



Analisis Kualitas Instrumen Pemahaman Konsep pada Materi Energi Terbarukan Menggunakan Model *Rasch*

Rapi Alwi Azhar^{1*}, Parlindungan Sinaga¹, Agus Danawan¹, Lina Aviyanti¹, Mohd Khairul Nizam Syah¹

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

DOI: <https://doi.org/10.29303/goescienceed.v7i3.2451>

Article Info:

Received : 17 Juni 2026
Revised : 26 Juni 2026
Accepted : 02 Juli 2026
Published : 07 Juli 2026

Correspondence:

Rapi Alwi Azhar

Phone: +6289513996962

Abstract: This study aimed to analyze the quality of a conceptual understanding test instrument on renewable energy using the *Rasch* measurement model. The instrument consisted of 19 multiple-choice items developed based on the conceptual understanding dimensions proposed by Anderson and Krathwohl (2001) and was administered to 75 tenth-grade senior high school students. Data were analyzed using MINISTEP version 5.10.3.0, including content validity (Aiken's V coefficient), construct validity (unidimensionality, outfit MNSQ, ZSTD, and Point Measure Correlation), reliability, item difficulty, Differential Item Functioning (DIF), and the Wright Map. The results showed that 14 of the 19 items were content valid ($V \geq 0.80$). The raw variance explained by measures was 36.5%, indicating that the instrument satisfied the unidimensionality assumption. The analysis also revealed a person reliability of 0.71 (acceptable), an item reliability of 0.94 (very good), and a Cronbach's alpha (KR-20) of 0.77 (very good). Three items (S7, S10, and S19) exhibited significant Differential Item Functioning (DIF) based on gender. The Wright Map indicated that the mean student ability was 1.14 logits, while item difficulty ranged from -2.15 to 3.33 logits. Overall, the instrument was considered suitable for measuring students' conceptual understanding of renewable energy after revisions were made to the problematic items.

Keywords: Conceptual Understanding; Renewable Energy; Assessment Instrument; *Rasch* Model; Validity.

Citation: Azhar, R. A., Sinaga, P., Danawan, A., Aviyanti, L., & Syah, M. K. N. (2026). Analisis Kualitas Instrumen Pemahaman Konsep pada Materi Energi Terbarukan Menggunakan Model *Rasch*. *Jurnal Pendidikan, Sains, Geologi, Dan Geofisika (GeoScienceEd Journal)*, 7(3), 3205-3215. <https://doi.org/10.29303/goescienceed.v7i3.2451>

Pendahuluan

Pemahaman konsep merupakan salah satu kemampuan mendasar yang harus dimiliki peserta didik dalam pembelajaran fisika. Kemampuan ini berperan penting dalam membantu peserta didik memahami hubungan antar konsep, menjelaskan fenomena alam, serta menerapkan prinsip-prinsip fisika dalam berbagai situasi kontekstual. Dalam Kurikulum Merdeka, pembelajaran fisika tidak hanya menekankan penguasaan fakta dan rumus, tetapi juga kemampuan peserta didik untuk membangun pemahaman konseptual secara bermakna (Kemendikbudristek (2022).

Menurut Anderson & Krathwohl (2001), pemahaman (understanding) merupakan salah satu

dimensi proses kognitif yang mencakup kemampuan menafsirkan, mencontohkan, mengklasifikasikan, merangkum, menyimpulkan, membandingkan, dan menjelaskan suatu konsep. Kemampuan tersebut tidak dapat diukur hanya melalui penguasaan hafalan, tetapi memerlukan instrumen yang dirancang sesuai dengan indikator pemahaman konsep yang hendak diukur (Mardian et al., 2023; Rafiah & Miyaqi, 2025).

Pemahaman konsep mengenai energi terbarukan memiliki kedudukan strategis dalam pendidikan sains karena berkaitan langsung dengan agenda transisi energi yang tengah menjadi prioritas kebijakan nasional di Indonesia (Setiawan & Rosid, 2024). Penguasaan konsep dasar energi terbarukan pada peserta didik tidak

berhenti pada ranah kognitif semata, melainkan menjadi fondasi bagi terbentuknya sikap positif, kebiasaan hemat energi, hingga niat mengadopsi teknologi energi bersih di masa depan (Festiana, Herlina, Zakiya, & Falamy, 2026). Kondisi ini sejalan dengan kerangka literasi energi kontekstual yang menempatkan pemahaman konsep sebagai salah satu pilar utama yang saling berkaitan dengan dimensi afektif dan perilaku (Nugraha & Suhandi, 2021). Meskipun demikian, tingkat pemahaman dan kesadaran peserta didik Indonesia terhadap energi terbarukan secara umum masih bervariasi dan belum sepenuhnya optimal, sehingga penguatan pemahaman konsep pada topik ini terus menjadi perhatian dalam penelitian pendidikan sains (Pambudi, Nanda, Alfina, & Syahrial, 2024).

Namun demikian, kualitas instrumen pemahaman konsep yang digunakan dalam pembelajaran dan penelitian masih menjadi perhatian. Sejumlah penelitian menunjukkan bahwa banyak instrumen yang dikembangkan hanya melalui validasi isi oleh para ahli tanpa disertai pengujian empiris yang memadai, sehingga kualitas psikometriknya belum dapat dipastikan secara menyeluruh (Bashooir & Supahar, 2018; Rohmatika et al., 2025). Kondisi serupa ditemukan secara spesifik pada pengembangan instrumen literasi energi terbarukan bagi siswa Indonesia, yang baru sampai pada tahap validasi isi oleh ahli, sementara pengujian validitas konstruk dan reliabilitas pada subjek penelitian sesungguhnya belum dilaporkan (Usman, Pramudawardani, Umam, & Takahashi, 2025). Pada konteks lain, analisis terhadap instrumen tes keterampilan berpikir kreatif pada materi energi terbarukan turut menunjukkan keterbatasan, antara lain ditemukannya butir soal dengan tingkat kesukaran ekstrem serta nilai reliabilitas yang hanya tergolong cukup, sehingga peneliti merekomendasikan pengembangan dan analisis butir soal yang lebih mendalam pada penelitian berikutnya. Berbagai temuan ini menegaskan bahwa ketersediaan instrumen pemahaman konsep energi terbarukan yang teruji secara empiris dan reliabel masih terbatas, padahal instrumen semacam itu sangat dibutuhkan untuk memetakan capaian pemahaman peserta didik secara akurat sekaligus mengevaluasi efektivitas intervensi pembelajaran yang dirancang.

