

IMPLEMENTASI *RASCH* MODEL DALAM UJI *INSTRUMENT* KEMAMPUAN *DESIGN THINKING* PADA MATERI PENCEMARAN LINGKUNGAN

Eka Maya Putri Anggraeni¹, Fenny Roshayanti^{2*}, M. Syaipul Hayat³

^{1,2,3}Universitas PGRI Semarang, Semarang, Indonesia

*Corresponding author email: fennyroshayanti@gmail.com

Received 3 March 2026; Received in revised form 2 April 2026; Accepted 10 April 2026

Abstrak

Penelitian ini bertujuan menganalisis validitas dan reliabilitas instrumen kemampuan *design thinking* pada materi pencemaran lingkungan menggunakan *Rasch Model*. Penelitian menggunakan pendekatan deskriptif kuantitatif dengan instrumen 30 soal pilihan ganda yang diujicobakan kepada siswa SMA Negeri 15 Semarang melalui teknik *purposive sampling*. Analisis data dilakukan menggunakan *software Winsteps 3.73* meliputi uji unidimensionalitas, tingkat kesukaran butir, *item fit*, serta reliabilitas person dan item. Hasil menunjukkan instrumen memenuhi kriteria unidimensionalitas dengan 80% butir soal dinyatakan fit, sedangkan 6 butir memerlukan revisi. Reliabilitas instrumen tergolong sangat baik dengan nilai Cronbach Alpha 0,86, person reliability 0,78, dan item reliability 0,97. Distribusi tingkat kesukaran didominasi butir mudah sehingga diperlukan penambahan soal berkesukaran tinggi. Secara keseluruhan, instrumen dinyatakan valid dan reliabel untuk mengukur kemampuan *design thinking* siswa SMA dalam konteks isu pencemaran lingkungan.

Kata Kunci: design thinking; rasch model; validitas instrumen; pencemaran lingkungan; asesmen pendidikan

Abstract

This study aims to analyze the validity and reliability of a design thinking ability instrument on environmental pollution material using the Rasch Model. A descriptive quantitative approach was employed using a 30-item multiple-choice test administered to students of SMA Negeri 15 Semarang through purposive sampling. Data analysis was conducted using Winsteps 3.73 software, including tests of unidimensionality, item difficulty, item fit, and person and item reliability. The results indicated that the instrument met the unidimensionality criteria, with 80% of the items categorized as fit, while six items required revision. The instrument demonstrated very good reliability with a Cronbach's Alpha of 0.86, person reliability of 0.78, and item reliability of 0.97. The difficulty distribution was dominated by easy items, suggesting the need for additional high-difficulty questions. Overall, the instrument was found to be valid and reliable for measuring senior high school students' design thinking abilities in the context of environmental pollution issues.

Keywords: design thinking; rasch model; instrument validity; environmental pollution; educational assessment



This is an open access article under the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

PENDAHULUAN

Pendidikan abad ke-21 menuntut pengembangan keterampilan yang melampaui kemampuan kognitif tradisional. Keterampilan berpikir tingkat tinggi,

DOI: <https://doi.org/10.26877/jp3.v12i1.531>

kreativitas, kolaborasi, dan kemampuan memecahkan masalah kompleks menjadi esensial dalam mempersiapkan siswa menghadapi tantangan global yang semakin dinamis (Trilling & Fadel, 2009; Partnership for 21st Century Skills, 2019). Survei World Economic Forum (2023) menunjukkan bahwa kemampuan *design thinking* termasuk dalam sepuluh keterampilan paling dibutuhkan di masa depan, seiring meningkatnya kebutuhan industri terhadap tenaga kerja yang mampu berinovasi dan berpikir solutif. Oleh karena itu, sistem pendidikan dituntut tidak hanya menyampaikan pengetahuan, tetapi juga menumbuhkan dan mengukur keterampilan tersebut secara terstandar dan berkelanjutan (OCED, 2019)

