



Kelayakan Instrumen Tes Berpikir Esensial Taksonomi Presseisen: Analisis Kualitas Tes dan Abilitas Peserta Didik Menggunakan Model Rasch

Azfa Fadhilah¹, Anisyah Yuniarti^{1*}, Afandi¹

¹ Program Studi S-1 Pendidikan Biologi, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Tanjungpura, Indonesia

*Email: anisyah.yuniarti@fkip.untan.ac.id

Info Artikel	Abstrak
<p>Diterima: 24 Juli 2024 Direvisi: 08 Agustus 2024 Diterima untuk diterbitkan: 30 November 2024</p> <p>Keywords: Analisis Tes, Berpikir Esensial, Model Rasch, Taksonomi Presseisen.</p>	<p>Instrumen tes yang dapat digunakan untuk mengukur keterampilan berpikir esensial pada pembelajaran biologi masih belum tersedia. Penelitian ini bertujuan untuk mengetahui kualitas instrumen tes berpikir esensial menggunakan taksonomi Presseisen dan tingkat abilitas peserta didik melalui analisis model Rasch. Penelitian ini menggunakan metode deskriptif dengan pendekatan kuantitatif. Sampel yang digunakan berjumlah 363 orang peserta didik. Instrumen yang digunakan adalah instrumen tes berpikir esensial berjumlah 10 butir soal berbentuk uraian. Pengumpulan data dilakukan dengan uji coba instrumen tes. Analisis data dilakukan dengan analisis model Rasch. Hasil penelitian menunjukkan seluruh butir soal <i>fit</i>, namun butir soal nomor 7 mengandung bias sehingga perlu dibuang. Tingkat kesukaran butir terdiri dari 2 soal sangat sulit, 4 soal sulit, 2 soal mudah, dan 2 soal sangat mudah. Nilai <i>Alpha Cronbach</i> sebesar 0.68 berkategori cukup, nilai <i>person reliability</i> sebesar 0.50 berkategori lemah, dan nilai <i>item reliability</i> sebesar 0.99 berkategori istimewa. <i>Person separation</i> sebesar 0.99 dengan nilai $H=1,65$ dan <i>item separation</i> sebesar 9.55 dengan nilai $H=13,06$. <i>Person measure</i> diperoleh 178 orang memiliki abilitas tinggi (49%), 18 orang memiliki abilitas sedang (5%), dan 167 orang memiliki abilitas rendah (46%). <i>Person fit</i> diperoleh 254 orang kategori <i>fit</i> (70%), 32 orang kategori tidak <i>fit</i> (9%), dan 77 orang kategori <i>outlier</i> (21%).</p>

© 2024 Azfa Fadhilah. This is an open-access article under the CC-BY license (<https://creativecommons.org/licenses/by/4.0>)

PENDAHULUAN

Berpikir esensial merupakan suatu keterampilan untuk memahami dan menganalisis inti atau esensi dari suatu permasalahan melalui analisis secara mendalam dan menyeluruh agar dapat terselesaikan dengan baik. (Chusni, Saputro, Suranto, & Rahardjo, 2020) menyatakan bahwa keterampilan berpikir esensial merupakan keterampilan berpikir yang harus dikuasai peserta didik



karena indikator dari berpikir esensial termasuk dalam keterampilan dasar yang diperlukan dalam proses berpikir sebelum mencapai tingkatan berpikir yang selanjutnya. Pernyataan tersebut juga diperkuat oleh (Presseisen, 1985) yang menyatakan bahwa keterampilan berpikir esensial merupakan dasar dalam proses berpikir yang diperlukan sebelum menuju tingkatan proses berpikir yang lebih tinggi atau lebih kompleks.

Instrumen tes yang sering dikembangkan untuk mengukur keterampilan berpikir peserta didik biasanya untuk mengukur keterampilan berpikir tingkat tinggi (*Higher Order Thinking Skills*), keterampilan pemecahan masalah, keterampilan berpikir kreatif, dan keterampilan berpikir kritis. Namun, pada penelitian ini dikembangkan instrumen tes untuk mengukur keterampilan berpikir esensial peserta didik yang masih jarang dilakukan. Instrumen tes berpikir esensial yang dikembangkan mengacu pada indikator berpikir esensial menurut taksonomi Presseisen.

Taksonomi Presseisen merupakan salah satu taksonomi yang dapat digunakan untuk mengidentifikasi keterampilan berpikir peserta didik dari level kognitif rendah ke tinggi, yaitu dengan menggunakan indikator berpikir esensial. Berpikir esensial menurut taksonomi Presseisen merupakan proses berpikir yang mencakup lima indikator, yaitu *qualification* (mengidentifikasi, mendefinisikan, mengumpulkan fakta, dan menggambarkan masalah), *classification* (membuat pengelompokan, mencari persamaan dan perbedaan, membandingkan), *relationships* (menemukan keterkaitan atau hubungan, menganalisis), *transformations* (transformasi atau mengubah, membuat analogi), dan *causation* (menentukan hubungan sebab akibat, memprediksi) (Presseisen, 1985).

Instrumen tes yang baik adalah instrumen tes yang sesuai dengan standar, yaitu instrumen yang telah teruji validitas dan reliabilitasnya (Camila, Afandi, Tenriawaru, Artika, & Siregar, 2023; Tarigan, Nilmarito, Islamiyah, Darmana, & Suyanti, 2022). Kriteria instrumen tes yang baik, meliputi validitas butir soal yang valid, reliabilitas yang baik, tingkat kesukaran butir soal yang beragam atau bervariasi, dan memiliki daya pembeda yang dapat membedakan abilitas peserta didik (Erfan, Maulyda, Hidayati, Astria, & Ratu, 2020). Semakin tinggi nilai validitas dan reliabilitas dari suatu instrumen tes, maka semakin akurat data yang diperoleh pada hasil penilaian yang dilakukan.

Kualitas dari suatu instrumen tes dapat diketahui dari analisis hasil uji coba tes yang dilakukan. Analisis butir soal tes dapat dilakukan dengan dua pendekatan, yaitu dengan pendekatan teori tes klasik (*Classical Test Theory/CTT*) dan pendekatan teori respons butir atau teori tes modern (*Item Response Theory/IRT*) (Erfan *et al.*, 2020, 2020; Jumini, Modnasri, Cahyono, & Parmin, 2023; Pratama, 2020). Instrumen tes yang dikembangkan pada penelitian ini dianalisis dengan menggunakan teori respons butir, yaitu dengan analisis model Rasch. Model Rasch merupakan salah satu metode analisis yang digunakan untuk mengetahui kelayakan dari suatu instrumen dengan bantuan aplikasi *Winsteps*.