Model *Rasch* merupakan salah satu pendekatan dalam *Item Response Theory* (IRT) yang memodelkan probabilitas peserta didik menjawab suatu butir soal dengan benar berdasarkan selisih antara kemampuan peserta didik (*person ability*) dan tingkat kesukaran butir (*item difficulty*). Model ini mengasumsikan bahwa semakin tinggi kemampuan peserta didik dibandingkan tingkat kesukaran suatu butir, semakin besar peluang peserta didik menjawab butir tersebut dengan benar. Berbeda dengan pendekatan *Classical Test Theory* (CTT),

Model *Rasch* menghasilkan estimasi kemampuan peserta didik (*person measure*) dan tingkat kesukaran item (*item measure*) pada skala logit yang sama serta bersifat *sample-free* dan *test-free*, sehingga hasil pengukuran menjadi lebih objektif, stabil, dan dapat dibandingkan pada kelompok maupun konteks yang berbeda (Sumintono & Widhiarso, 2015). Selain itu, Model *Rasch* memungkinkan evaluasi kualitas instrumen secara komprehensif melalui analisis unidimensionalitas, kesesuaian butir menggunakan indeks *outfit* MNSQ dan ZSTD, identifikasi keberpihakan butir melalui *Differential Item Functioning* (DIF), serta visualisasi hubungan antara kemampuan peserta didik dan tingkat kesukaran item menggunakan *Wright Map*.

Penelitian ini memiliki kebaruan dengan mengembangkan sekaligus menganalisis kualitas instrumen tes pemahaman konsep pada materi energi terbarukan berdasarkan dimensi pemahaman konsep Anderson dan Krathwohl menggunakan Model *Rasch*. Berbeda dengan penelitian sebelumnya yang umumnya berfokus pada pengembangan instrumen atau analisis kualitas instrumen pada materi lain, penelitian ini menyajikan evaluasi psikometrik yang komprehensif, meliputi validitas isi, validitas konstruk, reliabilitas, karakteristik butir, analisis keberpihakan item, serta pemetaan kemampuan peserta didik dan tingkat kesukaran butir (Chan, Ismail, & Sumintono, 2021). Hasil penelitian ini diharapkan dapat menghasilkan instrumen yang valid, reliabel, dan mampu memberikan informasi yang lebih akurat dalam mengukur pemahaman konsep peserta didik pada materi energi terbarukan, sehingga dapat digunakan sebagai dasar penyusunan strategi pembelajaran dan evaluasi yang lebih efektif pada pembelajaran fisika.

Berdasarkan kesenjangan tersebut, penelitian ini bertujuan untuk menganalisis kualitas instrumen pemahaman konsep pada materi energi terbarukan menggunakan Model *Rasch* secara komprehensif.

Metode

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan deskriptif yang bertujuan untuk menganalisis kualitas psikometrik instrumen pemahaman konsep pada materi energi terbarukan. Analisis kualitas instrumen dilakukan melalui validitas isi menggunakan koefisien Aiken's V dan analisis empiris menggunakan Model *Rasch* dengan bantuan perangkat lunak MINISTEP versi 5.10.3.0.

Instrumen yang dianalisis berupa tes pemahaman konsep berbentuk pilihan ganda yang terdiri atas 19 butir soal pada materi energi terbarukan. Penyusunan instrumen mengacu pada dimensi proses kognitif memahami (C2) berdasarkan Taksonomi Bloom

revisi (Anderson & Krathwohl, 2001). Distribusi indikator dan nomor butir soal disajikan pada Tabel 1.

Tabel 1. Distribusi Indikator dan Nomor Butir Soal

Indikator	Nomor Soal
Menjelaskan	S1, S2, S3, S4, S5, S13, S14, S15 dan S16
Mengklasifikasikan	S6 dan S7
Membandingkan	S8 dan S9
Menginterpretasikan	S10, S12 dan S17
Menyimpulkan	S11, S18 dan S19

Data penelitian diperoleh melalui uji coba instrumen kepada 75 peserta didik kelas X SMA yang telah mempelajari materi energi terbarukan. Responden terdiri atas 26 peserta didik laki-laki dan 49 peserta didik perempuan yang dipilih menggunakan teknik purposive sampling. Selain itu, validitas isi instrumen dinilai oleh lima orang ahli yang terdiri atas dua dosen Program Studi Pendidikan Fisika dan tiga guru fisika SMA.

Tahap pertama analisis kualitas instrumen adalah analisis validitas isi menggunakan koefisien validitas Aiken (Aiken's V). Koefisien ini dihitung berdasarkan penilaian lima orang ahli terhadap aspek materi, konstruksi, dan bahasa pada setiap butir soal menggunakan skala penilaian 1-5. Nilai Aiken's V dihitung menggunakan Persamaan.

$$V = \frac{\sum s}{[n(c-1)]} = \frac{\sum r - l_o}{[n(c-1)]}$$

dengan V adalah koefisien validitas Aiken, r adalah skor yang diberikan oleh penilai, l_o adalah skor penilaian terendah, c adalah skor penilaian tertinggi, dan n adalah jumlah penilai Aiken (1985). Butir soal dinyatakan memenuhi validitas isi apabila memperoleh nilai $V \geq 0,80$ (Taherdoost, 2018). Seluruh butir soal, baik yang memenuhi maupun yang belum memenuhi kriteria validitas isi, tetap diujicobakan kepada peserta didik untuk memperoleh bukti validitas empiris melalui analisis Model *Rasch*. Pendekatan ini dilakukan karena analisis *Rasch* tidak hanya mengevaluasi kesesuaian butir terhadap konstruk yang diukur, tetapi juga memberikan informasi mengenai karakteristik psikometrik setiap butir sebagai dasar penyempurnaan instrumen (Aditya et al., 2025; Irawan et al., 2025).

