Vol. 11, No. 1, 2025, pp. 92-106
DOI: https://doi.org/10.29210/1202525556

Daftar isi tersedia di Jurnal IICET

Jurnal EDUCATIO (Jurnal Pendidikan Indonesia)
ISSN: 2476-9886 (Cetak) ISSN: 2477-0302 (Elektronik)

Beranda jurnal: https://jurnal.iicet.org/index.php/jppi

Analisis butir soal instrumen tes keterampilan berpikir kreatif pada materi energi terbarukan menggunakan rasch model

Ria Rohmatika*), Sri Fatmawati, Suhartono Suhartono
Program Studi Tadris Fisika, Fakultas Tarbiyah dan Ilmu Keguruan, Institut Agama Islam Negeri Palangka Raya, Indonesia

Article Info

Article history:

Received Mar 04th, 2025
Revised Mar 20th, 2025
Accepted Apr 10th, 2025

Keywords:

Instrumen soal
Berpikir kreatif
Rasch model
Wright maps
Differential item function

ABSTRACT

Keterbatasan instrumen tes keterampilan berpikir kreatif pada materi energi terbarukan menjadi alasan perlunya pengembangan dan analisis instrumen secara mendalam. Penelitian ini bertujuan untuk menganalisis butir soal instrumen tes keterampilan berpikir kreatif pada materi energi terbarukan secara keseluruhan menggunakan Rasch Model. Pendekatan kuantitatif deskriptif diterapkan sebagai metode penelitian dengan mengambil subjek sebanyak 34 siswa kelas X SMA Negeri 1 Palangka Raya yang dipilih dengan teknik *purposive sampling.* Pengumpulan data menggunakan teknik tes berupa soal essay sebanyak 12 soal. Hasil data dianalisis menggunakan Rasch Model berbantuan *software Winsteps* 5.9.0. dengan beberapa parameter uji: *unidimensionality, item fit*, DIF, reliabilitas, *separation*, *wright maps*, dan tingkat kesukaran soal. Hasil penelitian menunjukkan *unidimensionality*, dengan nilai *raw variance explained by measures* sebesar 46,9% (baik), dari 12 soal yang dianalisis, 6 soal *fit* dengan model, sementara sisanya perlu direvisi. Tidak teramati bias terhadap butir soal berdasarkan hasil DIF (p>0,05). Namun, 2 butir soal dengan nilai logit ekstrem (2<logit<-2) tergolong sulit sekali dan mudah sekali sehingga perlu perhatian lebih lanjut. Nilai *Cronbach's Alpha* senilai 0,62 tergolong cukup dalam pengukuran. Hasil indeks *separation* (H) *item* (4), dan *person* (2) menunjukkan kemampuan instrumen dalam membedakan tingkat kemampuan siswa dan tingkat kesukaran soal secara jelas. Sebaran kemampuan siswa berdasarkan *wright maps,* berada pada logit 0 hingga -1. Instrumen ini memiliki kualitas yang cukup valid dan reliabel sebagai alat ukur keterampilan berpikir kreatif siswa pada materi energi terbarukan dan dapat digunakan oleh pendidik. Rekomendasi penelitian selanjutnya perlu mengembangkan instrumen dengan indikator berpikir kreatif secara merata, memperluas sampel uji coba instrumen, serta mengombinasikan pendekatan analisis lainnya agar kualitas instrumen lebih akurat.

© 2025 Para Penulis. Diterbitkan oleh IICET.
Ini adalah artikel akses terbuka di bawah lisensi CC BY-NC-SA
(https://creativecommons.org/licenses/by-nc-sa/4.0)

Corresponding Author:

Ria Rohmatika,
Institut Agama Islam Negeri Palangka Raya
Email: riarohmatika2111130052@iain-palangkaraya.ac.id

Pendahuluan

Seiring berkembangnya teknologi dan ilmu pengetahuan, keterampilan berpikir kreatif menjadi salah satu prospek yang berperan sangat penting dalam mengatasi berbagai tantangan. Keterampilan ini memungkinkan setiap individu agar menemukan solusi inovatif dalam meningkatkan efektivitas dan efisiensi dari berbagai bidang (Anindayati & Wahyudi, 2020). Dengan adanya keterampilan berpikir kreatif

ini, setiap individu tidak hanya mampu memecahkan masalah dengan pendekatan baru tetapi juga mampu beradaptasi terhadap perubahan lingkungan sekitar. Menurut Brandt (2024), berpikir kreatif merupakan proses berulang yang dilakukan seseorang dalam menghasilkan dan memanipulasi gagasan, menguji, menyempurnakan, dan memodifikasi gagasan tersebut melalui analisis dan evaluasi kritis, serta mengomunikasikan gagasan untuk memecahkan masalah, meningkatkan solusi masalah, atau memajukan pengetahuan dengan cara baru.

Dalam konteks pembelajaran, keterampilan berpikir kreatif berperan dalam meningkatkan pengetahuan siswa, baik dalam kehidupan sehari-hari maupun bidang akademik, termasuk fisika (Hermawati & Chen, 2023). Dalam pembelajaran fisika, keterampilan berpikir kreatif merupakan kemampuan mengemukakan ide secara kreatif guna menyelesaikan permasalahan fisika (Wulandari & Jumadi, 2023). Pada hakikatnya, pembelajaran fisika merupakan pembelajaran yang melibatkan konsep-konsep abstrak yang membutuhkan pemikiran tingkat lanjut dan kreatif. Oleh karena itu, pembelajaran fisika memerlukan strategi dan desain yang tepat dengan melibatkan siswa secara intelektual dan emosional serta memusatkan perhatian mereka pada pembelajaran mandiri (Rosha & Hidayat, 2023).

Namun, hasil *Program for International Student Assesment* (PISA) 2022, keterampilan berpikir kreatif siswa Indonesia teramati berada di bawah sebagian besar anggota *Organization for Economic Co-operation and Development* (OECD), menduduki peringkat ke 52 dari 64 negara dengan rata-rata 19 dari nilai rata-rata standar OECD sebesar 33 (OECD, 2024). Meskipun PISA lebih memfokuskan pada pengukuran literasi membaca, matematika, dan sains, namun melalui analisis terhadap kinerja siswa dalam ketiga domain tersebut dapat diambil kesimpulan bahwa terdapat potensi yang belum tergali secara optimal dalam mengembangkan keterampilan berpikir kreatif siswa Indonesia. Beberapa faktor seperti latar belakang sosial, ekonomi, dan kualitas pembelajaran di sekolah diduga menjadi faktor yang mempengaruhi kinerja siswa. Sejalan dengan temuan tersebut, hasil pra-penelitian yang dilakukan di salah satu sekolah menengah atas (SMA) dengan mengambil sampel sebanyak 1 kelas X yang berjumlah 38 siswa (19 laki-laki dan 19 perempuan). Pengumpulan data dilakukan melalui pemberian soal tes berpikir kreatif pada mata pelajaran fisika materi besaran dan satuan sebanyak 9 soal yang mencakup 3 indikator berpikir kreatif, yaitu berpikir lancar (*fluency*), berpikir luwes (*flexibility*), dan menghasilkan ide asli (*originality*). Hasil temuan tersebut mengindikasikan bahwa keterampilan berpikir kreatif siswa secara keseluruhan masih berada pada kriteria cukup dengan presentase sebesar 56,3%. Dari ketiga indikator yang diamati, berpikir luwes menjadi indikator terendah yang belum sepenuhnya dikuasai oleh sebagian besar siswa.

Merujuk pada temuan sebelumnya, pentingnya upaya untuk mengembangkan keterampilan berpikir kreatif siswa melalui strategi pembelajaran dan asesmen yang tepat. Dalam mengukur keterampilan berpikir kreatif, diperlukan rancangan asesmen yang sistematis dan sesuai dengan indikator berpikir kreatif. Sala satu teori yang dapat dijadikan acuan adalah *Torrance Test of Creative Thinking* (TTCT) yang dikembangkan oleh Torrance (1966). Teori ini mencakup empat indikator utama penilaian, yaitu berpikir luwes (*flexibility*), menghasilkan ide asli (*originality*), melakukan elaborasi (*elaboration*), dan berpikir lancar (*fluency*) (Afandi & Sajidan, 2017). Menurut Fitriyyah et al. (2024), asesmen/penilaian merupakan kegiatan penting untuk mengukur kemampuan siswa dalam proses pembelajaran. Sejalan dengan itu, Suprapto et al. (2020), mengatakan bahwa asesmen/penilaian memiliki fungsi sebagai alat untuk mengevaluasi ketercapaian tujuan pembelajaran, sehingga perlu mengacu pada rencana pembelajaran, sebagai umpan balik guru dan siswa guna memperbaiki strategi pengajaran maupun metode pembelajaran, serta sebagai laporan peningkatan belajar siswa kepada orang tua. Oleh karena itu, asesmen keterampilan berpikir kreatif perlu disusun dengan baik agar menjadi alat untuk meningkatkan berpikir kreatif siswa.

Sebuah instrumen asesmen yang akurat memerlukan sebuah pendekatan analisis yang tepat, salah satunya adalah Rasch Model yang merupakan bagian dari model terkenal dalam *Item Response Theory* (IRT) (Mawaddah et al., 2024). Menurut Sumintono & Widhiarso (2015), Rasch Model didefinisikan sebagai sebuah model analisis yang berperan untuk memeriksa validitas dan reliabilitas suatu instrumen, serta mengevaluasi tiap individu dan item dalam ruang yang sama. Pengukuran pada Rasch Model mengacu pada lima prinsip, yaitu pengukuran yang dihasilkan bersifat linier, mampu mengatasi masalah data yang tersembunyi, dapat menentukan keakuratan terhadap penilaian, mampu menyelediki ketidakakuratan model, serta menawarkan alat ukur yang tidak terpengaruh oleh parameter yang sedang diteliti (Ernawati, 2022; Suryani & Jaedun, 2022; Tanujaya & Yudiarso, 2023). Maka untuk mengetahui kualitas butir soal tes keterampilan berpikir kreatif khususnya pada materi energi terbarukan dapat memanfaatkan Rasch Model ini.

