

## PENGEMBANGAN INSTRUMEN PENILAIAN KETERAMPILAN BERNALAR KRITIS MATERI PEMBANGUNAN BERKELANJUTAN DAN KELANGKAAN BAGI SISWA SMP

**Azmho M. Sihombing<sup>1</sup>, Wasino<sup>2</sup>, Supriyadi<sup>3</sup>**

<sup>1</sup>Sekolah Pascasarjana, Universitas Negeri Semarang

<sup>2</sup>Sekolah Pascasarjana, Universitas Negeri Semarang

<sup>3</sup>Sekolah Pascasarjana, Universitas Negeri Semarang

Alamat [email: 1azmho.sihombing@gmail.com](mailto:1azmho.sihombing@gmail.com), [2wasino@mail.unnes.ac.id](mailto:2wasino@mail.unnes.ac.id),  
[3supriyadi@mail.unnes.ac.id](mailto:3supriyadi@mail.unnes.ac.id),

**Abstract:** *This study aims to develop a critical reasoning skills assessment instrument on sustainable development and scarcity materials for junior high school students. The method used was development research with the ADDIE model. The research subjects were seventh grade students of SMP Negeri 2 Lumbanjulu and SMP Negeri 2 Satu Atap Sibisa. The content validity of the instrument was analyzed using Aiken's V obtained a value of  $V > 0.79$  indicating validity in the aspects of content suitability, construction and language. Construct validity obtained an eigenvalue of 41.7% which was categorized as good. Empirical validity showed that 28 items met the criteria with an outfit mean square value  $< 1.2$  and 2 items did not meet the criteria with an outfit mean square value  $> 1.2$ . Rasch modeling analysis shows the item reliability value of 0.86 and is categorized as good and the person reliability value is 0.81 including in the good category. The level of difficulty of the question is divided into 4 namely very difficult (7.14%), difficult (39.29%), easy (46.43%), and very easy (7.14%). Analysis of the practicality of the instrument obtained an average value of 91.25% which was categorized as very practical.*

**Keyword:** *Critical Reasoning, Assessment Instrument, Sustainable Development and Scarcity*

**Abstrak:** Penelitian ini bertujuan untuk mengembangkan instrumen penilaian keterampilan bernalar kritis pada materi pembangunan berkelanjutan dan kelangkaan untuk siswa SMP. Metode yang digunakan adalah penelitian pengembangan dengan model ADDIE. Subjek penelitian adalah siswa kelas VII SMP Negeri 2 Lumbanjulu dan SMP Negeri 2 Satu Atap Sibisa. Validitas isi instrumen dianalisis menggunakan Aiken's V diperoleh nilai  $V > 0,79$  menunjukkan kevalidan dalam aspek kesesuaian isi, konstruksi dan bahasa. Validitas konstruk memperoleh nilai *eigenvalue* sebesar 41,7% yang dikategorikan baik. Validitas empiris menunjukkan bahwa 28 butir soal memenuhi kriteria dengan nilai *outfit mean square*  $< 1,2$  dan 2 butir soal tidak memenuhi kriteria dengan nilai *outfit mean square*  $> 1,2$ . Analisis pemodelan Rasch menunjukkan nilai reliabilitas item sebesar 0,86 dan dikategorikan baik dan nilai *person reliability* adalah 0,81 termasuk dalam kategori baik. Tingkat kesukaran soal terbagi menjadi 4 yaitu sangat sukar (7,14%), sukar (39,29%), mudah (46,43%), dan sangat mudah (7,14 %). Analisis kepraktisan instrumen memperoleh nilai rata-rata 91,25 % yang dikategorikan sangat praktis.

**Kata Kunci:** Bernalar Kritis, Instrumen Penilaian, Pembangunan Berkelanjutan dan Kelangkaan

## **A. Pendahuluan**

Asesmen merupakan elemen krusial dalam pembelajaran yang bertujuan mengevaluasi capaian kompetensi siswa. Menurut (Kemendikbudristek RI, 2022) penilaian merupakan suatu proses sistematis dalam mengumpulkan dan menganalisis data untuk mengukur tingkat pencapaian pembelajaran serta mengidentifikasi kebutuhan belajar peserta didik. Penilaian merupakan pernyataan untuk mengungkapkan karakteristik seseorang berdasarkan fakta. Proses penilaian dilakukan untuk memantau seberapa jauh siswa menguasai tujuan pembelajaran yang telah ditetapkan (Griffin & Nix, 1991). Kegiatan menilai merupakan bagian dari tanggung jawab yang wajib dilakukan, baik selama proses pembelajaran berlangsung maupun setelahnya. Kualitas pembelajaran sangat bergantung pada kualitas instrumen penilaian yang digunakan. Guru berperan penting dalam merancang instrumen yang mampu mengukur kemampuan peserta didik secara menyeluruh.

Kurikulum Merdeka belajar sebagai kurikulum nasional saat ini menetapkan 6 dimensi sebagai karakteristik Pelajar Pancasila, salah satunya adalah bernalar kritis. Pelajar dengan kemampuan bernalar kritis adalah individu yang mampu memproses informasi dan gagasan secara mendalam, baik yang bersifat kualitatif maupun kuantitatif. Pelajar dengan keterampilan bernalar kritis mampu menjelaskan gagasan atau informasi yang diperolehnya, serta memilah isi informasi dari pihak yang menyampaikan, memiliki keberanian untuk mencari data atau fakta yang mungkin bertentangan dengan keyakinan pribadinya, dan dapat mengambil keputusan yang tepat. Pengambilan keputusan oleh Pelajar Pancasila

mengandalkan penalaran berbasis logika dan kaidah ilmiah, disertai analisis dan evaluasi terhadap gagasan yang diterima. Bernalar kritis artinya mampu membuktikan argumennya, dan menarik kesimpulan. Adanya keterampilan bernalar kritis membuat peserta didik terbuka untuk mengubah opini pribadi jika berbeda dengan bukti yang ditemukan, sehingga kapasitasnya berkembang melalui refleksi dan usaha perbaikan berkelanjutan (Kemdikburistek RI, 2022, p. 30) Pengembangan kemampuan bernalar kritis pada peserta didik merupakan suatu keharusan dalam rangka mempersiapkan generasi muda untuk menghadapi dinamika kehidupan global. Kemampuan berpikir kritis memungkinkan seseorang mengambil keputusan yang lebih baik dengan analisis yang teliti, terstruktur, rasional, serta mempertimbangkan berbagai perspektif (Setiawati & Duran Corebima, 2017, p. 3521). Membiasakan diri berpikir kritis sejak remaja akan membantu peserta didik menggali potensi diri, memecahkan masalah dengan efektif, dan meraih kesuksesan dalam belajar. Keterampilan bernalar kritis akan membantu peserta didik dalam menyelesaikan berbagai tantangan dan masalah (Ernawati & Rahmawati, 2022, p. 6133). Tujuan pembelajaran HOTS adalah untuk menciptakan siswa yang mampu berpikir secara mendalam dan analitis, bukan hanya sekadar mengingat fakta (Nofrion & Wijayanto, 2018, p. 124). Keterampilan bernalar kritis membekali pelajar dengan kemampuan mengkaji dan menyelesaikan masalah kompleks, memungkinkan mereka tidak hanya memahami konsep ilmiah, tetapi juga menerapkannya dalam konteks nyata.

Belajar adalah proses berkelanjutan yang bertujuan membangun pemahaman yang kompleks. Menurut teori kognitif, belajar

merupakan aktivitas aktif di mana individu mengembangkan pengetahuan baru dengan menghubungkannya pada pengalaman dan pemahaman sebelumnya. Proses ini dapat diibaratkan seperti menyusun puzzle, di mana setiap informasi baru perlu selaras dengan potongan yang sudah ada untuk menciptakan gambaran yang menyeluruh.