Tahap kedua adalah analisis validitas konstruk, reliabilitas, tingkat kesukaran, *Differential Item Functioning* (DIF), dan peta variabel (*Wright Map*) menggunakan pemodelan *Rasch* melalui perangkat lunak MINISTEP 5.10.3.0. Validitas konstruk instrumen dianalisis menggunakan Model *Rasch* melalui pengujian

unidimensionalitas dan kesesuaian butir soal (*item fit*). Unidimensionalitas dievaluasi berdasarkan nilai *raw variance explained by measures* yang menunjukkan proporsi varians data yang dapat dijelaskan oleh konstruk yang diukur. Sementara itu, kesesuaian butir soal dianalisis menggunakan tiga indikator, yaitu *outfit mean square* (MNSQ), *outfit z-standard* (ZSTD), dan *point measure correlation* (Pt. Measure Corr). Kriteria interpretasi unidimensionalitas dan *item fit* disajikan pada Tabel 2.

Tabel 2. Kriteria Unidimensionalitas dan Item Fit

Aspek Analisis	Indikator	Kriteria
Unidimensionalitas	Raw variance explained by measure	>20% = terpenuhi; >40% = baik; >60% = istimewa
	<i>Outfit MNSQ</i>	$0,5 < \text{MNSQ} < 1,5$
	<i>Outfit ZSTD</i>	$-2,0 < \text{ZSTD} < +2,0$
<i>Item Fit</i>	<i>Pt. Measure Corr</i>	$0,4 < \text{Pt Measure Corr} < 0,85$

(Sumintono & Widhiarso, 2015; Dong et al., 2025)

Reliabilitas instrumen dievaluasi melalui tiga indeks, yaitu *person reliability*, *item reliability*, dan *Cronbach alpha* (KR-20). *Person reliability* menunjukkan konsistensi kemampuan peserta didik, sedangkan *item reliability* menunjukkan kestabilan tingkat kesukaran butir soal. Nilai *Cronbach alpha* digunakan untuk mengevaluasi konsistensi internal instrumen secara keseluruhan. Kriteria interpretasi reliabilitas disajikan pada Tabel 3.

Tabel 3. Kategori Nilai Person & Item Reliability serta Cronbach Alpha

Uji Reliabilitas	Hasil Uji	Kategori
Person & Item reliability	< 0,67	Rendah
	0,67 – 0,80	Cukup
	0,81 – 0,90	Baik
	0,91 – 0,94	Sangat Baik
	> 0,94	Luar Biasa
Cronbach Alpha (KR-20)	< 0,50	Rendah
	0,50 – 0,60	Cukup
	0,61 – 0,70	Baik

0,71 – 0,80	Sangat Baik
> 0,80	Luar Biasa

(Bond & Fox, 2007; Sumintono & Widhiarso, 2015)

Tingkat kesukaran setiap butir soal diestimasi berdasarkan nilai *JMLE Measure* dalam satuan logit. Kategori tingkat kesukaran ditentukan secara relatif menggunakan nilai rata-rata logit item dan simpangan baku (*standard deviation*) sebagaimana disajikan pada Tabel 4 (Boone, Staver, & Yale, 2014).

Tabel 4. Kategori Tingkat Kesukaran Butir Soal

Rentang Logit	Kategori
Logit < Mean - SD	Sangat Mudah
Mean - SD ≤ Logit < Mean	Mudah
Mean ≤ Logit < Mean + SD	Sulit

Logit ≥ Mean + SD	Sangat Sulit
-------------------	--------------

(Boone et al., 2014)

Analisis DIF dilakukan berdasarkan nilai probabilitas *Rasch-Welch* (Prob. t), di mana butir soal dinyatakan mengandung DIF apabila Prob. < 0,05. Wright Map digunakan untuk memvisualisasikan distribusi kemampuan peserta didik dan tingkat kesukaran butir soal pada skala logit yang sama.

Hasil dan Diskusi

Validitas Isi

Validitas isi instrumen dianalisis menggunakan koefisien Aiken (V) berdasarkan penilaian lima orang ahli terhadap aspek materi, konstruksi, dan bahasa. Hasil analisis disajikan pada Tabel 5.

Tabel 5. Hasil Analisis Validitas Isi Instrumen Tes Pemahaman Konsep

Butir Soal	Materi	Konstruksi	Bahasa	Rata-rata	Keterangan
S1	0,74	0,74	0,81	0,76	Tidak Valid
S2	0,76	0,76	0,81	0,78	Tidak Valid
S3	0,74	0,74	0,81	0,76	Tidak Valid
S4	0,74	0,76	0,81	0,77	Tidak Valid
S5	0,73	0,75	0,81	0,76	Tidak Valid
S6	0,83	0,85	0,88	0,85	Valid
S7	0,83	0,75	0,90	0,82	Valid
S8	0,92	0,90	0,91	0,91	Valid
S9	0,87	0,79	0,86	0,84	Valid
S10	0,74	0,79	0,88	0,80	Valid
S11	0,93	0,91	0,91	0,92	Valid
S12	0,85	0,88	0,86	0,86	Valid
S13	0,83	0,83	0,80	0,82	Valid
S14	0,79	0,85	0,81	0,82	Valid
S15	0,81	0,83	0,81	0,82	Valid
S16	0,79	0,81	0,83	0,81	Valid
S17	0,81	0,84	0,82	0,82	Valid
S18	0,85	0,84	0,88	0,86	Valid
S19	0,88	0,83	0,89	0,87	Valid

Berdasarkan Tabel 5, dari 19 butir soal yang divalidasi oleh lima orang ahli, sebanyak 14 butir soal dinyatakan valid secara isi dengan nilai rata-rata koefisien Aiken $V \geq 0,80$, yaitu S6 hingga S19. Nilai tertinggi diperoleh oleh S11 ($V = 0,92$) dan S8 ($V = 0,91$), menunjukkan bahwa kedua butir tersebut dinilai sangat sesuai oleh para ahli dari sisi materi, konstruksi,

maupun bahasa. Sementara itu, lima butir soal yaitu S1 ($V = 0,76$), S2 ($V = 0,78$), S3 ($V = 0,76$), S4 ($V = 0,77$), dan S5 ($V = 0,76$) dinyatakan tidak valid secara isi karena rata-rata koefisien Aiken-nya berada di bawah ambang batas 0,80.

