



Analisis butir soal pengukuran kemampuan berpikir kritis materi pencemaran lingkungan melalui teori tes klasik dan Rasch model

Lathifah Radha Istiyova ^{a,1,*}

^a SMP Islam Diponegoro Surakarta, Kota Surakarta, Jawa Tengah 57118, Indonesia.

¹ radhaistiyova@gmail.com*

* Corresponding author.

INFORMASI ARTIKEL

Lini Masa Artikel

Draft diterima : 2024-06-22
Revisi diterima : 2024-11-25
Diterbitkan : 2025-04-24

Keywords

Classical Test Theory;
Item Analysis;
Rasch Model;

ABSTRAK

Penelitian ini bertujuan untuk mengetahui kualitas butir soal pengukuran pemahaman konsep pencemaran lingkungan di salah satu SMA Negeri Kabupaten Sukoharjo Tahun Pelajaran 2023/2024. Metode penelitian yang digunakan yaitu deskriptif-kuantitatif dengan pelaksanaan penelitian pada bulan Februari 2024. Data diperoleh dengan metode dokumenter 25 soal pilihan ganda yang diujikan kepada 36 siswa. Analisis data menggunakan teori tes klasik dan Rasch Model. Kualitas soal berdasarkan validitasnya menggunakan teori tes klasik dihasilkan bahwa 25 soal valid, sedangkan pada Rasch Model 24 valid dan 1 tidak valid. Reliabilitas pada teori tes klasik yaitu 0,875 (tinggi) dan skor reliabilitas Rasch Model 0,85 (bagus). Tingkat kesukaran pada teori tes klasik, soal diklasifikasikan menjadi 6 soal sukar, 16 sedang, dan 3 mudah. Sedangkan Rasch Model butir soal terdistribusi menjadi 4 sangat sulit, 8 sulit, 7 mudah, dan 5 sangat mudah. Aspek daya beda menggunakan teori tes klasik diperoleh 6 soal kategori cukup dan 19 soal kategori baik, sedangkan pada Rasch Model diperoleh tiga kelompok butir soal berdasarkan pada indeks separasi butir soal $H = 3,34$.

ABSTRACT

This research aims to determine the quality of the items measuring understanding of environmental pollution in one of the Sukoharjo Regency State High Schools for the 2023/2024 academic year. The research method used was descriptive-quantitative with research carried out in February 2024. Data was obtained using a documentary method of 25 multiple-choice questions tested on 36 students. Data analysis uses classical test theory and the Rasch Model. The quality of the questions based on their validity using classical test theory resulted in 25 valid questions, while in the Rasch Model 24 were valid and 1 was invalid. The reliability of the classical test theory is 0.875 (high) and the Rasch Model reliability score is 0.85 (good). Difficulty level in classical test theory, questions are classified into 6 difficult questions, 16 medium and 3 easy. Meanwhile, the Rasch Model questions are distributed into 4 very difficult, 8 difficult, 7 easy, and 5 very easy. The different power aspects using classical test theory obtained 6 questions in the sufficient category and 19 questions in the good category. In contrast, in the Rasch Model three groups of items were obtained based on the item separation index $H = 3.34$.

Cara Sitasi Artikel Ini (APA Style):

Istiyova, L. R. (2025). Analisis butir soal pengukuran kemampuan berpikir kritis materi pencemaran lingkungan melalui teori tes klasik dan Rasch model. *Bio-Pedagogi*. 14(1), 36-46. <https://dx.doi.org/10.20961/bio-pedagogi.v14i1.88792>.

Artikel ini dapat diakses secara bebas dengan lisensi [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/).



PENDAHULUAN

Pendidikan yang berkualitas tidak terlepas pada proses evaluasi pembelajaran yang baik ([Tarigan et al., 2022](#)). Evaluasi adalah suatu bentuk kegiatan dalam mengumpulkan data untuk menguji sejauh mana tujuan dari pembelajaran sudah dicapai ([Murias et al., 2024](#)). Evaluasi meliputi mengukur yang bersifat kuantitatif dan menilai yang bersifat kualitatif ([Nuryanti et al., 2018](#)). Mengukur berupa aktivitas dalam membandingkan sesuatu dalam satu ukuran, sedangkan menilai merupakan pengambilan suatu keputusan dengan ukuran baik dan buruk ([Purniasari et al., 2021](#)). Pendidik sebagai ujung tombak sistem pendidikan di Indonesia dibekali kemampuan untuk mengevaluasi pembelajaran karena kegiatan belajar yang berhasil dapat dilihat dari hasil kegiatan suatu evaluasi yang telah dilakukan ([Polat, 2022](#)).

Pendidik salah satunya harus mempunyai kemampuan mengevaluasi tetapi juga harus dapat mengembangkan instrumen evaluasi yang digunakan. Instrumen yang baik selaras dengan capaian pembelajaran yang diharapkan ([Hayati, 2016](#)). Teknik dalam mengevaluasi dan memperoleh data harus disesuaikan, karena dapat berfungsi sebagai indikator tercapainya suatu tujuan pembelajaran. Teknik yang sering digunakan biasanya teknik tes dan non tes ([Susdelina et al., 2018](#)). Tes merupakan instrumen pengumpulan informasi tentang ketercapaian tujuan pembelajaran, sedangkan non tes merupakan teknik yang dilakukan tanpa menguji siswa, melainkan dengan observasi, wawancara, penyebaran angket, dan analisis dokumen ([Yuliawati et al., 2022](#)). Keterampilan seorang pendidik dalam menentukan suatu teknik yang tepat dan sesuai dalam aktivitas evaluasi ini merupakan kompetensi yang wajib untuk dikuasai.

