

Analisis Kualitas Instrumen Keterampilan Pemecahan Masalah Terintegrasi Metakognitif Menggunakan Model Rasch

Vira Veronika¹, Woro Setyarsih²

^{1,2} Pendidikan Fisika, Universitas Negeri Surabaya, Surabaya, Indonesia

* Email: vira.22079@mhs.unesa.ac.id, worosetyarsih@unesa.ac.id

Abstrak

Keterampilan pemecahan masalah merupakan salah satu kompetensi penting yang perlu dimiliki peserta didik dalam menghadapi tuntutan pembelajaran abad ke-21. Namun, pengukuran keterampilan tersebut masih cenderung berorientasi pada jawaban akhir sehingga belum sepenuhnya mengungkap proses berpikir peserta didik dalam menyelesaikan masalah. Penelitian ini bertujuan menganalisis kualitas empiris instrumen keterampilan pemecahan masalah berbasis *Socio-Scientific Issue* (SSI) terintegrasi metakognitif pada materi Usaha dan Energi menggunakan Model Rasch. Instrumen yang dianalisis berupa tes uraian sebanyak 16 butir soal yang dikembangkan berdasarkan empat tahapan pemecahan masalah Polya, yaitu memahami masalah, merencanakan penyelesaian, melaksanakan rencana, dan memeriksa kembali hasil penyelesaian, serta diintegrasikan dengan tiga aspek metakognitif, yaitu *planning*, *monitoring*, dan *evaluating*. Penelitian dilakukan terhadap 103 peserta didik kelas X yang telah mempelajari materi Usaha dan Energi. Data dianalisis menggunakan perangkat lunak Winsteps berdasarkan statistik kesesuaian butir, reliabilitas, dan *separation*. Hasil analisis menunjukkan bahwa seluruh 16 butir memenuhi kriteria *item fit* berdasarkan *Outfit Mean Square* (MNSQ), *Outfit Z-Standard* (ZSTD), dan *Point Measure Correlation* (Pt Measure Corr). Instrumen memperoleh nilai Cronbach Alpha sebesar 0,95, *Person Reliability* sebesar 0,94, *Item Reliability* sebesar 0,96, *Person Separation* sebesar 4,14, dan *Item Separation* sebesar 4,74. Hasil tersebut menunjukkan bahwa instrumen memiliki konsistensi internal yang sangat tinggi, mampu memberikan estimasi kemampuan peserta didik secara konsisten, serta memiliki kemampuan yang sangat baik dalam membedakan tingkat kemampuan peserta didik dan tingkat kesukaran butir. Dengan demikian, instrumen memiliki kualitas empiris yang sangat baik dan layak digunakan untuk mengukur keterampilan pemecahan masalah peserta didik.

Kata kunci: kualitas empiris, Model Rasch, keterampilan pemecahan masalah, *Socio-Scientific Issue*, metakognitif.

Quality Analysis (Validity, Reliability, and Discrimination Power) of Socio-Scientific Issue-Based Problem-Solving Skills Instrument Integrated with Metacognitive Using the Rasch Model

Abstract

Problem-solving skills are one of the important competencies that students need to develop to meet the demands of twenty-first-century learning. However, the assessment of these skills still tends to focus on final answers and therefore does not fully reveal students' thinking processes in solving problems. This study aims to analyze the empirical quality of a problem-solving skills instrument based on

Socio-Scientific Issues (SSI) integrated with metacognitive aspects on the topic of Work and Energy using the Rasch Model. The analyzed instrument consisted of 16 open-ended items developed based on Polya's four problem-solving stages, namely understanding the problem, devising a plan, carrying out the plan, and looking back, and integrated three metacognitive aspects, namely planning, monitoring, and evaluating. The study involved 103 tenth-grade students who had studied the topic of Work and Energy. The data were analyzed using Winsteps based on item fit statistics, reliability, and separation indices. The results showed that all 16 items met the item fit criteria based on Outfit Mean Square (MNSQ), Outfit Z-Standard (ZSTD), and Point Measure Correlation (Pt Measure Corr). The instrument obtained a Cronbach's Alpha of 0.95, Person Reliability of 0.94, Item Reliability of 0.96, Person Separation of 4.14, and Item Separation of 4.74. These results indicate that the instrument has very high internal consistency, provides consistent estimates of students' abilities, and has an excellent ability to distinguish students' ability levels and item difficulty levels. Therefore, the instrument demonstrates excellent empirical quality and is suitable for measuring students' problem-solving skills.

Keywords: Empirical quality, Rasch Model, problem solving skills, Socio-Scientific Issue, metacognitive.

Histori Naskah

Diserahkan: 28 Juli 2026

Direvisi: 19 Agustus 2026

Diterima: 22 Agustus 2026

How to cite:

Veronika, V., Setyarsih, W. (2026). Analisis Kualitas Instrumen Keterampilan Pemecahan Masalah Terintegrasi Metakognitif Menggunakan Model Rasch. *Inovasi Pendidikan Fisika*, **15**(1), 64-75. DOI: <https://doi.org/10.26740/ipf.vxx.nxx.xx-xx>.