Ditinjau lebih lanjut, kelima butir yang tidak valid tersebut secara konsisten memperoleh skor rendah pada

aspek materi dan konstruksi, sementara aspek bahasa relatif lebih baik dengan nilai berkisar 0,81. Kondisi ini mengindikasikan bahwa permasalahan utama terletak pada kesesuaian cakupan materi dan kejelasan konstruksi soal, bukan pada aspek kebahasaan. Temuan ini sejalan dengan Sujatmika et al. (2025) yang menyatakan bahwa kelemahan pada aspek konstruksi merupakan permasalahan umum dalam pengembangan instrumen kognitif, terutama pada butir soal yang menuntut pemahaman konseptual tingkat tinggi. Meskipun demikian, seluruh 19 butir soal tetap diujicobakan kepada peserta didik untuk memperoleh bukti validitas empiris melalui analisis Model *Rasch*, sebab validitas isi dan validitas konstruk merupakan dua dimensi yang saling melengkapi dalam evaluasi kualitas instrumen secara menyeluruh.

Validitas Konstruk

Validitas konstruk instrumen dianalisis melalui pengujian unidimensionalitas dan kesesuaian butir soal (*item fit*). Hasil uji unidimensionalitas disajikan pada Tabel 6.

Tabel 6. Hasil Uji Unidimensionalitas Instrumen Tes Pemahaman Konsep

Unidimensionalitas	Nilai Indeks	Keterangan
Raw variance explained by measures	36,5%	Terpenuhi

Berdasarkan Tabel 6, nilai *raw variance explained by measures* yang diperoleh adalah sebesar 36,5%, melebihi ambang batas minimum 20% yang dipersyaratkan untuk memenuhi asumsi unidimensionalitas. Nilai ini menunjukkan bahwa instrumen mengukur satu konstruk dominan secara memadai, yakni pemahaman konsep pada materi energi terbarukan. Selain itu, nilai *essential unidimensionality* yang mencapai 53,1% semakin mengonfirmasi bahwa sebagian besar varians respons peserta didik digerakkan oleh konstruk tunggal yang dimaksud (Linacre, 2016). Dengan demikian, dapat disimpulkan bahwa instrumen ini telah memenuhi asumsi unidimensionalitas dan layak dianalisis lebih lanjut menggunakan Model *Rasch*. Hasil analisis *item fit* selanjutnya disajikan pada Tabel 7.

Tabel 7. Hasil Analisis Item Fit Instrumen Tes Pemahaman Konsep

Butir Soal	Outfit MNSQ	Outfit ZSTD	Pt. Measure Corr	Keterangan
S1	0,88	-0,50	0,54	Ketiga kriteria terpenuhi
S2	1,49	2,42	0,25	Satu kriteria terpenuhi
S3	0,78	-0,85	0,59	Ketiga kriteria terpenuhi
S4	0,78	-1,17	0,61	Ketiga kriteria terpenuhi
S5	1,00	0,09	0,52	Ketiga kriteria terpenuhi
S6	0,79	-1,07	0,60	Ketiga kriteria terpenuhi
S7	1,86	3,72	0,15	Ketiga kriteria tidak terpenuhi
S8	0,57	-1,03	0,58	Ketiga kriteria terpenuhi
S9	0,80	-1,02	0,59	Ketiga kriteria terpenuhi
S10	1,12	0,50	0,41	Ketiga kriteria terpenuhi
S11	1,19	0,81	0,45	Ketiga kriteria terpenuhi
S12	0,42	-1,59	0,67	Dua kriteria terpenuhi
S13	1,10	0,53	0,46	Ketiga terpenuhi
S14	0,70	-1,11	0,59	Ketiga terpenuhi
S15	0,53	-1,29	0,61	Ketiga terpenuhi
S16	0,91	-0,52	0,55	Ketiga terpenuhi
S17	1,05	0,30	0,48	Ketiga terpenuhi
S18	0,72	-1,40	0,61	Ketiga terpenuhi
S19	2,14	3,20	0,06	Ketiga kriteria tidak terpenuhi

Berdasarkan Tabel 7, sebanyak 17 dari 19 butir soal memenuhi minimal dua dari tiga kriteria *item fit*

yang dipersyaratkan. Sebagian besar butir soal menunjukkan nilai *outfit* MNSQ yang berada dalam

rentang 0,5–1,5, nilai ZSTD dalam rentang –2,0 hingga +2,0, serta nilai *Pt. Measure Corr* yang berkisar antara 0,4 hingga 0,85, menandakan bahwa butir-butir tersebut berfungsi sesuai dengan model pengukuran *Rasch* dan berkontribusi secara konsisten dalam mengukur konstruk pemahaman konsep. Hasil ini konsisten dengan penelitian Aditya et al., (2025), Mardian et al., (2023), dan Irawan et al., (2025) yang juga memperoleh sebagian besar butir memenuhi kriteria item fit sehingga instrumen dinilai layak digunakan.

Butir S7 memperoleh nilai *outfit* MNSQ = 1,86 dan ZSTD = 3,72 dengan *Pt. Measure Corr* = 0,15, yang ketiganya berada di luar batas kriteria. Nilai MNSQ dan ZSTD yang sangat tinggi mengindikasikan adanya ketidaksesuaian yang bersifat acak (*noise*) pada pola respons peserta didik, sementara nilai *Pt. Measure Corr* yang sangat rendah menunjukkan bahwa butir ini tidak mampu membedakan peserta didik berkemampuan tinggi dari yang rendah. Kondisi ini mengindikasikan kelemahan pada konstruksi distraktor maupun ketidakjelasan *stem* soal (Kassiavera et al., 2024; Juandi et al., 2024). Butir S19 juga tidak memenuhi ketiga kriteria dengan *outfit* MNSQ = 2,14, ZSTD = 3,20, dan *Pt. Measure Corr* = 0,06, yang merupakan nilai paling ekstrem di antara seluruh butir. Kedua butir ini memerlukan revisi menyeluruh sebelum dapat digunakan kembali. Adapun butir S2 dipertahankan meskipun nilai *Pt. Measure Corr*-nya berada di bawah batas (0,25), karena kedua indikator lainnya, yakni *outfit* MNSQ = 1,49 dan ZSTD = 2,42, masih berada pada batas yang dapat ditoleransi, serta butir ini telah dinyatakan valid secara isi. Secara keseluruhan, hasil analisis validitas konstruk menunjukkan bahwa instrumen ini telah memenuhi standar kualitas psikometrik yang dipersyaratkan dan layak digunakan dalam pengukuran pemahaman konsep peserta didik.