Instrumen penilaian yang berkualitas harus dapat memenuhi beberapa syarat uji yang meliputi validitas, reliabilitas, tingkat kesukaran, dan daya beda/diskriminasi tiap butir soal ([Erfan et al., 2020](#)). Pengujian kualitas instrumen dilakukan dengan dua cara yaitu secara internal dan eksternal ([Azizah & Wahyuningsih, 2020](#)). Uji coba internal meliputi validitas isi dan validitas konstruk dimana dilakukan oleh para ahli untuk diberikan masukan terhadap struktur instrument yang telah dibuat. Sedangkan uji coba eksternal berupa uji coba lapangan ([Arikunto, 2017](#)), yang mana uji coba penelitian dilakukan pada subjek yang akan dilakukan penilaian. Hasil dari uji coba tersebut kemudian dianalisis validitas, reliabilitas, tingkat kesukaran, dan daya beda soal.

Validitas merupakan sejauh mana ketepatan suatu instrumen untuk mengukur/menguji apa yang seharusnya diukur ([Ida & Musyarofah, 2021](#)). Reliabilitas adalah suatu indeks untuk mengukur sejauh mana alat ukur tersebut dapat dipercaya atau diandalkan ([Magdalena et al., 2021](#)). Tingkat kesukaran butir soal merupakan sejauh mana tepatnya subyek penelitian dalam menjawab suatu soal ([Susilaningsih et al., 2021](#)). Soal dikatakan berkualitas yaitu soal yang tidak terlalu mudah tetapi juga tidak terlalu sulit. Sedangkan daya beda/diskriminasi merupakan kemampuan suatu soal dapat membedakan antara kemampuan peserta didik yang tinggi dan kemampuan peserta didik yang rendah ([Santosa & Badawi, 2022](#)). Untuk mengetahui kriteria penilaian instrumen diatas diperlukan analisis butir soal penilaian hasil evaluasi pembelajaran.

Pendekatan yang dilakukan untuk menganalisis butir soal berdasarkan hasil uji coba lapangan terdapat dua yaitu Teori Tes Klasik dan Rasch Model ([Sabekti, 2018](#)). Teori tes klasik disebut juga teori skor murni yang merupakan metode yang mendasarkan pada skor amatan yang berasal dari penjumlahan skor sebenarnya dan kesalahan dalam pengukuran ([Ramadhini & Nasution, 2022](#)). Kualitas butir soal pada pendekatan ini ditentukan dari tingkat kesukaran soal dan daya pembedanya, tetapi karakteristiknya tidak konsisten karena bergantung pada kemampuan subyek ([Rahayu, 2016](#)). Cara yang lebih modern dalam analisis butir soal yaitu Rasch Model yang dikemukakan oleh Dr. Georg Rasch ahli matematika yang berasal dari Denmark. Rasch Model mengatasi kelemahan dari teori tes klasik dengan memakai skor mentah yang diolah dengan metode yang berbeda untuk menghasilkan skala pengukuran dengan data interval yang sama ([Tarmizi et al., 2023](#)). Sehingga cara ini dapat memberikan informasi yang lebih akurat tentang kompetensi peserta tes maupun kualitas dari butir soal ([Mahamud & Othman, 2020](#)).

Artikel pada penelitian ini menganalisis perbandingan kualitas instrumental yang ditinjau dari aspek validitas, reliabilitas, tingkat kesukaran, dan daya beda butir soal menggunakan pendekatan teori tes klasik dan Rasch Model. Item soal yang dianalisis yaitu soal tes kemampuan berpikir kritis materi pencemaran lingkungan. Analisis butir soal yang dilakukan digunakan sebagai upaya untuk meningkatkan kualitas instrumen evaluasi sehingga dapat mengukur kemampuan berpikir kritis siswa dengan valid sebagai umpan balik terhadap kegiatan pembelajaran biologi.

METODE

Penelitian yang dilakukan menggunakan metode penelitian deskriptif kuantitatif. Tujuan dari metode ini adalah membuat suatu gambaran atau deskriptif mengenai keadaan secara nyata/objektif yang diinterpretasikan menggunakan angka, dimulai dari pengumpulan data, penafsiran data hingga penampilan serta hasilnya ([Arikunto, 2006](#)).

Populasi penelitian adalah siswa kelas X MIPA di salah satu SMA Negeri Kabupaten Sukoharjo. Jumlah siswa kelas X sebanyak 251 siswa. Teknik pengambilan sampel adalah sampling jenuh ([Nasution et al., 2022](#)), yaitu terdapat satu kelas yang digunakan sebagai sampel penelitian. Sampel penelitian yang terdapat pada uji coba lapangan terdiri atas satu kelas yang berjumlah 36 siswa.