Teori respons butir merupakan teori tes modern yang digunakan untuk menganalisis hasil tes yang dirancang oleh para ahli untuk memperbaiki kelemahan-kelemahan yang ada pada teori tes klasik (Mulyani, Efendi, & Ramalis, 2021). Analisis dengan teori respons butir dapat menghasilkan pengukuran yang lebih akurat dari teori tes klasik (Falani, 2023). Hal ini diperkuat dengan penelitian yang dilakukan oleh Erfan *et al.*, (2020) yang menunjukkan bahwa tingkat akurasi analisis item soal dengan pendekatan IRT lebih tinggi daripada analisis item soal dengan pendekatan CTT.

Model Rasch memiliki kelebihan dapat menguraikan keterkaitan antara subjek dan item tes yang membuat hasil pengukuran menjadi tepat dan lebih objektif (Sumintono & Widhiarso, 2015). Kelebihan dari analisis model Rasch adalah mampu melakukan identifikasi jawaban yang salah, mengidentifikasi penilaian yang tidak tepat, dan dapat memprediksi data yang hilang berdasarkan pola respons yang sistematis (Hamdu *et al.*, 2020). Model Rasch juga tidak bergantung pada sampel yang digunakan, dapat mengurutkan secara terstruktur dari soal yang tersulit sampai dengan soal yang termudah dan dapat mengurutkan peserta tes yang abilitasnya tinggi hingga rendah (Untary *et al.*, 2020).

Pada penelitian ini analisis model Rasch digunakan untuk menganalisis kualitas tes dan menganalisis abilitas (kemampuan) peserta didik. Analisis kualitas tes dengan model Rasch dilihat

dari hasil analisis tingkat kesesuaian butir soal (*item fit*), tingkat kesukaran butir soal (*item measure*), reliabilitas, dan daya pembeda (*separation*). Analisis tingkat abilitas atau kemampuan peserta didik dengan model Rasch dilihat dari hasil analisis tingkat abilitas individu (*person measure*) dan tingkat kesesuaian individu (*person fit*).

Berdasarkan permasalahan yang telah dijabarkan, peneliti tertarik untuk melakukan analisis kelayakan instrumen tes berpikir esensial taksonomi Presseisen pada materi ekosistem menggunakan analisis model Rasch. Penelitian ini bertujuan untuk mengetahui kualitas instrumen tes berpikir esensial dan tingkat abilitas peserta didik melalui analisis model Rasch. Analisis instrumen tes ini diharapkan dapat memberikan gambaran terkait kualitas soal tes yang digunakan dan abilitas peserta didik sehingga guru dapat mempertahankan dan meningkatkan kualitas instrumen tes serta merancang proses pembelajaran yang lebih baik. Selain itu, analisis dengan model Rasch diharapkan dapat menjadi alternatif bagi guru dalam melakukan analisis butir soal.

METODE

Metode penelitian yang digunakan pada penelitian ini adalah metode deskriptif dengan pendekatan kuantitatif. Penelitian ini berfokus pada analisis kualitas instrumen tes berpikir esensial dan analisis tingkat abilitas peserta didik dengan menggunakan analisis model Rasch. Populasi yang digunakan pada penelitian ini adalah seluruh peserta didik kelas X Sekolah Menengah Atas Negeri Pontianak. Teknik pengambilan sampel menggunakan teknik *cluster sampling*. Sampel yang digunakan berjumlah 363 orang peserta didik yang dihitung dengan menggunakan rumus Slovin yang terbagi ke dalam lima sekolah di setiap kecamatan yang ada di Kota Pontianak, yaitu SMA Negeri 3 Pontianak, SMA Negeri 5 Pontianak, SMA Negeri 8 Pontianak, SMA Negeri 9 Pontianak, dan SMA Negeri 11 Pontianak di mana masing-masing sekolah terdiri dari 72-73 orang peserta didik. Instrumen yang digunakan adalah instrumen tes yang dikembangkan menggunakan metode *Research and Development* (R&D) dengan model pengembangan 4D yang dimodifikasi menjadi 3D yang terdiri dari tahap *define*, tahap *design*, dan tahap *develop* yang dikombinasi dengan tahapan pengembangan tes dari Mardapi (2008) yang terdiri dari 9 tahapan dimodifikasi menjadi 6 tahapan, yaitu menyusun spesifikasi tes, menulis soal tes, menelaah soal tes, melakukan uji coba soal tes, menganalisis soal tes, dan memperbaiki tes. Hal ini disebabkan karena keterbatasan waktu, tenaga, dan biaya yang dimiliki serta fokus penelitian yang bertujuan untuk menghasilkan produk berupa instrumen tes yang memenuhi standar. Instrumen tes disusun dengan indikator berpikir esensial menurut taksonomi Presseisen yang terdiri dari 10 butir soal berbentuk tes uraian. Instrumen tes divalidasi oleh 7 orang ahli yang terdiri dari 2 orang Dosen Pendidikan Biologi FKIP UNTAN dan 5 orang guru Biologi SMA. Hasil analisis dari validitas isi menggunakan perhitungan rumus *Aiken's V* dan reliabilitas *interrater* dengan uji *Intraclass Correlation Coefficient* menggunakan *Statistical Package for the Social Sciences* versi 27 menunjukkan bahwa instrumen tes valid dan reliabel untuk dilakukan uji coba soal tes. Pengumpulan data dilakukan dengan uji coba instrumen tes berpikir esensial pada sampel yang telah ditentukan. Data yang diperoleh merupakan data mentah yang akan dianalisis dengan model Rasch. Analisis data kuantitatif berupa data mentah hasil uji coba soal tes dilakukan dengan pendekatan IRT model Rasch dengan bantuan program *Winsteps* versi 3.73. Analisis model Rasch dilakukan untuk mengetahui tingkat kesesuaian butir soal (*item fit*), reliabilitas, tingkat kesukaran butir soal (*item measure*), daya pembeda (*separation*), tingkat abilitas peserta didik (*person measure*), dan tingkat kesesuaian peserta didik (*person fit*).

HASIL DAN PEMBAHASAN

Instrumen tes berpikir esensial yang dianalisis berjumlah 10 butir soal berbentuk tes uraian. Jawaban yang diperoleh dari uji coba instrumen tes pada 363 sampel dikoreksi menggunakan rubrik penskoran soal tes yang telah dibuat untuk memperoleh data mentah. Data mentah selanjutnya dianalisis dengan model Rasch berbantuan *software Winsteps* versi 3.73. Analisis yang diperoleh pada model Rasch, yaitu tingkat kesesuaian butir soal (*item fit*), reliabilitas, tingkat kesukaran butir

soal (*item measure*), daya pembeda (*separation*), tingkat abilitas peserta didik (*person measure*), dan tingkat kesesuaian peserta didik (*person fit*) yang dijelaskan sebagai berikut.