PENDAHULUAN

Kemajuan ilmiah dan teknologi pada abad ke-21 menuntut pengembangan sejumlah bakat penting, termasuk kemampuan memecahkan masalah. Kemampuan ini sangat penting dalam pendidikan fisika karena siswa harus bisa mengaitkan konsep fisika dengan fenomena nyata dengan memahami masalah, mencari solusi, menerapkan konsep, dan menilai hasilnya. Menurut kebutuhan ini, Kurikulum Merdeka juga menekankan profil Pelajar Pancasila dan pembelajaran kontekstual, yang mendorong perkembangan kemampuan berpikir tingkat tinggi (Juraidah & Hartoyo, 2022). Namun demikian, hasil *Programme for International Student Assessment* (PISA) menunjukkan bahwa kemampuan peserta didik Indonesia dalam menyelesaikan masalah berbasis sains masih berada di bawah rata-rata negara OECD. Peserta didik masih mengalami kesulitan menghubungkan konsep ilmiah dengan permasalahan kehidupan sehari-hari serta memberikan alasan ilmiah terhadap solusi yang dipilih (OECD, 2023). Salah satu penyebabnya adalah instrumen yang masih didominasi soal rutin dan berorientasi pada jawaban akhir sehingga belum mampu menggambarkan proses berpikir peserta didik selama memecahkan masalah (Kurniawan & Taqwa, 2018).

Pengembangan instrumen keterampilan pemecahan masalah memerlukan konteks yang autentik agar peserta didik dapat menerapkan konsep fisika dalam situasi nyata. Salah satu pendekatan yang relevan adalah *Socio-Scientific Issue* (SSI), yaitu penyajian isu-isu sosial yang berkaitan dengan sains sehingga peserta didik terdorong menganalisis informasi, mengevaluasi alternatif solusi, dan mengambil keputusan berdasarkan bukti ilmiah (Sadler, 2011; Zeidler *et al.*, 2019). Selain itu, keberhasilan pemecahan masalah juga dipengaruhi oleh kemampuan metakognitif yang meliputi

aktivitas merencanakan, memantau, dan mengevaluasi strategi penyelesaian (Flavell, 1979). Oleh karena itu, integrasi aspek metakognitif ke dalam instrumen diharapkan mampu memberikan informasi yang lebih komprehensif mengenai proses berpikir peserta didik.

Beberapa penelitian telah mengembangkan instrumen keterampilan pemecahan masalah pada berbagai materi fisika maupun kimia dan menunjukkan bahwa instrumen yang dihasilkan memiliki validitas serta reliabilitas yang baik (Fathurrahman *et al.*, 2016; Kurniawan & Taqwa, 2018; Karira *et al.*, 2023). Penelitian lain juga menunjukkan bahwa Model Rasch mampu memberikan informasi kualitas instrumen secara lebih komprehensif dibandingkan teori tes klasik, seperti kesesuaian butir, reliabilitas, dan daya beda instrumen (Bond *et al.*, 2021; Boone *et al.*, 2014). Meskipun demikian, penelitian yang mengembangkan instrumen keterampilan pemecahan masalah berbasis *Socio-Scientific Issue* yang terintegrasi metakognitif pada materi Usaha dan Energi serta menganalisis kualitasnya menggunakan Model Rasch masih sangat terbatas. Dengan demikian, masih terdapat celah penelitian berupa belum tersedianya instrumen yang mampu mengukur keterampilan pemecahan masalah sekaligus regulasi metakognitif peserta didik dalam konteks isu-isu sosial berbasis sains.

Berdasarkan uraian tersebut, diperlukan analisis kualitas empiris terhadap instrumen keterampilan pemecahan masalah berbasis SSI terintegrasi metakognitif pada materi Usaha dan Energi. Rumusan masalah dalam penelitian ini adalah bagaimana kualitas empiris instrumen keterampilan pemecahan masalah berbasis *Socio-Scientific Issue* terintegrasi metakognitif pada materi Usaha dan Energi berdasarkan validitas empiris, reliabilitas, dan daya beda menggunakan Model Rasch?. Berdasarkan rumusan masalah tersebut, penelitian ini bertujuan untuk menganalisis kualitas empiris instrumen berdasarkan kesesuaian butir soal, reliabilitas, serta kemampuan instrumen dalam membedakan tingkat kemampuan peserta didik.

METODE

Penelitian ini merupakan bagian dari penelitian dan pengembangan (*Research and Development*). Analisis dilakukan pada tahap uji coba luas terhadap instrumen final untuk memperoleh informasi mengenai validitas empiris, reliabilitas, dan daya beda instrumen keterampilan pemecahan masalah berbasis *Socio-Scientific Issue* (SSI) terintegrasi metakognitif pada materi Usaha dan Energi. Uji coba dilakukan setelah instrumen dinyatakan layak berdasarkan hasil validasi ahli. Penelitian dilaksanakan pada semester genap tahun ajaran 2025/2026 di SMA YPM 2 Sukodono Sidoarjo. Subjek penelitian terdiri atas 103 peserta didik kelas X yang telah mempelajari materi Usaha dan Energi. Pemilihan subjek menggunakan teknik *purposive sampling* dengan pertimbangan bahwa seluruh peserta didik telah memperoleh materi yang sesuai dengan kompetensi yang diukur.

Instrumen penelitian berupa tes uraian yang terdiri atas 16 butir soal keterampilan pemecahan masalah berbasis *Socio-Scientific Issue* terintegrasi metakognitif. Penyusunan butir soal mengacu pada empat tahapan pemecahan masalah menurut Polya (1973), yaitu memahami masalah (*understanding the problem*), merencanakan penyelesaian (*devising a plan*), melaksanakan rencana (*carrying out the plan*), dan memeriksa kembali hasil penyelesaian (*looking back*). Selain itu, setiap butir soal diintegrasikan dengan aspek metakognitif yang meliputi *planning*, *monitoring*, dan *evaluating*. Adapun kisi-kisi yang dirancang dapat diketahui dari Tabel 1 berikut.