Design thinking merupakan pendekatan *human-centered* yang melibatkan lima tahap iteratif: *empathize, define, ideate, prototype, dan test* (Brown, 2008), yang dinilai efektif dalam menumbuhkan kemampuan berpikir kreatif dan solutif dalam pembelajaran (Razzouk & Shute, 2012). Relevansinya semakin menguat dalam konteks implementasi Kurikulum Merdeka di Indonesia, yang menekankan pembelajaran berbasis proyek dan penguatan profil pelajar Pancasila (Kemendikbudristek, 2022). Sejumlah penelitian telah mengembangkan instrumen pengukuran kemampuan *design thinking*. Carroll et al. (2010) mengidentifikasi empat dimensi utama melalui *Scale of Design Thinking for Teaching* (SDTT), yakni *problematizing, ideating, prototyping, dan testing*. Dosi et al. (2018) mengembangkan *Design Thinking Mindset Questionnaire* dengan 71 item yang mengukur 22 konstruk pola pikir *design thinking*. Di Indonesia, Patria et al. (2024) mengembangkan instrumen khusus untuk mahasiswa desain grafis, sementara Karyanto et al. (2024) menegaskan bahwa evaluasi berbasis keterampilan lunak pada siswa sekolah menengah merupakan komponen penting dalam pengembangan keterampilan abad ke-21.

Validasi instrumen asesmen merupakan tahapan krusial untuk memastikan bahwa proses pengukuran berlangsung secara konsisten dan akurat (Azwar, 2017). Salah satu pendekatan yang banyak digunakan dalam validasi instrumen pendidikan adalah *Rasch Model*, yang merupakan bagian dari teori respons butir. Model ini memiliki keunggulan dalam mereduksi kompleksitas data pengukuran, memisahkan parameter kesulitan butir dari kemampuan peserta, serta mendeteksi butir yang tidak sesuai (*misfit*) (Boone et al., 2014; Wright & Stone, 2004). Berbeda dengan teori tes klasik, *Rasch Model* mendasarkan probabilitas jawaban benar pada selisih antara tingkat kemampuan peserta dan tingkat kesulitan butir, sehingga menghasilkan interpretasi yang lebih objektif dan bermakna (Bond & Fox, 2015; Linacre, 2002). Penerapan *Rasch Model* telah dilaporkan efektif tidak hanya untuk asesmen kognitif, tetapi juga untuk pengukuran aspek afektif seperti sikap, persepsi, dan motivasi (Engelhard, 2013). Penerapannya telah terbukti efektif tidak hanya pada asesmen kognitif, tetapi juga pada pengukuran aspek afektif seperti sikap, persepsi, dan motivasi belajar (Aryadoust et al., 2019)

Meskipun penelitian-penelitian tersebut telah memberikan kontribusi penting, terdapat beberapa kesenjangan yang belum terpenuhi. Pertama, instrumen yang telah dikembangkan umumnya belum disertai validasi

DOI: <https://doi.org/10.26877/jp3.v12i1.531>

psikometrik yang robust menggunakan *Rasch Model*, sehingga kualitas pengukurannya belum dapat dijamin secara optimal (Boone, 2016; Aryadoust et al., 2019). Kedua, sebagian besar instrumen dikembangkan untuk konteks pendidikan tinggi atau negara lain, sehingga belum tentu tepat digunakan untuk siswa sekolah menengah atas di Indonesia (Karyanto et al., 2024). Ketiga, belum ada instrumen *design thinking* yang dikembangkan dalam konteks pembelajaran biologi berbasis isu lingkungan lokal, khususnya permasalahan pencemaran sungai (Sasongko et al., 2021; KLHK, 2020).. Kesenjangan ini menunjukkan perlunya pengembangan dan validasi instrumen kemampuan *design thinking* yang kontekstual, terstandar, dan berbasis pendekatan psikometrik modern untuk jenjang sekolah menengah atas.

SMA Negeri 15 Semarang merupakan salah satu sekolah menengah atas yang telah mengimplementasikan Kurikulum Merdeka, termasuk pembelajaran berbasis proyek yang menuntut keterampilan *design thinking*. Namun, berdasarkan studi pendahuluan yang dilakukan peneliti, sekolah ini belum memiliki instrumen yang terstandarisasi dan tervalidasi secara psikometrik untuk mengukur kemampuan tersebut. Kondisi ini menyebabkan guru kesulitan dalam mengevaluasi perkembangan keterampilan siswa secara akurat (Mardapi, 2017). Peserta didik kelas XI dipilih karena pada usia 16–17 tahun telah memasuki tahap operasional formal yang memungkinkan berpikir abstrak dan hipotetis (Inhelder & Piaget, 1958), yang relevan dengan proses *design thinking*.