Instrumen penelitian tes kemampuan berpikir kritis siswa materi pencemaran lingkungan menggunakan 25 soal pilihan ganda. Aspek-aspek kemampuan berpikir kritis siswa diambil menurut [Ennis \(2011\)](#) terdapat basic support, inference, advance clarification, strategies and tactics, and elementary clarification. Soal tes kemampuan berpikir kritis dilakukan validitas isi dan validitas konstruk ([Rashwan, 2024](#)). Validitas isi dilakukan dengan diskusi kepada dosen yang berguna untuk memperbaiki kekurangan dalam penyusunan instrumen soal dan sudah dinyatakan valid ([Novia et al., 2019](#)), sedangkan validitas konstruk menggunakan analisis butir soal teori tes klasik dan Rasch Model.

Penelitian ini dilaksanakan dengan tiga tahap, yaitu tahap analisis kebutuhan (defining), tahap pengembangan instrumen, dan tahap validasi. Tahap analisis kebutuhan (defining), pada tahap ini peneliti menganalisis profil dari kemampuan siswa yang akan dijadikan dasar dalam merancang instrumen penelitian, termasuk menganalisis kebutuhan sekolah dalam kurikulum yang perlu ditingkatkan dalam kemampuan siswa ([Febriana et al., 2022](#)). Analisis ini dilakukan melalui wawancara dengan guru serta studi dokumentasi terhadap kurikulum dan hasil belajar sebelumnya. Hasil analisis kebutuhan digunakan sebagai dasar untuk menyusun indikator kemampuan berpikir kritis yang relevan dengan konteks dan kebutuhan peserta didik.

Tahap pengembangan diawali dengan penyusunan naskah soal tes kemampuan berpikir kritis berdasarkan indikator yang menjadi rujukan. Indikator yang digunakan pada instrumen kemampuan berpikir kritis yaitu menurut [Ennis \(2011\)](#) berupa basic support, inference, advance clarification, strategies and tactics, and elementary clarification. Setiap butir soal disusun dengan mengacu pada konteks pembelajaran yang autentik agar mampu mengukur kemampuan berpikir kritis secara mendalam. Draft soal yang telah disusun kemudian direview secara internal oleh tim ahli sebelum masuk ke tahap validasi lebih lanjut.

Tahap validasi instrumen soal dilakukan dengan dua cara yaitu validitas isi dan validitas konstruk. Validitas isi yaitu kualitas soal yang telah dibuat dengan didiskusikan bersama ahli yaitu dosen. Proses diskusi ini melibatkan penilaian terhadap kesesuaian soal dengan indikator berpikir kritis dan keterkaitannya dengan materi pembelajaran. Sedangkan validitas konstruk yaitu analisis butir soal menggunakan teori tes klasik dan Rasch Model ([Khalid et al., 2023](#)).

Analisis butir soal pada instrumen tes dilakukan melalui dua pendekatan, yaitu Teori Tes Klasik dan Rasch Model ([Krebs et al., 2023](#)). Teori Tes Klasik merupakan model pengukuran skor murni. Analisis ini dilakukan untuk menilai tingkat kesukaran, daya beda, dan fungsi distraktor dari masing-masing butir soal. Sedangkan Rasch Model yaitu pengujian instrumen yang berbantu dengan software Winstep yang lebih akurat ([Ramadhani et al., 2024](#)). Model ini memungkinkan peneliti

untuk mengevaluasi kelayakan butir secara lebih objektif dengan mempertimbangkan interaksi antara responden dan soal.

HASIL DAN PEMBAHASAN

Hasil Uji Validitas

Validitas merupakan suatu pengujian tingkat kesahihan dari instrumen, yang berarti kemampuan suatu butir tes dalam mengukur apa yang seharusnya diukur ([Arikunto, 2017](#)). Tipe dari validitas terdapat tiga, yaitu validitas isi, validitas konstruk, dan validitas kriteria. Analisis validitas soal melalui teori tes klasik dilakukan berdasarkan rumus korelasi product moment atau menggunakan teknik korelasi point biserial (rpbi). Soal dapat dikatakan valid apabila koefisien korelasi $r_{xy} > r$ tabel, dan apabila $r_{xy} \leq r$ tabel maka soal dikatakan tidak valid ([Wolfs et al., 2023](#)). Butir soal yang dianalisis dengan model rasch untuk melihat kualitasnya yaitu dengan beberapa kriteria berikut.

- Nilai Outfit MNSQ (Mean Square) yang diterima adalah: $0,5 < \text{Outfit} - \text{MNSQ} < 1,5$
- Nilai Outfit ZSTD (Z-Standard) yang diterima adalah: $-2,0 < \text{ZSTD} < +2,0$
- Nilai Pt Measure Corr (Point Measure Correlation): $0,4 < \text{Point Measure Corr} < 0,85$

Butir soal digolongkan valid apabila memenuhi minimal 2 kriteria diatas dan perlu diperbaiki apabila hanya memenuhi salah satu kriteria saja ([Amalina & Vidakovich, 2023](#)). Penggolongan soal tidak valid apabila soal tersebut tidak memenuhi ketiga kriteria diatas.