Tabel 1. Rancangan Kisi-Kisi Instrumen

Indikator	Soal dan Jawaban	Ranah Kognitif	Aspek Pemecahan Masalah	Aspek Metakognitif
	 Gambar 3 Kereta Cepat Jakarta-Bandung (Sumber: finance.detik.com) Menurut berita dari detikFinance berjudul “Kereta Cepat Ngebut hingga 350 Km/Jam Saat Beroperasi, Jarak 142 Km” yang diterbitkan pada 22 Juni 2023, kereta cepat Whoosh mampu mencapai kecepatan hingga 350 km/jam saat uji coba. Kereta cepat ini beroperasi pada lintasan sepanjang sekitar 142,3 km yang menghubungkan Jakarta dan Bandung. Dalam uji coba tersebut, perjalanan dari Stasiun Halim menuju Padalarang dapat ditempuh dalam waktu sekitar 32 menit, sedangkan perjalanan dari Tegal luar menuju Halim memakan waktu sekitar 44 menit. Pengujian tersebut menunjukkan bahwa kereta cepat mampu bergerak dengan kecepatan tinggi pada lintasan yang panjang sehingga waktu tempuh menjadi lebih singkat. Hal ini menjadi salah satu alasan dikembangkannya kereta cepat sebagai solusi transportasi modern. Secara fisika, pergerakan kereta cepat tidak terlepas dari adanya gaya dorong mesin yang bekerja selama kereta bergerak sepanjang lintasan. Gaya ini menyebabkan kereta dapat terus melaju dan menempuh jarak yang jauh. Dalam suatu analisis sederhana, gaya dorong rata-rata mesin kereta diperkirakan sebesar 13.200 N, dengan mengabaikan hambatan udara dan gesekan. Namun demikian, keberadaan kereta cepat juga memunculkan berbagai tanggapan di masyarakat. Sebagian pihak menilai bahwa kereta cepat memberikan manfaat dalam meningkatkan efisiensi waktu dan mobilitas. Di sisi lain, terdapat pandangan yang mempertanyakan biaya pembangunan dan operasional serta dampak lingkungan yang mungkin ditimbulkan.			
Peserta didik mampu mengidentifikasi hubungan usaha dan energi pada pergerakan kereta cepat berdasarkan informasi yang tersedia serta menentukan	Konsep usaha dan energi apa yang terlibat pada pergerakan kereta cepat? Jelaskan alasan memilih konsep tersebut. Jawaban: - Konsep yang terlibat pada pergerakan kereta cepat adalah usaha, energi kinetik, dan energi listrik. Energi listrik digunakan sebagai sumber tenaga untuk menggerakkan motor listrik. Motor menghasilkan gaya dorong yang	C2	Memahami Masalah	Planning Skill

Indikator	Soal dan Jawaban	Ranah Kognitif	Aspek Pemecahan Masalah	Aspek Metakognitif
konsep fisika yang relevan.	melakukan usaha sehingga kereta bergerak dan memiliki energi kinetik. - Saya memilih konsep tersebut karena pergerakan kereta melibatkan gaya yang menyebabkan perpindahan (usaha) dan peningkatan kecepatan kereta (energi kinetik). Energi listrik menjadi sumber energi utama yang diubah untuk menggerakkan kereta.			

Pengumpulan data dilakukan melalui pemberian instrumen kepada seluruh responden secara langsung di kelas sesuai dengan waktu yang telah ditentukan. Skor hasil pengerjaan kemudian dikodekan ke dalam format data Winsteps untuk dianalisis menggunakan Model Rasch dengan bantuan perangkat lunak Winsteps. Analisis data difokuskan pada tiga aspek kualitas instrumen, yaitu validitas empiris, reliabilitas, dan daya beda instrumen. Validitas empiris dianalisis berdasarkan statistik kesesuaian butir (*item fit statistics*), yaitu Outfit Mean Square (MNSQ), Outfit Z-Standard (ZSTD), dan *Point Measure Correlation* (Pt Measure Corr). Butir soal dinyatakan memenuhi kriteria valid apabila memenuhi sedikitnya dua dari tiga indikator, yaitu nilai Outfit MNSQ berada pada rentang 0,50–1,50, Outfit ZSTD berada pada rentang –2,00 hingga +2,00, dan Point Measure Correlation berada pada rentang 0,40–0,85 (Bond *et al.*, 2021; Linacre, 2024).

Reliabilitas instrumen dianalisis berdasarkan nilai *Cronbach Alpha* (KR-20), *Person Reliability*, dan *Item Reliability* yang diperoleh dari keluaran *Summary Statistics* Winsteps. Nilai *Cronbach Alpha* digunakan untuk menunjukkan konsistensi internal instrumen, sedangkan *Person Reliability* dan *Item Reliability* digunakan untuk menunjukkan konsistensi kemampuan peserta didik dan kestabilan tingkat kesukaran butir. Interpretasi nilai reliabilitas mengacu pada kriteria Sumintono dan Widhiarso (2015). Daya beda instrumen dianalisis menggunakan indeks *Person Separation* dan *Item Separation*. Nilai *Person Separation* digunakan untuk menunjukkan kemampuan instrumen dalam membedakan tingkat kemampuan peserta didik, sedangkan *Item Separation* digunakan untuk menunjukkan kemampuan instrumen dalam membedakan tingkat kesukaran butir soal. Semakin tinggi nilai separation, semakin baik kemampuan instrumen dalam mengelompokkan kemampuan peserta didik maupun tingkat kesukaran butir (Bond *et al.*, 2021).