Analisis Rasch Model telah banyak diterapkan pada penelitian terdahulu, sebagaimana penelitian yang dilakukan oleh Hasanah & Purwanto (2023) yang menganalisis kualitas instrumen tes fisika berbasis *High*

Order Thinking Skills (HOTS) pada topik elastisitas dan Hukum Hooke. Penelitian tersebut menunjukkan bahwa dengan analisis Rasch Model dapat diketahui kevalidan butir soal melalui parameter uji kesesuaian butir soal (*item fit*), dan distribusi tingkat kemampuan siswa dan tingkat kesukaran butir soal yang digambarkan dalam parameter uji variabel (*wright*) *maps.* Selain itu, penelitian Irmayanti et al. (2023), juga fokus menganalisis instrumen tes fisika berbasis *High Order Thinking Skills* (HOTS) dengan parameter uji reliabilitas (*person reliability, item reliability* dan *Cronbach's Alpha*), kesesuaian butir soal (*item fit*), tingkat kesukaran tiap butir soal, dan variabel (*wright*) *maps*. Dengan parameter uji yang digunakan, menghasilkan instrumen yang valid dan reliabel, serta mampu membedakan tingkat kesukaran butir soal serta kemampuan siswa. Selanjutnya, penelitian yang serupa dengan topik penelitian ini telah diteliti oleh Rosha & Hidayat (2023), yang menerapkan Rasch Model untuk menganalisis instrumen tes berpikir kreatif siswa SMA pada topik energi terbarukan. Penelitian tersebut menerapkan parameter uji *unidimensionality,* reliabilitas (*person reliability, item reliability* dan *Cronbach's Alpha*), kesesuaian butir soal (*item fit*), tingkat kesukaran tiap butir soal, dan variabel (*wright*) *maps*. Hasil penelitian tersebut menunjukkan bahwa analisis Rasch Model mampu menghasilkan instrumen tes keterampilan berpikir kreatif yang valid secara keseluruhan melalui parameter uji *unidimensionality* dan reliabel.

Meskipun demikian, beberapa penelitian tersebut belum sepenuhnya menyajikan hasil analisis secara menyeluruh. Belum banyak ditemukan penelitian yang menganalisis serta menggambarkan butir soal instrumen tes keterampilan berpikir kreatif dalam pembelajaran fisika khususnya materi energi terbarukan dengan pendekatan kuantitatif melalui analisis Rasch Model dari berbagai parameter uji yang lebih luas, seperti *Differential Item Function* (DIF) yang menguji adanya bias terhadap instrumen yang digunakan. Dengan demikian, penelitian ini akan melengkapi kekurangan parameter uji yang belum digunakan secara umum oleh penelitian sebelumnya. Sehingga, rumusan masalah dalam penelitian ini adalah bagaimana kualitas butir soal instrumen tes keterampilan berpikir kreatif pada materi energi terbarukan secara menyeluruh berdasarkan analisis Rasch Model dari berbagai parameter uji Rasch Model tersebut?. Adapun tujuan dilakukannya penelitian ini adalah untuk menganalisis serta menggambarkan kualitas butir soal instrumen tes keterampilan berpikir kreatif pada materi energi terbarukan secara menyeluruh dengan berbagai parameter uji Rasch Model agar memperoleh instrumen yang valid serta reliabel.

Penelitian ini akan menerapkan pendekatan kuantitatif melalui analisis Rasch Model dari berbagai parameter uji, seperti *unidimensionality,* kesesuaian butir soal (*item fit*), *Differential Item Function* (DIF) , reliabilitas, *separation* (indeks pemisahan), variabel (*wright*) *maps*, dan tingkat kesukaran tiap butir soal. Hasil penelitian ini diharapkan mampu memberikan kontribusi terhadap pengembangan instrumen tes keterampilan berpikir kreatif yang lebih objektif dan sesuai dengan kebutuhan pembelajaran abad ke-21, khususnya pada pembelajaran fisika.

Metode

Penelitian ini menerapkan metode deskriptif dengan pendekatan kuantitatif yang bertujuan untuk menganalisis serta menggambarkan kualitas butir soal instrumen tes keterampilan berpikir kreatif pada materi energi terbarukan menggunakan Rasch Model dari berbagai aspek, seperti *unidimensionality,* kesesuaian butir soal (*item fit*), *Differential Item Function* (DIF) , reliabilitas, *separation* (indeks pemisahan), variabel (*wright*) *maps*, dan tingkat kesukaran tiap butir soal. Populasi penelitian mencakup siswa kelas X di SMA Negeri 1 Palangka Raya yang berjumlah 531 siswa. Sampel dipilih menggunakan teknik *purposive sampling,* yaitu 1 kelas berjumlah 34 siswa yang memperoleh materi energi terbarukan. Jumlah sampel yang digunakan cukup memadai untuk uji coba awal instrumen sesuai dengan pernyataan Linacre (1994), bahwa untuk memperoleh tingkat kepercayaan 99% dengan kalibrasi item yang cukup akurat (±1 logit) cukup membutuhkan 30-60 responden sebagai uji coba awal.

Instrumen soal yang akan digunakan telah melalui tahap pengembangan 4D, salah satu model *Research and Development* (R&D) yang dikembangkan oleh Thiagarajan & Sivasailan (1976). Tahap 4D meliputi, *define, design, develop,* dan *disseminate.* Tahapan model 4D dalam pengembangan instrumen dimulai dari tahap *define* yang meliputi studi kepustakaan terkait indikator berpikir kreatif, instrumen soal dan materi fisika energi terbarukan. Setelah itu, lanjut ke tahap *design* yang meliputi penyusunan soal merujuk pada indikator berpikir kreatif yang dikembangkan oleh Torrance, kisi-kisi soal yang terdiri dari indikator soal, indikator berpikir kreatif, level kognitif, bentuk soal dan kunci jawaban. Kemudian, lanjut ke tahap *develop* yang meliputi pengembangan instrumen tes berpikir kreatif, melakukan validasi instrumen, dan menganalisis hasil validasi dari para ahli. Tahap terakhir, yaitu *disseminate,* yang meliputi tahap uji coba instrumen tes berpikir kreatif pada siswa, serta mengolah hasil uji coba menggunakan Rasch Model.

Tabel 1 <Rangkuman Kriteria Parameter Uji dan Sumber *Output Winsteps*>

Parameter Uji	Kriteria	Sumber *Output*
Unidimensionality	1. Nilai *raw variance explained by measure* tidak kurang dari 20%. 2. Nilai *unexplned variance in 1st constract* tidak lebih dari 15%	23. *Item: dimensionality*
Item Fit (Tingkat Kesesuaian Butir Soal)	0,5 < MNSQ < 1,5 -2,0 < ZSTD < +2,0 0,4 < PT Mean Corr < 0,85	10. *Item* (*Column*): *fit order*
DIF (*Differential Item Functioning*)	Prob > 0,05 (Tidak bias)	30. *Item*: *DIF, between/within*
Person Reliability dan *Item Reliability*	>0,94 (Istimewa) 0,91 – 0,94 (Baik Sekali) 0,81 – 0,90 (Baik) 0,67 – 0,80 (Cukup) <0,67 (Lemah)	3.1 *Summary Statistics*
Cronbach's Alpha	>0,8 (Baik Sekali) 0,7 – 0,8 (Baik) 0,6 – 0,7 (Cukup) 0,5 – 0,6 (Jelek) <0,5 (Buruk)	3.1 *Summary Statistics*
Separation (indeks pemisahan)	$H_{separation} \geq 2,0$	3.1 *Summary Statistics*
Tingkat Kesukaran Butir Soal	*Measure logit* > +1 SD (Sulit Sekali) 0,00 ≤*Measure logit* ≤ +1 SD (Sulit) 0,00≥*Measure logit*≥–1SD (Mudah) *Measure logit* < – 1 SD (Mudah Sekali)	13. *Item*: *measure*
Wright Maps (Peta variabel)	-2,0 ≤*Measure logit* ≤ +2,0	1. *Variabel maps*

Sumber: Sumintono & Widhiarso, 2014

Merujuk pada kriteria pada Tabel 1, analisis butir soal instrumen tes dapat dilakukan secara sistematis dan menyeluruh. Hasil analisis yang memenuhi kriteria yang telah ditetapkan mencerminkan kualitas instrumen yang baik, valid dan reliabel untuk mengukur keterampilan berpikir kreatif siswa pada materi energi terbarukan.

Teknik pengumpulan data berupa tes tertulis berbentuk essay sebanyak 12 soal yang disusun berdasarkan indikator berpikir kreatif yang dikembangkan oleh Torrance (1966). Uji coba instrumen dilakukan secara *luring* di kelas pada jam pelajaran fisika selama 2 jam pelajaran (JP) yang berkisar selama 90 menit. Selama pengerjaan soal, siswa tidak diperkenankan menggunakan alat elektronik guna memastikan kejujuran serta kemampuan tiap siswa secara objektif. Sebelum soal tes digunakan, telah dilakukan tahap validasi ahli dengan melibatkan 2 validator ahli di bidang fisika yang meliputi guru dan dosen. Validasi ahli dilakukan agar dapat dinyatakan bahwa dengan estimasi soal tersebut sesuai dengan indikator berpikir kreatif dan layak diuji coba. Untuk melihat hasil kelayakan dari validasi ahli, digunakan kriteria kategorisasi yang perlu ditentukan sebelumnya dengan memuat data statistik seperti, jumlah subjek (N), mean (M), skor maksimum (X_{maks}), skor minimum (X_{min}), standar deviasi (SD), dan range. Hasil dari data tersebut kemudian digunakan untuk menentukan kriteria kategorisasi sesuai dengan pedoman yang dikemukakan oleh (Azwar, 2012) pada Tabel 2.