Analisis butir soal pengukuran pemahaman konsep materi pencemaran lingkungan SMA N 1 Sukoharjo menggunakan teori tes klasik dan model rasch ditunjukkan pada **Tabel 1** sebagai berikut.

Tabel 1. Perbandingan Hasil Analisis Butir Soal

N	Hasil	Nomor Soal - Teori Tes Klasik	Nomor Soal - Model Rasch
o			
1	Valid	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25	1,2,3,4,5,6,7,8,9,10,11,13,14,15,16,17,18,19,20,21,22,23,24,25
2	Tidak valid	-	12

Berdasarkan hasil analisis butir soal pemahaman konsep pencemaran lingkungan diperoleh hasil pada teori tes klasik terdapat 25 soal valid. Butir soal pada analisis menggunakan Rasch Model dinyatakan 24 soal valid, sedangkan 1 soal tidak valid yaitu nomor 12 karena tidak memenuhi persyaratan *outfit ZSTD* dan *Point Measure Correlation*.

Analisis tersebut menyatakan bahwa terdapat perbedaan hasil validitas soal menggunakan metode tes klasik dan Rasch Model. Teori tes klasik menyatakan bahwa 25 soal valid, sedangkan pada Rasch Model 24 soal valid. Kualitas butir soal terdapat yang dinyatakan tidak valid pada Rasch Model karena analisis butir soal lebih akurat dengan adanya 3 syarat yaitu nilai *Outfit ZSTD*, *Outfit MNSQ*, dan *Point Measure Correlation*. Hal ini menunjukkan bahwa Rasch Model mampu mengidentifikasi butir soal yang kurang sesuai meskipun secara klasik dianggap valid.

Reliabilitas Butir Soal

Analisis reliabilitas butir soal melalui pendekatan teori tes klasik menggunakan rumus Alpha Cronbach (KR-20). Kriteria butir soal dinyatakan reliabel ditunjukkan pada Tabel 2 ([Erfan et al., 2020](#)). Nilai koefisien reliabilitas yang diperoleh menunjukkan tingkat konsistensi internal antarbutir soal dalam mengukur kemampuan yang sama. Semakin tinggi nilai koefisien reliabilitas, maka semakin baik pula kualitas instrumen dalam menghasilkan data yang stabil dan dapat dipercaya.

Tabel 2. Kriteria Koefisien Korelasi Reliabilitas Teori Tes Klasik

Koefisien Korelasi	Interpretasi
$0,90 \leq r \leq 1,00$	Sangat Tinggi
$0,70 \leq r < 0,90$	Tinggi
$0,40 \leq r \leq 0,70$	Sedang
$0,20 \leq r \leq 0,40$	Rendah
$r < 0,20$	Sangat Rendah

Rasch Model menentukan reliabilitas butir soal dengan kriteria nilai Item Reliability dan Person Reliability berdasarkan kriteria pada Tabel 3.

Tabel 3. Kriteria Reliabilitas Rasch Model

Nilai Reliability	Interpretasi
$>0,94$	Istimewa
$0,91 - 0,94$	Bagus Sekali
$0,81 - 0,90$	Bagus
$0,67 - 0,80$	Cukup
$< 0,67$	Lemah

Perbandingan hasil analisis reliabilitas butir soal pemahaman konsep pencemaran lingkungan melalui pendekatan teori tes klasik dan Rasch Model diuraikan pada Tabel 4.

Tabel 4. Perbandingan Analisis Reliabilitas

Teori Tes Klasik	Kategori 1	Rasch Model - Person Reliability	Kategori 2	Rasch Model - Item Reliability	Kategori 3
Alpha Cronbach					
0,87	Tinggi	0.85	Bagus	0.84	Bagus

Berdasarkan hasil analisis pada Tabel 4, diperoleh Alpha Cronbach 0,87 sehingga pada Teori Tes Klasik butir soal dinyatakan reliabel tinggi. Nilai Person Reliability pada Rasch Model 0,85 yang berkategori bagus, sedangkan pada Item Reliability mendapatkan hasil 0,84 yang juga berkategori bagus. Kedua hasil analisis menunjukkan interpretasi yang sama, sehingga dapat disimpulkan kemampuan siswa dapat menjawab soal berpikir kritis materi pencemaran lingkungan termasuk konsisten. Butir-butir soal dalam instrumen pengukuran ini memiliki aspek reliabilitas yang bagus.

Tingkat Kesukaran Butir Soal

Aspek ini menunjukkan besarnya kemungkinan seberapa banyak responden menjawab benar suatu butir soal. Kriteria pengukuran tingkat kesukaran butir soal pada teori tes klasik ditunjukkan pada Tabel 5. Semakin tinggi proporsi siswa yang menjawab benar, maka soal tersebut dikategorikan semakin mudah, dan sebaliknya semakin rendah proporsinya maka soal dikategorikan semakin sukar. Informasi ini penting untuk memastikan bahwa distribusi soal mencakup berbagai tingkat kesulitan agar mampu membedakan kemampuan siswa secara lebih tepat dan adil.