Penelitian ini menggunakan konteks pencemaran sungai sebagai isu autentik, mengingat Indonesia merupakan salah satu penyumbang sampah plastik terbesar ke laut melalui aliran sungai (Jambeck et al., 2015). Data BPS Kota Semarang (2022) dan KLHK (2020) juga menunjukkan adanya penurunan kualitas lingkungan yang berdampak pada kesehatan masyarakat. Konteks ini relevan dengan materi biologi dan memberikan peluang bagi siswa untuk mengembangkan solusi berbasis *design thinking*. Oleh karena itu, penelitian ini bertujuan mengembangkan dan memvalidasi instrumen kemampuan *design thinking* berbasis konteks pencemaran sungai menggunakan Rasch Model.

METODE PENELITIAN

Penelitian ini menerapkan metode deskriptif dengan pendekatan kuantitatif yang bertujuan menganalisis validitas dan reliabilitas instrumen kemampuan *design thinking* pada materi pencemaran lingkungan menggunakan *Rasch Model* yaitu menggunakan aplikasi *Winsteps 3.73* (<https://www.winsteps.com/index.htm>) meliputi uji unidimensionalitas, tingkat kesukaran butir, kesesuaian butir (*item fit*), serta reliabilitas instrumen (Sumintono & Widhiarso, 2015). Populasi penelitian adalah seluruh siswa kelas XI SMA Negeri 15 Semarang yang berjumlah 360 siswa. Sampel penelitian ditentukan menggunakan teknik *simple random sampling*, sehingga setiap anggota populasi memiliki peluang yang sama untuk terpilih sebagai responden (Sugiyono, 2019). Berdasarkan teknik tersebut, diperoleh jumlah sampel sebanyak 313 responden, yang dinilai telah representatif untuk analisis menggunakan Rasch Model.

DOI: <https://doi.org/10.26877/jp3.v12i1.531>

Instrumen penelitian berupa tes pilihan ganda sebanyak 30 butir soal yang disusun berdasarkan lima indikator kemampuan *design thinking*, yaitu *empathize*, *define*, *ideate*, *prototype*, dan *test*. Setiap butir soal dirancang untuk mengukur aspek kognitif siswa dalam menyelesaikan permasalahan kontekstual terkait pencemaran lingkungan (Tabel 1). Sebelum digunakan, instrumen terlebih dahulu melalui proses validasi isi oleh tiga ahli di bidang pendidikan biologi dan asesmen. Hasil validasi dianalisis menggunakan teknik Aiken's V, dengan seluruh butir memperoleh nilai $V > 0,80$ sehingga dinyatakan valid secara isi (Azwar, 2017). Selanjutnya, instrumen digunakan dalam pengumpulan data yang dilakukan satu kali (*one-shot data collection*) sesuai dengan prinsip dalam Rasch Model (Tabel 2).

Tabel 1. Kisi-kisi Instrumen

Indikator <i>Design Thinking</i>	Deskripsi Kemampuan	Nomor Soal	Jumlah
<i>Empathize</i>	Mengidentifikasi permasalahan berdasarkan kondisi lingkungan	1,2,3,4,5,6	6
<i>Define</i>	Merumuskan masalah secara jelas dan terarah	7,8,9,10,11,12	6
<i>Ideate</i>	Menghasilkan berbagai alternatif solusi	13,14,15,16,17,18	6
<i>Prototype</i>	Merancang solusi dalam bentuk model/konsep	19,20,21,22,23,24	6
<i>Test</i>	Mengevaluasi dan menguji solusi yang dibuat	25,26,27,28,29,30	6
Total		30 soal	30