HASIL DAN PEMBAHASAN

Validitas Empiris

Validitas empiris instrumen dianalisis menggunakan Model Rasch melalui statistik *item fit* yang terdiri atas *Outfit Mean Square* (MNSQ), *Outfit Z-Standard* (ZSTD), dan *Point Measure Correlation* (Pt Measure Corr). Ketiga parameter tersebut digunakan untuk mengevaluasi kesesuaian setiap butir soal terhadap model pengukuran sehingga dapat diketahui apakah setiap butir mampu mengukur konstruk yang sama secara konsisten (Bond *et al.*, 2021; Linacre, 2024). Instrumen yang dianalisis merupakan instrumen final yang terdiri atas 16 butir soal uraian keterampilan pemecahan masalah berbasis *Socio-Scientific Issue* (SSI) terintegrasi metakognitif pada materi Usaha dan Energi yang telah diujikan kepada 103 peserta didik kelas X SMA YPM 2 Sukodono Sidoarjo. Hasil analisis validitas empiris disajikan pada Tabel 2.

Tabel 2. Hasil Validitas Instrumen Tes

No. Soal	MNSQ	ZSTD	Pt. Mcorr	Keterangan
1	1,20	1,3	0,71	Valid
2	1,05	0,4	0,78	Valid
3	1,05	0,4	0,74	Valid
4	1,04	0,3	0,77	Valid

5	0,97	-0,1	0,78	Valid
6	1,02	0,2	0,76	Valid
7	1,02	0,2	0,73	Valid
8	1,00	0,0	0,77	Valid
9	1,00	0,0	0,73	Valid
10	0,99	0,0	0,80	Valid
11	0,90	-0,5	0,73	Valid
12	0,98	-0,1	0,73	Valid
13	0,95	-0,4	0,77	Valid
14	0,94	-0,4	0,79	Valid
15	0,88	-0,9	0,80	Valid
16	0,86	-1,0	0,78	Valid

Hasil analisis menunjukkan bahwa seluruh 16 butir soal memenuhi kriteria *item fit* berdasarkan nilai Outfit MNSQ, Outfit ZSTD, dan Pt Measure Corr. Temuan tersebut menunjukkan bahwa setiap butir memiliki tingkat kesesuaian yang baik terhadap Model Rasch sehingga mampu mengukur konstruk keterampilan pemecahan masalah secara konsisten. Menurut Bond *et al.* (2021), suatu butir dikatakan *fit* apabila pola respons peserta didik sesuai dengan probabilitas yang diprediksi oleh model, sehingga estimasi kemampuan yang dihasilkan menjadi lebih objektif. Boone *et al.* (2014) menambahkan bahwa kesesuaian butir terhadap Model Rasch menunjukkan bahwa setiap item bekerja secara stabil dalam mengukur konstruk yang sama tanpa dipengaruhi karakteristik sampel, sedangkan Linacre (2024) menjelaskan bahwa kombinasi statistik Outfit MNSQ, Outfit ZSTD, dan Pt Measure Corr merupakan indikator penting untuk mengevaluasi kualitas empiris suatu instrumen.

Seluruh butir yang memenuhi kriteria *item fit* menunjukkan bahwa instrumen memiliki representasi konstruk yang baik. Hal tersebut tidak terlepas dari proses penyusunan butir yang mengacu pada empat tahapan pemecahan masalah menurut Polya, yaitu memahami masalah (*understanding the problem*), merencanakan penyelesaian (*devising a plan*), melaksanakan rencana (*carrying out the plan*), dan memeriksa kembali hasil penyelesaian (*looking back*) (Polya, 1973). Tahapan tersebut memungkinkan peserta didik menunjukkan proses berpikir secara sistematis mulai dari mengidentifikasi informasi yang tersedia hingga mengevaluasi solusi yang diperoleh. Docktor dan Mestre (2014) menyatakan bahwa pengukuran keterampilan pemecahan masalah yang baik harus mampu mengungkap keseluruhan proses penyelesaian masalah, bukan hanya jawaban akhir.

Kualitas empiris instrumen juga dipengaruhi oleh penggunaan konteks *Socio-Scientific Issue* (SSI) dalam setiap butir soal. Penyajian permasalahan yang berasal dari isu-isu sosial berbasis sains memberikan pengalaman belajar yang lebih autentik sehingga peserta didik terdorong untuk menghubungkan konsep fisika dengan fenomena nyata. Zeidler *et al.* (2005) menyatakan bahwa SSI mampu mengembangkan kemampuan berpikir kritis, argumentasi ilmiah, dan pengambilan keputusan berbasis bukti. Selanjutnya, Sadler (2011) menjelaskan bahwa pembelajaran berbasis SSI memberikan kesempatan kepada peserta didik untuk mempertimbangkan berbagai alternatif solusi terhadap suatu persoalan yang bersifat multidimensional. Dengan demikian, penggunaan konteks SSI dalam instrumen ini diduga menjadi salah satu faktor yang mendukung tingginya kesesuaian butir terhadap Model Rasch.

Selain konteks SSI, instrumen ini juga mengintegrasikan aspek metakognitif, yaitu *planning*, *monitoring*, dan *evaluating*. Integrasi ketiga aspek tersebut memungkinkan instrumen tidak hanya mengukur hasil penyelesaian masalah, tetapi juga mengidentifikasi proses berpikir peserta didik selama menyelesaikan permasalahan. Flavell (1979) menjelaskan bahwa metakognitif merupakan kemampuan individu dalam mengendalikan proses berpikir melalui aktivitas perencanaan, pemantauan, dan evaluasi. Schraw dan Dennison (1994) menambahkan bahwa peserta didik dengan kemampuan metakognitif yang baik cenderung lebih mampu memilih strategi penyelesaian yang sesuai dan mengevaluasi efektivitas strategi tersebut. Temuan tersebut diperkuat oleh Veenman *et al.* (2006), Zimmerman (2008), dan Panadero (2017) yang menyatakan bahwa kemampuan metakognitif

berkontribusi secara signifikan terhadap keberhasilan seseorang dalam menyelesaikan permasalahan kompleks karena mampu meningkatkan kesadaran terhadap proses berpikir yang dilakukan.