Tabel 5. Kriteria Tingkat Kesukaran Butir Soal pada Teori Tes Klasik

Nilai Indeks Kesukaran	Interpretasi
$IK = 0,00$	Sangat Sukar
$0,00 < IK \leq 0,30$	Sukar
$0,30 < IK \leq 0,70$	Sedang
$0,70 \leq IK \leq 1,00$	Mudah
$IK = 1$	Terlalu Mudah

Rasch Model menggunakan analisis pengujian tingkat kesukaran butir soal berdasarkan pada dua indikator yaitu Measure logit dan nilai Simpangan Baku (SD) logit item yang dikategorikan menjadi 4 ditunjukkan pada **Tabel 6**.

Tabel 6. Kriteria Tingkat Kesukaran Butir Soal Rasch Model

Nilai Measure (logit)	Interpretasi
Measure logit $< -SD$ logit	Sangat Mudah
$-SD \text{ logit} \leq \text{Measure logit} \leq 0$	Mudah
$0 \leq \text{Measure logit} \leq SD \text{ logit}$	Sulit
Measure logit $> SD \text{ logit}$	Sangat Sulit

Perbandingan hasil analisis tingkat kesukaran butir soal pemahaman konsep pencemaran lingkungan melalui pendekatan teori tes klasik dan Rasch Model diuraikan pada **Tabel 7** ([Ramli & Utari, 2023](#)).

Tabel 7. Hasil Analisis Tingkat Kesukaran

Teori Tes Klasik - Interpretasi	Nomor Soal	Rasch Model - Interpretasi	Nomor Soal.1
Sukar	6,7,10,13,18,24	Sangat Sulit	10,13,15,18
Sedang	1,2,3,4,5,8,9,11,12,14,15,16,17,20,21,23	Sulit	5,6,7,8,9,11,16,24
Mudah	19,22,25	Mudah	1,22,3,4,20,21,23
		Sangat Mudah	12,17,19,22,25
		Mudah	

Berdasarkan pada **Tabel 6** dan **Tabel 7** dapat diketahui perbedaan antara hasil analisis butir soal pada aspek tingkat kesukaran melalui pendekatan teori tes klasik dan Rasch Model. Hasil analisis kedua table tersebut menunjukkan bahwa pada Teori Tes Klasik Sebagian besar soal termasuk kategori sedang, 6 berkategori sukar, dan 3 termasuk mudah.

Hasil yang berbeda didapatkan setelah menganalisis melalui pendekatan Rasch Model. Tingkat kesulitan butir soal dinilai dari nilai measure pada satuan logit tiap butir soal. **Tabel 7** menjelaskan bahwa nilai measure (logit) item dikelompokkan pada kategori sangat mudah, mudah, sulit, dan sangat sulit terdistribusi secara merata.

SUMMARY OF 36 MEASURED Person									
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		
					MNSQ	ZSTD	MNSQ	ZSTD	
MEAN	10.5	25.0	-.45	.54	1.01	.02	.96	-.06	
SEM	1.0	.0	.25	.02	.04	.19	.07	.17	
P.SD	5.8	.0	1.49	.13	.24	1.12	.41	1.03	
S.SD	5.9	.0	1.52	.13	.25	1.14	.41	1.05	
MAX.	24.0	25.0	3.60	1.03	1.51	2.38	2.01	1.68	
MIN.	2.0	25.0	-2.95	.45	.60	-2.49	.28	-2.26	
REAL RMSE	.58	TRUE SD	1.38	SEPARATION	2.36	Person	RELIABILITY	.85	
MODEL RMSE	.56	TRUE SD	1.39	SEPARATION	2.50	Person	RELIABILITY	.86	
S.E. OF Person MEAN = .25									
Person RAW SCORE-TO-MEASURE CORRELATION = .99									
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .88 SEM = 2.02									
SUMMARY OF 25 MEASURED Item									
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		
					MNSQ	ZSTD	MNSQ	ZSTD	
MEAN	15.1	36.0	.00	.43	.99	.04	.96	.09	
SEM	1.3	.0	.22	.01	.03	.15	.05	.11	
P.SD	6.4	.0	1.09	.03	.15	.74	.26	.53	
S.SD	6.6	.0	1.11	.03	.15	.76	.26	.54	
MAX.	28.0	36.0	1.47	.49	1.28	1.44	1.43	1.23	
MIN.	7.0	36.0	-2.19	.39	.71	-1.28	.49	-1.05	
REAL RMSE	.44	TRUE SD	1.00	SEPARATION	2.26	Item	RELIABILITY	.84	
MODEL RMSE	.43	TRUE SD	1.00	SEPARATION	2.33	Item	RELIABILITY	.84	
S.E. OF Item MEAN = .22									

Gambar 1. Hasil *summary statistic* 25 butir soal Rasch Model

Daya Pembeda Soal

Daya pembeda soal merupakan pengujian kemampuan suatu soal dalam membedakan siswa yang memiliki kemampuan akademik tingkat tinggi atau mampu menjawab soal dengan siswa yang mempunyai kemampuan tingkat rendah. Analisis butir soal ini dapat dapat membedakan siswa yang pandai dan kurang pandai (Magdalena et al., 2021). Pada teori tes klasik analisis daya pembeda dikategorikan pada Tabel 8 (Isnaeni et al., 2022).