Tabel 2. Rangkuman Kriteria Parameter Uji dan Sumber Output Winsteps

Parameter Uji	Kriteria	Sumber Output
<i>Unidimensionality</i>	1. Nilai <i>raw variance explained by measure</i> tidak kurang dari 20%. 2. Nilai <i>unexplained variance in 1st construct</i> tidak lebih dari 15%	23. <i>Item: dimensionality</i>
Tingkat Kesukaran Butir Soal	$Measure\ logit > +1\ SD$ (Sulit Sekali) $0,00 \leq Measure\ logit \leq +1\ SD$ (Sulit) $0,00 \geq Measure\ logit \geq -1SD$ (Mudah)	13. <i>Item: measure</i>
<i>Item Fit</i> (Tingkat Kesesuaian Butir Soal)	$0,5 < MNSQ < 1,5$ $-2,0 < ZSTD < +2,0$ $0,4 < PT\ Mean\ Corr < 0,85$ >0,94 (Istimewa)	10. <i>Item (Column): fit order</i>
<i>Person Reliability</i> dan <i>Item Reliability</i>	0,91 – 0,94 (Baik Sekali) 0,81 – 0,90 (Baik) 0,67 – 0,80 (Cukup) <0,67 (Lemah) >0,8 (Baik Sekali) 0,7 – 0,8 (Baik)	3.1 <i>Summary Statistics</i>
<i>Cronbach's Alpha</i>	0,6 – 0,7 (Cukup) 0,5 – 0,6 (Jelek) <0,5 (Buruk)	3.1 <i>Summary Statistics</i>

HASIL DAN PEMBAHASAN

1. Analisis *Unidimensionality* Instrumen

Nilai tersebut dapat diperoleh berdasarkan *output tabel 23. Item: dimensionality* pada *Winsteps* yang tertera pada Tabel 3.

Tabel 3. Uji *Unidimensionality*

Keterangan		<i>Empirical</i>	<i>Modeled</i>	
Total raw variance in observations	39.2	100.0%		100.0%
Raw variance explained by measures	9.2	23.5%		22.8%
Raw variance explained by persons	1.1	2.9%		2.8%
Raw variance explained by items	8.1	20.6%		19.9%
Raw unexplained variance (total)	30.0	76.5%	100.0%	77.2%
Unexplained variance in 1 st contrast	1.9	4.8%	6.3%	
Unexplained variance in 2 nd contrast	1.9	4.7%	6.2%	
Unexplained variance in 3 rd contrast	1.7	4.2%	5.5%	
Unexplained variance in 4 th contrast	1.5	3.8%	5.0%	
Unexplained variance in 5 th contrast	1.4	3.6%	4.7%	

Validitas instrumen keterampilan *design thinking* dalam *Rasch Model* ditinjau melalui uji unidimensionalitas yang menunjukkan kemampuan instrumen dalam mengukur konstruk yang seharusnya diukur secara akurat. Instrumen dinyatakan memenuhi unidimensionalitas apabila nilai *raw variance explained by measures* minimal 20%, dengan kategori cukup baik pada rentang 20%–40%, baik pada 40%–60%, dan sangat baik jika melebihi 60%. Selain itu, nilai *unexplained variance in 1st contrast* tidak boleh melebihi 15% (Muntazhimah *et al.*, 2020; Sumintono & Widhiarso, 2015).

Berdasarkan hasil analisis *item dimensionality* pada Tabel 3, diperoleh nilai *raw variance explained by measures* sebesar 23,5% yang berada pada kategori cukup baik (20%–40%). Nilai *unexplained variance in 1st contrast* sebesar 4,8% serta kontras ke-2 hingga ke-5 yang masing-masing bernilai 4,7%, 4,2%, 3,8%, dan 3,6% menunjukkan seluruh nilai berada di bawah batas 15%. Hasil tersebut mengindikasikan bahwa instrumen keterampilan *design thinking* pada materi pencemaran lingkungan memenuhi kriteria unidimensionalitas sehingga mampu mengukur konstruk yang dituju secara tepat. Temuan ini sejalan dengan penelitian Maulana *et al.* (2021) yang juga memperoleh nilai *raw variance explained by measures* sebesar 26,8% dengan seluruh nilai *unexplained variance* di bawah 15%, sehingga instrumen dinyatakan memenuhi syarat unidimensionalitas. Terpenuhinya kriteria unidimensionalitas menunjukkan bahwa seluruh butir soal secara konsisten mengukur satu konstruk, yaitu kemampuan *design thinking* siswa dalam konteks pencemaran lingkungan. Hal ini relevan karena permasalahan pencemaran sungai merupakan isu kompleks yang menuntut kemampuan empati, identifikasi masalah, dan perancangan solusi secara terpadu. Dengan demikian, instrumen yang bersifat unidimensional mampu merepresentasikan kemampuan siswa secara holistik dalam merespons permasalahan lingkungan sebagai konteks pembelajaran yang nyata (Fadhilah *et al.*, 2024; Sasongko *et al.*, 2021).