Pentingnya pembelajaran IPS untuk peserta didik dikarenakan mereka akan menghadapi tantangan global terkait isu lingkungan dan sosial ekonomi yang terus berubah (Aulia & Wandini, 2023, pp. 4034–4035). Ilmu pengetahuan sosial berguna untuk memberikan pengalaman kepada peserta didik sehingga mudah dalam memahami fenomena sosial, bernalar kritis, memecahkan masalah, serta mengambil keputusan yang bertanggung jawab dalam kehidupan bermasyarakat. Salah satu materi pembahasan IPS adalah pembangunan berkelanjutan dan kelangkaan. Menurut (Sya'ban, 2018, p. 34) mengajarkan materi Pembangunan berkelanjutan dan kelangkaan memiliki dampak positif kepada peserta didik supaya mampu memanfaatkan sumberdaya alam dengan efektif dan peka terhadap isu lingkungan. Materi kelangkaan menjadi dasar ilmu yang perlu diberikan kepada peserta didik untuk memahami masalah ekonomi secara umum.

Peneliti melakukan observasi langsung di lingkungan sekolah dan menemukan bahwa sebanyak 80% sampah yang ada di lingkungan sekolah adalah kertas. Lembar soal dan lembar jawaban peserta didik yang sudah tidak digunakan lagi menjadi yang paling dominan jumlahnya. Setelah kertas selesai digunakan, dilakukan pembakaran di bak sampah yang menimbulkan pencemaran udara. Isu pemanasan global dan cuaca ekstrim perlu disikapi dengan serius oleh

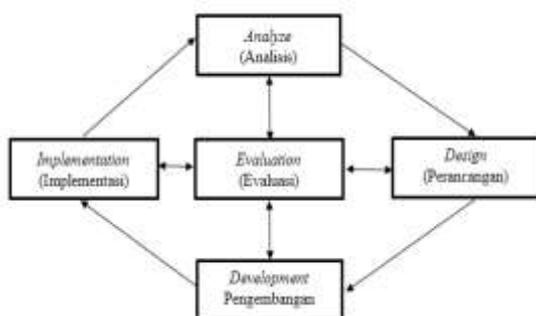
semua elemen masyarakat. Sekolah dapat memberikan partisipasi melalui pembelajaran yang menumbuhkan kesadaran peserta didik terhadap kondisi lingkungan sekitarnya. Menurut (Reffhaug & Lysgaard, 2024, p. 1519) kurikulum pendidikan di Norwegia saat ini mengutamakan pembangunan berkelanjutan dan pengasahan kemampuan berpikir kritis. Konsep menjaga lingkungan dan menciptakan masyarakat yang lebih baik diintegrasikan ke dalam semua mata pelajaran, mulai dari tingkat dasar hingga menengah. Berpikir kritis tidak hanya melibatkan kemampuan kognitif, tetapi juga dimensi afektif dan psikomotorik. Penelitian memberikan suatu rekomendasi bahwa kolaborasi diperlukan dalam mengatasi kompleksitas permasalahan lingkungan dan sosial.

Berdasarkan observasi langsung, peneliti menemukan buku IPS Kurikulum Merdeka belum menyediakan instrumen penilaian yang memadai dan mampu mengukur keterampilan bernalar kritis peserta didik. Instrumen penilaian berbentuk pilihan ganda yang secara spesifik membahas tentang Pembangunan berkelanjutan dan kelangkaan hanya sebanyak 2 butir soal. Jumlah ini tentu sangat sedikit dibandingkan dengan materi pembangunan berkelanjutan dan kelangkaan yang banyak. Setelah melakukan observasi terhadap buku teks, peneliti melakukan wawancara kepada guru IPS di SMP Negeri 2 Lumbanjulu. Guru menyatakan mengalami kesulitan dalam mengembangkan instrumen penilaian keterampilan bernalar kritis materi Pembangunan berkelanjutan dan kelangkaan. Kesulitan guru dalam mengembangkan instrumen didasari karena penerapan kurikulum Merdeka Belajar baru dimulai sejak tahun ajaran 2024/2025. Guru juga masih belum memahami secara penuh elemen-

elemen yang terkandung dalam Dimensi Bernalar Kritis sebagai bagian dari Profil Pelajar Pancasila. Guru IPS lainnya menyatakan keterampilan bernalar kritis merupakan istilah baru baginya, walaupun dalam penjabarannya bernalar kritis hampir sama dengan berpikir kritis. Guru mengaku lebih banyak menggunakan instrumen yang tersedia dalam buku paket guru dan peserta didik sebagai alat untuk mengukur keterampilan bernalar kritis. Keterbatasan waktu merupakan kendala yang sering dialami oleh guru dalam penyusunan instrumen penilaian keterampilan bernalar kritis yang secara holistik. Guru tidak memiliki panduan yang praktis dalam menuntun pengembangan suatu instrumen penilaian keterampilan bernalar kritis peserta didik materi pembangunan berkelanjutan dan kelangkaan. Respon guru terhadap ide dan inovasi ini juga positif karena dapat mempersingkat waktu dan meningkatkan efektivitas penilaian.

### **B. Metode Penelitian**

Penelitian ini menggunakan pendekatan *Research and Development (R&D)*, yaitu suatu metodologi yang bertujuan untuk menghasilkan produk tertentu sekaligus menguji tingkat keefektifannya (Sugiyono, 2013, pp. 297–298). Pengembangan instrumen dalam penelitian ini menggunakan model ADDIE seperti gambar berikut ini:



Gambar 2.1 Alur Penelitian Pengembangan Model ADDIE

Gambar 2.1 menunjukkan bahwa penelitian pengembangan model ADDIE menekankan evaluasi pada setiap tahap pengembangan yang bertujuan untuk memastikan kekurangan dan kelemahan produk yang dikembangkan dapat diperbaiki sebelum dilakukan langkah pengembangan selanjutnya. Subjek penelitian adalah 30 orang peserta didik kelas VII SMP Negeri 2 Lumbanjulu untuk uji skala kecil dan 100 orang peserta didik kelas VII SMP Negeri 2 Lumbanjulu dan SMP Negeri 2 Satu Atap Sibisa untuk uji coba skala besar. Alat pengumpulan data yang digunakan meliputi tes objektif berbentuk pilihan ganda, panduan wawancara, angket, dan studi dokumen. Pengolahan data dilakukan dengan pendekatan kualitatif dan kuantitatif.

### **C. Hasil Penelitian dan Pembahasan Validitas Instrumen**

Validitas isi dalam penelitian ini diperoleh melalui penilaian ahli (*expert judgment*) terhadap butir-butir instrumen yang dikembangkan. Penilaian dilakukan oleh 6 orang validator yang terdiri dari dosen ahli instrumen, evaluasi dan guru sebagai ahli praktisi dan materi pembelajaran IPS. Prosedur penilaian menggunakan lembar validasi yang dilengkapi dengan skala penilaian serta ruang untuk saran perbaikan. Skor validitas isi kemudian dianalisis dengan formula *Aiken's V* untuk mengetahui sejauh mana tingkat kesepakatan para ahli terhadap kelayakan isi instrumen yang dikembangkan. Validitas isi instrumen terdiri dari aspek kesesuaian materi, aspek konstruk dan aspek bahasa.