Tabel 8. Kategori Daya Pembeda Soal dengan Teori Tes Klasik

Daya Pembeda (DP)	Interpretasi
$DP \geq 0,70$	Baik Sekali
$0,40 \leq DP < 0,70$	Baik
$0,20 \leq DP < 0,40$	Cukup
$DP < 0,20$	Buruk

(Jumini, et al., 2023)

Berdasarkan hasil dari analisis melalui pendekatan Teori Tes Klasik diperoleh 19 soal berkategori baik dan 6 soal termasuk kategori cukup. Analisis menggunakan pendekatan ini menunjukkan bahwa Sebagian besar soal sudah termasuk memiliki daya beda yang baik sehingga dapat membedakan siswa yang memiliki tingkat kemampuan tinggi dengan rendah.

Berbeda dengan Rasch Model, analisis pada pendekatan ini menggunakan analisis pada tingkat kemampuan individu sebagai alat yang dapat menganalisis perbedaan kompetensi dari peserta didik yang tinggi yaitu mampu menjawab soal dan yang rendah yaitu tidak mampu menjawab soal (Priyayi et al., 2022). Cara yang kedua juga dapat menggunakan identifikasi kelompok responden yang berdasarkan indeks separasi responden. Semakin besar nilai separasi item maka memiliki kualitas alat ukur dalam hal keseluruhan responden dan butir soal semakin bagus, dikarenakan mampu mengidentifikasi kelompok responden dan butir soal (Erfan et al., 2020).

Persamaan lain untuk mengelompokkan lebih teliti pada pemodelan Rasch menggunakan persamaan strata (H) (Kurli at al., 2022):

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

Hasil dari perhitungan nilai persamaan strata (H) pada separasi butir soal sebesar 2,26 memperoleh nilai $H = 3,34$, yang berarti terdapat 3 kelompok butir soal. Sedangkan nilai separasi responden yaitu 2,36 yang memperoleh $H = 3,48$ menunjukkan bahwa responden dibedakan menjadi 4 kelompok. Hal ini mengindikasikan bahwa butir soal yang dikembangkan memiliki kemampuan yang baik dalam membedakan tingkat kesukaran antarsoal. Selain itu, tingginya nilai separasi responden mencerminkan bahwa instrumen yang digunakan cukup sensitif dalam membedakan tingkat kemampuan siswa secara jelas dan terstruktur.

KESIMPULAN

Hasil analisis butir soal instrumen kemampuan berpikir kritis materi pencemaran lingkungan yang dianalisis menggunakan pendekatan Teori Tes Klasik dan Rasch Model menunjukkan hasil yang berbeda. Pada teori Tes Klasik dihasilkan 25 soal valid sedangkan pada Rasch Model hanya 24 soal tergolong valid dan 1 soal tidak valid. Analisis reliabilitas melalui Teori Tes Klasik diperoleh nilai Alpha Cronbach sebesar 0,87 dengan kategori tinggi dan Rasch Model yang dilihat dari Person Reliability sebesar 0,85 dan Item Reliability 0,84 tergolong pada kategori bagus. Tingkat kesukaran butir soal pada Teori Tes Klasik dihasilkan 6 soal sukar, 16 soal sedang, dan 3 soal mudah. Berbeda dengan Rasch Model yang lebih merata diperoleh hasil 4 soal sangat sulit, 8 soal sulit, 7 soal mudah, dan 5 soal sangat mudah. Aspek terakhir yaitu daya beda diperoleh hasil 19 butir soal baik dan 6 soal cukup pada pendekatan Teori Tes Klasik, sedangkan pada Rasch Model diperoleh tiga kelompok butir soal berdasarkan pada indeks separasi butir soal $H = 3,34$.

DAFTAR PUSTAKA

- Andini, N. A., & Saifuddin, M. F. (2023). Learning media of digital booklets for class VIII junior high school. Biosfer: Jurnal Pendidikan Biologi, 16(1), 1-12. <https://doi.org/10.21009/biosferjpb.25457>
- Amalina, I. K., & Vidákovich, T. (2023). Assessment of domain-specific prior knowledge: A development and validation of mathematical problem-solving test. *International Journal of Evaluation and Research in Education*, 12(1), 468-476.
- Arikunto. (2006). *Prosedur Penelitian Suatu Pendekatan Praktek*. Jakarta: PT Rineka Cipta.
- Azizah, & Wahyuningsih, S. (2020). Penggunaan Model Rasch untuk Analisis Instrumen Tes pada Mata Kuliah Matematika Aktuaria. *Jupitek: Jurnal Pendidikan Matematika*, 3(1), 45-50. <https://doi.org/10.30598/jupitekv3iss1ppx45-50>
- Ennis, R. H. (2011). *The Nature of Critical Thinking: An Outline of Critical Thinking Dispostions and Abilities*. Chicago: University of Illinois.
- Erfan, M., Maulyda, 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. <https://doi.org/10.23887/ijerr.v3i1.24080>