Reliabilitas Instrumen

Reliabilitas instrumen dievaluasi melalui tiga indeks, yaitu *person reliability*, *item reliability*, dan *Cronbach alpha* (KR-20). Hasil uji reliabilitas disajikan pada Tabel 8.

Tabel 8. Hasil Uji Reliabilitas Instrumen Tes Pemahaman Konsep

Indeks	Nilai	Kategori
Person Reliability	0,71	Cukup

Tabel 9. Hasil Analisis Tingkat Kesukaran Instrumen Tes Pemahaman Konsep

Butir Soal	JMLE Measure	SD	Kategori
S1	0,19	0,66	Sulit
S2	0,88	0,66	Sangat Sulit
S3	-0,15	0,66	Mudah

Item Reliability	0,94	Sangat Baik
Cronbach Alpha KR-20	0,77	Sangat Baik

Berdasarkan Tabel 8, nilai *person reliability* yang diperoleh adalah sebesar 0,71 yang tergolong dalam kategori cukup. Nilai *item reliability* tercatat sebesar 0,94 dan berada dalam kategori sangat baik. Adapun nilai *Cronbach alpha* KR-20 yang dihasilkan sebesar 0,77 juga termasuk dalam kategori sangat baik. Berdasarkan keseluruhan perolehan tersebut, dapat disimpulkan bahwa instrumen tes pemahaman konsep yang dikembangkan telah memenuhi standar reliabilitas yang dipersyaratkan dan layak digunakan dalam kegiatan penelitian (Noperi et al., 2025).

Ditinjau dari aspek reliabilitas, instrumen menunjukkan performa yang kuat pada level item dengan nilai *item reliability* sebesar 0,94 (Sangat Baik), yang mengindikasikan bahwa hierarki tingkat kesukaran butir soal bersifat sangat stabil dan tidak bergantung pada sampel yang digunakan. Nilai *Cronbach alpha* KR-20 sebesar 0,77 (Sangat Baik) turut mencerminkan konsistensi internal instrumen yang solid, menandakan bahwa seluruh butir soal secara bersama-sama mengukur konstruk yang sama secara ajeg. Sementara itu, nilai *person reliability* sebesar 0,71 yang berada pada kategori cukup menunjukkan bahwa instrumen mampu membedakan kemampuan peserta didik pada tingkat moderat. Kondisi ini kemungkinan besar dipengaruhi oleh kondisi *mistargeting*, di mana sebagian besar peserta didik terkonsentrasi pada rentang kemampuan tertentu yang belum sepenuhnya terwakili oleh distribusi kesukaran item yang tersedia. Dengan menambahkan butir soal pada rentang kesukaran yang belum terwakili, nilai *person reliability* diperkirakan akan meningkat secara signifikan pada pengembangan instrumen selanjutnya (Boone et al., 2014). Temuan ini memperkuat hasil penelitian sebelumnya yang menunjukkan bahwa penerapan Model *Rasch* berkontribusi pada peningkatan kualitas reliabilitas instrumen dalam penelitian pendidikan (Sumintono and Widhiarso 2015; Rohmatika et al., 2025).

Tingkat Kesukaran Butir Soal

Tingkat kesukaran setiap butir soal diestimasi berdasarkan nilai *JMLE Measure* dalam satuan logit. Hasil analisis disajikan pada Tabel 9.

S4	0,43	0,66	Sulit
S5	1,09	0,66	Sangat Sulit
S6	0,43	0,66	Sulit
S7	0,58	0,66	Sulit
S8	-1,11	0,66	Sangat Mudah
S9	0,35	0,66	Sulit
S10	-0,34	0,66	Mudah
S11	-0,06	0,66	Mudah
S12	-1,11	0,66	Sangat Mudah
S13	0,11	0,66	Sulit
S14	-0,34	0,66	Mudah
S15	-0,98	0,66	Sangat Mudah
S16	-0,53	0,66	Mudah
S17	0,23	0,66	Sulit
S18	-0,11	0,66	Mudah
S19	1,47	0,66	Sangat Sulit

Berdasarkan Tabel 9, distribusi tingkat kesukaran instrumen mencakup seluruh spektrum kemampuan peserta didik. Terdapat 3 butir soal berkategori sangat sulit, yaitu S2 (0,88 logit), S5 (1,09 logit), dan S19 (1,47 logit); 6 butir soal berkategori sulit, yaitu S1, S4, S6, S7, S9, S13, dan S17; 6 butir soal berkategori mudah, yaitu S3, S10, S11, S14, S16, dan S18; serta 3 butir soal berkategori sangat mudah, yaitu S8, S12, dan S15 (keduanya pada -1,11 logit). Rentang keseluruhan nilai logit membentang dari -1,11 hingga 1,47. Distribusi tingkat kesukaran yang mencakup beberapa kategori menunjukkan bahwa instrumen memiliki kemampuan yang baik dalam membedakan kemampuan peserta didik pada berbagai tingkat kemampuan, sebagaimana juga ditemukan oleh Rafiah & Miyaqi (2025) serta (Irawan et al., 2025).