Tabel 3.1 Analisis Validitas Isi Instrumen

No. Butir	Aspek Isi		Aspek Konstruksi		Aspek Bahasa		Ket
	Nilai V	Kategori Validitas	Nilai V	Kategori Validitas	Nilai V	Kategori Validitas	
1	0,7916	Baik	0,75	Kurang	0,875	Baik	Direvisi
2	0,975	Baik	0,8611	Baik	0,902	Baik	Dipakai
3	0,7604	Kurang	0,7638	Kurang	0,8055	Baik	Direvisi
4	0,8541	Baik	0,8888	Baik	0,8611	Baik	Dipakai
5	0,9375	Baik	0,9583	Baik	0,9166	Baik	Dipakai
6	0,8333	Baik	0,875	Baik	0,861	Baik	Dipakai
7	0,8645	Baik	0,8888	Baik	0,8333	Baik	Dipakai
8	0,7708	Kurang	0,7916	Kurang	0,9166	Baik	Direvisi
9	0,9375	Baik	0,9722	Baik	0,9027	Baik	Dipakai
10	0,9062	Baik	0,9166	Baik	0,8472	Baik	Dipakai
11	0,7604	Kurang	0,8611	Baik	0,8333	Baik	Dipakai
12	0,9375	Baik	0,9166	Baik	0,9166	Baik	Dipakai
13	0,9479	Baik	0,9444	Baik	0,9583	Baik	Dipakai
14	0,9479	Baik	0,9166	Baik	0,9583	Baik	Dipakai
15	0,8958	Baik	0,9166	Baik	0,9166	Baik	Dipakai
16	0,9375	Baik	0,9027	Baik	0,9583	Baik	Dipakai
17	0,9479	Baik	0,9588	Baik	0,9444	Baik	Dipakai
18	0,7812	Kurang	0,7638	Kurang	0,8888	Baik	Direvisi
19	0,8541	Baik	0,9583	Baik	0,8888	Baik	Dipakai
20	0,6875	Kurang	0,8333	Baik	0,75	Kurang	Direvisi
21	0,9533	Baik	0,9722	Baik	0,9305	Baik	Dipakai
22	0,8854	Baik	0,9027	Baik	0,8888	Baik	Dipakai
23	0,9479	Baik	0,9722	Baik	0,9305	Baik	Dipakai
24	0,9062	Baik	0,9444	Baik	0,9444	Baik	Dipakai
25	0,9270	Baik	0,9722	Baik	0,9444	Baik	Dipakai
26	0,9375	Baik	0,9583	Baik	0,8888	Baik	Dipakai
27	0,9270	Baik	0,9444	Baik	0,9027	Baik	Dipakai
28	0,9270	Baik	0,9583	Baik	0,8888	Baik	Dipakai
29	0,8958	Baik	0,9583	Baik	0,9444	Baik	Dipakai
30	0,8958	Baik	0,9722	Baik	0,9305	Baik	Dipakai

Berdasarkan tabel 3.1, instrumen yang divalidasi oleh 6 orang ahli, menunjukkan 4 soal memiliki nilai *Aiken's V* dibawah nilai kritis yang sudah ditetapkan. Soal nomor 1, 3, 18, 20 perlu dilakukan revisi dalam aspek kesesuaian isi, konstruksi dan bahasa. Berdasarkan penilaian validator instrumen harus diperbaiki untuk memperjelas tujuan dan indikator yang ingin diukur dan menyesuaikan setiap butir soal berdasarkan kisi-kisi, materi dan indikator.

Validitas konstruk mengacu pada kecermatan suatu tes dalam mengukur secara menyeluruh suatu variabel laten atau kemampuan tertentu yang menjadi fokus penilaian (Umar & Faella Nisa, 2020). Validitas konstruk digunakan untuk mengonfirmasi apakah instrumen penelitian mengukur konsep yang ingin diteliti secara tepat (Maulana et al., 2023, p. 751) Hasil analisis

validitas konstruk instrumen pada gambar berikut ini:

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

	Eigenvalue	Observed	Expected
Total raw variance in observations =	50.2077	100.0%	100.0%
Raw variance explained by measures =	20.2077	40.2%	41.7%
Raw variance explained by persons =	7.3542	14.6%	15.2%
Raw Variance explained by items =	12.8535	25.6%	26.5%
Raw unexplained variance (total) =	30.0000	59.8%	100.0%
Unexplained variance in 1st contrast =	3.8755	7.7%	12.9%
Unexplained variance in 2nd contrast =	3.2612	6.5%	10.9%
Unexplained variance in 3rd contrast =	2.9802	5.9%	9.9%
Unexplained variance in 4th contrast =	2.5363	5.1%	8.5%
Unexplained variance in 5th contrast =	2.4298	4.8%	8.1%

Gambar 3.1 Uji Validitas Konstruk

Berdasarkan gambar 3.1, menunjukkan bahwa nilai *eigenvalue* adalah 20.2077 (41,7%) artinya validitas konstruk dalam kategori baik. Data ini mengindikasikan bahwa instrumen yang dikembangkan tepat dalam mengukur keterampilan bernalar kritis.

Validitas empiris merupakan validitas yang didasarkan pada hubungan statistik antara skor tes dengan kriteria eksternal terkait, menunjukkan sejauh mana instrumen tersebut terbukti akurat dalam mengukur variabel target berdasarkan bukti lapangan. Validitas empiris dapat diperoleh dari hasil uji coba skala besar kepada responden. Validitas empiris suatu instrumen dapat dilakukan menggunakan teori respon butir (Sumintono & Widhiarso, 2015, pp. 12–47). Hasil uji validitas empiris dituangkan pada tabel 3.2 berikut ini:

Tabel 3.2 Analisis Validitas Empiris Instrumen

Nomor Butir	Nilai Outfit MNSQ	Nilai Outfit ZSTD	Nilai Pt. Measure Correlation	Kategori
1.	0,89	- 0,75	0,40	Valid
2.	0,83	- 0,83	0,39	Valid
3.	0,83	- 0,96	0,43	Valid
4.	0,89	- 0,63	0,40	Valid
5.	1,10	0,76	0,30	Valid
6.	0,87	- 0,85	0,42	Valid
7.	0,98	-0,14	0,42	Valid
8.	0,83	-0,67	0,34	Valid
9.	0,93	-0,52	0,43	Valid
10.	1,30	1,27	0,37	Valid
11.	1,06	0,45	0,34	Valid
12.	1,15	0,99	0,31	Valid
13.	0,95	-0,30	0,45	Valid
14.	0,88	-0,96	0,47	Valid
15.	0,92	-0,56	0,42	Valid

16.	0,79	-1,41	0,48	Valid
17.	1,53	3,50	0,21	Tidak Valid
18.	0,97	-0,13	0,33	Valid
19.	1,02	0,17	0,37	Valid
20.	1,22	1,34	0,29	Valid
21.	1,00	0,05	0,40	Valid
22.	0,91	-0,68	0,46	Valid
23.	0,84	-1,02	0,43	Valid
24.	0,90	-0,71	0,43	Valid
25.	0,93	-0,46	0,40	Valid
26.	1,02	0,18	0,42	Valid
27.	1,23	1,45	0,32	Valid
28.	1,33	2,17	0,26	Tidak Valid
29.	0,95	-0,30	0,39	Valid
30.	0,86	-1,10	0,46	Valid

Menurut (Boone et al., 2014) terdapat 3 kriteria agar item dikatakan valid atau sesuai:

- Nilai Outfit Mean Square ( $0,5 < MNSQ < 1,5$ ).
- Nilai Outfit Z-Standard ( $-2 < ZSTD < 2$ ).
- Nilai Pt-Measure Correlation ( $0,4 < Pt\ Measure\ Corr < 0,85$ ).