Tabel 2 <Pedoman Interpretasi Kelayakan>

Interval Koefisien	Kategori
X ≥ (Mean + 1,5SD)	Sangat Layak
(Mean +0,5SD)< X ≤ (Mean + 1,5SD)	Layak
(Mean - 0,5SD)< X ≤ (Mean +0,5SD)	Cukup Layak
(Mean - 1,5SD) < X ≤ (Mean - 0,5SD)	Kurang Layak
X ≤ (Mean - 1,5SD)	Tidak Layak

Merujuk pada pedoman tersebut, diperoleh kriteria kategorisasi untuk melihat kelayakan instrumen tes keterampilan berpikir kreatif pada materi energi terbarukan merujuk pada skor penilaian dari 2 validator yang disajikan pada Tabel 3.

Setelah instrumen memenuhi kriteria kelayakan pada Tabel 3 dan telah dilakukan perbaikan berdasarkan komentar dari masing-masing validator, instrumen tes kemudian diuji coba ke sampel yang telah ditentukan

guna mengetahui kualitas dari tiap butir soal instrumen tes. Sebelum pengumpulan data uji coba dilakukan, peneliti telah menyampaikan permohonan izin kepada pihak sekolah dan menyampaikan tujuan penelitian kepada siswa yang menjadi sampel penelitian. Penelitian dilakukan dengan menjaga kerahasiaan identitas siswa dan data yang diperoleh hanya digunakan untuk keperluan penelitian.

Tabel 3 <Kriteria Kelayakan Instrumen>

Interval Koefisien	Kategori
$X \leq 24$	Tidak Layak
$24 < X \leq 35$	Kurang Layak
$35 < X \leq 45$	Cukup Layak
$45 < X \leq 56$	Layak
$X \geq 56$	Sangat Layak

Teknik analisis data uji coba instrumen tes menggunakan Rasch Model berbantuan *software Winsteps* 5.9.0 untuk menguji dari berbagai parameter uji, antara lain *unidimensionality,* kesesuaian butir soal (*item fit*), DIF (*Differential Item Function*), reliabilitas, *separation* (indeks pemisahan), variabel (*wright*) *maps*, dan tingkat kesukaran tiap butir soal. Rasch Model dipilih karena mampu mengevaluasi kualitas instrumen secara menyeluruh, serta mampu memberikan informasi mengenai kualitas respon siswa dan butir soal dalam kerangka pengukuran yang sama. Selain itu, Rasch Model mampu menggambarkan kualitas butir soal dari berbagai parameter uji yang tidak sepenuhnya dapat dilakukan melalui pendekatan teori klasik (Sumintono & Widhiarso, 2015). Setiap parameter uji memiliki kriteria tertentu yang dijadikan sebagai acuan penilaian instrumen agar memenuhi syarat validitas dan reliabilitas. Tabel 3 berikut menyajikan rangkuman kriteria dari berbagai parameter uji, serta tabel *output* pada *software Winsteps* 5.9.0.

Hasil dan Diskusi

Hasil data dari penelitian merupakan hasil analisis uji coba instrumen tes keterampilan berpikir kreatif pada materi energi terbarukan sebanyak 12 soal dengan merujuk pada 4 indikator berpikir kreatif, yaitu elaborasi (*elaboration*), keaslian (*originality*), keluwesan (*flexibility*), dan kelancaran (*fluency*). Soal tersebut telah divalidasi oleh 2 validator, yaitu dosen dan guru yang ahli dalam bidang instrumen tersebut. Terdapat 6 aspek penilaian untuk mengevaluasi instrumen tes berpikir kreatif, meliputi kejelasan (*clarity*), relevansi (*relevenace*), keluwesan (*flexibility*), kelancaran (*fluency*), keaslian (*originality*), dan elaborasi (*elaboration*). Dari enam aspek tersebut, diuraikan menjadi 8 butir kriteria penilaian dengan 5 skala penilaian, yang meliputi 1 (sangat kurang baik), 2 (kurang baik), 3 (cukup baik), 4 (baik), dan 5 (sangat baik). Hasil skor penilaian dari 2 validator diperoleh skor sebesar 65 dengan kategori sangat layak yang merujuk pada Tabel 2. Kemudian, instrumen diuji coba ke siswa kelas 10 SMA Negeri 1 Palangka Raya yang berada di dalam satu kelas dengan jumlah siswa sebanyak 34 orang (15 laki-laki dan 19 perempuan). Hasil data uji coba kemudian dianalisis menggunakan *software Winsteps* 5.9.0. dengan berbagai parameter uji, seperti *unidimensionality,* kesesuaian butir soal (*item fit*), *Differential Item Function* (DIF), reliabilitas, *separation* (indeks pemisahan), variabel (*wright*) *maps*, dan tingkat kesukaran butir soal.

Unidimensionality

Nilai tersebut dapat diperoleh berdasarkan *output* tabel 23. *Item: dimensionality* pada *Winsteps* yang tertera pada Gambar 1.

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units

	Eigenvalue	Observed	Expected
Total raw variance in observations =	22.5885	100.0%	100.0%
Raw variance explained by measures =	10.5885	46.9%	49.1%
Raw variance explained by persons =	3.5153	15.6%	16.3%
Raw Variance explained by items =	7.0732	31.3%	32.8%
Raw unexplained variance (total) =	12.0000	53.1% 100.0%	50.9%
Unexplained variance in 1st contrast =	2.6724	11.8% 22.3%	
Unexplained variance in 2nd contrast =	2.0210	8.9% 16.8%	
Unexplained variance in 3rd contrast =	1.7729	7.8% 14.8%	
Unexplained variance in 4th contrast =	1.4644	6.5% 12.2%	
Unexplained variance in 5th contrast =	.9733	4.3% 8.1%	

Essential Unidimensionality (Rasch/Common variance) = 57.2%

Gambar 1 <Hasil *Output Item: Dimensionality*>

Validitas instrumen tes keterampilan berpikir kreatif dapat dilihat pada Rasch Model menggunakan istilah *item unidimensionality* yang dapat menggambarkan kemampuan instrumen tes dalam mengukur apa yang seharusnya diukur, yaitu keterampilan berpikir kreatif siswa, serta mampu mewakili data dari variabel secara akurat sehingga butir soal dapat dikatakan valid (Putri & Khusna, 2020; Sumintono & Widhiarso, 2015). Suatu instrumen dapat dianggap *unidimensionality* jika memenuhi dua persyaratan, yaitu nilai *raw variance explained by measure* tidak kurang dari 20%, yang berarti jika berada pada rentang 20%–40%, maka instrumen tes keterampilan berpikir kreatif pada materi energi terbarukan dinyatakan cukup baik. Sementara itu, jika berada pada rentang 40%–60% instrumen dianggap baik, dan jika nilai lebih dari 60%, maka instrumen dianggap baik sekali. Dan syarat yang kedua yaitu nilai *unexplned variance in 1st constract* tidak lebih dari 15% (Sumintono & Widhiarso, 2015).

Merujuk pada hasil *output item: dimensionality* yang tertera pada Gambar 1. Terlihat bahwa nilai *raw variance explained by measures* sebesar 46,9%, yang berarti nilai tersebut berada pada rentang 40%–60% sehingga instrumen dianggap baik. Sedangkan, nilai *unexplained variance in 1st contrast* teramati bahwa tidak lebih dari 15%, yaitu sebesar 11,8%. Begitu juga dengan nilai *unexplained variance in 2st contrast* hingga 5th tidak lebih dari 15% yang masing-masing bernilai, 8,9%, 7,8%, 6,5%, dan 4,3%. Hasil uji *unidimensionality* ini menunjukkan bahwa instrumen tes keterampilan berpikir kreatif pada materi energi terbarukan memenuhi syarat *unidimensionality* dan dianggap dapat mengevaluasi apa yang harusnya diukur. Temuan ini serupa dengan temuan Muhammad et al. (2023), yang memperoleh nilai *raw variance explained by measures* dengan kriteria baik dengan nilai sebesar 57,2% yang termasuk pada rentang 40%–60%. Begitupun dengan nilai *unexplained variance in 1st contrast* hingga 5th teramati bahwa berada di bahwa 15%, yang masing-masing bernilai 10,1%, 8,0%, 6,1%, 4,9%, dan 4,1%. Instrumen penelitian tersebut juga dianggap *unidimensionality* karena telah memenuhi kedua syarat yang ditentukan.

***Item Fit* (Tingkat Kesesuaian Butir Soal)**

Setelah diperoleh hasil uji validasi instrumen, selanjutnya yaitu mengamati kesesuaian tiap butir soal dari instrumen. Merujuk pada tabel *output* 10. Item: *fit order,* terlihat bahwa kesesuaian butir soal ditentukan berdasarkan 3 kriteria, yaitu *Outfit Means Square* (MNSQ), *Outfit Z-Standard* (ZTSD), dan *Point Measure Correlation* (Pt Measure Corr) yang tersaji dalam Tabel 4. Nilai *Outfit Means Square* (MNSQ), *Outfit Z-Standard* (ZTSD), dan *Point Measure Correlation* merupakan kriteria acuan untuk menyelidiki kesesuaian butir soal. Apabila suatu butir soal tidak sesuai dengan ketiga kriteria tersebut, maka dapat dianggap butir soal tersebut kurang layak, sehingga soal tersebut perlu direvisi atau digantikan dengan soal yang baru. Hal ini dilakukan untuk memastikan bahwa kemampuan siswa benar-benar teruji dengan butir soal yang berkualitas (Suryani & Jaedun, 2022).