- Febriana, R., Feranie, S., & Kaniawati, I. (2022, December). Profile analysis of student's concept mastery using Rasch model on Newton's law subject. *In AIP Conference Proceedings (Vol. 2468, No. 1). AIP Publishing*.
- Hayati, S., & Lailatussaadah, L. (2016). Validitas Dan Reliabilitas Instrumen Pengetahuan Pembelajaran Aktif, Kreatif Dan Menyenangkan (PAKEM) Menggunakan Model Rasch. *Jurnal Ilmiah Didaktika*, 16(2), 169. <https://doi.org/10.22373/jid.v16i2.593>
- Ida, F. F., & Musyarofah, A. (2021). Validitas dan Reliabilitas dalam Analisis Butir Soal. *Al-Mu'arrib: Jurnal Pendidikan Bahasa Arab*, 1(1), 34-44.
- Isnaeni, B., Aprilia, N., & Saifuddin, M. F. (2022). Rasch model: Quality of final semester assessment items for class x on biology subject. *Biosfer: Jurnal Pendidikan Biologi*, 15(2), 355-367.
- Jumini, et al. (2023). Analisis Kualitas Butir Soal Pengukuran Literasi Sains Melalui Teori Tes Klasik dan Rasch Model. *Prosiding Seminar Nasional Pascasarjana*, 6 (1): 758-765
- Khalid, N. H. M., Yusof, I. J., Latif, A. A., & Jani, M. D. M. (2023). Psychometric properties of teacher classroom assessment literacy instrument using Rasch model analysis. *Int J Eval & Res Educ ISSN*, 2252(8822), 8822.
- Krebs, R. E., Rost, M., & Lembens, A. (2023). Developing and evaluating a multiple-choice knowledge test about Brønsted-Lowry acid-base reactions for upper secondary school students. *Chemistry Teacher International*, 5(2), 177-188. <https://doi.org/10.1515/cti-2022-0038>
- Kurli, H., Haliqah, N., & Anekawati, A. (2021). Analisis Kemampuan Kognitif Siswa pada Mata Pelajaran IPA SMP Menggunakan Rasch Model. *Prosiding SNAPP: Sosial Humaniora, Pertanian, Kesehatan dan Teknologi*, 1(1), 74-79.
- Magdalena, I., Anggraini, I. A., & Khoiriah, S. (2021). Analisis Daya Pembeda dan Taraf Kesukaran pada Soal Bilangan Romawi Kelas 4 SDN Tobat 1 Balaraja. *NUSANTARA*, 3(1), 151-158. <https://doi.org/10.36088/nusantara.v3i1.1284>
- Magdalena, I., Fauziah, S. N., Faziah, S. N., & Nupus, F. S. (2021). Analisis Validitas, Reliabilitas, Tingkat Kesulitan dan Daya Beda Butir Soal Ujian Akhir Semester Tema 7 Kelas III SDN Karet 1 Sepatan. *Jurnal Pendidikan Sains*, 3(2): 198-214. <https://doi.org/10.36088/bintang.v3i2.1291>
- Mahamud, S., & Othman, N. (2020). Kesahan dan Kebolehpercayaan Instrumen Kecerdasan Menghadapi Cabaran Teknologi Menggunakan Model Rasch (Validity and Reliability the Instruments Adversity Quotient for Technological Challenges Using the Rasch Model). *Akademika*, 90(S3). <https://doi.org/10.17576/akad-2020-90IK3-03>
- Murias Quintana, E., Rodríguez Castro, J., Sánchez Lasheras, F., Vega Villar, J., Curbelo García, J. J., Cadenas Rodríguez, M., & Baladrón Romero, J. (2024). Improving the ability to discriminate medical multiple-choice questions through the analysis of the competitive examination to assign residency positions in Spain. *BMC Medical Education*, 24(1), 367. <https://doi.org/10.1186/s12909-024-05324-2>