Tabel 3.2 menunjukkan bahwa butir soal nomor 17 tidak valid secara empiris dengan nilai MNSQ, ZSTD dan Pt Measure Correlation diluar kriteria yang ditetapkan, sedangkan butir soal nomor 28 juga tidak valid karena nilai ZSTD dan nilai *Pt Mearson Correlation* tidak memenuhi kriteria, namun nilai *MNSQ* memenuhi kriteria yang ditetapkan.

### Reliabilitas Instrumen

Instrumen dapat dinyatakan memiliki reliabilitas apabila hasil pengukurannya stabil dan menunjukkan tingkat keajegan yang tinggi saat dipakai secara berulang (Sari & Mahmudi, 2024, p. 10). Nilai KR-20 yang diperoleh harus  $>0,70$ , sehingga nilai reliabilitasnya dikatakan baik. Semakin tinggi nilai KR-20, maka semakin baik reliabilitasnya (Fraenkel et al., 2012, p. 157). Analisis menggunakan pemodelan Rasch, tingkat reliabilitas butir instrumen diindikasikan melalui nilai item separation (separasi butir). Separasi butir merefleksikan sejauh mana sampel yang diukur terdistribusi secara merata pada skala interval linier. Hasil uji reliabilitas butir instrumen sebagai berikut:

SUMMARY OF 30 MEASURED Item

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MEAN	45.7	100.0	.00	.23	1.00	-.08	1.00	-.02
SEM	2.2	.0	.11	.00	.01	.17	.03	.20
P.SD	12.0	.0	.61	.01	.08	.93	.17	1.00
S.SD	12.3	.0	.62	.01	.08	.94	.17	1.10
MAX.	72.0	100.0	1.39	.28	1.16	2.16	1.53	3.50
MIN.	21.0	100.0	-1.30	.22	.87	-2.12	.79	-1.41

REAL RMSE	.23	TRUE SD	.57	SEPARATION	2.45	Item	RELIABILITY	.86
MODEL RMSE	.23	TRUE SD	.57	SEPARATION	2.50	Item	RELIABILITY	.86
S.E. OF Item MEAN	= .11							

Gambar 3.1 Uji Reliabilitas Item

Berdasarkan tabel 3.1 menginformasikan bahwa instrumen yang dikembangkan memiliki reliabilitas butir (item reliability) yang dikategorikan baik, dengan nilai 0.86. Menurut (Sari & Mahmudi, 2024, p. 8) nilai reliabilitas pada rentang 0,80–0,90 dikategorikan bagus. Nilai ini menunjukkan konsistensi internal yang tinggi, mengindikasikan bahwa butir-butir dalam instrumen mampu mengukur konstruk yang sama secara ajeg.

*Person reliability* dalam pemodelan Rasch mengacu pada konsistensi pengukuran kemampuan responden (person ability) ketika diuji dengan instrumen yang sama. Nilai ini menunjukkan sejauh mana hasil pengukuran dapat dipercaya untuk membedakan antara individu dengan tingkat kemampuan yang berbeda. Hasil uji *person reliability* dapat dilihat pada gambar berikut ini:

SUMMARY OF 100 MEASURED Person

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MEAN	13.7	30.0	-.15	.43	1.00	-.01	1.00	.00
SEM	.6	.0	.10	.01	.01	.09	.02	.09
P.SD	5.7	.0	1.04	.10	.11	.86	.18	.92
S.SD	5.7	.0	1.04	.10	.11	.87	.18	.92
MAX.	29.0	30.0	3.54	1.02	1.31	2.00	1.47	2.18
MIN.	3.0	30.0	-2.34	.30	.73	-2.05	.65	-2.65

REAL RMSE	.45	TRUE SD	.94	SEPARATION	2.10	Person	RELIABILITY	.81
MODEL RMSE	.44	TRUE SD	.94	SEPARATION	2.15	Person	RELIABILITY	.82
S.E. OF Person MEAN	= .10							

Person RAW SCORE-TO-MEASURE CORRELATION = .99  
 CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .81 SEM = 2.40  
 STANDARDIZED (50 ITEM) RELIABILITY = .88

Gambar 3.2 Uji *person reliability*

Nilai *person reliability* dinyatakan dalam skala 0 hingga 1, semakin tinggi nilai reliabilitas maka akan semakin baik (Sumintono & Widhiarso, 2014, p. 112). Hasil analisis pemodelan Rasch menunjukkan bahwa instrumen keterampilan bernalar kritis memiliki *person reliability* yang tergolong baik, dengan nilai 0.81 (REAL) dan 0.82 (MODEL). Nilai *person reliability* yang berada dalam rentang 0,80 – 0,90 dikategorikan dalam reliabilitas yang baik (Sari & Mahmudi, 2024, p. 8). Hasil analisis yang lain didukung nilai *person separation index* sebesar 2.10–2.15, yang mengindikasikan kemampuan instrumen dalam membedakan responden ke dalam minimal tiga kelompok kemampuan (rendah, sedang, tinggi). Nilai *true standard deviation* (SD = 0.94 logit) dan RMSE (0.44–0.45) menunjukkan variasi kemampuan responden yang nyata dengan kesalahan pengukuran yang relatif kecil. Infit dan outfit MNSQ rata-rata 1.00 (ZSTD mendekati 0, mengkonfirmasi bahwa jawaban yang diperoleh dari responden sesuai dengan model Rasch, tanpa indikasi deviasi yang signifikan. Korelasi *raw-score-to-measure* (0.99) dan *Cronbach's Alpha* (0.81) memperkuat konsistensi internal instrumen, sementara *standardized reliability* (0.88) menunjukkan stabilitas yang lebih tinggi ketika disetarakan dengan 50 butir.

Selanjutnya dilakukan uji *Intraclass Correlation Coefficient (ICC)*. *Intraclass Correlation Coefficient* merupakan reliabilitas instrumen yang dianalisis berdasarkan hasil pengamatan yang dilakukan oleh beberapa ahli (*expert judgement*) (Bashooir & Supahar, 2018, p. 224). ICC adalah alat statistik yang esensial untuk menilai reliabilitas pengukuran, terutama dalam penelitian yang melibatkan multi-rater. Tipe reliabilitas ICC yang digunakan adalah *ICC two-way mixed effects model* karena pertimbangan penilai memiliki

karakteristik yang sama (Koo & Li, 2016, p. 3). Hasil analisis ICC sebagai berikut:

	Intraclass Correlation Coefficient			
	Intraclass Correlation <sup>b</sup>	95% Confidence Interval		Value
		Lower Bound	Upper Bound	
Single Measures	.303 <sup>a</sup>	.161	.490	3.607
Average Measures	.723 <sup>c</sup>	.535	.852	3.607

Gambar 3.3 Uji Reliabilitas ICC

Berdasarkan gambar 3.4, reliabilitas antar-ahli dianalisis menggunakan *ICC two-way mixed effects model (Type C, consistency)* dengan skala 1–5. Hasil menunjukkan reliabilitas tunggal (*single measures ICC* = 0.303, 95% CI: 0.161–0.490) yang rendah, tetapi reliabilitas rata-rata (*average measures ICC* = 0.723, 95% CI: 0.535–0.852) tergolong baik ( $p < 0.001$ ). Hasil ini mengindikasikan bahwa meskipun penilaian individu ahli kurang konsisten, agregasi penilaian dari seluruh ahli dapat dijadikan acuan yang andal untuk validasi instrumen.