Tabel 4 <Hasil *Output Item: fit order* (Outfit MNSQ, ZTSD, dan Pt. Measure Corr)>

No. Item	Outfit		Pt. Measure Corr	Keterangan
	MNSQ	ZSTD		
S1	0,47	-2,52	0,51	*Misfit*
S2	0,52	-2,06	0,55	*Misfit*
S3	1,48	0,97	-0,06	*Misfit*
S4	2,20	3,19	0,14	*Misfit*
S5	0,81	-0,60	0,54	*Fit*
S6	0,77	-0,98	0,52	*Fit*
S7	1,06	0,33	0,45	*Fit*
S8	0,71	-0,78	0,61	*Fit*
S9	0,47	-1,16	0,69	*Misfit*
S10	0,98	0,14	0,54	*Fit*
S11	1,35	0,90	0,42	*Fit*
S12	0,41	-0,61	0,44	*Misfit*

Merujuk pada Tabel 4. Nilai *Outfit* MNSQ yang *fit* sebanyak 8 butir soal, yaitu S11, S10, S3, S7, S5, S8, S6, dan S2, sedangkan nilai yang *misfit* sebanyak 4 butir soal, yaitu S4, S9, S12, dan S1. Adapun nilai *outfit* ZTSD yang *fit* diperoleh sebanyak 9 butir soal, yaitu S11, S10, S3, S7, S5, S9, S8, S6, dan S12, sedangkan nilai yang *misfit* diperoleh sebanyak 3 butir soal, yaitu S4, S2, dan S1. Nilai *outfit* Pt Measure Corr juga nilai *fit* dan *misfit.* Butir soal yang *fit* pada *outfit* Pt Measure Corr lebih banyak daripada *outfit* MNSQ dan *outfit* ZTSD, yaitu sebanyak 10 butir, yang terdiri dari S11, S10, S7, S5, S9, S8, S6, S12, S2, dan S1, sedangkan butir soal yang *misfit* hanya ada 2 butir soal, yaitu S4, dan S3. Berdasarkan perolehan nilai *fit* dan *misfit* butir soal pada masing-masing kriteria, ditemukan bahwa hanya sebagian soal yang *fit* dan *misfit* terhadap 3 kriteria *Outfit* MNSQ, ZTSD, dan Pt Measure Corr. yang didistribusikan ke dalam Tabel 5.

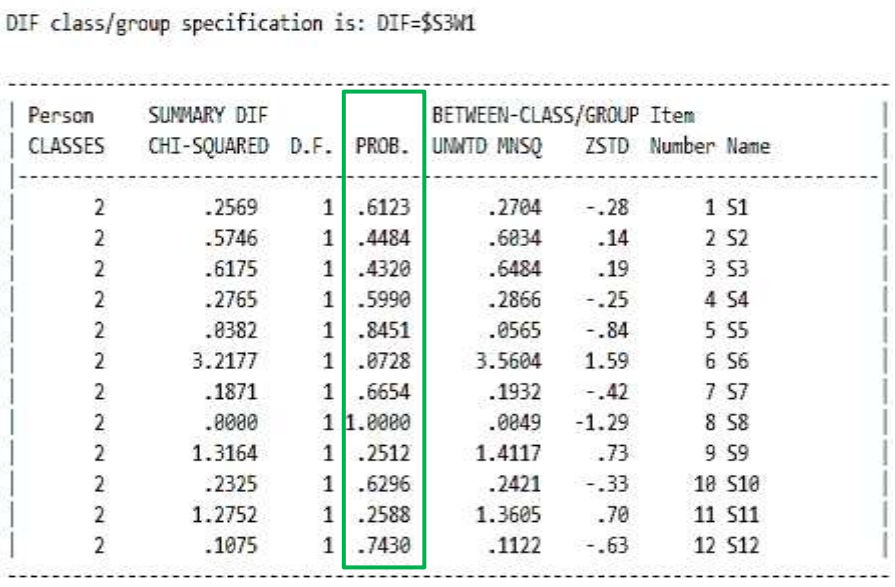
Tabel 5 <Distribusi Item *Fit* dan *Misfit*>

Kriteria	Nomor butir soal	Jumlah
Fit	S5, S6, S7, S8, S10, dan S11	6
Misfit	S1, S2, S3, S4, S9, dan S12	6

Merujuk pada Tabel 5, terlihat bahwa terdapat 6 butir soal yang *fit.* Butir soal ini dianggap *fit* karena memenuhi 3 kriteria *outfit* yang telah ditetapkan, yaitu 0,5 < MNSQ < 1,5, -2,0 < ZSTD < +2,0, dan 0,4 < PT Mean Corr < 0,85 (Sumintono & Widhiarso, 2015). Sedangkan, 6 soal lainnya dianggap *misfit* karena tidak memenuhi 3 kriteria tersebut. Hal ini mencerminkan pola yang tidak konsisten mengenai respon siswa terhadap butir soal dengan tingkat kemampuan siswa dalam berpikir kreatif. Butir soal S1, menuntut siswa untuk menjelaskan beberapa aspek dalam satu pertanyaan, seperti bentuk energi, contoh penerapannya dalam kehidupan sehari-hari dan sumber energi yang digunakan. Soal tersebut mencakup materi dasar, sehingga besar kemungkinan mayoritas siswa memahami soal tersebut dan mampu menjawab walaupun dengan beragam jawaban. Begitu juga dengan butir soal S2, yang menuntut siswa untuk menjelaskan cara memanfaatkan energi alternatif dalam kehidupan sehari-hari. Kemudian, butir Soal S3, S4, dan S12 merupakan soal perhitungan mengenai jumlah panel surya yang digunakan, konsumsi listrik yang dibutuhkan di bundaran besar, serta kebutuhan energi surya yang digunakan dalam sebulan. Soal-soal tersebut cenderung mengukur keterampilan numerasi daripada keterampilan berpikir kreatif siswa. Sehingga, besar kemungkinan siswa menganggap soal tersebut sulit diselesaikan. Hal ini sejalan dengan Boone (2016), yang menyatakan bahwa suatu item dikatakan *misfit* karena soal yang sulit justru dapat diselesaikan dengan benar oleh siswa berkemampuan rendah. Begitupun sebaliknya, suatu butir soal yang mudah justru dijawab dengan salah oleh siswa yang memiliki kemampuan tinggi.

***Differential Item Function* (DIF)**

Kevalidan suatu instrumen juga dapat dilihat dari ada tidaknya bias pada butir soal. Butir soal dapat dianggap valid jika butir soal tersebut tidak teridentifikasi bias (Ummah et al., 2022). Analisis Rasch Model juga dapat digunakan untuk mendeteksi bias pada butir soal yang disebut dengan uji *Differential Item Function* (DIF). Butir soal dianggap teridentifikasi bias apabila nilai probabilitas dari *output* DIF yang diperoleh < 0,05 (Sumintono & Widhiarso, 2015). Analisis DIF dilakukan berdasarkan jenis kelamin, apakah ada butir soal yang dominan terhadap siswa laki-laki daripada siswa perempuan, dan sebaliknya, berdasarkan tingkat kesulitan butir soal dan kemampuan rata-rata setiap kelompok jenis kelamin (Yim et al., 2024). Untuk melihat nilai probabilitas dapat diperoleh dari *output* tabel 30. DIF, *between/within* pada *Winsteps* yang tersaji pada Gambar 2.



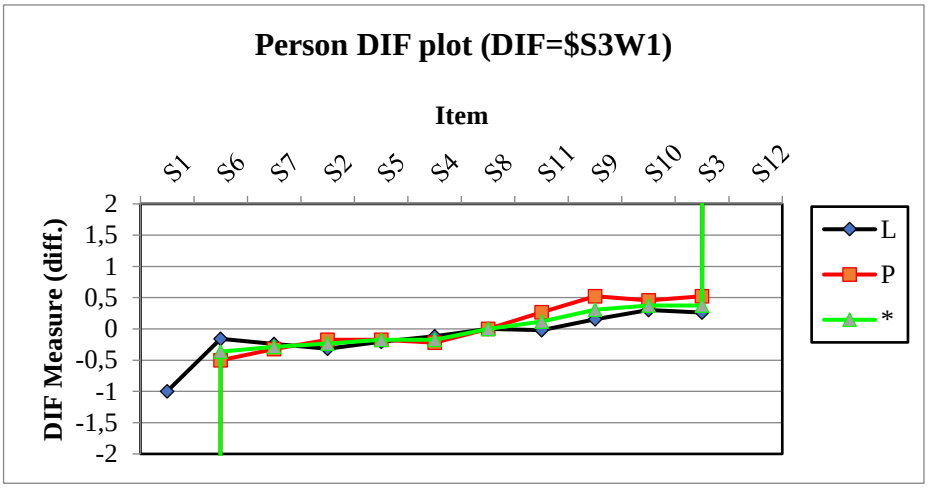
DIF class/group specification is: DIF=$S3W1

Person CLASSES	SUMMARY DIF CHI-SQUARED	D.F.	PROB.	BETWEEN-CLASS/GROUP Item UNWTD MNSQ	ZSTD	Number	Name
2	.2569	1	.6123	.2704	-.28	1	S1
2	.5746	1	.4484	.6834	.14	2	S2
2	.6175	1	.4320	.6484	.19	3	S3
2	.2765	1	.5990	.2866	-.25	4	S4
2	.0382	1	.8451	.0565	-.84	5	S5
2	3.2177	1	.0728	3.5604	1.59	6	S6
2	.1871	1	.6654	.1932	-.42	7	S7
2	.0000	1	1.0000	.0049	-1.29	8	S8
2	1.3164	1	.2512	1.4117	.73	9	S9
2	.2325	1	.6296	.2421	-.33	10	S10
2	1.2752	1	.2588	1.3605	.70	11	S11
2	.1075	1	.7430	.1122	-.63	12	S12

Gambar 2 <Hasil *Output* DIF *between/within*>

Merujuk pada nilai probabilitas yang tersaji dalam Gambar 2, terlihat bahwa tidak terdapat nilai yang berada dibawah 0,05, yang artinya tidak terdapat butir soal yang mempengaruhi perbedaan signifikan terhadap kelompok siswa laki-laki dan perempuan. Namun, jika melihat dari hasil grafik pada Gambar 3,

terdapat 2 soal yang memiliki perbedaan DIF yang cukup tinggi dibandingkan soal yang lainnya, yaitu S1 dan S12.