1. Analisis Item Fit

Analisis *item fit* pada pemodelan Rasch menjelaskan apakah butir soal tes berfungsi secara normal atau tidak untuk melakukan suatu pengukuran (Kirom & Hasyim, 2021; Ridwan *et al.*, 2023; Rusilowati, 2018), artinya instrumen tes tersebut dapat mengukur variabel yang diteliti secara tepat. Apabila pada hasil analisis terdapat butir soal yang tidak *fit*, hal ini mengindikasikan bahwa terdapat miskonsepsi pada peserta didik terhadap butir soal tersebut (Sumintono & Widhiarso, 2015).

Kriteria yang digunakan untuk melihat diterima atau tidak suatu butir soal atau tingkat kesesuaian butir soal (*item fit*) dalam pemodelan Rasch dilihat dari nilai yang dihasilkan pada *outfit mean-square*, *outfit z-standard*, dan *point measure correlation* sebagai berikut:

- Nilai *Outfit mean square* (MNSQ) yang diterima: $0.5 < MNSQ < 1.5$.
- Nilai *Outfit Z-standard* (ZSTD) yang diterima: $-2.0 < ZSTD < +2.0$.
- Nilai *Point Measure Correlation* (*Pt Measure Corr*) yang diterima: $0.4 < Pt Measure Corr < 0.85$.

Butir soal dapat dikategorikan *fit* apabila memenuhi satu dari tiga kriteria yang digunakan. Namun, apabila terdapat butir soal yang tidak memenuhi ketiga kriteria, maka butir soal memiliki kualitas yang kurang baik sehingga perlu direvisi, diganti, atau dibuang (Sumintono & Widhiarso, 2015). Hasil analisis *item fit* dapat dilihat pada menu *Output Tables* pada *software Winsteps*, kemudian pilih Tabel 10. *Item Fit Order* yang dapat dilihat pada Gambar 1.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	TOTAL MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	EXACT EXP.	MATCH OBS%	MATCH EXP%	Item
2	1214	352	-.89	.08	1.39	4.4	1.38	3.6	A .47	.41	54.0	55.2	E2
1	740	330	1.01	.07	1.12	1.7	1.10	1.4	B .43	.47	46.4	44.1	E1
4	960	339	.15	.07	1.07	1.1	1.06	.9	C .42	.48	38.6	39.7	E4
5	922	339	.31	.06	1.06	1.0	1.05	.7	D .61	.49	31.9	39.3	E5
7	862	338	.58	.06	1.03	.5	1.04	.6	E .40	.48	40.8	41.2	E7
8	757	332	.95	.07	1.00	.1	1.02	.3	e .44	.47	41.9	43.9	E8
6	1219	355	-.86	.08	.92	-1.0	1.00	.1	d .39	.42	54.9	53.6	E6
10	1139	341	-.68	.07	.91	-1.2	.95	-.6	c .43	.44	53.1	50.1	E10
3	982	343	.10	.07	.79	-3.5	.78	-3.5	b .46	.48	48.7	39.0	E3
9	1175	351	-.68	.07	.77	-3.3	.76	-3.0	a .52	.44	51.9	50.1	E9
MEAN	997.0	342.0	.00	.07	1.01	.0	1.01	.1			46.2	45.6	
S.D.	172.4	7.9	.70	.01	.17	2.2	.16	2.0			7.3	5.8	

Gambar 1. Output Item Fit.

Berdasarkan Gambar 1, butir soal tes berpikir esensial yang dianalisis diterima secara keseluruhan, namun terdapat beberapa butir soal yang tidak memenuhi (*misfit*) salah satu dari ketiga kriteria. Butir soal dengan kode *item* E2, E3, dan E9 termasuk dalam kategori *misfit* karena memiliki nilai *outfit ZSTD* di luar kriteria yang sesuai ($-2.0 < ZSTD < +2.0$). Butir soal dengan kode *item* E6 dan E7 termasuk dalam kategori belum sesuai (*misfit*) karena memiliki nilai *outfit Pt Measure Corr* di luar kriteria yang sesuai ($0.4 < Pt Measure Corr < 0.85$). Butir-butir soal tersebut tetap dipertahankan karena hanya tidak memenuhi kriteria (*misfit*) pada salah satu kriteria.

Kevalidan suatu instrumen juga dapat dilihat dari ada tidaknya bias pada butir soal. Butir soal dapat dikatakan valid apabila butir soal tersebut tidak mengandung bias (Ummah, Mardhiya, & Mulyanti, 2022). Analisis model Rasch juga dapat digunakan untuk mendeteksi bias pada butir soal yang disebut dengan deteksi DIF (*Differential Item Functioning*). Butir soal dikatakan mengandung bias apabila nilai probabilitas dari *output* DIF yang diperoleh $< 5\%$ (< 0.05) (Sumintono & Widhiarso, 2015). Analisis bias pada instrumen tes berpikir esensial pada

penelitian ini dilakukan terhadap karakteristik jenis kelamin. Nilai probabilitas dapat pada menu *Output Tables* pada *software Winsteps*, kemudian pilih Tabel 30. *Item DIF* yang dapat dilihat pada Gambar 2.

30-377WS - Notepad
File Edit Format View Help
▲TABLE 30.4 C:\Users\User\Desktop\POLITOMI Data M ZOU377WS.TXTc Jun 27 18:02 2024
INPUT: 363 Person 10 Item REPORTED: 363 Person 10 Item 4 CATS WINSTEPS 3.73

DIF class specification is: DIF=\$S4W1

Person CLASSES	SUMMARY DIF CHI-SQUARE	D.F.	PROB.	BETWEEN-CLASS MEAN-SQUARE	t=ZSTD	Item Number Name
2	3.5460	1	.0597	1.7292	.8963	1 E1
2	.0000	1	1.0000	.0050	-1.2880	2 E2
2	.0000	1	1.0000	.0001	-1.5388	3 E3
2	.0000	1	1.0000	.0014	-1.4114	4 E4
2	.0000	1	1.0000	.0035	-1.3278	5 E5
2	1.1396	1	.2857	.5683	.1072	6 E6
2	3.8987	1	.0483	1.9292	.9909	7 E7
2	.0000	1	1.0000	.0066	-1.2521	8 E8
2	.3285	1	.5666	.1605	-.4970	9 E9
2	.0000	1	1.0000	.0023	-1.3680	10 E10

Gambar 2. Output Item DIF.

Berdasarkan Gambar 2, butir soal yang terdeteksi mengandung bias terdapat pada butir soal nomor 7 yang memiliki nilai probabilitas $0.0483 < 0.05$ sehingga butir soal nomor 7 dinyatakan tidak valid dan harus dibuang.

2. Analisis Item Measure

Item measure menunjukkan tingkat kesulitan butir soal. Analisis tingkat kesulitan butir soal dilakukan untuk mengetahui pengelompokkan butir soal mana yang dikategorikan sangat sukar, sukar, mudah, dan sangat mudah. Tingkat kesukaran butir soal dapat dilihat dari kemampuan peserta didik dalam menjawab soal (Imania & Bariah, 2019). Pengelompokkan tingkat kesukaran butir soal didasarkan pada nilai standar deviasi yang diperoleh dari *output item measure*. Hasil analisis *item measure* dapat dilihat pada menu *Output Tables* pada *software Winsteps*, kemudian pilih Tabel 13. *Item Measure* yang dapat dilihat pada Gambar 3.