### Kepraktisan Instrumen

Uji kepraktisan instrumen merupakan bagian dari uji kelayakan instrumen penelitian yang mengevaluasi kemudahan penggunaan, efisiensi waktu, dan kesesuaian instrumen dengan konteks responden/pengguna. Uji kepraktisan instrumen dilakukan untuk memastikan produk yang dikembangkan praktis dan efektif digunakan untuk melakukan asesmen terhadap kemampuan peserta didik khususnya keterampilan bernalar kritis. Hasil analisis kepraktisan dituangkan dalam tabel 3.3 berikut ini:

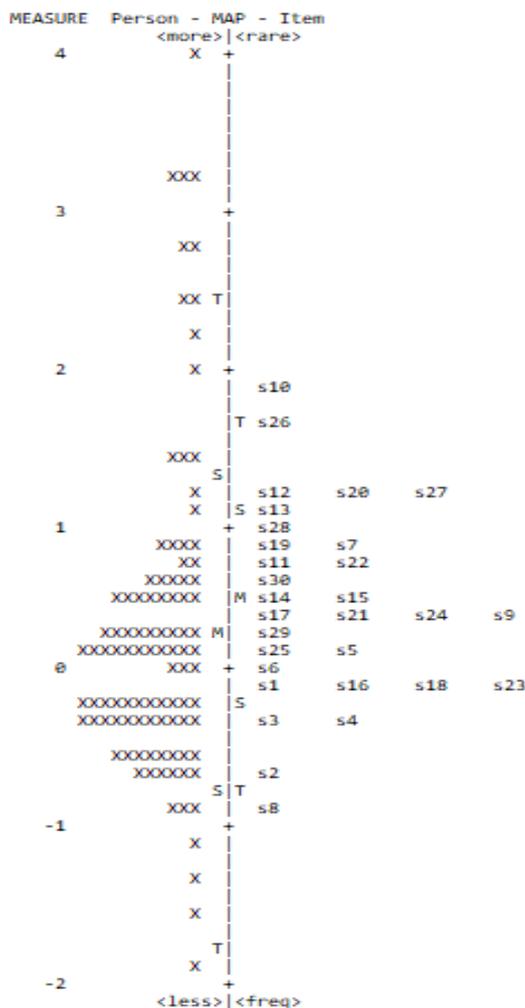
Tabel 3.3 Hasil Uji Kepraktisan Instrumen

Aspek Kepraktisan	Total Skor	Nilai Maksimum	Kepraktisan (Total skor/Nilai Maksimum) x 100 %
Kemudahan Penggunaan	38	40	95
Waktu Pengerjaan	36	40	90
Kualitas Teknis	35	40	87,5
Ketersediaan dan Kebermanfaatan	37	40	92,5
Kesesuaian Konten	37	40	92,5

Berdasarkan tabel 3.3, seluruh aspek menunjukkan tingkat kepraktisan yang sangat praktis dengan persentase di atas 81%. Aspek kemudahan penggunaan memiliki kepraktisan 95%, aspek ketersediaan dan kebermanfaatan serta kesesuaian konten masing-masing 92,5%, aspek waktu pengerjaan 90%, dan aspek keterbacaan soal 87,5%. Secara keseluruhan hasil ini mengindikasikan bahwa instrumen telah memenuhi kriteria ideal dalam hal kemudahan, kesesuaian isis, dan kejelasan, sehingga sangat layak digunakan untuk mengukur keterampilan bernalar kritis materi pembangunan berkelanjutan dan kelangkaan bagi siswa SMP.

### Tingkat Kesukaran Soal

Indeks kesukaran butir soal dalam evaluasi pendidikan dan pengukuran menunjukkan persentase responden yang mampu memberikan jawaban benar terhadap suatu item tes. Analisis menggunakan pemodelan Rasch, tingkat kesukaran dinyatakan dalam logit (log-odds unit) yang menempatkan soal dan kemampuan peserta pada skala interval yang sama (Sumintono & Widhiarso, 2015). Indeks kesukaran butir soal dapat dilihat pada analisis pemodelan Rasch diperoleh dari output *item: wright map*, seperti gambar berikut ini:



Gambar 3.4 Item Wright Map

Indeks kesukaran butir soal dari hasil uji coba skala kecil menggunakan aplikasi winstep menggambarkan kategori kesukaran soal yang dituangkan dalam gambar berikut ini:

Tabel 3.4 Tingkat Kesukaran Soal

Nilai Logit	Nomor Butir	Tingkat Kesukaran	Jumlah
Measure >1	10, 26	Sangat Sukar	2
0 < Measure ≤ 1	7, 11, 12, 13, 14, 15, 19, 20, 22, 27, 28, 30	Sukar	12
-1 < Measure ≤ 0	1, 3, 4, 5, 6, 9, 16, 17, 18, 21, 23, 24, 25, 29,	Mudah	14
Measure ≤ -1	2, 8	Sangat Mudah	2
<b>Total</b>			<b>30</b>

Berdasarkan hasil analisis pemodelan Rasch, tingkat kesukaran soal (*item difficulty*) yang diukur dalam satuan logit menunjukkan

variasi yang cukup luas, mulai dari -1.30 hingga +1.39. Soal nomor 10 (logit = 1.39) dan soal nomor 26 (logit = 1.17) tergolong dalam kategori sangat sukar, karena hanya dapat dijawab dengan benar oleh peserta yang memiliki kemampuan di atas 1.0 logit. Sebanyak 12 soal lainnya, masuk dalam kategori sukar dengan nilai logit antara 0 hingga 1. Sebanyak 14 soal berada dalam kategori mudah dengan nilai logit antara -1 sampai 0. Sementara soal nomor 8 (logit = -1.30) dan soal nomor 2 (logit = -1.04) termasuk dalam kategori sangat mudah, karena dapat dijawab benar bahkan oleh peserta dengan kemampuan rendah.

**Daya Pembeda Soal**

Indeks daya beda menunjukkan sejauh mana suatu soal dapat membedakan secara signifikan kelompok responden unggul (*upper group*) dari kelompok kurang (*lower group*) berdasarkan indikator penilaian yang ditetapkan seperti skor keseluruhan ujian.. Analisis daya beda menggunakan pemodelan Rasch dilakukan dengan melihat nilai Model S.E pada luaran tabel *item measure* (Ramadhan & Hidayatullah, 2023, p. 7). Daya beda suatu item dianggap baik apabila nilai Model SE yang diperoleh <0,5, karena hal ini menunjukkan kemampuannya yang efektif dalam membedakan kemampuan antarresponden. Jika nilai Model SE melebihi 1, hal ini mengindikasikan bahwa daya beda item tersebut lemah atau tidak memadai dalam membedakan tingkat kemampuan responden (Purniasari et al., 2021, p. 211).