Gambar 3 <Grafik Hasil DIF *plot* Berdasarkan Jenis Kelamin>

Dari grafik terlihat bahwa garis hijau S1 dan S12 yang menunjukkan adanya perbedaan sangat signifikan terhadap kelompok siswa laki-laki dan perempuan, karena kedua soal tersebut masing-masing jauh dari -2 logit dan 2 logit dan merupakan soal yang mudah sekali dan sulit sekali. Sementara itu, 10 soal lainnya memiliki nilai DIF yang hampir sama antara kelompok siswa laki-laki dan perempuan, yang menyatakan bahwa soal tersebut tidak bias terhadap kelompok mana pun dan dianggap cukup adil untuk setiap kelompok. Hasil ini serupa dengan penelitian Susongko et al. (2021) yang juga memperoleh soal bias terhadap jenis kelamin. Soal tes yang digunakan adalah soal tes *Scientific Literacy Skills with Integrated Science* (SLiSIS) sebanyak 42 butir soal dan diperoleh sebanyak 8 butir soal terdeteksi bias.

Reliabilitas

Reliabilitas suatu tes merupakan sejauh mana suatu instrumen menghasilkan hasil yang konsisten dan stabil meskipun dilakukan pengulangan pada kondisi dan subjek yang sama (Gunawan et al., 2023; Muqaffi et al., 2024). Reliabilitas diukur dengan mengamati nilai *Cronbach's Alpha* dan nilai reliabilitas butir soal. Dalam Rasch Model, reliabilitas dinilai melalui dua indikator utama: *person reliability* dan *item reliability*, yang memberikan wawasan tentang reliabilitas pengukuran pada tingkat siswa dan item. Rasch Model juga menggunakan indeks pemisahan orang untuk mengelompokkan responden berdasarkan kemampuan mereka, yang memberikan dimensi tambahan dalam menilai konsistensi pengukuran. Dalam analisis ini, nilai reliabilitas Rasch yang lebih tinggi dibandingkan dengan CTT menunjukkan keunggulan dalam konsistensi pengukuran instrumen (Królikowska et al., 2023). Adapun untuk nilai reliabilitas terlihat dari hasil *Summary Statistics* yang disajikan pada Tabel 6.

Tabel 6 <Hasil *Summary Statistics*>

	Kriteria	Nilai
Reliabilitas	*Cronbach's Alpha*	0,62
	Item Reliability	0,90
	Person Reliability	0,49
Separation	*Item Separation*	2,97
	Person Separation	0,98

Merujuk pada hasil *Summary Statistics,* nilai *person reliability* dianggap lemah, karena berada pada rentang <0,67 dengan nilai sebesar 0,49 dan nilai *item reliability* berada pada rentang 0,81 – 0,90 dengan kriteria baik, yaitu sebesar 0,90. Sedangkan, nilai *Cronbach's Alpha* (KR-20) yang menyatakan hubungan antara siswa dan butir soal diperoleh sebesar 0,62 yang berada pada rentang 0,6 – 0,7 dengan kriteria cukup. Hasil ini menunjukkan bahwa instrumen yang digunakan kurang mampu membedakan kemampuan berpikir kreatif siswa secara efektif, sehingga hal tersebut mengindikasikan perlunya perbaikan pada butir soal agar dapat lebih akurat dalam mengukur keterampilan berpikir kreatif. Merujuk pada nilai *Cronbach's Alpha* dengan kriteria cukup menunjukkan bahwa ada korelasi yang cukup antara siswa dan butir soal. Hal ini Meskipun nilai ini menyatakan bahwa instrumen memiliki konsistensi yang dapat diterima, namun masih ada peluang untuk perbaikan agar memperoleh instrumen dengan tingkat reliabilitas yang tinggi. Hal ini merujuk pada

pernyataan Plummer & Ozcelik (2015), bahwa ambang batas nilai *Cronbach's Alpha* yang diterima hanya sebagai aturan praktis dan tidak menyiratkan bahwa nilai yang lebih rendah dianggap sebagai instrumen yang tidak memuaskan atau tidak dapat diterima. Dengan demikian, hasil perolehan nilai *Cronbach's Alpha* 0,62 pada penelitian ini masih dapat diterima sebagai hasil uji coba awal instrumen dan dapat ditingkatkan secara bertahap agar mencapai reliabilitas yang lebih tinggi (Taber, 2018; Tavakol & Dennick, 2011).

Hasil ini serupa dengan penelitian sebelumnya oleh Erfan et al. (2020), yang juga memperoleh nilai *person reliability* dengan kriteria lemah, yaitu sebesar 0,60, dan *Cronbach's Alpha* dengan kategori cukup, yaitu sebesar 0,64. Namun, nilai *item reliability* pada penelitian tersebut tidak serupa dengan hasil penelitian ini. *Item reliability* pada penelitian tersebut memiliki kriteria yang istimewa dengan nilai sebesar 0,97. Penelitian lainnya seperti Fadhilah et al. (2024), menghasilkan nilai *Cronbach's Alpha* dengan kategori cukup, yaitu sebesar 0,68, dan *person reliability* juga memiliki kriteria lemah dengan nilai sebesar 0,55. Namun, nilai *item reliability* penelitian tersebut tidak serupa, yaitu sebesar 0,99 dengan kriteria istimewa.

***Separation* (Indeks Pemisahan)**

Hasil dari *Summary Statistics* juga dapat diketahui daya pembeda antara siswa dan butir soal. Pada Rasch Model dapat diketahui dari nilai *separation* yang tertera pada Tabel 6 hasil *Summary Statistics.* Menurut Sumintono & Widhiarso (2014), nilai *separation* yang semakin tinggi, mengindikasikan suatu instrumen secara keseluruhan dianggap memiliki kualitas yang baik secara *person* dan *item*. Hal ini dapat dianggap mampu membedakan kelompok *person* dan *item.* Merujuk pada hasil analisis, diperoleh *person separation* senilai 0,98 serta *item separation* senilai 2,97. Agar hasil pengelompokkan perlu menghitung indeks strata (H) separation menggunakan persamaan 1.

$$H = \frac{[(4\ x\ SEPARATION) + 1\ ]}{3} \qquad (1)$$

Setelah dilakukan perhitungan, diperoleh nilai indeks strata *person separation* ($H_{person\ separation}$) sebesar 1,64 dibulatkan menjadi 2. Sementara itu, nilai indeks strata *item separation* ($H_{item\ separation}$) sebesar 4,29 dibulatkan menjadi 4. Merujuk pada hasil tersebut, mengindikasikan bahwa siswa yang terlibat dalam penelitian ini dapat diklasifikasikan menjadi 2 kelompok, yang artinya siswa memiliki kemampuan yang berbeda dan dapat diklasifikasikan berdasarkan tiap kemampuan siswa. Sedangkan, tingkat kesulitan butir soal dapat diklasifikasikan menjadi 4 kriteria mulai dari tingkat yang mudah hingga sulit. Hasil ini menunjukkan bahwa secara keseluruhan instrumen memiliki kualitas yang baik dari siswa maupun butir soal, karena mampu membedakan kelompok siswa dan butir soal (Sumintono & Widhiarso, 2014). Hasil ini sejalan dengan Bond & Fox (2015), bahwa nilai indeks *separation* yang di atas 2,0 mencerminkan bahwa instrumen yang dikembangkan efektif dalam mengidentifikasi keragaman kemampuan siswa dan tingkat kesukaran butir soal.

Hasil indeks strata (H) yang diperoleh dalam penelitian ini memiliki kesamaan dengan penelitian sebelumnya oleh Prayoga et al. (2024), yang memperoleh $H_{item\ separation}$ sebesar 3,70 yang dibulatkan menjadi 4. Namun, $H_{person\ separation}$ penelitian tersebut berbeda dengan penelitian ini, yaitu sebesar 2,50 yang dibulatkan menjadi 3, sehingga menunjukkan kriteria tingkat kesulitan butir soal penelitian tersebut lebih banyak daripada penelitian ini. Sementara itu, penelitian oleh Nur et al. (2022) ditemukan bahwa terdapat nilai indeks strata yang serupa, yaitu $H_{person\ separation}$ dengan nilai 1,5 yang dibulatkan menjadi 2 sehingga klasifikasi siswa terbagi menjadi 2. Sedangkan, tingkat kesulitan butir soal pada penelitian tersebut diklasifikasikan ke dalam 2 kriteria merujuk pada $H_{item\ separation}$ yang diperoleh sebesar 1,93 yang dibulatkan menjadi 2. Hal tersebut menunjukkan bahwa kriteria tingkat kesulitan penelitian ini lebih banyak daripada penelitian tersebut.

Tingkat Kesukaran Butir Soal

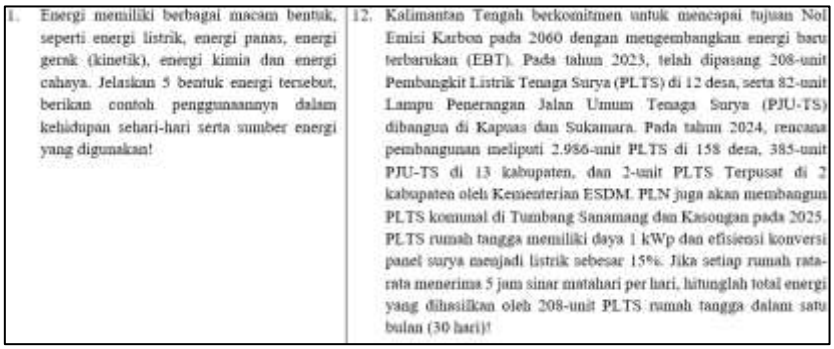
Untuk mengamati tingkat kesukaran tiap butir soal secara efektif dapat diamati dengan merujuk pada hasil *output* tabel 13. *Item measure* yang disajikan pada Gambar 4.