Ditinjau dari distribusi kesukaran secara keseluruhan, profil instrumen ini menunjukkan dominasi butir pada kategori sulit dan mudah yang relatif seimbang, sehingga instrumen mampu

mendiskriminasi kemampuan peserta didik di berbagai level secara lebih akurat (Boone et al., 2014). Namun demikian, perlu dicatat bahwa terdapat celah distribusi item pada rentang logit di atas 1,5, yang menunjukkan ketiadaan butir soal pada rentang kesukaran sangat tinggi tersebut. Celah ini berpotensi membatasi kemampuan instrumen dalam mengukur peserta didik dengan kemampuan sangat tinggi secara presisi. Oleh karena itu, penambahan 2-3 butir baru pada rentang kesukaran yang belum terwakili direkomendasikan untuk meningkatkan cakupan dan ketepatan pengukuran pada pengembangan instrumen selanjutnya.

Differential Item Functioning (DIF)

Analisis DIF dilakukan berdasarkan nilai probabilitas *Rasch-Welch* (Prob. t) untuk mendeteksi keberpihakan item berdasarkan gender. Hasil analisis disajikan pada Tabel 10.

Tabel 10. Hasil Analisis DIF Instrumen Tes Pemahaman Konsep

Item	Logit (L)	Logit (P)	DIF Contrast	Prob. (t)	Keterangan
S1	0,40	0,07	0,33	0,582	Tidak signifikan
S2	0,62	1,01	-0,39	0,501	Tidak signifikan
S3	-0,59	0,07	-0,67	0,306	Tidak signifikan
S4	-0,07	0,68	-0,74	0,221	Tidak signifikan
S5	1,26	1,01	0,25	0,663	Tidak signifikan
S6	0,84	0,20	0,64	0,280	Tidak signifikan
S7	1,26	0,20	1,27	0,036	Signifikan

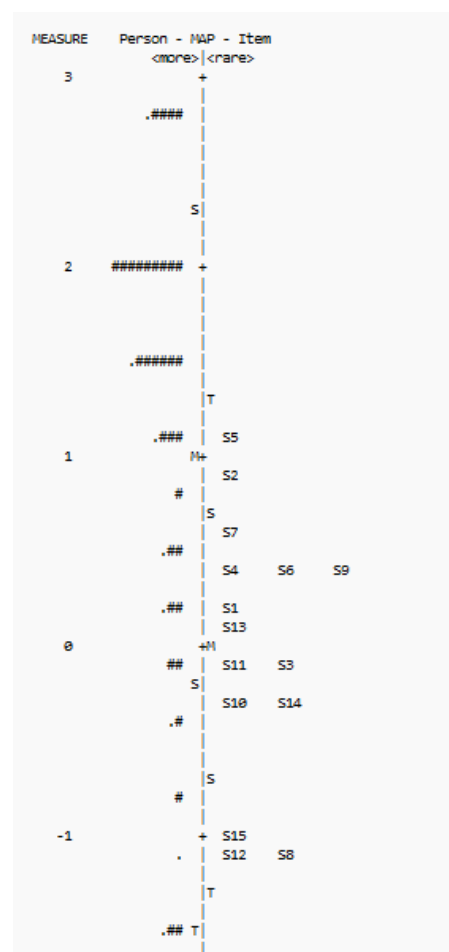
S8	-1,60	-0,85	-0,76	0,332	Tidak signifikan
S9	0,62	0,20	0,42	0,476	Tidak signifikan
S10	-0,07	-0,50	0,44	0,029	Signifikan
S11	-0,89	0,33	-1,22	0,071	Tidak signifikan
S12	-0,89	-1,26	0,37	0,620	Tidak signifikan
S13	-0,06	0,20	-0,27	0,665	Tidak signifikan
S14	-0,59	-0,20	-0,39	0,550	Tidak signifikan
S15	-0,59	-1,26	0,67	0,357	Tidak signifikan
S16	-0,07	-0,50	0,44	0,496	Tidak signifikan
S17	0,23	-0,07	0,30	0,627	Tidak signifikan
S18	-0,89	-0,20	-0,69	0,268	Tidak signifikan
S19	1,47	-0,43	1,90	0,009	Signifikan

Berdasarkan Tabel 10, dari 19 butir soal yang dianalisis, sebanyak 16 butir soal tidak menunjukkan DIF yang signifikan (Prob. > 0,05), yang berarti butir-butir tersebut berfungsi secara setara bagi peserta didik laki-laki maupun perempuan pada tingkat kemampuan yang sama. Terdapat tiga butir soal yang menunjukkan DIF signifikan berdasarkan gender, yaitu S7 (Prob. = 0,036), S10 (Prob. = 0,029), dan S19 (Prob. = 0,009).

Ditinjau lebih lanjut, butir S7 memiliki nilai logit sebesar 1,26 untuk peserta didik laki-laki dan 0,20 untuk peserta didik perempuan, dengan DIF Contrast = 1,27, mengindikasikan bahwa butir ini relatif lebih sulit bagi peserta didik laki-laki dibandingkan perempuan. Butir S10 menunjukkan pola sebaliknya dengan nilai logit -0,07 (laki-laki) dan -0,50 (perempuan) serta DIF Contrast = 0,44, yang berarti butir ini sedikit lebih menantang bagi peserta didik laki-laki. Butir S19 menunjukkan DIF paling kuat di antara ketiganya, dengan DIF Contrast = 1,90 logit dan nilai probabilitas paling kecil (Prob. = 0,009), mengindikasikan keberpihakan yang paling nyata terhadap salah satu kelompok gender. Keberadaan DIF pada ketiga butir tersebut mengindikasikan bahwa perbedaan peluang menjawab benar tidak semata-mata dipengaruhi oleh kemampuan peserta didik, melainkan juga oleh faktor konteks, penggunaan bahasa, maupun karakteristik stimulus soal yang secara tidak langsung lebih familiar bagi kelompok gender tertentu. Temuan ini sejalan dengan Susongko et al., (2021) dan Halim et al. (2026) yang menekankan pentingnya analisis DIF sebagai bagian dari prosedur standar pengembangan instrumen untuk menjamin keadilan pengukuran lintas kelompok. Ketiga butir ini perlu dikaji ulang dan direvisi sebelum digunakan dalam penilaian sumatif yang melibatkan kelompok peserta didik yang beragam secara gender.