DOI: <https://doi.org/10.26877/jp3.v12i1.531>

2. Analisis Kesukaran Butir Item (Analisis *Item Measure*)

Tingkat kesukaran butir item dapat diklasifikasikan ke dalam empat kategori berdasarkan nilai standar deviasi (SD) dan mean pada kolom measure. Kategori tersebut meliputi: sangat sulit (lebih dari +1 SD), sulit (antara 0,00 logit hingga +1 SD), mudah (antara 0,00 logit hingga -1 SD), dan sangat mudah (kurang dari -1 SD) (Sumintono & Widhiarso, 2015).

ITEM STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	TOTAL MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM
18	138	313	.59	.06	.97	-.2	1.01	.1	.56	.55	80.6	80.3	S18
29	180	313	.43	.06	1.04	.4	1.09	.6	.48	.51	62.2	64.4	S29
27	207	313	.35	.05	1.18	1.8	1.63	3.7	.36	.50	47.4	52.3	S27
19	213	313	.34	.05	1.08	.9	1.18	1.2	.44	.49	49.0	52.0	S19
4	222	313	.31	.05	1.05	.6	1.19	1.4	.45	.49	43.4	45.7	S4
11	234	313	.28	.05	1.19	2.1	1.31	2.2	.37	.48	40.1	45.3	S11
5	276	313	.18	.05	.90	-1.4	.83	-1.7	.52	.46	30.6	30.0	S5
24	285	313	.16	.05	1.13	1.9	1.17	1.6	.38	.46	28.3	29.8	S24
30	300	313	.13	.05	.98	-.2	1.10	1.0	.44	.45	24.0	25.5	S30
16	303	313	.13	.05	.95	-.8	.89	-1.1	.48	.45	26.3	25.4	S16
8	312	313	.11	.05	1.16	2.5	1.24	2.3	.35	.44	23.4	25.3	S8
3	336	313	.06	.04	.98	-.4	1.20	2.1	.43	.43	17.8	18.5	S3
17	336	312	.06	.04	1.18	3.1	1.25	2.5	.32	.43	16.5	18.2	S17
13	357	313	.02	.04	.88	-2.4	.89	-1.2	.49	.43	12.8	13.0	S13
25	378	313	-.02	.04	.92	-1.7	.86	-1.6	.47	.42	13.5	12.9	S25
28	381	313	-.03	.04	.97	-.6	.93	-.8	.44	.42	13.5	12.9	S28
26	399	313	-.06	.04	.98	-.5	.97	-.3	.42	.41	12.5	12.5	S26
20	405	313	-.07	.04	.91	-1.9	.88	-1.4	.46	.41	8.9	8.9	S20
2	426	313	-.11	.04	.99	-.1	.98	-.2	.40	.40	8.6	9.2	S2
1	444	313	-.14	.04	1.04	1.0	1.04	.5	.37	.39	8.9	9.1	S1
9	453	313	-.15	.04	.99	-.2	.96	-.4	.40	.39	5.6	5.5	S9
10	456	313	-.16	.04	.88	-3.0	1.01	.1	.46	.39	6.3	6.1	S10
7	486	313	-.21	.04	.94	-1.5	.91	-1.0	.41	.38	6.3	6.1	S7
22	489	313	-.22	.04	.92	-1.9	.88	-1.4	.42	.38	6.6	6.1	S22
14	510	313	-.25	.04	1.07	1.6	1.12	1.3	.32	.37	7.9	8.3	S14
12	534	313	-.29	.04	.94	-1.4	.91	-1.0	.39	.36	8.6	8.3	S12
23	552	313	-.33	.04	1.08	1.8	1.35	3.1	.29	.35	7.6	8.4	S23
15	558	313	-.34	.04	.89	-2.8	.89	-1.0	.41	.35	11.2	10.6	S15
6	567	313	-.35	.04	.83	-4.1	1.13	1.2	.44	.35	10.9	10.7	S6
21	612	313	-.43	.04	.89	-2.3	.85	-1.2	.39	.33	12.8	12.5	S21
MEAN	378.3	313.0	.06	.05	1.00	-.3	1.06	.4			21.7	22.5	
S.D.	125.9	.2	.25	.01	.10	1.8	.18	1.5			18.4	19.2	