Hasil penelitian ini sejalan dengan berbagai penelitian pengembangan instrumen berbasis Model Rasch. Kurniawan dan Taqwa (2018) melaporkan bahwa instrumen keterampilan pemecahan masalah fisika yang dikembangkan berdasarkan indikator yang jelas memiliki karakteristik butir yang baik dan memenuhi kriteria validitas empiris. Hasil serupa juga dilaporkan oleh Karira *et al.* (2023) yang menyatakan bahwa instrumen yang memenuhi kriteria *item fit* menghasilkan estimasi kemampuan yang lebih akurat dan objektif. Selain itu, penelitian Subhaktiyasa *et al.* (2024) menunjukkan bahwa instrumen yang memenuhi parameter *Rasch* memiliki kualitas pengukuran yang lebih baik sehingga layak digunakan untuk mengukur keterampilan berpikir tingkat tinggi pada pembelajaran sains.

Temuan penelitian ini juga memperkuat pandangan bahwa analisis menggunakan Model Rasch memberikan keunggulan dibandingkan pendekatan teori tes klasik. Dalam teori tes klasik, kualitas butir sangat dipengaruhi oleh karakteristik sampel, sedangkan Model Rasch menghasilkan estimasi parameter yang lebih stabil karena kemampuan peserta didik dan tingkat kesukaran butir diestimasi secara simultan pada skala logit yang sama (Boone, 2016). Oleh karena itu, instrumen yang memenuhi kriteria *item fit* tidak hanya menunjukkan kualitas butir yang baik pada kelompok responden tertentu, tetapi juga memiliki peluang yang lebih besar untuk digunakan pada populasi yang memiliki karakteristik serupa.

Secara keseluruhan, hasil analisis menunjukkan bahwa seluruh 16 butir instrumen memiliki validitas empiris yang sangat baik berdasarkan Model Rasch. Hal ini menunjukkan bahwa setiap butir mampu mengukur konstruk keterampilan pemecahan masalah berbasis *Socio-Scientific Issue* terintegrasi metakognitif secara konsisten dan akurat. Temuan tersebut menunjukkan bahwa instrumen layak digunakan sebagai alat ukur keterampilan pemecahan masalah pada pembelajaran fisika di tingkat SMA serta berpotensi dimanfaatkan dalam penelitian maupun evaluasi pembelajaran yang berorientasi pada pengembangan keterampilan abad ke-21.

Reliabilitas

Selain validitas empiris, kualitas suatu instrumen juga ditentukan oleh tingkat reliabilitasnya. Reliabilitas menunjukkan konsistensi instrumen dalam menghasilkan skor pengukuran ketika digunakan pada kelompok responden yang memiliki karakteristik serupa. Dalam Model Rasch, reliabilitas dianalisis melalui tiga parameter, yaitu *Cronbach Alpha* (KR-20), *Person Reliability*, dan *Item Reliability*. Ketiga parameter tersebut saling melengkapi dalam memberikan informasi mengenai konsistensi internal instrumen, konsistensi respons peserta didik, serta kestabilan tingkat kesukaran butir soal (Boone *et al.*, 2014; Linacre, 2024). Hasil reliabilitas secara lebih rinci disajikan pada Gambar 1.

SUMMARY OF 103 MEASURED Person								
	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ ZSTD		OUTFIT MNSQ ZSTD	
MEAN	35.5	16.0	-1.01	.48	.99	.0	.99	.0
S.D.	9.8	.0	2.19	.04	.31	.9	.32	.9
MAX.	55.0	16.0	3.32	.78	1.89	2.2	1.89	2.2
MIN.	18.0	16.0	-5.68	.46	.27	-2.9	.27	-2.9
REAL RMSE	.51	TRUE SD	2.13	SEPARATION	4.14	Person RELIABILITY		.94
MODEL RMSE	.49	TRUE SD	2.14	SEPARATION	4.39	Person RELIABILITY		.95
S.E. OF Person MEAN = .22								
Person RAW SCORE-TO-MEASURE CORRELATION = 1.00								
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .95								
SUMMARY OF 16 MEASURED Item								
	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ ZSTD		OUTFIT MNSQ ZSTD	
MEAN	228.8	103.0	.00	.19	1.00	.0	.99	.0
S.D.	26.2	.0	.93	.01	.07	.5	.08	.5
MAX.	274.0	103.0	1.70	.20	1.18	1.3	1.20	1.3
MIN.	183.0	103.0	-1.58	.19	.88	-.9	.86	-1.0
REAL RMSE	.19	TRUE SD	.91	SEPARATION	4.74	Item RELIABILITY		.96
MODEL RMSE	.19	TRUE SD	.91	SEPARATION	4.81	Item RELIABILITY		.96
S.E. OF Item MEAN = .24								

Gambar 1. Hasil Reliabilitas Instrumen

Berdasarkan Gambar 1, instrumen memiliki nilai *Cronbach Alpha* sebesar 0,95, yang menunjukkan bahwa instrumen memiliki konsistensi internal yang tinggi. Nilai tersebut menunjukkan bahwa keseluruhan butir soal mampu mengukur konstruk keterampilan pemecahan masalah secara konsisten sehingga hasil pengukuran memiliki tingkat kepercayaan yang baik. Menurut Bond *et al.* (2021), nilai *Cronbach Alpha* yang mendekati 1,00 menunjukkan bahwa interaksi antara peserta didik dan butir soal berlangsung secara konsisten. Boone *et al.* (2014) juga menjelaskan bahwa nilai reliabilitas yang tinggi mencerminkan bahwa variasi skor lebih banyak dipengaruhi oleh perbedaan kemampuan peserta didik daripada kesalahan pengukuran (*measurement error*).

