapat terlaksana dengan baik. Dan penulis juga turut menyampaikan terima kasih kepada seluruh pihak yang ikut terlibat dalam pelaksanaan penelitian ini.

Referensi

- Andriani, W., Kaniwati, I., & Suwarma, I. R. (2026). Jurnal Pendidikan Progresif. 12, 1-19. <https://doi.org/10.23960/jpp.v>
- Boone, W. J. (2016). *Rasch* Analysis for Instrument Development: Why , When , and How ? <https://doi.org/10.1187/cbe.16-04-0148>
- Fryda, C., Ayudha, H., & Setyarsih, W. (2021). Studi Literatur : Analisis Praktik Pembelajaran Fisika. 11(1), 15-28.
- Heller, P., Keith, R., & Anderson, S. (1992). Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving. *American Journal of Physics*, 60(7), 627-636. <https://doi.org/10.1119/1.17117>
- Jayadi, A., Putri, D. H., & Johan, H. (2020). Identifikasi Pembekalan Keterampilan Abad 21 Pada Aspek Keterampilan Pemecahan Masalah Siswa SMA Kota Bengkulu Dalam Mata Pelajaran Fisika. 3(1), 25-32.
- Ningsih, P. W., Putri, D. H., & Purwanto, A. (2024). Development of Student Worksheets based on Problem-Solving of Temperature and Heat to Improve the Students Problem-Solving Ability. 10(2), 940-950. <https://doi.org/10.29303/jppipa.v10i2.6524>
- Qirom, M. S., & Nurlaelah, E. (2020). *Rasch* Model : Analyzing the Items Quality of Mathematics Higher- Order Thinking Skill Instrument.
- Ratnaningdyah, D. (2017). Upaya Melatihkan Kemampuan Pemecahan Masalah Melalui Pembelajaran Fisika Dengan Model Cooperative Problem Solving. 2(1), 1-3.
- Sari, I. P. M., Jatmiko, B., & Suprpto, N. (2023). Students' Physics Problem-Solving Skills in Daily Life Context: Between Confession and Fact. *Jurnal Penelitian Pendidikan IPA*, 9(1), 231-241. <https://doi.org/10.29303/jppipa.v9i1.2561>
- Smiley, J. (2015). Classical test theory or *Rasch* : A personal account from a novice user.
- Sumintono, B., & Widhiarso, W. (2015). Aplikasi Pemodelan *Rasch* pada Assessment Pendidikan. *Trim Komunika*.
- Susdelina, Perdana, S. A., & Febrian. (2018). Analisis kualitas instrumen pengukuran pemahaman konsep persamaan kuadrat melalui teori tes klasik dan *Rasch* model. 1, 41-48.
- Tanjung, Y. I., Festiyed, Diliarosta, S., Jubaidah, & Sanjaya, A. P. (2024). Analysis of Students ' Problem Solving Skills based on Heller ' s Indicator. 14(4), 421-433. <https://doi.org/10.47750/pegegog.14.04.40>
- Walidain, S. N., Haris, A., & Fitriyanto, S. (2023). Analisis Kemampuan Pemecahan Masalah Fisika dalam Pembelajaran Diferensiasi. 3(1), 50-57.