Peta Variabel (Wright Map)

Selain analisis butir soal dan reliabilitas, Model *Rasch* juga menghasilkan *Wright Map* yang memvisualisasikan distribusi kemampuan peserta didik dan tingkat kesukaran butir soal secara bersamaan pada skala logit yang sama. *Wright Map* hasil analisis instrumen disajikan pada Gambar 1 berikut.



Gambar 1. Wright Map Instrumen Tes Pemahaman Konsep

Berdasarkan Gambar 1, rata-rata kemampuan peserta didik berada pada 1,14 logit, sementara rata-rata tingkat kesukaran item berada pada 0,00 logit sebagaimana ditetapkan dalam Model *Rasch*. Selisih sebesar 1,14 logit ini mengindikasikan bahwa secara umum kemampuan peserta didik berada di atas rata-rata tingkat kesukaran instrumen (Nurani, Rokhman, & Sarwi, 2025), atau dengan kata lain instrumen cenderung lebih mudah dari yang dibutuhkan untuk mengukur kelompok peserta didik ini secara optimal. Butir S19 merupakan butir tersulit dengan nilai 1,47 logit, sedangkan butir S8 dan S12 merupakan butir termudah dengan nilai masing-masing -1,11 logit. Rentang kesukaran item secara keseluruhan membentang dari -1,11 hingga 1,47 logit, dengan nilai simpangan baku sebesar 0,66 logit.

Ditinjau dari keselarasan antara distribusi kemampuan peserta didik dan tingkat kesukaran item, kondisi *mistargeting* yang tergambar pada *Wright Map* ini merupakan hal yang wajar pada tahap uji coba awal sebuah instrumen (Boone et al., 2014). Meskipun demikian, peta variabel juga mengungkapkan adanya celah distribusi item yang cukup signifikan pada rentang 1,5–2,5 logit, di mana tidak terdapat satu pun butir soal yang mampu mengukur peserta didik dengan kemampuan tinggi secara akurat. Kondisi ini berpotensi menurunkan presisi pengukuran bagi peserta didik yang berada pada rentang kemampuan tersebut. Oleh karena itu, penambahan 2–3 butir baru dengan tingkat kesukaran pada rentang logit yang belum terwakili sangat direkomendasikan, agar instrumen mampu mengukur kemampuan peserta didik di seluruh spektrum secara lebih merata dan presisi pada pengembangan selanjutnya.

Kesimpulan

Instrumen tes pemahaman konsep pada materi energi terbarukan yang dikembangkan dalam penelitian ini secara keseluruhan telah memenuhi kriteria kualitas psikometrik yang memadai dan layak digunakan sebagai alat ukur pemahaman konsep peserta didik. Instrumen terbukti memenuhi asumsi unidimensionalitas, memiliki reliabilitas item yang sangat baik, serta konsistensi internal yang sangat baik, meskipun reliabilitas person masih berada pada kategori cukup dan perlu ditingkatkan.

Dari sisi validitas isi, sebagian besar butir soal telah dinyatakan valid, sementara sebagian kecil butir soal pada indikator menjelaskan belum memenuhi standar validitas isi dan memerlukan revisi pada aspek konstruksi dan kesesuaian materi. Dari sisi validitas konstruk, hampir seluruh butir soal memenuhi kriteria item fit, dengan pengecualian pada beberapa butir yang menunjukkan pola respons tidak sesuai model dan perlu direvisi secara menyeluruh. Selain itu, beberapa butir

teridentifikasi mengandung keberpihakan (bias) berdasarkan gender sehingga perlu dikaji ulang dari aspek kebahasaan dan konteks stimulus soal sebelum digunakan dalam penilaian lintas kelompok.

Untuk mengoptimalkan kualitas instrumen, disarankan agar butir-butir yang bermasalah direvisi dan dilengkapi dengan penambahan butir baru pada rentang kesukaran yang belum terwakili, sehingga instrumen mampu mengukur kemampuan peserta didik di seluruh level secara lebih merata dan presisi pada penelitian selanjutnya.

Ucapan Terimakasih

Penulis mengucapkan terima kasih kepada Universitas Pendidikan Indonesia atas dukungan fasilitas dan sumber daya akademik yang diberikan selama pelaksanaan penelitian ini. Terima kasih juga disampaikan kepada seluruh validator ahli, yaitu dosen Program Studi Pendidikan Fisika dan guru fisika SMA yang telah meluangkan waktu untuk memberikan penilaian terhadap instrumen penelitian. Penghargaan yang tulus disampaikan pula kepada Kepala Sekolah, guru fisika, serta seluruh peserta didik kelas X yang telah memberikan izin, dukungan, dan partisipasi aktif selama proses pengambilan data berlangsung. Tanpa keterbukaan dan kerja sama dari seluruh pihak di sekolah, penelitian ini tidak dapat terlaksana dengan baik.