13-254WS - Notepad
File Edit Format View Help
TABLE 13.1 C:\Users\User\Desktop\POLITOMI Data M ZOU254WS.TXTc Jul 16 10:50 2024
INPUT: 363 Person 10 Item REPORTED: 363 Person 10 Item 4 CATS WINSTEPS 3.73

Person: REAL SEP.: .99 REL.: .50 ... Item: REAL SEP.: 9.55 REL.: .99

Item STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PT-MEASURE CORR.	EXP.	EXACT MATCH OBS%	EXP%	Item
1	740	330	1.01	.07	1.12	1.7	1.10	1.4	.43	.47	46.4	44.1	E1
8	757	332	.95	.07	1.00	.1	1.02	.3	.44	.47	41.9	43.9	E8
7	862	338	.58	.06	1.03	.5	1.04	.6	.40	.48	40.8	41.2	E7
5	922	339	.31	.06	1.06	1.0	1.05	.7	.61	.49	31.9	39.3	E5
4	960	339	.15	.07	1.07	1.1	1.06	.9	.42	.48	38.6	39.7	E4
3	982	343	.10	.07	.79	-3.5	.78	-3.5	.46	.48	48.7	39.0	E3
10	1139	341	-.68	.07	.91	-1.2	.95	-.6	.43	.44	53.1	50.1	E10
9	1175	351	-.68	.07	.77	-3.3	.76	-3.0	.52	.44	51.9	50.1	E9
6	1219	355	-.86	.08	.92	-1.0	1.00	.1	.39	.42	54.9	53.6	E6
2	1214	352	-.89	.08	1.39	4.4	1.38	3.6	.47	.41	54.0	55.2	E2
MEAN	997.0	342.0	.00	.07	1.01	.0	1.01	.1			46.2	45.6	
S.D.	172.4	7.9	.70	.01	.17	2.2	.16	2.0			7.3	5.8	

Gambar 3. Output Item Measure.

Berdasarkan Gambar 3, hasil analisis diperoleh nilai standar deviasi sebesar 0.70 sehingga pengelompokkan tingkat kesukaran butir soal dapat dikategorikan sebagai berikut: $logit > 0.70$ berkategori sangat sukar; $0.0 < logit < (+0.70)$ berkategori sukar; $(-0.70) < logit <$

0.0 berkategori mudah; dan $logit < (-0.70)$ berkategori sangat mudah. Berdasarkan pengelompokan tingkat kesukaran dan Gambar 3 dari hasil *logit measure* maka butir soal instrumen tes berpikir esensial dapat dikelompokkan seperti pada Tabel 1.

Tabel 1.

Tingkat Kesukaran Butir Soal.

Kategori	Butir Soal	Jumlah	Persentase (%)
Sangat sukar	1 & 8	2	20%
Sukar	3, 4, 5, & 7	4	40%
Mudah	9 & 10	2	20%
Sangat mudah	2 & 6	2	20%
Total		10	100%

(Sumber: Data peneliti)

Hasil analisis menunjukkan tingkat kesukaran butir soal memiliki sebaran yang tidak konsisten. Jika dikaitkan dengan indikator berpikir esensial maka tingkat kesukaran secara berurutan dari mudah hingga sulit, yaitu *qualification*, *classification*, *relationship*, *transformation*, dan *causation*. Data yang digunakan untuk menganalisis tingkat kesukaran butir soal adalah jumlah skor yang diperoleh peserta didik setelah butir soal diujicobakan (data mentah) sehingga tingkat kesukaran butir soal bergantung pada jawaban yang diberikan oleh peserta didik. Hal ini sejalan dengan penelitian yang dilakukan oleh Feranda, Ningsih, & Afandi (2020) bahwa sebaran tingkat kesukaran butir soal pada hasil analisis bergantung pada jawaban peserta didik. Hasil penelitian oleh Rahmat, Hamdu, & Nur'aeni (2020) juga menyatakan bahwa meskipun butir soal memiliki level indikator yang sama, setiap butir soal akan menghasilkan hasil analisis tingkat kesukaran yang berbeda-beda karena dipengaruhi oleh jawaban peserta didik.

3. Ringkasan Summary Statistics

Hasil analisis dari *output summary statistics* digunakan untuk melihat nilai reliabilitas dan daya pembeda (*separation*). Reliabilitas adalah uji yang dilakukan untuk menunjukkan sejauh mana suatu alat ukur dapat dipercaya atau diandalkan (Amanda, Yanuar, & Devianto, 2019); Sugiono, Noerdjanah, & Wahyu, 2020). Reliabilitas pada pemodelan Rasch ditunjukkan dengan nilai *alpha Cronbach*, *person reliability*, dan *item reliability*. *Alpha Cronbach* digunakan untuk mengukur interaksi antara *person* dan *item* secara keseluruhan. *Person reliability* menunjukkan tingkat konsistensi subjek penelitian atau peserta didik dalam memberikan jawaban. *Item reliability* atau reliabilitas item digunakan untuk menentukan suatu reliabilitas item soal. Kriteria yang digunakan untuk melihat nilai dari *alpha Cronbach* dapat dilihat pada Tabel 2.

Tabel 2.

Kriteria Nilai *Alpha Cronbach*.

Nilai <i>Alpha Cronbach</i> (<i>a</i>)	Keterangan
$a < 0.5$	Buruk
$0.5 \leq a < 0.6$	Jelek
$0.6 \leq a < 0.7$	Cukup
$0.7 \leq a < 0.8$	Baik
$a \geq 0.8$	Sangat Baik

(Sumber: Sumintono & Widhiarso, 2015)

Kriteria yang digunakan untuk melihat nilai *person reliability* dan *item reliability* dapat dilihat pada Tabel 3.

Tabel 3.

Kriteria Nilai *person reliability* dan *Item Reliability*.

Nilai <i>Person Reliabilitas</i> dan <i>Item Reliability</i> (P)	Keterangan
$P < 0.67$	Lemah
$0.67 \leq P < 0.80$	Cukup
$0.81 \leq P < 0.90$	Baik
$0.91 \leq P < 0.94$	Sangat Baik
$P \geq 0.94$	Istimewa

(Sumber: Sumintono & Widhiarso, 2015)

Daya pembeda butir soal merupakan kemampuan butir soal untuk membedakan peserta didik yang dapat memiliki tingkat kemampuan yang tinggi dengan peserta didik yang memiliki tingkat kemampuan rendah (Erfan *et al.*, 2020). Daya pembeda pada pemodelan Rasch dapat dilihat pada *person separation* dan *item separation*. nilai dari daya pembeda kemudian dihitung dengan persamaan pemisahan strata yang dapat digunakan untuk melihat pengelompokan secara lebih teliti dengan rumus sebagai berikut (Sumintono & Widhiarso, 2015).