Nilai *Cronbach Alpha* sebesar 0,95 menunjukkan bahwa instrumen memiliki konsistensi internal yang kuat meskipun terdiri atas butir-butir soal yang mengukur tahapan keterampilan pemecahan masalah yang berbeda. Hal ini menunjukkan bahwa setiap butir saling mendukung dalam mengukur konstruk yang sama, yaitu keterampilan pemecahan masalah berbasis *Socio-Scientific Issue* terintegrasi metakognitif. Temuan ini sejalan dengan pendapat Messick (1995) yang menyatakan bahwa reliabilitas merupakan salah satu bukti penting dalam mendukung validitas suatu instrumen karena instrumen yang konsisten akan menghasilkan interpretasi skor yang lebih dapat dipercaya.

Analisis juga menunjukkan bahwa *Person Reliability* sebesar 0,94 berada pada kategori baik. Nilai tersebut menunjukkan bahwa pola respons peserta didik terhadap instrumen relatif konsisten sehingga kemampuan peserta didik dapat diestimasi secara stabil. Dalam Model Rasch, *Person Reliability* menggambarkan sejauh mana instrumen mampu membedakan kemampuan peserta didik berdasarkan pola respons yang diberikan (Linacre, 2024). Semakin tinggi nilai *Person Reliability*, semakin kecil kemungkinan perubahan urutan kemampuan peserta didik apabila instrumen diberikan kembali kepada kelompok dengan karakteristik yang sama (Bond *et al.*, 2021).

Nilai *Person Reliability* yang tinggi diduga dipengaruhi oleh karakteristik instrumen yang disusun secara sistematis berdasarkan tahapan pemecahan masalah Polya dan disajikan dalam konteks *Socio-Scientific Issue*. Penyusunan soal yang bertahap memungkinkan peserta didik memberikan respons sesuai dengan kemampuan sebenarnya, sehingga pola jawaban menjadi lebih konsisten.

Selain itu, penggunaan rubrik analitik pada setiap tahapan penyelesaian masalah juga membantu meminimalkan subjektivitas dalam proses penskoran sehingga meningkatkan konsistensi hasil pengukuran.

Sementara itu, *Item Reliability* sebesar 0,96 berada pada kategori sangat baik, yang menunjukkan bahwa tingkat kesukaran keenam belas butir soal dapat diestimasi secara sangat stabil. Menurut Boone *et al.* (2014), *Item Reliability* yang tinggi menunjukkan bahwa jumlah responden dan variasi kemampuan peserta didik telah memadai untuk mengonfirmasi hierarki tingkat kesukaran butir. Linacre (2024) menjelaskan bahwa nilai *Item Reliability* di atas 0,90 menunjukkan bahwa posisi setiap butir pada skala logit akan relatif tetap apabila instrumen digunakan pada sampel lain yang memiliki karakteristik serupa.

Perbedaan antara *Person Reliability* (0,94) dan *Item Reliability* (0,96) merupakan kondisi yang umum dijumpai dalam analisis Rasch. Nilai *Item Reliability* yang lebih tinggi menunjukkan bahwa penyebaran tingkat kesukaran butir lebih konsisten dibandingkan variasi kemampuan peserta didik. Hal tersebut menunjukkan bahwa instrumen telah tersusun dengan baik dan mampu merepresentasikan rentang tingkat kesukaran secara jelas, sedangkan variasi kemampuan peserta didik masih dipengaruhi oleh faktor-faktor individual, seperti penguasaan konsep, pengalaman belajar, dan kemampuan berpikir tingkat tinggi (Sumintono & Widhiarso, 2015).

Hasil penelitian ini sejalan dengan berbagai penelitian sebelumnya yang melaporkan bahwa instrumen berbasis Model Rasch umumnya memiliki reliabilitas tinggi apabila disusun berdasarkan indikator yang jelas dan melalui proses pengembangan yang sistematis. Kurniawan dan Taqwa (2018) melaporkan bahwa instrumen keterampilan pemecahan masalah fisika yang dikembangkan memiliki nilai reliabilitas tinggi sehingga layak digunakan sebagai alat ukur. Penelitian Biyan dan Setyarsih (2020) juga menunjukkan bahwa analisis Rasch memberikan informasi reliabilitas yang lebih komprehensif dibandingkan teori tes klasik karena mampu mengevaluasi konsistensi peserta didik dan butir soal secara terpisah. Temuan serupa dilaporkan oleh Karira *et al.* (2023), yang menyatakan bahwa instrumen dengan *Person Reliability* dan *Item Reliability* yang tinggi menghasilkan pengukuran yang lebih akurat serta mampu memberikan estimasi kemampuan peserta didik secara lebih objektif.

Secara keseluruhan, hasil analisis reliabilitas menunjukkan bahwa instrumen keterampilan pemecahan masalah berbasis *Socio-Scientific Issue* terintegrasi metakognitif memiliki tingkat konsistensi yang tinggi. Nilai *Cronbach Alpha*, *Person Reliability*, dan *Item Reliability* yang diperoleh menunjukkan bahwa instrumen mampu menghasilkan pengukuran yang stabil, konsisten, dan dapat dipercaya. Oleh karena itu, instrumen ini memenuhi salah satu karakteristik utama instrumen pengukuran yang baik dan layak digunakan dalam kegiatan evaluasi pembelajaran maupun penelitian pada materi Usaha dan Energi.