Berdasarkan hasil analisis data menggunakan Winstep, data yang diperoleh melalui uji coba kepada 100 orang siswa diperoleh nilai Model SE berada dibawah 0,5, artinya sebanyak 30 butir item yang dikembangkan memiliki daya pembeda yang baik. Analisis pemodelan Rasch

menggunakan tingkat kemampuan individu (*ability level*) untuk membedakan antara peserta didik yang mampu dan tidak mampu menjawab suatu soal dengan benar. Pengelompokan responden juga dapat diidentifikasi melalui indeks separasi. Skor separasi item yang besar menunjukkan instrumen yang lebih baik, karena memiliki kemampuan membedakan responden dan tingkat kesulitan soal secara lebih jelas (Sumintono & Widhiarso, 2015). Analisis pengelompokan yang lebih presisi, dapat digunakan persamaan strata (H), yang memberikan informasi mengenai sejauh mana instrumen mampu mengklasifikasikan responden dan butir ke dalam strata yang berbeda (Erfan et al., 2020, pp. 17–18). Nilai persamaan strata dapat dihitung menggunakan rumus sebagai berikut:

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

Hasil uji daya beda soal dengan pemodelan Rasch dapat dilihat pada gambar berikut ini:

TABLE 15.1 Uji Skala Besar 200934HS.TXT Jun 4 2025 16:38  
 INPUT: 100 Person 30 Item REPORTED: 100 Person 30 Item 2 CATS WINSTEPS 4.8.1  
 Person: REAL SEP.: 2.10 REL.: .61 ... Item: REAL SEP.: 2.45 REL.: .66

Item STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFO	IMFIT	OUTFIT	PTMEASURE-ALL	EXACT MATCH	DESN	EXPN	Item
10	21	100	1.39	.20	1.85	.22	1.30	1.27	.57	.45	85.0	s10
26	24	100	1.17	.26	1.02	.22	1.02	.18	.42	.44	82.0	s10
30	31	100	.74	.24	1.15	1.24	1.22	1.34	.29	.43	75.0	s10
12	32	100	.66	.24	1.14	1.18	1.13	.99	.31	.43	76.0	s10
27	32	100	.66	.24	1.10	.89	1.23	1.45	.32	.43	74.0	s10
13	34	100	.57	.23	.97	-.29	.95	-.30	.45	.42	73.0	s10
38	35	100	.51	.23	1.12	1.13	1.33	2.17	.28	.42	72.0	s10
7	37	100	.43	.23	.99	-.04	.96	-.14	.42	.41	70.0	s10
19	38	100	.36	.23	1.06	.67	1.02	.17	.37	.41	67.0	s10
11	39	100	.31	.23	1.09	1.00	1.06	.45	.34	.41	64.0	s10
22	38	100	.31	.23	.95	-.51	.91	-.60	.46	.41	70.0	s10
50	42	100	.16	.22	.93	-.51	.88	-1.10	.48	.40	71.0	s10
14	44	100	.06	.22	.93	-.97	.88	-.96	.47	.39	74.0	s10
15	45	100	.01	.22	.97	-.56	.92	-.56	.42	.39	67.0	s10
11	46	100	-.04	.22	.99	-.15	1.00	.05	.40	.39	74.0	s10
9	47	100	-.06	.22	.95	-.75	.93	-.52	.43	.39	69.0	s10
17	47	100	-.06	.22	1.16	2.16	1.53	3.50	.21	.39	59.0	s10
34	47	100	-.06	.22	.96	-.52	.98	-.71	.43	.39	67.0	s10
29	50	100	-.23	.22	.99	-.19	.95	-.30	.39	.38	63.0	s10
5	51	100	-.27	.22	1.07	1.14	1.10	.76	.30	.37	58.0	s10
25	51	100	-.27	.22	.98	-.37	.93	-.46	.40	.37	66.0	s10
6	54	100	-.41	.22	.95	-.61	.87	-.85	.42	.36	79.0	s10
1	59	100	-.46	.22	.97	-.52	.89	-.75	.40	.36	68.0	s10
16	56	100	-.51	.22	.87	-2.12	.79	-1.41	.48	.36	75.0	s10
23	56	100	-.51	.22	.93	-1.09	.84	-1.02	.43	.36	65.0	s10
18	57	100	-.55	.22	1.03	.53	.97	.13	.33	.35	62.0	s10
3	61	100	-.74	.22	.91	-1.34	.83	-.96	.43	.34	75.0	s10
4	61	100	-.74	.22	.94	-.68	.89	-.63	.40	.34	75.0	s10
2	67	100	-1.04	.23	.93	-.79	.83	-.83	.39	.32	66.0	s10
8	73	100	-1.30	.23	.98	-.18	.93	-.67	.34	.30	79.0	s10
MEAN	45.7	100.0	.00	.23	1.00	-.1	1.00	.0			69.6	s10
P.50	12.0	.0	.01	.01	.86	.0	.17	1.1			5.7	s10

TABLE 15.3 Uji Skala Besar 200934HS.TXT Jun 4 2025 16:38  
 INPUT: 100 Person 30 Item REPORTED: 100 Person 30 Item 2 CATS WINSTEPS 4.8.1

Gambar 3.4 Daya Pembeda Soal

Berdasarkan gambar 3.4, nilai item separation adalah sebesar 2,45 maka nilai H adalah 3,6 dan dibulatkan menjadi 4, sehingga terdapat 4 kelompok butir soal. Nilai person separation adalah 2,10 maka nilai H adalah 3,13 dan dibulatkan menjadi 3, sehingga responden terbagi 3 kelompok berdasarkan nilai separasi.

**Profil Keterampilan Bernalar Kritis Siswa**

Bernalar kritis adalah kompetensi yang harus dikuasai peserta didik sebagai bagian dari dimensi Profil Pelajar Pancasila. Visi pendidikan Indonesia melalui Profil Pelajar Pancasila berfokus pada pengembangan siswa yang menguasai kompetensi global sekaligus mencerminkan sikap dan perilaku yang selaras dengan nilai Pancasila (Anggrarina et al., 2025, pp. 122–123). Profil kemampuan bernalar kritis siswa dituangkan dalam tabel berikut ini:

Tabel 3.5 Profil Keterampilan Bernalar Kritis Siswa

Tujuan Pembelajaran	Domain Kognitif	Nomor Butir	Total Skor	Presentase
Peserta didik mampu menganalisis konsep pembangunan berkelanjutan, mengkritisi karakteristiknya, dan membuat ide berbasis pembangunan berkelanjutan.	C4	1	55	55 %
	C4	2	67	67 %
	C4	3	61	61 %
	C4	4	61	61 %
	C4	5	51	51 %
	C4	6	44	44 %
	C4	7	37	37 %
	C4	8	72	72 %
	C4	9	46	46 %
	C5	10	21	21 %
	C4	11	39	39 %
Peserta didik mampu membuat langkah pencegahan kelangkaan sumber daya dan menganalisis kebutuhan manusia yang tidak terbatas.	C4	12	31	31 %
	C4	13	34	34 %
	C5	14	44	44 %
	C5	15	45	45 %
	C5	16	56	56 %
Peserta didik mampu membuat langkah pencegahan kelangkaan sumber daya dan mengevaluasi masalah pokok ekonomi.	C5	17	47	47 %
	C5	18	57	57 %
	C5	19	38	38 %
	C5	20	30	30 %
	C5	21	46	46 %
	C5	22	39	39 %
	C5	23	56	56 %
	C5	24	47	47 %
Peserta didik mampu mengevaluasi faktor penyebab kelangkaan dan upaya mengatasinya secara berkelanjutan.	C5	25	51	51 %
	C5	26	22	22 %
	C5	27	31	31 %
	C5	28	35	35 %
	C5	29	50	50 %
	C5	30	42	42 %

Hasil uji coba instrumen yang dilakukan kepada 100 siswa SMP kelas VII yang berasal

dari SMP Negeri 2 Lumbanjulu sebanyak 2 kelas berjumlah 53 orang dan SMP Negeri 2 Satu Atap Sibisa sebanyak 2 kelas berjumlah 47 orang. Data yang dituangkan dalam tabel diperoleh dari analisis pemodelan Rasch menggunakan software Winstep pada output tabel item measure. Hasil yang diperoleh menunjukkan rata-rata kemampuan siswa sebesar 51,27% sudah menguasai tujuan pembelajaran yaitu menganalisis konsep pembangunan berkelanjutan, mengkritisi karakteristiknya, dan membuat ide berbasis pembangunan berkelanjutan. Kemampuan bernalar kritis siswa diukur dengan 11 butir instrumen yang memiliki tingkat kesulitan beragam. Butir soal nomor 10 adalah yang paling sukar dan hanya bisa dijawab oleh 21 % peserta didik, sedangkan butir soal yang paling mudah adalah nomor 8 dengan presentase siswa menjawab benar sebesar 72%. Data ini menunjukkan kemampuan siswa dalam memahami konsep pembangunan berkelanjutan dapat dikategorikan sedang.