Merujuk pada Gambar 4, tertera nilai standar deviasi (SD) item sebesar 0,50. Dari nilai SD tersebut, dapat digunakan untuk menentukan kriteria tingkat kesukaran butir soal dengan merujuk pada Tabel 4. Kriteria dapat diklasifikasikan menjadi 4 tingkat, yaitu Sulit Sekali (*Measure logit* > +0,50), Sulit (0,00 ≤ *Measure logit* ≤ +0,50), Mudah (0,00 ≥ *Measure logit* ≥ –0,50), dan Mudah Sekali (*Measure logit* < – 0,50). Terlihat pada Gambar 6, bahwa butir soal S1 terbukti bahwa soal tersebut mudah sekali dengan nilai *Measure logit* -1,05. Sebaliknya, butir soal dengan nilai *Measure logit* tertinggi adalah S12 dengan nilai sebesar 1,11 tergolong ke dalam kriteria butir soal yang sulit sekali. Butir soal ini perlu dianalisis lebih lanjut faktor apa saja yang menyebabkan kedua soal tersebut merupakan soal yang mudah sekali dan sulit sekali. Gambar 5 menyajikan bentuk soal S1 dan S12 untuk memperjelas karakteristik butir soal dengan tingkat kesukaran yang ekstrem.

Item STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
12	5	34	1.11	.35	.58	-.31	.41	-.61	.44	.17	88.2	88.4	S12
3	20	34	.38	.16	1.25	.69	1.48	.97	-.06	.35	44.1	48.4	S3
10	20	34	.38	.16	1.60	1.32	.98	.14	.54	.35	52.9	48.4	S10
9	23	34	.31	.15	.84	-.29	.47	-1.16	.69	.37	47.1	42.1	S9
11	34	34	.12	.12	1.61	1.68	1.35	.90	.42	.42	20.6	27.0	S11
8	43	34	.00	.11	.83	-.54	.71	-.78	.61	.45	32.4	24.1	S8
4	60	34	-.17	.10	2.17	3.83	2.20	3.19	.14	.48	14.7	15.3	S4
5	60	34	-.17	.10	.86	-.54	.81	-.60	.54	.48	11.8	15.3	S5
2	67	34	-.24	.09	.57	-2.20	.52	-2.06	.55	.49	38.2	14.9	S2
7	71	34	-.29	.09	1.19	.91	1.06	.33	.45	.50	5.9	17.0	S7
6	83	34	-.36	.09	.82	-.84	.77	-.98	.52	.51	17.6	14.1	S6
1	175	34	-1.05	.09	.42	-2.90	.47	-2.52	.51	.48	32.4	20.2	S1
MEAN	55.3	34.0	.00	.13	1.06	.07	.94	-.26			33.8	31.3	
P.SD	43.1	.0	.50	.07	.50	1.71	.50	1.46			21.8	21.3	

Gambar 4 <*Item Measure*>



(a) **(b)**

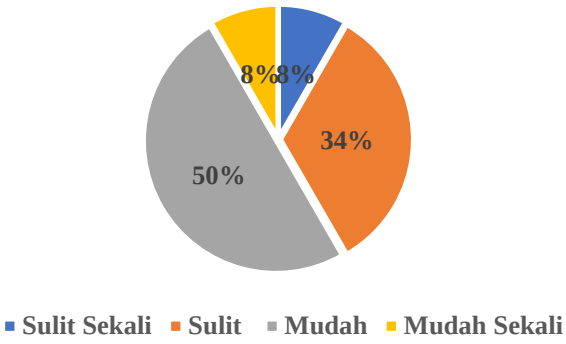
Gambar 5 <(a) Butir Soal S1: Mudah Sekali; (b) Butir Soal S12: Sulit Sekali>

Merujuk pada bentuk soal S1 dan S2 yang disajikan pada Gambar 5, terlihat bahwa pertanyaan yang digunakan pada soal S1 merupakan materi yang relatif mudah dipahami oleh sebagian besar siswa, karena berkaitan dengan sumber energi yang digunakan dalam kehidupan sehari-hari, sehingga sebagian besar siswa mampu menjawab soal tersebut. Karakteristik soal yang tidak asing memungkinkan siswa untuk lebih mudah memahami pertanyaan dan menyelesaikannya dengan tepat. Sementara itu, bentuk soal S12 terlihat memang cukup rumit, karena siswa dituntut untuk menyelesaikan perhitungan dari total energi yang dihasilkan dari PLTS rumah tangga dalam sebulan. Karakteristik soal ini memerlukan pemahaman konsep dan numerasi yang lebih kompleks. Hal ini memungkinkan sebagian besar siswa belum sepenuhnya menguasai materi tersebut, sehingga menyebabkan soal ini tergolong dalam kriteria sulit sekali. Untuk memperoleh gambaran secara menyeluruh mengenai sebaran tingkat kesukaran butir soal instrumen tes keterampilan berpikir kreatif pada materi energi terbarukan, berikut disajikan Tabel 7.

Tabel 7 <Sebaran Tingkat Kesukaran Butir Soal>

Kriteria	Rentang	Nomor butir soal	Nilai Logit	Jumlah
Sulit Sekali	*Measure logit* > +0,50	S12	1,11	1
Sulit	0,00 ≤ *Measure logit* ≤ +0,50	S3, S10, S9, dan S11	0,38, 0,38, 0,31, dan 0,12	4
Mudah	0,00 ≥ *Measure logit* ≥ –0,50	S6, S7, S2, S5, S8 dan S4	-0,36, -0,29, -0,24, -0,17, 0,00, dan -0,17	6
Mudah Sekali	*Measure logit* < – 0,50	S1	-1,05	1

Merujuk pada Tabel 7, menunjukkan bahwa sebagian besar butir soal tersebar secara merata dalam kriteria mudah dan sulit, sehingga instrumen ini menggambarkan tingkat kesukaran yang beragam dan mampu mengukur keterampilan berpikir kreatif siswa dari berbagai tingkat kemampuan dengan perolehan nilai *Measure logit* dalam rentang -1,05 hingga 1,11. Hasil ini sesuai dengan teori Hambleton et al. (1992), bahwa jika nilai *Measure logit* berada dalam rentang -2,0 hingga +2,0, maka tingkat kesukaran dianggap baik. Sebaran tingkat kesukaran butir soal instrumen tes keterampilan berpikir kreatif pada materi energi terbarukan divisualisasikan dalam bentuk diagram pie pada Gambar 6.

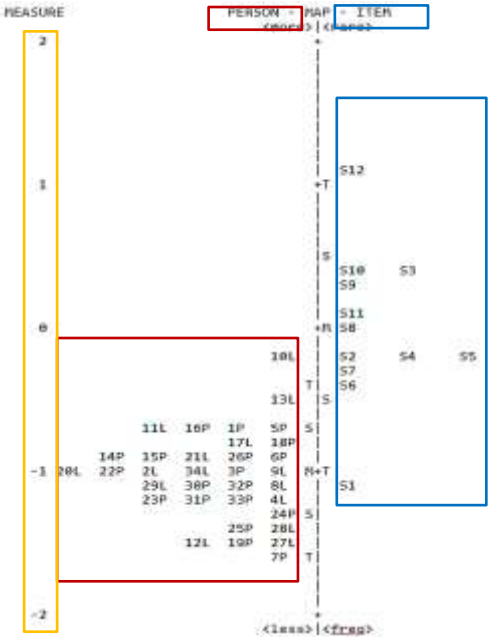


Gambar 6 <Sebaran Tingkat Kesukaran Butir Soal>

Merujuk pada Gambar 6, presentase tertinggi tingkat kesukaran butir soal berada pada kriteria mudah, yaitu sebesar 50%. Kemudian, urutan kedua diduduki oleh kriteria mudah sekali dengan presentase sebesar 34%. Adapun kriteria sulit sekali dan kriteria mudah sekali memiliki presentase yang sama, yaitu sebesar 8%. Merujuk pada penelitian Jumini et al. (2023), juga memperoleh sebaran tingkat kesukaran berdasarkan analisis Rasch Model, yaitu soal dengan kriteria sulit sekali sebanyak 2 soal, kriteria sulit sebanyak 3 soal, kriteria mudah sebanyak 7 soal, dan kriteria mudah sekali sebanyak 3 soal. Sebaran tingkat kesukaran yang merata ini memungkinkan instrumen mampu mengukur keterampilan berpikir kreatif siswa dengan lebih akurat (Sumintono & Widhiarso, 2014).

***Wright Maps* (Peta Variabel)**

Hasil analisis selanjutnya menunjukkan distribusi kemampuan siswa dan tingkat kesulitan butir soal instrumen tes keterampilan berpikir kreatif pada materi energi terbarukan dapat diamati pada *output* Tabel 1. *Variabel (Wright) maps* yang tersaji pada Gambar 7. Peta Variabel atau biasa disebut peta Wright menunjukkan sebaran siswa dan butir soal pada skala logit yang sama memberikan peluang untuk membantu peneliti dalam mengevaluasi kualitas butir soal terhadap kemampuan siswa (Saidi & Siew, 2019). Jika ditemukan kesenjangan besar antara butir soal, maka butir soal baru perlu ditambahkan dengan menyesuaikan tingkat kesukarannya. Jika ditemukan terlalu banyak butir soal dengan tingkat kesukaran yang sama, beberapa di antaranya lebih baik dihilangkan agar instrumen terjaga keseimbangannya (Planinic et al., 2019). Peta Wright terbagi menjadi dua area, yaitu kanan dan kiri (Sumintono & Widhiarso, 2014). Area kanan menunjukkan tingkat kesukaran butir soal, dan area kiri menunjukkan tingkat kemampuan siswa. Siswa berkemampuan tinggi akan ditempatkan di area kiri atas, dan siswa berkemampuan rendah akan ditempatkan di area kiri bawah. Soal dengan tingkat kesulitan tinggi akan ditempatkan di area kanan atas, dan soal dengan tingkat kesulitan rendah akan ditempatkan di area kanan bawah (Hikmah et al., 2021).