Daya Beda

Daya beda instrumen dianalisis menggunakan parameter *Person Separation* dan *Item Separation* pada Model Rasch. Kedua parameter tersebut menunjukkan kemampuan instrumen dalam membedakan tingkat kemampuan peserta didik serta mengelompokkan tingkat kesukaran butir soal (Bond *et al.*, 2021). Hasil analisis validitas empiris disajikan pada Tabel 3.

Tabel 3. Hasil Analisis Daya Beda

Parameter	Separation	Strata	Kategori
Person	4.14	6	Sangat Baik
Item	4.74	7	Sangat Baik

Hasil analisis menunjukkan bahwa nilai *Person Separation* sebesar 4,14 dan *Item Separation* sebesar 4,74. Kedua nilai tersebut berada di atas kriteria minimum ($>2,00$), yang menunjukkan bahwa instrumen memiliki daya beda yang sangat baik (Linacre, 2024). Nilai *Person Separation*

menunjukkan bahwa instrumen mampu membedakan peserta didik ke dalam beberapa tingkat kemampuan, sedangkan nilai *Item Separation* menunjukkan bahwa butir soal memiliki tingkat kesukaran yang bervariasi dan tersusun secara stabil (Bond *et al.*, 2021; Sumintono & Widhiarso, 2015).

Tingginya nilai *Person Separation* menunjukkan bahwa setiap butir soal mampu mengidentifikasi perbedaan kemampuan peserta didik dalam menyelesaikan masalah fisika. Hal ini didukung oleh penyusunan soal berdasarkan tahapan pemecahan masalah Polya dan penggunaan konteks *Socio-Scientific Issue* (SSI), sehingga peserta didik dituntut memahami masalah, merencanakan strategi, melaksanakan penyelesaian, dan mengevaluasi hasil secara sistematis (Polya, 1973). Sementara itu, tingginya *Item Separation* menunjukkan bahwa instrumen memiliki rentang tingkat kesukaran yang baik sehingga mampu mengukur peserta didik dengan kemampuan yang beragam (Engelhard, 2013).

Hasil penelitian ini sejalan dengan penelitian Kurniawan dan Taqwa (2018) dan Karira *et al.* (2023) yang melaporkan bahwa nilai separation yang tinggi menunjukkan kemampuan instrumen dalam membedakan kemampuan peserta didik dan menghasilkan estimasi parameter butir yang stabil. Temuan serupa juga dilaporkan oleh Subhaktiyasa *et al.* (2024), yang menyatakan bahwa instrumen dengan separation tinggi lebih efektif digunakan untuk mengukur keterampilan berpikir tingkat tinggi karena mampu mengidentifikasi variasi kemampuan peserta didik secara lebih akurat.

Secara keseluruhan, nilai *Person Separation* sebesar 4,14 dan *Item Separation* sebesar 4,74 menunjukkan bahwa instrumen memiliki daya beda yang sangat baik. Dengan demikian, instrumen mampu mengelompokkan peserta didik berdasarkan tingkat kemampuan serta menyusun hierarki tingkat kesukaran butir secara konsisten, sehingga layak digunakan sebagai instrumen pengukuran keterampilan pemecahan masalah pada materi Usaha dan Energi.

KESIMPULAN

Penelitian ini bertujuan menganalisis kualitas instrumen keterampilan pemecahan masalah berbasis *Socio-Scientific Issue* (SSI) terintegrasi metakognitif pada materi Usaha dan Energi menggunakan Model Rasch. Hasil analisis menunjukkan bahwa seluruh 16 butir soal memenuhi kriteria kesesuaian (*item fit*), sehingga seluruh butir dapat digunakan untuk mengukur keterampilan pemecahan masalah peserta didik. Instrumen juga memiliki reliabilitas yang sangat baik, ditunjukkan oleh nilai *Cronbach Alpha* sebesar 0,95, *Person Reliability* sebesar 0,94, dan *Item Reliability* sebesar 0,96. Selain itu, nilai *Person Separation* sebesar 4,14 dan *Item Separation* sebesar 4,74 menunjukkan bahwa instrumen mampu membedakan peserta didik berdasarkan tingkat kemampuan pemecahan masalah serta membedakan butir soal berdasarkan tingkat kesukarannya. Dengan demikian, instrumen keterampilan pemecahan masalah berbasis SSI terintegrasi metakognitif pada materi Usaha dan Energi memiliki kualitas empiris yang sangat baik dan layak digunakan sebagai instrumen penilaian dalam pembelajaran fisika di SMA maupun sebagai instrumen penelitian untuk mengukur keterampilan pemecahan masalah peserta didik. Instrumen ini dapat membantu guru memperoleh informasi mengenai tingkat keterampilan pemecahan masalah peserta didik sehingga dapat menjadi dasar dalam menentukan tindak lanjut pembelajaran yang sesuai serta dapat digunakan oleh peneliti sebagai alternatif instrumen untuk mengukur keterampilan pemecahan masalah peserta didik.

KONTRIBUSI PENULIS

Vira Veronika: Data Curation, Project Administration, Methodology, Formal Analysis, Resources, dan Writing -Original Draft. **Woro Setyarsih:** Conceptualization, Methodology, dan Validation. Semua penulis telah membaca dan menyetujui versi akhir dari naskah ini.

PERNYATAAN BEBAS KONFLIK KEPENTINGAN

Para penulis menyatakan bahwa tidak ada konflik kepentingan finansial maupun hubungan pribadi yang dapat mempengaruhi hasil yang dilaporkan dalam naskah ini.