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

Keterangan:

H : Pemisah strata

Separation : Nilai *separation* pada *summary statistics*

Output dari *summary statistics* dapat dilihat pada menu *Output Tables* pada *software Winsteps*, kemudian pilih Tabel 3.1 *Summary Statistics* yang dapat dilihat pada Gambar 4.

03-254WS - Notepad

File Edit Format View Help

TABLE 3.1 C:\Users\User\Desktop\POLITOMI Data Me ZOU254WS.TXT Jul 16 10:50 2024

INPUT: 363 Person 10 Item REPORTED: 363 Person 10 Item 4 CATS WINSTEPS 3.73

SUMMARY OF 363 MEASURED Person

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	27.5	9.4	.70	.42	1.00	.0	1.02	.0
S.D.	4.5	.8	.65	.04	.48	1.2	.56	1.2
MAX.	36.0	10.0	2.46	.72	2.70	3.0	3.83	3.5
MIN.	14.0	6.0	-1.06	.38	.04	-4.8	.04	-4.6

REAL RMSE	.46	TRUE SD	.46	SEPARATION	.99	Person RELIABILITY	.50
MODEL RMSE	.42	TRUE SD	.49	SEPARATION	1.16	Person RELIABILITY	.58
S.E. OF Person MEAN = .03							

VALID RESPONSES: 94.2% (APPROXIMATE)

Person RAW SCORE-TO-MEASURE CORRELATION = .85 (approximate due to missing data)

CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .68 (approximate due to missing data)

SUMMARY OF 10 MEASURED Item

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	997.0	342.0	.00	.07	1.01	.0	1.01	.1
S.D.	172.4	7.9	.70	.01	.17	2.2	.16	2.0
MAX.	1219.0	355.0	1.01	.08	1.39	4.4	1.38	3.6
MIN.	740.0	330.0	-.89	.06	.77	-3.5	.76	-3.5

REAL RMSE	.07	TRUE SD	.69	SEPARATION	9.55	Item RELIABILITY	.99
MODEL RMSE	.07	TRUE SD	.69	SEPARATION	9.90	Item RELIABILITY	.99
S.E. OF Item MEAN = .23							

Gambar 4. *Output Summary Statistics.*

Berdasarkan Gambar 4, reliabilitas yang diperoleh dari nilai *alpha Cronbach* sebesar 0.68 yang berada pada rentang 0,6 – 0,7 (cukup), artinya interaksi antara peserta didik (*person*) dan butir soal (*item*) secara keseluruhan termasuk dalam kategori cukup baik. Nilai *person*

reliability yang diperoleh sebesar $0.50 < 0.67$, artinya konsistensi jawaban peserta didik termasuk dalam kategori lemah. Nilai *item reliability* yang diperoleh sebesar $0.99 > 0.94$, artinya kualitas butir soal istimewa.

Nilai *person reliability* yang diperoleh dari hasil analisis termasuk dalam kategori lemah, hal ini dikarenakan jawaban peserta didik yang beragam dilihat dari sampel penelitian yang digunakan berasal dari sekolah yang berbeda. Suryanto, Indriyani, & Sofyani (2017) menyatakan bahwa terdapat dua kemungkinan yang menyebabkan rendahnya nilai *person reliability*. Pertama, jawaban peserta didik memiliki keragaman (varians) yang cukup tinggi sehingga hasil analisis menilai adanya kemungkinan pemahaman peserta didik atas butir soal tes tidak seragam. Artinya, kemungkinan instrumen kurang mudah dipahami secara umum sehingga dinilai memiliki *person reliability* yang lemah. Kedua, jawaban memang menjurus pada keseragaman yang rendah mengindikasikan beragamnya sampel yang digunakan. Berdasarkan pernyataan tersebut, hasil analisis mengarah pada kemungkinan kedua dimana nilai *person reliability* yang diperoleh sebesar 0.50 berkategori lemah, sedangkan nilai *item reliability* yang diperoleh sebesar 0.99 berkategori istimewa. Hal ini disebabkan karena sampel penelitian yang digunakan dan jawaban yang diperoleh dari peserta didik memiliki tingkat keseragaman yang tinggi. Hasil penelitian ini didukung dengan hasil penelitian oleh Suryanto *et al.*, (2017) dimana hasil analisis dari *person reliability* yang diperoleh berkategori lemah (0.60), sedangkan nilai *item reliability* berkategori baik (0.85) yang artinya tidak terdapat masalah pada instrumen yang digunakan dan rendahnya nilai *person reliability* disebabkan oleh sampel penelitian yang relatif tinggi dan beragam.

Hasil analisis dari *output summary statistics* pada nilai *person separation* dan *item separation* yang diperoleh selanjutnya dihitung dengan persamaan pemisah strata. Nilai *person separation* diperoleh nilai sebesar 0.99 maka nilai $H = 1,65$ yang kemudian dibulatkan menjadi 2. Nilai *item separation* diperoleh nilai sebesar 9.55 maka nilai $H = 13,06$ yang kemudian dibulatkan menjadi 13. Pengelompokan *person* dapat dikelompokkan menjadi 2 kelompok, sedangkan kelompok *item* dapat dikelompokkan menjadi 13 kelompok. Nilai *separation* yang lebih tinggi menunjukkan kualitas instrumen dalam hal jumlah responden dan butir soal semakin bagus karena lebih mudah untuk mengidentifikasi kelompok responden dan kelompok butir (Sumintono & Widhiarso, 2015). Hasil analisis yang diperoleh pada model Rasch hanya dapat mengetahui jumlah pengelompokannya saja, namun tidak diketahui pasti pembagian kelompok yang diperoleh.

4. Analisis Person Measure

Analisis dengan menggunakan pemodelan Rasch, selain dapat digunakan untuk menganalisis butir soal tes juga dapat digunakan untuk menganalisis tingkat abilitas (kemampuan) peserta didik dalam mengerjakan butir soal tes yang diujikan. Tingkat abilitas individu (*person measure*) merupakan analisis yang bertujuan untuk mengetahui tingkat kemampuan peserta didik dalam menyelesaikan butir soal (Rohmah, Susilaningih, Haryani, & Kasmui, 2022).