Gambar 7. Hasil *Variabel (Wright) maps*

Hasil Peta *Wright* pada Gambar 7 memberikan gambaran bahwa terdapat perbedaan tingkat kesulitan butir soal pada area sebelah kanan. Butir soal S12 berada jauh di atas melebihi nilai logit 1 yang terlihat pada *measure*. Sementara itu, butir soal S1 berada jauh di bawah melewati nilai logit -1. Hal itu mengindikasikan bahwa butir soal S12 dianggap sebagai butir soal yang sangat sulit oleh siswa karena tidak ada siswa yang berhasil mengerjakan butir soal tersebut, dan butir soal S1 dianggap mudah, karena mayoritas siswa berhasil mengerjakan butir soal tersebut. Jika dilihat dari tingkat kemampuan siswa yang berada di area kiri, sebagian besar berada pada logit 0 hingga -1, yang menggambarkan bahwa mayoritas siswa memiliki kemampuan tingkat rendah. Distribusi kemampuan siswa terhadap tingkat kesukaran soal penelitian ini hampir serupa dengan penelitian Handayani & Iba (2020), yang memperoleh distribusi kemampuan siswa berada pada logit 1 hingga -1 yang mencerminkan sebagian besar kemampuan siswa masih rendah karena sebagian besar berada pada logit 0 hingga -1.

Penelitian ini memiliki kelebihan dalam penerapan parameter uji Rasch Model, seperti *unidimensionality,* kesesuaian butir soal (*item fit*), *Differential Item Function* (DIF), reliabilitas, *separation* (indeks pemisahan), variabel (*wright*) *maps*, dan tingkat kesukaran tiap butir soal. Berbagai parameter uji yang digunakan mampu menggambarkan hasil analisis secara lebih objektif dan akurat. Jika dibandingkan dengan penelitian Yuliadinda et al. (2019), yang menganalisis validitas tes keterampilan berpikir kreatif pada materi tekanan zat dengan Rasch Model, penelitian ini lebih unggul karena menerapkan berbagai parameter uji yang mampu menghasilkan validitas dan reliabilitas secara menyeluruh, serta mampu menggambarkan kemampuan siswa dan tingkat kesukaran butir soal yang digunakan. Sementara itu, penelitian tersebut hanya menerapkan 1 parameter uji saja, yaitu *item fit* yang hanya menggambarkan kesesuaian butir soal tanpa menggambarkan reliabilitas secara keseluruhan, tingkat kemampuan siswa, serta tingkat kesukaran butir soal. Penelitian lainnya seperti Lidinillah et al. (2020), menunjukkan kesamaan dalam parameter uji yang diterapkan dalam analisis instrumen tes keterampilan berpikir kreatif, seperti *unidimensionality, item fit,* reliabilitas, variabel *Wright maps, Differential Item Function* (DIF), dan tingkat kemampuan siswa. Penelitian tersebut menghasilkan instrumen yang lebih tinggi (*Cronbach's Alpha* = 0,8) dibandingkan dengan penelitian ini (0,62), dan menggunakan sampel yang lebih luas, yakni 100 siswa dari 4 sekolah dasar.

Hasil analisis penelitian ini memiliki relevansi penting dalam dunia pendidikan, yaitu dalam pengembangan instrumen yang lebih objektif dan adil untuk menguji keterampilan berpikir kreatif. Selain memastikan tingkat kesulitan yang beragam, uji *Differential Item Function* (DIF) mengungkapkan kecenderungan untuk membiaskan beberapa item terhadap kelompok tertentu, baik laki-laki maupun perempuan. Oleh karena itu, penyesuaian diperlukan agar instrumen dapat digunakan secara lebih netral oleh semua siswa. Selain sebagai alat ukur, instrumen ini juga dapat berfungsi sebagai alat diagnostik dalam pembelajaran, membantu pendidik mengidentifikasi aspek-aspek keterampilan berpikir kreatif yang masih perlu ditingkatkan.

Namun, penelitian ini memiliki keterbatasan pada jumlah sampel yang hanya mencakup 1 kelas dengan jumlah 34 siswa. Hal ini memungkinkan hasil yang diperoleh kurang efektif jika diaktualisasikan ke populasi dengan ukuran yang lebih besar. Selain itu, penelitian ini dilakukan hanya di sekolah tertentu. Dalam merumuskan pertanyaan, perlu memperhatikan kemungkinan bias dengan adanya faktor non-kognitif, seperti gaya bahasa atau konteks pertanyaan yang dapat mempengaruhi respon siswa terhadap pertanyaan. Oleh sebab itu, penelitian lebih lanjut perlu direalisasikan untuk mengatasi keterbatasan ini.

Merujuk pada terbatasnya hasil penelitian ini, maka dihimbau untuk penelitian yang akan datang agar dapat mengembangkan serta menguji instrumen serupa pada sampel yang cukup luas dan lebih beragam agar instrumen dapat dipastikan kelayakannya dalam konteks pendidikan yang beragam pula. Kemudian, adanya bias butir soal terhadap gender agar lebih diperhatikan dan perlu diperbaiki agar dapat menunjukkan kesetaraan terhadap berbagai kelompok siswa. Selain itu, butir soal perlu dikembangkan secara inovatif, yang dapat mencakup pengetahuan berbasis teknologi dan proyek mengenai energi terbarukan, sehingga dapat menjadi alternatif dalam mengukur keterampilan berpikir kreatif siswa secara efektif. Analisis kualitatif melalui wawancara dengan siswa atau guru perlu dilakukan agar dapat memberikan wawasan lebih mendalam tentang faktor-faktor yang memengaruhi hasil data. Langkah-langkah ini memungkinkan penyempurnaan instrumen secara berkelanjutan untuk menilai dan meningkatkan keterampilan berpikir kreatif siswa secara lebih akurat di berbagai lingkungan belajar.

Simpulan

Hasil penelitian menggambarkan bahwa instrumen tes keterampilan berpikir kreatif pada materi energi terbarukan dinyatakan memiliki kualitas yang baik dilihat dari beberapa parameter uji Rasch Model. Hal ini dibuktikan dengan nilai *raw variance explained by measures* sebesar 46,9% dengan kriteria baik dan

unexplained variance in 1st contrast sebesar 11,8% yang mengindikasikan bahwa instrumen dinyatakan *unidimensionality* atau instrumen secara objektif mengukur keterampilan berpikir kreatif sesuai harapan. Butir soal yang digunakan sebagian besar memiliki kesesuaian dengan Rasch Model, yaitu sebanyak 6 soal dari 12 soal yang dianalisis, yaitu karena memenuhi 3 kriteria *outfit* (MNSQ, ZTSD, dan Pt Measure Corr) yang mengindikasikan tiap butir soal tersebut layak digunakan dalam mengukur keterampilan berpikir kreatif siswa. Instrumen keterampilan berpikir kreatif ini memperoleh nilai *Cronbach's Alpha* sebesar 0,62, *person reliability* 0,49, dan *item reliability* 0,90 yang menunjukkan konsistensi instrumen yang cukup baik dalam mengukur keterampilan berpikir kreatif siswa dengan kualitas butir soal yang baik, meskipun konsistensi kemampuan siswa masih dikatakan lemah. Selain itu, hasil indeks *separation* (H) menggambarkan bahwa instrumen mampu membedakan kemampuan siswa menjadi 2 kelompok, serta butir soal terbagi menjadi 4 tingkat kesukaran. Butir soal instrumen tidak terdapat bias yang signifikan setelah dilakukan uji *Differential Item Function* (DIF). Meskipun terdapat 2 soal dengan nilai logit ekstrem, yaitu S1 dan S12 yang masing-masing mewakili indikator berpikir kreatif kelancaran (*fluency*), serta elaborasi (*elaboration*), sehingga perlu dilakukan tinjauan ulang terhadap kedua butir soal tersebut. Berdasarkan hasil analisis, instrumen ini cukup valid dan reliabel untuk digunakan oleh pendidik sebagai alat ukur keterampilan berpikir kreatif siswa pada materi energi terbarukan dan diharapkan mampu memberikan gambaran mengenai keandalan Rasch Model dalam menganalisis instrumen tes. Rekomendasi untuk penelitian selanjutnya adalah mengembangkan lebih banyak butir soal yang mencakup indikator berpikir kreatif secara merata, melakukan uji coba instrumen pada sampel yang lebih luas dan beragam guna memperoleh hasil yang lebih akurat, serta mengombinasikan Rasch Model dengan pendekatan analisis lainnya, seperti analisis kualitatif terhadap respon siswa.

Referensi

Afandi, & Sajidan. (2017). *Stimulasi Keterampilan Berpikir Kreatif Tinggi. (Konsep dan Implementasinya dalam Pembelajaran Abad 21)* (2nd ed.). UNS Press.

Anindayati, A. T., & Wahyudi. (2020). Kajian Pendekatan Pembelajaran STEM dengan Model PjBL dalam Mengasah Kemampuan Berpikir Kreatif Matematis Siswa. *EKSAKTA : Jurnal Penelitian Dan Pembelajaran MIPA*, *5*(2), 217. https://doi.org/10.31604/eksakta.v5i2.217-225

Azwar, S. (2012). *Reliabilitas dan Validitas* (4th ed.). Pustaka Belajar.

Bond, T., & Fox, C. (2015). *Applying the Rasch Model; Fundamental Measurement in the Human Sciences*.

Boone, W. J. (2016). Rasch analysis for instrument development: Why,when,and how? *CBE Life Sciences Education*, *15*(4). https://doi.org/10.1187/cbe.16-04-0148

Brandt, W. C. (2024). *Measuring Student Success Skills : a Review of the Literature on Creative Thinking* (Issue May).