Tujuan pembelajaran kedua, yaitu mampu membuat langkah pencegahan kelangkaan sumber daya dan menganalisis kebutuhan manusia yang tidak terbatas diperoleh data rata-rata kemampuan siswa adalah 42 %. Data ini mengindikasikan bahwa kemampuan peserta didik dalam membuat langkah pencegahan kelangkaan dan menganalisis kebutuhan manusia yang tidak terbatas masih rendah. Tujuan pembelajaran ketiga yang berfokus pada mengevaluasi masalah pokok ekonomi menunjukkan rata-rata pencapaian siswa dalam indikator ini adalah 44,7 %. Kondisi ini menunjukkan tingkat kemampuan evaluasi siswa masih rendah. Tujuan pembelajaran keempat diukur dengan 7 butir instrumen dan diperoleh rata-rata kemampuan siswa yang sudah

menguasai adalah 39,7 %. Kemampuan siswa pada indikator ini adalah yang paling rendah. Namun pada tujuan pembelajaran ini, instrumen yang disusun mengukur kemampuan kognitif C5 sehingga beberapa butir soal masuk dalam kategori sukar.

Profil keterampilan bernalar kritis siswa pada materi pembangunan berkelanjutan dan kelangkaan dapat dilihat dari hasil output analisis pemodelan Rasch yaitu person measure. Siswa dengan kemampuan tertinggi diurutkan sampai siswa dengan kemampuan terendah. Nilai yang dilihat adalah nilai logit pada kolom measure. Tingkat kemampuan bernalar kritis siswa dapat dilihat pada tabel berikut ini:

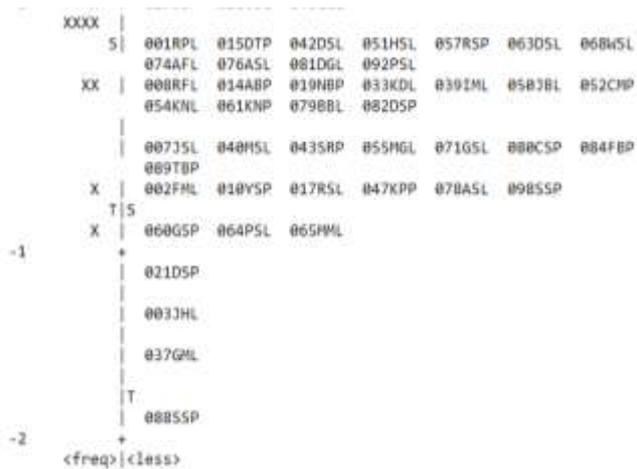
Tabel 3.6 Tingkat Abilitas Siswa

Nomor Urut Siswa	Nilai Logit Abilitas Siswa	Kriteria
048HSP	3,54	Tinggi
024JSL, 099RHL, 100ETL	2,80	Tinggi
036GIL, 091MSL	2,34	Tinggi
005ITP, 087SSP	2,00	Tinggi
011FBL	1,73	Sedang
044ABP	1,49	Sedang
013MSP, 034DHL, 085GSP	0,92	Sedang
018SSP	0,75	Sedang
094PSL	0,59	Sedang
012NVP, 029ATP, 030JTL, 031ATP	0,44	Sedang
032IML, 045LSP	0,29	Sedang
035NPL, 049MNP, 067RSL, 073RNL, 075NML	0,14	Sedang
006ESP, 020ATL, 022RSL, 053VMP, 062SNP, 069DSL, 086GSP, 097RPL	0,00	Sedang
004ZGL, 009CSL, 023JSL, 026LTP, 066ABP, 077AML, 090VML, 093PBL, 096RTP	-0,15	Sedang
016JSL, 025VML, 038CMP, 041MLL, 056AMP, 058ANP, 059YSP, 070RML, 072ESL, 083FNL, 095RSP	-0,29	Sedang
027JSP, 028JSL, 046GLL	-0,44	Sedang
001RPL, 015DTP, 042DSL, 051HSL, 057RSP, 063DSL, 068WSL, 074AFL, 076ASL, 081DGL, 092PSL	-0,59	Sedang
008RFL, 014ABP, 019NBP, 033KDL, 039IML, 050JBL, 052CMP, 054KNL, 061KNP, 079BBL, 082DSP	-0,75	Sedang
007JSL, 040MSL, 043SRP, 055MGL, 071GSL, 080CSP, 084FBP, 089TBP	-0,92	Sedang
002FML, 010YSP, 017RSL, 047KPP, 078ASL, 098SSP	-1,09	Rendah
060GSP, 064PSL, 065MML	-1,28	Rendah
021DSP	-1,49	Rendah
003JHL	-1,73	Rendah
037GML	-2,00	Rendah
088SSP	-2,34	Rendah

Berdasarkan tabel 3.6, siswa dengan tingkat abilitas paling tinggi adalah 048HSP. Abilitas siswa dengan kategori tinggi sebanyak 8 orang yaitu 048HSP, 024JSL, 099RHL, 100ETL, 036GIL, 091MSL, 005ITP, 087SSP yang semuanya memiliki nilai logit  $\geq 2$ . Siswa dengan abilitas paling rendah adalah 088SSP dengan nilai logit -2,34. Secara keseluruhan siswa dengan kriteria abilitas rendah ada 13 orang yaitu 002FML, 010YSP, 017RSL, 047KPP, 078ASL, 098SSP, 060GSP, 064PSL, 065MML, 021DSP, 003JHL, 037GML, 088SSP.

Analisis kemampuan siswa menggunakan pemodelan Rasch juga dapat dilihat dari output analisis wright map. Tingkat abilitas siswa hasil analisis pemodelan Rasch dapat dilihat pada gambar berikut ini:





Gambar 3.5 *Person Wright Map*

Hasil uji coba instrumen pada skala besar kepada 100 orang siswa diperoleh sebanyak 8 siswa (8%) memiliki tingkat abilitas dengan kategori tinggi. Sebanyak 79 orang (79%) memiliki tingkat abilitas sedang, dan sebanyak 13 orang (13%) memiliki tingkat abilitas rendah. Siswa dengan inisial 048HSP memiliki abilitas yang sangat tinggi, menunjukkan bahwa siswa ini dapat menjawab soal bahkan dengan tingkat kesukaran tinggi. Sementara siswa dengan inisial 088HSP memiliki nilai abilitas terendah, yang menunjukkan bahwa siswa ini bahkan tidak dapat menjawab soal dengan tingkat kesulitan yang mudah.

Hasil yang diperoleh menunjukkan sebagian besar siswa berada pada tingkat abilitas sedang. Data ini mengindikasikan bahwa keterampilan bernalar kritis siswa pada materi pembangunan berkelanjutan dan kelangkaan perlu ditingkatkan. Sementara siswa pada kategori abilitas rendah memerlukan pengenalan konsep awal pembangunan berkelanjutan dan kelangkaan dan lebih sering melatih siswa dengan soal-soal studi kasus terkhusus pada materi pembangunan berkelanjutan dan kelangkaan.