-
- Nasution, I., Harahap, H. T., Nurfadillah, L., & Purba, S. L. B. (2022). Evaluasi Program Ekstrakurikuler pada Sekolah MIS Nur Al Amin Medan. *Jurnal Pendidikan dan Konseling (JPDK)*, 4(3), 1638-1646. <https://doi.org/10.31004/jpdk.v4i3.4931>
- Novia, R., Ramalis, T. R., & Efendi, R. (2019). Pengembangan dan Karakterisasi Tes Keterampilan Berpikir Kritis Materi Tekanan berdasarkan Teori Respon Butir. *WaPFI (Wahana Pendidikan Fisika)*, 4(2), 155-162. <https://doi.org/10.17509/wapfi.v4i2.20181>.
- Nuryanti, S., Masykuri, M., & Susilowati, E. (2018). Analisis Iteman dan Model Rasch pada Pengembangan Instrumen Kemampuan Berpikir Kritis Peserta Didik Sekolah Menengah Kejuruan. *Jurnal Inovasi Pendidikan IPA*, 4(2), 224-233. <https://doi.org/10.21831/jipi.v4i2.21442>
- Polat, M., Turhan, N. S., & Toraman, Ç. (2022). Comparison of Classical Test Theory vs. Multi-Facet Rasch Theory in Writing Assessment. *Pegem Journal of Education and Instruction*, 12(2), 213-225. <https://doi.org/10.47750/pegegog.12.02.21>
- Priyayi, D. F., Upa, D. R., & Hastuti, S. P. (2022). Keterampilan Proses Sains (KPS) Siswa Melalui Asesmen Daring Menggunakan Analisis Model Rasch. *BIOEDUSAINS: Jurnal Pendidikan Biologi dan Sains*, 5(2), 330-341. <https://doi.org/10.22219/raden.v3i1>
- Purniasari, L., Masykuri, M., & Ariani, S. R. D. (2021). Analisis Butir Soal Ujian Sekolah Mata Pelajaran Kimia SMA N 1 Kutowinangun Tahun Pelajaran 2019/2020 Menggunakan Model Iteman Dan Rasch. *Jurnal Pendidikan Kimia*, 10(2), 205-214. <https://doi.org/10.20961/jpkim.v10i2.48244>
- Rahayu, R., & Djazari, M. (2016). Analisis Kualitas Soal Pra Ujian Nasional Mata Pelajaran Ekonomi Akuntansi. *Jurnal Pendidikan Akuntansi Indonesia*, 14(1), 85-94. <https://doi.org/10.21831/jpai.v14i1.11370>
- Ramadhini, F., & Nasution, L. A. A. (2022). Pengukuran Kesiapan Sekolah: Analisis Empirik Berdasar Teori Tes Klasik. *BUHUTS AL ATHFAL: Jurnal Pendidikan dan Anak Usia Dini*, 2(1), 16-35. <https://doi.org/10.24952/alathfal.v2i1.5274>
- Ramadhani, R., Syahputra, E., dan Simamora, E. (2024). The construct validity of self-regulated learning questionnaire for senior high school students: a Rasch model analysis. *International Journal of Evaluation and Research in Education*, 13 (3): 2023-2038.
- Ramli, M., & Utami, B. (2023). Analisis Instrumen Order Multiple Choice Konsep Nutrisi Berbasis Learning Progression Menggunakan Model Rasch. *Bioedusiana: Jurnal Pendidikan Biologi*, 8(2). <https://doi.org/10.37058/bioed.v8i2.10096>
- Rashwan, N. I., Aref, S. R., Nayel, O. A., & Rizk, M. H. (2024). Postexamination item analysis of undergraduate pediatric multiple-choice questions exam: implications for developing a validated question Bank. *BMC Medical Education*, 24(1), 168. <https://doi.org/10.1186/s12909-024-05153-3>
- Sabekti, A. W., & Khoirunnisa, F. (2018). Penggunaan Rasch Model untuk Mengembangkan Instrumen Pengukuran Kemampuan Berikir Kritis Siswa pada Topik Ikatan Kimia. *Jurnal Zarah*, 6(2), 68-75. <https://doi.org/10.31629/zarah.v6i2.724>
-

- Santosa, S., & Badawi, J. A. (2022). Analisis Butir Soal Pilihan Ganda Tema Pertumbuhan dan Perkembangan Makhluk Hidup Kelas III Madrasah Ibtidaiyah. *Jurnal Basicedu*, 6(2), 1678-1686. <https://doi.org/10.31004/basicedu.v6i2.2206>
- Sibarani, R. A. M., Afandi, A., Tenriawaru, A. B., & Bilanti, E. (2023, January). Comparison of TOSLS-B instrument analysis result using Iteman and Rasch model. In *AIP Conference Proceedings* (Vol. 2540, No. 1). AIP Publishing. <https://doi.org/10.1063/5.0105736>
- Susdelina, Perdana, S. A., & Febrian. (2018). Analisis Kualitas Instrumen Pengukuran Pemahaman Konsep Persamaan Kuadrat Melalui Teori Tes Klasik Dan Rasch Model. *Jurnal Kiprah*, 1(1), 41â€“48. <https://doi.org/10.31629/kiprah.v6i1.574>
- Susilaning Sih, et al. (2021). Dissemination of test instruments as product of the development research to measure the problem-solving ability of class X students by online in the pandemic period. *Journal of Physics: Conference Series*, 1918(3):023024
- 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), 92â€“96. <https://doi.org/10.15294/jipk.v16i2.30530>
- Tarmizi, P., Setiono, P., Amaliyah, Y., & Agrian, A. (2020). Analisis butir soal pilihan ganda tema sehat itu penting kelas V SD Negeri 04 Kota Bengkulu. *ELSE (Elementary School Education Journal): Jurnal Pendidikan dan Pembelajaran Sekolah Dasar*, 4(2), 124-132. <https://doi.org/10.30651/else.v4i2.7090>
- Wolfs, Z. G., Brand-Gruwel, S., & Boshuizen, H. P. (2023). Assessing Tonal Abilities in Elementary School Children: Testing Reliability and Validity of the Implicit Tonal Ability Test Using Rasch Measurement Model. *SAGE Open*, 13(3), 21582440231199041.
- Yulawati, D. D., Bintang, K. A., Fitrotul, R., Gheafitri, Z., & Hanatan, A. (2022). Analisis Kemampuan Berpikir Kritis Peserta Didik dengan Penilaian Tes dan Non Tes. *Jurnal Pembelajaran Fisika*, 11(2), 65-68.