Tingkat abilitas peserta didik dapat dilihat pada kolom *measure* berdasarkan pada nilai *logit* yang telah diurutkan dari peserta didik dengan abilitas tertinggi hingga terendah. Nilai *logit* dari *person* dengan nilai tertinggi kemampuan peserta didik yang tinggi dalam menyelesaikan butir soal dengan tingkat kesulitan tertinggi. Pengelompokan tingkat abilitas peserta didik pada model Rasch menggunakan standar deviasi dan nilai rata-rata *logit* yang dihasilkan dari *output Person Measure* (Lestari, Hamdu, & Saputra, 2023, 2023; Tyas, Hamdu, & Pranata, 2020). Kriteria pengelompokan tingkat abilitas peserta didik dikategorikan ke dalam tiga kategori, yaitu abilitas tinggi, abilitas sedang, dan abilitas rendah (Sumintono & Widhiarso, 2015). Nilai standar deviasi yang diperoleh sebesar 0.65 dan nilai rata-rata *logit* 0.70 sehingga pengelompokan tingkat abilitas peserta didik dapat dikategorikan sebagai berikut: *logit* > 0.70 memiliki tingkat abilitas tinggi; $0.65 < \text{logit } t < 0.70$ memiliki tingkat abilitas sedang; dan *logit* < 0.65 memiliki tingkat abilitas rendah.

Analisis tingkat abilitas individu dapat dilihat dari hasil *Output Tables* pada *software Winsteps*, pada Tabel 17. *Person Measure*. Rangkuman hasil analisis *person measure* yang dapat dilihat pada Tabel 4.

Tabel 4.

Rangkuman Hasil Analisis *Person Measure*.

Kategori	Jumlah	Presentase (%)
Tinggi	178	49%
Sedang	18	5%
Rendah	167	46%
Total	363	100%

(Sumber: Data peneliti)

Berdasarkan Tabel 4.15 diketahui bahwa peserta didik dengan tingkat abilitas tinggi memiliki persentase 49% berjumlah 178 orang. Peserta didik dengan tingkat abilitas sedang memiliki persentase 5% berjumlah 18 orang. Peserta didik dengan tingkat abilitas rendah memiliki persentase 46% berjumlah 167 orang.

Hasil analisis menunjukkan peserta didik dengan kode *person* 123LB memiliki nilai *logit* tertinggi dengan nilai sebesar 2.46, peserta didik ini dapat menyelesaikan 8 dari 10 butir soal dengan total skor mentah yang diperoleh sebesar 30 poin. Nilai *logit* tertinggi yang diperoleh peserta didik 123LB disebabkan oleh kemampuannya dalam menjawab butir soal yang sangat sulit dengan benar, sedangkan peserta didik dengan tingkat abilitas terendah dimiliki oleh peserta didik dengan kode *person* 068LA dengan nilai *logit* sebesar -1.06 yang dapat menyelesaikan keseluruhan butir soal, namun hanya memperoleh total skor mentah 18 poin. Nilai *logit* terendah yang diperoleh peserta didik 068LA disebabkan oleh kemampuannya yang rendah dalam menjawab butir soal, yakni poin yang diperoleh tiap butir soal rendah. Hal ini sejalan dengan penelitian yang dilakukan oleh Rohmah *et al.*, (2022) yang menyatakan bahwa nilai *logit* yang tinggi mengindikasikan kemampuan peserta didik untuk menyelesaikan atau menjawab butir soal dengan benar juga tinggi. Hasil analisis juga menunjukkan bahwa terdapat beberapa peserta didik yang memiliki nilai *logit* yang sama, seperti peserta didik dengan kode *person* 322PE, 015PA, dan 129LB, yaitu sebesar 0.37. Hal ini menunjukkan peserta didik memiliki tingkat abilitas yang sama dengan memperoleh skor mentah (*total score*) yang sama (Sumintono & Widhiarso, 2015).

5. Analisis *Person Fit*

Tingkat kesesuaian individu (*person fit*) dalam pemodelan Rasch digunakan untuk mendeteksi adanya peserta didik yang memiliki pola respons jawaban yang tidak sesuai. Pola respons yang tidak sesuai artinya terdapat ketidaksesuaian antara abilitas peserta didik dengan jawaban yang diberikan dalam menjawab butir soal (Aprilia, Susilaningih, Priatmoko, & Kasmui, 2020; Lestari *et al.*, 2023; Rohmah *et al.*, 2022). Pola respons peserta didik yang dimaksud seperti peserta didik melakukan kecurangan (*cheating*), menebak jawaban (*lucky guess*), atau ceroboh (*careless*) dalam menjawab butir soal (Rohmah *et al.*, 2022). Selain itu, analisis tingkat kesesuaian individu juga dapat digunakan guru untuk mengetahui konsistensi kemampuan berpikir peserta didik (Sumintono & Widhiarso, 2015).

Tingkat kesesuaian individu dilihat dari kriteria yang sama dengan tingkat kesesuaian butir soal, yaitu dengan melihat *outfit mean-square*, *outfit z-standard*, dan *point measure correlation*, di mana peserta didik dapat dikatakan *fit* apabila memenuhi satu dari ketiga kriteria yang digunakan (Sumintono & Widhiarso, 2015). Analisis tingkat kesesuaian individu pada menu *Output Tables* pada *software Winsteps*, pada Tabel 16. *Person Fit Order*. Rangkuman hasil analisis *person fit* yang dapat dilihat pada Tabel 5.

Tabel 5.Rangkuman Hasil Analisis *Person Fit*.

Kategori	Jumlah	Presentase (%)
Fit	254	70%
Tidak Fit	32	9%
<i>Outlier</i>	77	21%
Total	363	100%

Berdasarkan Tabel 3 diketahui bahwa peserta didik yang termasuk ke dalam kategori *fit* memiliki persentase sebesar 70% berjumlah 254 orang. Peserta didik yang termasuk ke dalam kategori tidak *fit* memiliki persentase sebesar 9% berjumlah 32 orang. Peserta didik yang termasuk kategori *outlier* memiliki persentase 21% berjumlah 77 orang.

Peserta didik yang memiliki hasil analisis tidak *fit* diduga karena peserta didik tersebut memberikan respons jawaban yang tidak sesuai dengan model ideal (Lestari *et al.*, 2023). Informasi yang dapat digunakan untuk mengetahui penyebab lebih lanjut penyebab terjadinya *misfit* pada peserta didik dapat dilihat melalui *output scalograms* (Sumintono & Widhiarso, 2015) yang dapat dilihat pada Gambar 5.

```

GUTTMAN SCALOGRAM OF RESPONSES:
Person | Item
      | 1
      | 2690345781
      | -----
      | 123 +43443 444 123LB
      | 329 + 4443 34 4 329LE
      | 358 + 43344 434 358LE|
      | 9 +2444 4 434 009LA
      | 27 +4444344333 027PA
      | 32 +4334334442 032PA
      | 52 +2343344434 052LA

```

Gambar 5. *Output Scalograms.*

Gambar 5 merupakan potongan hasil analisis *scalograms* di mana peserta didik dengan kode *person* 009LA dan 052LA termasuk peserta didik yang tidak *fit*. Peserta didik 009LA termasuk dalam kategori peserta didik unik karena kurang cermat (*careless*) dalam menjawab butir soal, butir soal dengan tingkat kesulitan rendah (mudah) tidak mendapatkan skor yang maksimal, sedangkan butir soal dengan tingkat kesulitan tinggi (sulit) mendapatkan skor maksimal. Pada peserta didik 052LA merupakan peserta didik yang menunjukkan pola respons (jawaban) yang tidak konsisten. Sumintono & Widhiarso (2015) menjelaskan bahwa pola respons yang ideal adalah butir soal dengan tingkat kesulitan yang rendah memiliki skor yang tinggi, sedangkan butir soal dengan tingkat kesulitan tinggi akan memiliki skor yang rendah. Hal ini yang menyebabkan peserta didik 009LA dan 052LA termasuk ke dalam kategori peserta didik yang tidak *fit*.