### Kesimpulan

Validitas isi instrumen yang dikembangkan menggunakan rumus Aiken's

V dan membandingkan nilai V yang diperoleh dengan tabel Aiken menunjukkan bahwa instrumen yang dikembangkan valid dan memadai dengan nilai V yang diperoleh diatas 0,79 sebagai *cut off value*. Hasil uji analisis pemodelan Rasch menunjukkan bahwa instrumen valid secara konstruk dengan nilai *raw variance explained by measure* pada kolom eigenvalue adalah 20,2077 (40,2 %) yang masuk dalam kategori baik. Hasil analisis pemodelan Rasch menunjukkan bahwa sebanyak 28 butir soal dinyatakan valid secara empiris. Butir soal nomor 17 dan nomor 28 tidak memenuhi kriteria validitas empiris dengan nilai *Outfit Mean Square* >1,2. Kedua butir soal dihapus pada revisi produk untuk mengurangi resiko bahwa butir soal akan mengganggu proses pengukuran.

Analisis pemodelan Rasch menghasilkan dua jenis reliabilitas dalam pengembangan instrumen, yaitu *item reliability* dan *person reliability*. Nilai reliabilitas item adalah sebesar 0,86 yang masuk dalam kategori baik, sedangkan nilai person reliability adalah sebesar 0,81 yang juga masuk dalam kategori baik. Uji reliabilitas dengan uji klasik (KR-20) juga diperoleh nilai sebesar 0,81 dan dikategorikan baik. Kepraktisan buku pedoman guru yang memuat soal pilihan ganda untuk mengukur keterampilan bernalar kritis siswa SMP kelas VII pada mata pelajaran IPS dianalisis dari data angket kepraktisan yang diberikan kepada guru mata pelajaran IPS di lima sekolah berbeda. Hasil analisis uji kepraktisan menunjukkan nilai rata-rata sebesar 91,5 % yang dikategorikan sangat praktis.

Siswa yang memiliki tingkat abilitas yang tinggi sebanyak 8% artinya memiliki tingkat keterampilan bernalar kritis yang tinggi pula. Siswa dengan tingkat abilitas sedang sebanyak 71 % artinya sebagian besar siswa

belum memiliki keterampilan bernalar kritis pada materi pembangunan berkelanjutan dan kelangkaan. Siswa dengan tingkat abilitas rendah sebanyak 13% artinya perlu perhatian khusus dalam upaya meningkatkan kemampuan bernalar kritis.

Guru sebagai evaluator pembelajaran memerlukan alat ukur yang valid dan reliabel. Pengembangan instrumen ini sangat membantu guru dalam mengukur kemampuan bernalar kritis siswa dalam materi pembangunan berkelanjutan dan kelangkaan. Siswa sebagai subjek penelitian juga memberikan tanggapan positif atas pengembangan instrumen yang dilakukan karena dapat mengasah keterampilan bernalar kritis, bukan sekedar hafalan saja.

#### **DAFTAR PUSTAKA**

- Anggrarina, Rakhmawati, D., & Purnamasari, I. (2025). The effect of emotional intelligence and communication skills on critical reasoning dimensions in realizing the profile of Pancasila students. *Primary: Jurnal Pendidikan Guru Sekolah Dasar*, 14(2), 122–133. <https://doi.org/10.33578/jpfkip-v14i2.p122-133>
- Aulia, R., & Wandini, R. R. (2023). Karakteristik Mata Pelajaran IPS. *Jurnal Pendidikan Dan Konseling*, 5, 4034–4040.
- Bashooir, K., & Supahar, S. (2018). Validitas dan reliabilitas instrumen asesmen kinerja literasi sains pelajaran fisika berbasis STEM. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 22(2), 219–230. <https://doi.org/10.21831/pep.v22i2.19590>
- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. Springer.
- Erfan, M., Archi Mauliyda, M., 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. <https://doi.org/https://doi.org/10.23887/ijer.v3i1.24080>
- Ernawati, Y., & Rahmawati, F. P. (2022). Analisis Profil Pelajar Pancasila Elemen Bernalar Kritis dalam Modul Belajar Siswa Literasi dan Numerasi Jenjang Sekolah Dasar. *Jurnal Basicedu*, 6(4), 6132–6144. <https://doi.org/10.31004/basicedu.v6i4.3181>
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *HowTo Design and Evaluate Reseach in Education* (S. Kiefer, Ed.; 8th ed.). The McGraw’Hill Companies.
- Griffin, P. E., & Nix, P. (1991). *Educational Assessment and Reporting: A New Approach*. Harcourt Brace Javanovich.
- Kemdikburistek RI. (2022). *Dimensi, Elemen, dan Subelemen Profil Pelajar Pancasila pada Kurikulum Merdeka 2*.
- Kemendikbudristek RI. (2022). *Peraturan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia*.
- Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Maulana, S., Rusilowati, A., Nugroho, S. E., & Susilaningsih, E. (2023). *Implementasi*

- Rasch Model dalam Pengembangan Instrumen Tes Diagnostik.* 748–757.  
<http://pps.unnes.ac.id/pps2/prodi/prosidin-g-pascasarjana-unnes748>
- Nofrion, N., & Wijayanto, B. (2018). Learning Activities In Higher Order Thinking Skill (HOTS) Oriented Learning Context. *Geosfera Indonesia*, 3(2), 122.  
<https://doi.org/10.19184/geosi.v3i2.8126>
- 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>
- Ramadhan, A. F., & Hidayatullah, R. S. (2023). Analisis Kualitas Butir Soal Ujian Satuan Pendidikan (USP) Materi C2 Teknik Pemesinan Kelas XII di SMK PGRI 1 Lamongan Melalui Model Rasch. *JPTM*, 12(03), 1–10.
- Reffhaug, M. B. A., & Lysgaard, J. A. (2024). Conceptualisations of ‘critical thinking’ in environmental and sustainability education. *Environmental Education Research*, 30(9), 1519–1534.  
<https://doi.org/10.1080/13504622.2024.2363848>
- Sari, E. D. K. S., & Mahmudi, I. (2024). *Analisis Pemodelan Rasch Pada Assessment Pendidikan*. PT. Pena Persada Kerta Utama.  
<https://www.researchgate.net/publication/378210331>
- Setiawati, H., & Duran Corebima, A. (2017). Empowering Critical Thinking Skills Of The Students Having Different Academic Ability in Biology Learning of Senior High School through PQ4R-TPS Strategy. *The International Journal of Social Sciences and Humanities Invention*, 4(5), 3521–3526.  
<https://doi.org/10.18535/ijsshi/v4i5.09>
- Sugiyono. (2013). *METODE PENELITIAN KUANTITATIF, KUALITATIF, DAN R&D* (19th ed.). ALFABETA.
- Sumintono, B., & Widhiarso, W. (2014). *Aplikasi Model Rasch untuk Penelitian Ilmu-Ilmu Sosial (edisi revisi)*. Trim Komunikata Publishing House.  
<https://www.researchgate.net/publication/268688933>
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan RASCH pada Assessment Pendidikan*. Tim Komunikata Publishing House.
- Sya’ban, Moh. B. A. (2018). Tinjauan Mata Pelajaran IPS SMP Pada Penerapan Pendidikan Lingkungan Hidup Untuk Peduli Akan Tanggung Jawab Lingkungan. *Jurnal Geografi, Edukasi Dan Lingkungan (JGEL)*, 2(1), 32–44.
- Umar, J., & Faela Nisa, Y. (2020). Uji Validitas Konstruk dengan CFA dan Pelaporannya. *Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia*, 9(2), 1–11.  
<https://doi.org/10.15408/jp3i.v9i2.XXX>