Gambar 1. Uji Tingkat Kesukaran Butir Item

Tabel 4. Analisis *Item Measure*

Butir Soal	Jumlah	Kategori	Persentase (%)
18, 29, 27, 19, 4, 11	6	Sangat Sulit	20%
5, 24, 30, 16, 8, 3, 17, 13	8	Sulit	26,6%
25, 28, 26, 20, 2, 1, 9, 10, 7, 22, 14	11	Mudah	36,6%
12, 23, 15, 6, 2	5	Sangat Mudah	16,6%

Berdasarkan Gambar 1, nilai SD sebesar 0,25 digunakan sebagai batas kategori kesulitan, yaitu: sangat sulit ($>0,25$), sulit ($0,00-0,25$), mudah ($-0,25- <0,00$), dan sangat mudah ($<-0,25$). Hasil analisis pada 30 butir (Gambar 3) menunjukkan bahwa 6 butir (20%) tergolong sangat sulit, 8 butir (26,6%) sulit, 11 butir (36,6%) mudah, dan 5 butir (16,7%) sangat mudah. Rincian lengkap disajikan pada Tabel 4. Dalam perspektif *Rasch Model*, distribusi tingkat kesukaran butir yang tidak seimbang dapat memengaruhi ketepatan targeting antara kemampuan responden dan tingkat kesukaran item. Dominasi butir mudah bahwa instrumen lebih sensitif dalam mengukur kemampuan rendah hingga sedang, sehingga kurang optimal dalam membedakan responden berkemampuan tinggi. Namun, hasil penelitian menunjukkan distribusi yang belum proporsional, dengan dominasi butir mudah hingga sangat mudah sebesar

DOI: <https://doi.org/10.26877/jp3.v12i1.531>

53,3% mengindikasikan bahwa instrumen cenderung lebih dominan mengukur kemampuan pada level rendah hingga sedang. Kondisi ini berpotensi mengurangi sensitivitas instrumen dalam membedakan responden dengan kemampuan tinggi, sehingga diperlukan penyesuaian pada konstruksi butir soal. Oleh karena itu, diperlukan revisi dan penyeimbangan tingkat kesukaran butir agar distribusi mendekati normal serta meningkatkan daya diskriminasi instrumen (Mardapi, 2017 Sudaryono, 2012).

Tabel 5. Kategorisasi Tingkat Kemampuan Responden

Kategori	Rentang Person	Jumlah Responden	Persentase	Interpretasi
Sangat Tinggi	≥ 2.00 logit	9	2.9%	<i>Maximum measure</i> , menjawab sempurna semua item
Tinggi	0.97 – 1.22 logit	8	2.6%	Kemampuan di atas rata-rata
Sedang Tinggi	0.26 – 0.69 logit	45	14.4%	Kemampuan sedikit di atas rata-rata
Sedang	-0.26 – 0.20 logit	143	45.7%	Kemampuan rata-rata
Sedang Rendah	-0.045 – -0.32 logit	68	21.7%	Kemampuan sedikit di bawah rata-rata
Rendah	-0.69 – -0.52 logit	33	10.5%	Kemampuan di bawah rata-rata
Sangat Rendah	< -0.80 logit	7	2.2%	Kemampuan sangat rendah

Berdasarkan *Rasch Model*, kemampuan responden berada pada rentang $-1,20$ hingga $+2,00$ logit dengan rata-rata $-0,13$ logit, menunjukkan kemampuan siswa sedikit di bawah tingkat kesukaran item. Hal ini mengindikasikan bahwa instrumen cukup menantang namun tetap sesuai dengan karakteristik responden (Sumintono & Widhiarso, 2015).

Hasil analisis distribusi kemampuan pada Tabel 5 menunjukkan bahwa distribusi kemampuan responden menunjukkan bahwa 9 responden (2,9%) mencapai *maximum measure* sebesar 2,00 logit, menandakan kemampuan yang melampaui tingkat kesukaran seluruh item. Sebagian besar responden ($\pm 60\%$) berada pada kategori kemampuan sedang ($-0,26$ hingga $0,69$ logit), yang menunjukkan distribusi relatif normal dan kesesuaian instrumen dengan populasi sasaran. Sebanyak 25% responden berada pada kategori kemampuan rendah ($-0,80$ hingga $-0,26$ logit), sementara 6 responden (1,9%) tergolong sangat rendah dengan nilai di bawah $-0,80$ logit, di mana responden nomor 108 memiliki kemampuan terendah ($-1,20$ logit). Variasi ini mengindikasikan bahwa instrumen mampu mengukur kemampuan responden secara efektif pada berbagai tingkat, sesuai dengan prinsip pengukuran yang baik (Bond & Fox, 2015).