PERNYATAAN ETIKA PENELITIAN DAN PUBLIKASI

Para penulis menyatakan bahwa penelitian dan penulisan naskah ini telah mematuhi standar etika penelitian dan publikasi, sesuai dengan prinsip ilmiah, serta bebas dari plagiasi.

PERNYATAAN PEMANFAATAN TEKNOLOGI ASISTIF

Para penulis menyatakan bahwa Kecerdasan Buatan Generatif (*Generative Artificial Intelligence*) dan teknologi asistif lainnya tidak digunakan secara berlebihan dalam proses penelitian dan penulisan naskah ini. Secara khusus, ChatGPT digunakan untuk Brainstromingide, Grammarly untuk koreksi tata dan gaya bahasa, serta Quillbot untuk parafrase. Para penulis telah meninjau dan menyunting semua konten yang dihasilkan AI guna memastikan ketepatan, kelengkapan, serta kepatuhan terhadap standar etika dan ilmiah, dan bertanggung jawab penuh atas naskah versi akhir. Para penulis telah meninjau dan menyunting semua konten yang dihasilkan AI guna memastikan ketepatan, kelengkapan, serta kepatuhan terhadap standar etika dan ilmiah, dan bertanggung jawab penuh atas naskah versi akhir.

DAFTAR PUSTAKA

- Belland, B. R., Kim, N. J., & Hannafin, M. J. (2019). A framework for designing scaffolds that improve motivation and cognition. *Educational Psychologist*, 54(4), 243–270.
- Biyan, M. B., & Setyarsih, W. (2020). Pengembangan instrumen keterampilan pemecahan masalah fisika menggunakan Model Rasch. *Jurnal Inovasi Pendidikan Fisika*, 9(3), 193–201.
- Bond, T. G., Yan, Z., & Heene, M. (2021). *Applying the Rasch model: Fundamental measurement in the human sciences* (4th ed.). Routledge. <https://doi.org/10.4324/9780429030499>
- Boone, W. J. (2016). Rasch analysis for instrument development: Why, when, and how? *CBE—Life Sciences Education*, 15(4), rm4. <https://doi.org/10.1187/cbe.16-04-0148>
- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Springer.
- Docktor, J. L., & Mestre, J. P. (2014). Synthesis of discipline-based education research in physics. *Physical Review Special Topics—Physics Education Research*, 10(2), 020119.
- Engelhard, G. (2013). *Invariant measurement: Using Rasch models in the social, behavioral, and health sciences*. Routledge.
- Evagorou, M., & Osborne, J. (2013). Exploring young students' collaborative argumentation within a socioscientific issue. *Journal of Research in Science Teaching*, 50(2), 209–237.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring. *American Psychologist*, 34(10), 906–911.
- Ilahi, D. P., Festiyed, Utami, L., & Ratih, A. (2024). Pengembangan dan Validitas Instrumen Metakognitif Peserta didik pada Pembelajaran Fisika SMA. *Jurnal Pendidikan Biologi*, 9(2), 1087–1098. <https://doi.org/https://doi.org/10.30605/biogenerasi.v9i2.3903>
- Jonassen, D. H. (2011). *Learning to solve problems: A handbook for designing problem-solving learning environments*. Routledge.
- Junaidi, I., & Fitriani, W. (2025). Pemecahan Masalah dan Metakognisi dalam Proses Transfer Pengetahuan. *J-CEKI: Jurnal Cendikia Ilmiah*, 4(2), 1587–1592. <https://doi.org/https://doi.org/10.56799/jceki.v4i2.6910>
- Karira, N. F., Sunarti, T., Niswati, M. R. J., & Setyarsih, W. (2023). Validitas instrumen tes berbasis literasi sains untuk mengukur keterampilan berpikir kritis siswa SMA pada materi energi terbarukan. *Inovasi Pendidikan Fisika*, 12(2), 118–125. <https://ejournal.unesa.ac.id/index.php/inovasi-pendidikan-fisika/article/view/52957>
- Kurniawan, B. R., & Taqwa, M. R. A. (2018). Pengembangan instrumen tes kemampuan pemecahan masalah fisika pada materi listrik dinamis. *Jurnal Pendidikan: Teori, Penelitian, dan Pengembangan*, 3(11), 1451–1457. <https://doi.org/10.17977/jptpp.v3i11.11761>
- Linacre, J. M. (2024). *Winsteps Rasch Measurement Computer Program User's Guide*.

- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749. <https://doi.org/10.1037/0003-066X.50.9.741>
- OECD. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing.
- Polya, G. (1973). *How to solve it: A new aspect of mathematical method* (2nd ed.). Princeton University Press.
- Sadler, T. D. (Ed.). (2011). *Socio-scientific issues in the classroom: Teaching, learning and research*. Springer. <https://doi.org/10.1007/978-94-007-1159-4>
- Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, 19(4), 460–475. <https://doi.org/10.1006/ceps.1994.1033>
- Subhaktiyasa, P. G. (2024). Evaluasi validitas dan reliabilitas instrumen penelitian kuantitatif: Sebuah studi pustaka. *Journal of Education Research*, 5(4), 5599–5609. <https://doi.org/10.37985/jer.v5i4.1747>
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan Rasch pada assessment pendidikan*. Trim Komunika.
- Zeidler, D. L. (2019). Advancing socioscientific issues as a curriculum emphasis. *Journal of Research in Science Teaching*, 56(7), 935–956. <https://doi.org/10.1002/tea.21548>
- Zeidler, D. L., Sadler, T. D., Simmons, M. L., & Howes, E. V. (2005). Beyond STS: A research-based framework for socioscientific issues education. *Science Education*, 89(3), 357–377. <https://doi.org/10.1002/sce.20048>