Pada hasil analisis *person fit* juga terdeteksi adanya data yang *outlier* dilihat dari hilangnya data *person* (peserta didik) pada hasil *output person fit* sebanyak 77 orang peserta didik. Data *outlier* (pencilan) merupakan data yang berbeda dengan data lainnya (Febriyansyah, Fadhil, & Zulfahrizal, 2019). Data pencilan disebabkan karena kesalahan dalam memasukkan data, kesalahan pengukuran, analisis, dan kesalahan lainnya (Gusriani, Firdaniza, & Octavianti, 2017). Hilangnya data *person* disebabkan karena data memiliki kecenderungan berbeda dengan data lainnya atau tidak sesuai dengan pemodelan Rasch sehingga tidak dapat terbaca pada proses analisis. Hasil analisis *person fit* yang dilakukan Sari & Mahmudi (2024) menunjukkan terdapat data *outlier* yang merusak data sebanyak 110 orang dengan persentase 50% dari total *person* 216 orang. *Person* yang termasuk dalam kategori *outlier* tidak mempengaruhi kualitas *item*. Hal ini sejalan dengan pernyataan (Widodo & Chotimah, 2023) yang menyatakan bahwa pada analisis model Rasch penilaian kemampuan *person* tidak terpengaruh oleh *item* dan kualitas *item* tidak terpengaruh kemampuan *person*.

KESIMPULAN

Berdasarkan hasil analisis instrumen tes berpikir esensial menggunakan model Rasch dari 10 butir soal diketahui bahwa: 1) Kualitas instrumen tes dari analisis *item fit* menunjukkan seluruh butir soal *fit* atau memenuhi kriteria, namun terdapat 1 butir soal yang terdeteksi mengandung bias, yaitu butir soal nomor 7 sehingga butir soal dinyatakan tidak valid. 2) Tingkat kesukaran butir soal menunjukkan sebanyak 2 *item* termasuk kategori sangat sulit (20%), 4 *item* termasuk kategori sulit (40%), 2 *item* termasuk kategori mudah (20%), dan 2 *item* termasuk kategori sangat mudah (20%). 3) Ringkasan *summary statistics* diperoleh nilai *alpha Cronbach* sebesar 0.68 dengan kategori cukup, nilai *person reliability* sebesar 0.50 dengan kategori lemah, dan nilai *item reliability* sebesar 0.99 dengan kategori istimewa. Sedangkan nilai *person separation* diperoleh nilai sebesar 0.99 dengan nilai $H=1,65$ dan nilai *item separation* diperoleh nilai sebesar 9.55 dengan nilai $H=13,06$. 4) Analisis *person measure* diperoleh 178 orang memiliki abilitas tinggi (49%), 18 orang memiliki abilitas sedang (5%), dan 167 orang memiliki abilitas rendah (46%). 5) Analisis *person fit* diperoleh 254 orang kategori *fit* (70%), 32 orang kategori tidak *fit* (9%), dan 77 orang kategori *outlier* (21%). Hasil analisis kualitas butir soal dengan model Rasch menunjukkan 9 butir soal dinyatakan valid dan layak digunakan, sedangkan analisis abilitas peserta didik menunjukkan tingkat abilitas yang merata. Dengan demikian, analisis model Rasch sangat disarankan bagi guru untuk dijadikan sebagai metode analisis yang dapat menganalisis kualitas butir soal dan abilitas peserta didik secara bersamaan.

DAFTAR PUSTAKA

- Amanda, L., Yanuar, F., & Devianto, D. (2019). Uji validitas dan reliabilitas tingkat partisipasi politik masyarakat Kota Padang. *Jurnal Matematika UNAND*, 8(1), 179–188.
- Aprilia, N., Susilaningih, E., Priatmoko, S., & Kasmui. (2020). Desain instrumen tes pemahaman konsep berbasis HOT dengan analisis model Rasch. *Chemistry in Education*, 9(2), 1–8. Diambil dari <https://journal.unnes.ac.id/sju/chemined/article/view/39068/17404>.
- Camila, C. G., Afandi, A., Tenriawaru, A. B., Artika, W., & Siregar, N. (2023). Development of higher order thinking skill questions using Stahl and Murphy's taxonomy on excretion system topic. *Assimilation: Indonesian Journal of Biology Education*, 6(2), 97–108. doi: 10.17509/aijbe.v6i2.60632.
- Chusni, M. M., Saputro, S., Suranto, & Rahardjo, S. B. (2020). The conceptual framework of designing a discovery learning modification model to empower students' essential thinking skills. *Journal of Physics: Conference Series*, 1467(1), 1-9. doi: 10.1088/1742-6596/1467/1/012015.
- Erfan, M., Maulida, M. A., Hidayati, V. R., Astria, F. P., & Ratu, T. (2020). Analisis kualitas soal kemampuan membedakan rangkaian seri dan paralel melalui teori tes klasik dan model Rasch. *Indonesian Journal of Educational Research and Review*, 3(1), 11–19. doi: 10.23887/ijerr.v3i1.24080.
- Falani, I. (2023). Desain dan validasi aplikasi tes literasi matematika berbasis komputer dengan pendekatan *item response theory*. *Edu-Sains: Jurnal Pendidikan Matematika dan Ilmu Pengetahuan Alam*, 12(2), 16–28. doi: 10.22437/jmpmpipa.v12i2.26716.
- Febriyansyah, I. A., Fadhil, R., & Zulfahrizal. (2019). Pengembangan metode klasifikasi biji kopi beras arabika gayo dan robusta gayo dengan metode PCA (*Principal Component Analysis*) berdasarkan pengolahannya. *Jurnal Ilmiah Mahasiswa Pertanian*, 4(4), 472–481. doi: 10.17969/jimfp.v4i4.12776.
- Feranda, E., Ningsih, K., & Afandi. (2020). Penggunaan taksonomi Marzano dalam membuat soal HOTS: sebuah kajian literatur. *Prosiding Seminar Nasional Pendidikan 2020*, 1053–1061. Pontianak: Universitas Tanjungpura. Diambil dari <https://snpfkip.untan.ac.id/wp-content/uploads/2022/12/1-Prosiding-SemNas-Pendidikan-FKIP-Untan-Tahun-2020-Revisi-Pertama.pdf>.
- Gusriani, N., Firdaniza, & Octavianti, N. (2017). Regresi TELBS untuk mengatasi masalah pencilan. *Industrial Research Workshop and National Seminar*, 580–584. Bandung: Politeknik Negeri Bandung. Diambil dari <https://jurnal.polban.ac.id/ojs-3.1.2/proceeding/article/view/670/523>.
- Hamdu, G., Fuadi, F. N., Yulianto, A., & Akhirani, Y. S. (2020). Items quality analysis using Rasch model to measure elementary school students' critical thinking skill on STEM learning. *JPI (Jurnal Pendidikan Indonesia)*, 9(1), 61-74. doi: 10.23887/jpi-undiksha.v9i1.20884.