Erfan, M., Maulyda, M. A., Hidayati, V. R., Astria, F. P., & Ratu, T. (2020). Analisis Kualitas Soal Kemampuan Membedakan Ragkaian Seri dan Paralel Melalui Teori Klasik dan Model Rasch. *Indonesian Journal Of Educational Research and Review*, *3*(1), 11. https://doi.org/10.23887/ijerr.v3i1.24080

Ernawati. (2022). Analisis Soal Ujian Matematika Berkategori Higher Order Thinking Skills (Hots) Siswa SMP dengan AnalisisRasch Model. *Jurnal Pendidikan Matematika Universitas Lampung*, *10*(4), 375–391. https://doi.org/10.23960/mtk/v10i4.pp375-391

Fadhilah, A., Yuniarti, A., & Afandi. (2024). Kelayakan Instrumen Tes Berpikir Esensial Taksonomi Presseisen: Analisis Kualitas Tes dan Abilitas Peserta Didik Menggunakan Model Raschfile:///C:/Users/Asus/Downloads/citation-309190697.txt. *Diklabio: Jurnal Pendidikan Dan Pembelajaran Biologi*, *8*(2), 261–272. https://doi.org/10.33369/diklabio.8.2.261-272

Fitriyyah, A., Anwar, L., & Hafiizh, M. (2024). Assessment of Creative Thinking Skills through Imitation, Modification, and Construction of Creative Thinking Models. *PRISMA*, *13*(2), 175. https://doi.org/10.35194/jp.v13i2.4342

Gunawan, F. M., Suryana, D., & Rahmanadia, H. (2023). Analysis of Validity and Reliability of Self-Acceptance Scale Using Rasch Model. *Jurnal Pendidikan*, *24*(2), 71–80. https://doi.org/10.33830/jp.v24i2.4604.2023

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1992). Fundamentals of Item Response Theory. In *Contemporary Sociology* (Vol. 21, Issue 2). https://doi.org/10.2307/2075521

Handayani, S. L., & Iba, K. (2020). Karakteristik Tes Keterampilan Proses Sains: Validitas, Reliabilitas, Tingkat Kesukaran dan Daya Pembeda Soal. *Publikasi Pendidikan*, *10*(2), 100. https://doi.org/10.26858/publikan.v10i2.13051

Hasanah, D., & Purwanto, J. (2023). Rasch Model Analysis of Physics Test of HOTS on Topic of Elasticity and Hooke.s Law. *Thabiea*.

Hermawati, R., & Chen, D. (2023). Increasing Student Cooperation Through A Project-Based Learning Model with the Theme of Ecosystems. *Journal of Basic Education Research*, *4*(1), 45–49. https://doi.org/10.37251/jber.v4i1.299

Hikmah, F. N., Sukarelawan, M. I., Nurjannah, T., & Djumati, J. (2021). Elaboration of High School Student's Metacognition Awareness on Heat and Temperature Material: Wright Map in Rasch Model. *Indonesian Journal of Science and Mathematics Education*, *4*(2), 172–182. https://doi.org/10.24042/ijsme.v4i2.9488

Irmayanti, R., Rusdi, M., & Yusnaidar, Y. (2023). The Rasch Model: Implementation of Physics Learning Evaluation Instrument Based on Higher Order Thinking Skills. *Integrated Science Education Journal*, *4*(2), 62–68. https://doi.org/10.37251/isej.v4i2.325

Jumini, S., Madnasri, S., Cahyono, E., & Parmin, P. (2023). Analisis kualitas butir soal pengukuran literasi sains melalui teori tes klasik dan rasch model. *Prosiding Seminar Nasional Pascasarjana UNNES*, *6*, 758–765.

Królikowska, A., Reichert, P., Karlsson, J., Mouton, C., Becker, R., & Prill, R. (2023). Improving the Reliability of Measurements in Orthopaedics and Sports Medicine. *Knee Surgery, Sports Traumatology, Arthroscopy*, *31*(12), 5277–5285. https://doi.org/10.1007/s00167-023-07635-1

Lidinillah, D. A. M., Aprilia, M., Suryana, D., & Ahmad, A. B. (2020). Development of creativity instrument through Rasch model analysis. *Universal Journal of Educational Research*, *8*(4), 1620–1627. https://doi.org/10.13189/ujer.2020.080455

Linacre, J. M. (1994). Sample Size and Item Calibration or Person Measure Stability. *Rasch Measurement Transactions*, *7*(4), 328. http://www.rasch.org/rmt/rmt74m.htm

Mawaddah, D. A., Faradillah, A., & Alyani, F. (2024). Validity of the Rasch Model-Based Mathematical Critical Thinking Ability Instrument. *PATTIMURA PROCEEDING: Conference of Science and Technology*, *5*(1), 41–48. https://doi.org/10.30598/ppcst.2024.knmxxii.41-48

Muhammad, I., Angraini, L. M., Darmayanti, R., Sugianto, R., & Usmiyatun. (2023). Students' Interest in Learning Mathematics Using Augmented Reality: Rasch Model Analysis. *Edutechnium Journal of Educational Technology*, *1*(1), 89–99. https://www.edutechnium.com/journal

Muqaffi, Rusmana, N., & Ilfiandra. (2024). Emotional Regulation Scale: A Rasch Model Approach. *G-Couns: Jurnal Bimbingan Dan Konseling*, *9*(1), 451–459. https://doi.org/10.31316/gcouns.v9i1.6453

Nur, L., Yulianto, A., Suryana, D., Malik, A. A., Al Ardha, M. A., & Hong, F. (2022). An Analysis of the Distribution Map of Physical Education Learning Motivation through Rasch Modeling in Elementary School. *International Journal of Instruction*, *15*(2), 815–830. https://doi.org/10.29333/iji.2022.15244a

OECD. (2024). *PISA 2022 Results Creative Minds, Creative Schools*. OECD Publishing. https://doi.org/https://doi.org/10.1787/765ee8c2-en.

Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch Analysis in Physics Education Research: Why Measurement Matters. *Physical Review Physics Education Research*, *15*(2). https://doi.org/10.1103/PhysRevPhysEducRes.15.020111

Plummer, J. D., & Ozcelik, A. T. (2015). Preservice Teachers Developing Coherent Inquiry Investigations in Elementary Astronomy. *Science Education*, *99*(5), 932–957. https://doi.org/https://doi.org/10.1002/sce.21180

Prayoga, K. P., Suryana, D., Supriatna, M., & Budiman, N. (2024). Penggunaan Rasch Model Untuk Menganalisis Konstruk Instrumen Kontrol Diri Pada Siswa Sekolah Menengah. *G-Couns: Jurnal Bimbingan Dan Konseling*, *9*(1), 367–381. https://doi.org/10.31316/gcouns.v9i1.4459

Putri, S., & Khusna, H. (2020). Rasch Model untuk Memvalidasi Instrumen Resiliensi Matematis Mahasiswa Calon Guru Matematika. *JKPM (Jurnal Kajian Pendidikan Matematika)*, *1*(6), 65–74. http://journal.lppmunindra.ac.id/index.php/jkpm/

Rosha, J. M., & Hidayat, A. (2023). Analysis of Creative Thinking Skill Instrument Test (CreaTSIT) on renewable energy topic for senior high school student using Rasch model. *Journal of Physics: Conference Series*, *2596*(1), Rosha, J. M., Hidayat, A. (2023). Analysis of Cr. https://doi.org/10.1088/1742-6596/2596/1/012066

Saidi, S. S., & Siew, N. M. (2019). Reliability and Validity Analysis of Statistical Reasoning Test Survey Instrument using the Rasch Measurement Model. *International Electronic Journal of Mathematics Education*, *14*(3). https://doi.org/10.29333/iejme/5755

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi Model Rasch Untuk Penelitian Ilmu-ilmu Sosial*. Trim Komunitas Publishing House.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. Trim Komunikata. https://core.ac.uk/download/pdf/162015993.pdf

Suprapto, E., Saryanto, S., Sumiharsono, R., & Ramadhan, S. (2020). The Analysis of Instrument Quality to Measure the Students' Higher Order Thinking Skill in Physics Learning. *Journal of Turkish Science Education*, *17*(4), 520–527. https://doi.org/10.36681/tused.2020.42

Suryani, Y. E., & Jaedun, A. (2022). An Application of the Rasch Model in Quality Testing of Polytomic Instruments. *Proceedings of the 5th International Conference on Current Issues in Education (ICCIE 2021)*. https://doi.org/10.2991/assehr.k.220129.006

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

Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, *48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Tanujaya, L., & Yudiarso, A. (2023). Uji Validitas dan Reliabilitas Multicultural Attitude Scale Questionnaire (MASQUE) Versi Indonesia dengan Model Rasch. *Humanitas*, *7*(2), 139–150.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Thiagarajan, & Sivasailan. (1976). Instructional development for training teachers of exceptional children: A sourcebook. *Journal of School Psychology*, *14*(1), 75. https://doi.org/10.1016/0022-4405(76)90066-2

Torrance, E. P. (1966). *Torrance tests of creative thinking : norms-technical manual*. Scholastic Testing Services.

Ummah, K., Mardhiya, J., & Mulyanti, S. (2022). Pengembangan Instrumen Tes Penguasaan Konsep Representasi Kimia Pada Lima Indikator Asam Basa Dari Alam : Analisis Dengan Rasch Model. *Jurnal Tarbiyah*, *29*(2), 212. https://doi.org/10.30829/tar.v29i2.1706

Wulandari, M., & Jumadi, J. (2023). Analisis Penggunaan E-Modul Untuk Mendukung Kemampuan Pembelajaran Fisika Siswa Di Sma: Tinjauan Pustaka Sistematik. *EduFisika: Jurnal Pendidikan Fisika*, *8*(1), 37–45. https://doi.org/10.59052/edufisika.v8i1.23317

Yim, L. W. K., Lye, C. Y., & Koh, P. W. (2024). A psychometric evaluation of an item bank for an English reading comprehension tool using Rasch analysis. *REID (Research and Evaluation in Education)*, *10*(1), 18–34. https://doi.org/10.21831/reid.v10i1.65284

Yuliadinda, N., Muslim, M., & Efendi, R. (2019). Validitas Tes Keterampilan Berpikir Kreatif Materi Tekanan Zat Dengan Analisis Rasch Model. *Konstan - Jurnal Fisika Dan Pendidikan Fisika*, *4*(2), 89–97. https://doi.org/10.20414/konstan.v4i2.40