Referensi

- Aditya, R., Suhendi, E., Aviyanti, L., Nawas, A., & Rusnayati, H. (2025). Analysis of critical thinking test instruments on the light waves concept (CT-LiWa): *Rasch* model. *Momentum: Physics Education Journal*, 9(2), 337–349. <https://doi.org/10.21067/mpej.v9i2.11662>
- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45(1), 131–142. <https://doi.org/10.1177/0013164485451012>
- Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.
- Bashooir, K., & Supahar, S. (2018). Validitas dan reliabilitas instrumen asesmen kinerja literasi sains pelajaran fisika berbasis STEM. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 22(2), 219–230. <https://doi.org/10.21831/pep.v22i2.19590>
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. Dordrecht: Springer.
- Chan, S. W., Ismail, Z., & Sumintono, B. (2021). *Rasch analysis for instrument development: Why, when, and how?* International Journal of Academic Research in Business and Social Sciences, 11(9), 585-599. <https://doi.org/10.6007/IJARBS/v11-i9/10802>
- Dong, Y., Xu, W., Huang, J., & Yann, K. (2025). Validating and refining a multi-dimensional scale for measuring AI literacy in education using the *Rasch* Model. Humanities and Social Sciences Communications, 12, 1317. <https://doi.org/10.1057/s41599-025-05670-6>
- Festiana, I., Herlina, K., Zakiya, H., & Falamy, R. A. (2026). Analisis Literasi Energi Terbarukan Mahasiswa Pendidikan Fisika: Sikap, Perilaku Hemat Energi, dan Niat Adopsi Teknologi. JIPFRI (Jurnal Inovasi Pendidikan Fisika Dan Riset Ilmiah), 10(1), 139-147. <https://doi.org/10.30599/jipfri.v10i1.5227>
- Halim, M. G., Saepuzaman, D., Aviyanti, L., Utama, J. A., & Nawas, A. (2026). Measuring critical thinking in physics: A *Rasch* analysis of instrument quality and gender equivalence. Jurnal Pendidikan MIPA, 27(1), 452-475. <https://doi.org/10.23960/jpmipa.v27i1.pp452-475>
- Irawan, I. D. A., Indraloka, R. M., Basri, N. A., Salmah, U., & Parno. (2025). Analysis of concept understanding test items on static fluid material using *Rasch* model. Jurnal Pendidikan Fisika, 13(1), 1-13. <https://doi.org/10.26618/jpf.v13i1.15687>
- Juandi, T., Kaniawati, I., Samsudin, A., & Riza, L. S. (2024). The application of *Rasch* model to analyse the validity and reliability of an instrument for reflective thinking skills on topic of wave-particle dualism. Kappa Journal, 8(2), 270-277. <https://doi.org/10.29408/kpj.v8i2.27049>
- Kassiavera, S., Suparmi, A., Cari, C., & Sukarmin. (2024). Application of *Rasch* model in two-tier test for assessing critical thinking in physics education. Journal of Baltic Science Education, 23(6), 1227-1242. <https://doi.org/10.33225/jbse/24.23.1227>
- Kemendikbudristek. (2022). Keputusan Kepala Badan Standar, Kurikulum, dan Asesmen Pendidikan Nomor 008/H/KR/2022 tentang Capaian Pembelajaran pada Pendidikan Anak Usia Dini, Jenjang Pendidikan Dasar, dan Jenjang Pendidikan Menengah pada Kurikulum Merdeka. Jakarta: Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi.
- Linacre, J. M. (2016). Winsteps® *Rasch* measurement computer program user's guide (Version 3). Version 3. Winsteps.com. Retrieved from <https://www.winsteps.com>
- Mardian, V., Samsudin, A., Utama, J. A., Suwarma, I. R., & Costu, B. (2023). Validity and reliability of elasticity multiple-choice items (EMCI) using *Rasch* model. Jurnal Ilmiah Pendidikan Fisika Al-Biruni, 12(2). <https://doi.org/10.24042/jipfalbiruni.v12i2.17037>
- Noperi, H., Diani, R., Irwandani, Suryani, Y., Sodikin, S., & Saputro, T. (2025). Psychometric analysis of the critical thinking skills instrument in physics learning using the *Rasch* model. Indonesian Journal of Science and Mathematics Education, 8(3), 627-641. <https://doi.org/10.24042/ijsme.v8i3.28770>
- Nugraha, I., & Suhandi, A. (2021). Developing a context-based energy literacy framework for secondary school students. Journal of Physics: Conference Series, 1806(1), 12005. <https://doi.org/10.1088/1742-6596/1806/1/012005>
- Nurani, N. S., Rokhman, F., & Sarwi. (2025). Measuring the unseen: Digitalization of physics diagnostic assessment based on *Rasch* model. Journal of Educational Sciences, 9(6), 4950-4963. <https://doi.org/10.31258/jes.9.6.p.4950-4963>
- Pambudi, N. A., Nanda, I. R., Alfina, F. T., & Syahrial, A. Z. (2024). Renewable energy education and awareness among Indonesian students: Exploring challenges and opportunities for a sustainable future. Sustainable Energy Technologies and Assessments, 63, 103631. <https://doi.org/10.1016/j.seta.2024.103631>
- Rafiah, J. A., & Miyaqi, I. A. (2025). Analyze the students' conceptual understanding of elasticity and Hooke's law using the *Rasch* model. Impulse: Journal of Research and Innovation in Physics Education, 5(1). <https://doi.org/10.14421/impulse.2025.51-06>
- Rohmatika, R., Fatmawati, S., & Suhartono, S. (2025). Analisis butir soal instrumen tes keterampilan berpikir kreatif pada materi energi terbarukan menggunakan *Rasch* model. Jurnal EDUCATIO: Jurnal Pendidikan Indonesia, 11(1), 92-106. <https://doi.org/10.29210/1202525556>
- Setiawan, B., & Rosid, M. S. (2024). Indonesia's energy transition path: Challenges in public awareness and education. International Journal of Energy Economics and Policy, 14(1), 12-25.
- Sujatmika, S., Sutarno, Masykuri, M., & Prayitno, B. A. (2025). Applying the *Rasch* model to measure students' critical thinking skills on the science

- topic of the human circulatory system. *Eurasia Journal of Mathematics, Science and Technology Education*, 21(4), em2622.
<https://doi.org/10.29333/ejmste/16221>
- Sumintono, B., & Widhiarso, W. (2015). Aplikasi Pemodelan *Rasch* pada Asesmen Pendidikan. Cimahi: Trim Komunikata.
- Susongko, P., Arfiani, Y., & Kusuma, M. (2021). Determination of gender differential item functioning in Tegal students' scientific literacy skills with integrated science (SLiSIS) test using *Rasch* model. *Jurnal Pendidikan IPA Indonesia*, 10(2), 270–281.
<https://doi.org/10.15294/jpii.v10i2.26775>
- Taherdoost, H. (2018). Validity and reliability of the research instrument: How to test the validation of a questionnaire/survey in a research. *International Journal of Academic Research in Management*, 5(3), 28–36.
- Usman, M., Pramudawardani, H., Umam, R., & Takahashi, H. (2025). Synthesizing a framework and establishing content validity of an energy literacy instrument for Indonesian students. *Islamic Journal of Integrated Science Education (IJISE)*, 4(3), 175–185.
<https://doi.org/10.30762/ijise.v4i3.6993>