- Imania, K. A. N., & Bariah, S. K. (2019). Rancangan pengembangan instrumen penilaian pembelajaran berbasis daring. *PETIK: Jurnal Pendidikan Teknologi Informasi dan Komunikasi*, 5(1), 31–47. Diambil dari <https://journal.institutpendidikan.ac.id/index.php/petik/article/view/1534/1258>.
- Jumini, S., Modnasri, S., Cahyono, E., & Parmin. (2023). Analisis kualitas butir soal pengukuran literasi sains melalui teori tes klasik dan Rasch model. *Prosiding Seminar Nasional Pascasarjana*, 758–765. Semarang: Universitas Negeri Semarang. Diambil dari <https://proceeding.unnes.ac.id/snpasca/article/view/2215/1698>.
- Kirom, A., & Hasyim, M. (2021). Analisis butir soal sebagai standarisasi mutu Sekolah Dasar pada mata pelajaran PAI dengan menggunakan pendekatan Rasch model di SD Ma'arif Nu Kecamatan Pandaan Pasuruan. *AL MURABBI: Jurnal Pendidikan Agama Islam*, 6(2), 92–98. Diambil dari <https://jurnal.yudharta.ac.id/v2/index.php/pai/article/view/2631/1881>.
- Lestari, T. D., Hamdu, G., & Saputra, E. R. (2023). Analisis soal literasi numerasi menggunakan pemodelan Rasch konteks pemanasan global berbasis ESD untuk Sekolah Dasar. *Pendas: Jurnal Ilmiah Pendidikan Dasar*, 8(2), 2489–2503. doi: 10.23969/jp.v8i2.9270.
- Mulyani, S., Efendi, R., & Ramalis, T. R. (2021). Karakterisasi tes keterampilan pemecahan masalah fisika berdasarkan teori respon butir. *Jurnal Pendidikan dan Ilmu Fisika*, 1(1), 1–14. doi: 10.52434/jpif.v1i1.1006.
- Pratama, D. (2020). Analisis kualitas tes buatan guru melalui pendekatan Item Response Theory (IRT) model Rasch. *Tarbawy: Jurnal Pendidikan Islam*, 7(1), 61–70. doi: 10.32923/tarbawy.v7i1.1187.
- Preseisen, B. Z. (1985). *Thinking skills throughout the curriculum: a conceptual design*. Washington, DC: Research for Better Schools, Philadelphia.
- Rahmat, A. A., Hamdu, G., & Nur'aeni, E. (2020). Pengembangan soal tes tertulis berbasis STEM dengan pemodelan Rasch di Sekolah Dasar. *Metodik Didaktik: Jurnal Pendidikan Ke-SD-an*, 16(1), 29–40. doi: 10.17509/md.v16i1.25099.
- Ridwan, I. M., Kaniawati, I., Suhandi, A., Ramalis, T. R., Rizal, R., & Sujarwanti, E. (2023). Instrumen kemampuan pemecahan masalah pada tema perubahan iklim: analisis Rasch model. *DIFFRACTION: Journal for Physics Education and Applied Physics*, 5(1), 37–46. doi: 10.37058/diffraction.v5i1.7738.
- Rohmah, N. F., Susilaningsih, E., Haryani, S., & Kasmui. (2022). Desain asesmen kompetensi minimum literasi membaca bermuatan High Order Thinking Skills untuk menganalisis kompetensi minimum siswa materi asam-basa. *Chemistry in Education*, 11(2), 117–125. doi: 10.15294/chemined.v11i2.58847.
- Rusilowati, A. (2018). Asesmen literasi sains: analisis karakteristik instrumen dan kemampuan siswa menggunakan teori tes modern Rasch model. *Prosiding Seminar Nasional Fisika*, 2–15. Pekanbaru: Universitas Riau. Diambil dari <https://snf.fmipa.unri.ac.id/wp-content/uploads/2019/03/0.-300B-2-15NI.pdf>.
- Sari, E. D. K., & Mahmudi, I. (2024). *Buku analisis pemodelan rasch pada asesmen pendidikan*. Jawa Tengah: PT. Pena Persada Kerta Utama.
- Sugiono, Noerdjanah, & Wahyu, A. (2020). Uji validitas dan reliabilitas alat ukur SC posture evaluation. *Jurnal Keterampilan Fisik*, 5(1), 55–61. doi: 10.37341/jkf.v5i1.167.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan Rasch pada assessment pendidikan*. Cimahi: Trim Komunikata.
- Suryanto, R., Indriyani, Y., & Sofyani, H. (2017). Determinan kemampuan auditor dalam mendeteksi kecurangan. *Jurnal Akuntansi dan Investasi*, 18(1), 102–118. doi: 10.18196/jai.18163.
- Tarigan, E. F., Nilmarito, S., Islamiyah, K., Darmana, A., & Suyanti, R. D. (2022). Analisis instrumen tes menggunakan Rasch model dan software SPSS 22.0. *Jurnal Inovasi Pendidikan Kimia*, 16(2), 92–96. doi: 10.15294/jipk.v16i2.30530.
- Tyas, E. H., Hamdu, G., & Pranata, O. H. (2020). Analisis soal pilihan ganda dengan menggunakan pemodelan Rasch untuk mengukur kemampuan siswa dalam mengurutkan bilangan pecahan di Sekolah Dasar. *PEDADIDAKTIKA: Jurnal Ilmiah Pendidikan Guru Sekolah Dasar*, 7(2), 1–12. doi: 10.17509/pedadidaktika.v7i2.24773.
- 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), 212–225. doi: 10.30829/tar.v29i2.1706.
- Untary, H., Risdianto, E., & Kusen. (2020). *Analisis data penelitian dengan model Rasch dan Winstep*. Bogor: Halaman Moeka Publishing.
- Widodo, W., & Chotimah, C. (2023). Adaptasi dan analisis psikometri skala kompetensi multikultural calon guru menggunakan pemodelan Rasch. *Jurnal Kependidikan dan Kebudayaan*, 8(2), 158–172. doi: 10.24832/jpnk.v8i2.4228.