Rasch Model mengukur kemampuan responden dan tingkat kesukaran item pada skala logit yang sama dengan distribusi ideal berbentuk kurva normal. Distribusi kemampuan responden menunjukkan mayoritas berada pada kategori sedang (60%), diikuti rendah (25%), sangat rendah (1,9%), dan *maximum measure* (2,9%), dengan rentang pengukuran 3,20 logit. Distribusi ini

DOI: <https://doi.org/10.26877/jp3.v12i1.531>

mengindikasikan *targeting* instrumen yang cukup sesuai dengan populasi, namun proporsi responden pada *maximum measure* mengindikasikan *ceiling effect* yakni keterbatasan instrumen dalam membedakan responden berkemampuan tinggi (Wright & Stone, 2004), sementara sedikitnya responden pada kategori sangat rendah menunjukkan bahwa *floor effect* tidak menjadi masalah signifikan. Meskipun proporsi tersebut masih dalam batas wajar (Sumintono & Widhiarso, 2015; Aryadoust et al., 2019), penambahan item berkesukaran tinggi tetap diperlukan guna meningkatkan presisi pengukuran (Linacre, 2002; Wolfe & Smith, 2007). Ketidakseimbangan tingkat kesukaran juga dipengaruhi oleh kompleksitas konteks pencemaran lingkungan sebagai stimulus soal, sehingga penyeimbangan gradasi kompleksitas diperlukan agar instrumen mampu mengukur kemampuan *design thinking* secara proporsional pada seluruh rentang kemampuan.

3. Analisis Kesesuaian Butir *Item* dan *Person*

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL		INFIT		OUTFIT		PT-MEASURE		EXACT MATCH		ITEM
				S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%		
18	138	313	.59	.06	.97	-.2	1.01	.1	.56	.55	80.6	80.3	S18	
29	180	313	.43	.06	1.04	.4	1.09	.6	.48	.51	62.2	64.4	S29	
27	207	313	.35	.05	1.18	1.8	1.63	3.7	.36	.50	47.4	52.3	S27	
19	213	313	.34	.05	1.08	.9	1.18	1.2	.44	.49	49.0	52.0	S19	
4	222	313	.31	.05	1.05	.6	1.19	1.4	.45	.49	43.4	45.7	S4	
11	234	313	.28	.05	1.19	2.1	1.31	2.2	.37	.48	40.1	45.3	S11	
5	276	313	.18	.05	.98	-1.4	.83	-1.7	.52	.46	30.6	30.0	S5	
24	285	313	.16	.05	1.13	1.9	1.17	1.6	.38	.46	28.3	29.8	S24	
30	300	313	.13	.05	.98	-.2	1.10	1.0	.44	.45	24.0	25.5	S30	
16	303	313	.13	.05	.95	-.8	.89	-1.1	.48	.45	26.3	25.4	S16	
8	312	313	.11	.05	1.16	2.5	1.24	2.3	.35	.44	23.4	25.3	S8	
3	336	313	.06	.04	.98	-.4	1.20	2.1	.43	.43	17.8	18.5	S3	
17	336	312	.06	.04	1.18	3.1	1.25	2.5	.32	.43	16.5	18.2	S17	
13	357	313	.02	.04	.88	-2.4	.89	-1.2	.49	.43	12.8	13.0	S13	
25	378	313	-.02	.04	.92	-1.7	.86	-1.6	.47	.42	13.5	12.9	S25	
28	381	313	-.03	.04	.97	-.6	.93	-.8	.44	.42	13.5	12.9	S28	
26	399	313	-.06	.04	.98	-.5	.97	-.3	.42	.41	12.5	12.5	S26	
20	405	313	-.07	.04	.91	-1.9	.88	-1.4	.46	.41	8.9	8.9	S20	
2	426	313	-.11	.04	.99	-.1	.98	-.2	.40	.40	8.6	9.2	S2	
1	444	313	-.14	.04	1.04	1.0	1.04	.5	.37	.39	8.9	9.1	S1	
9	453	313	-.15	.04	.99	-.2	.96	-.4	.40	.39	5.6	5.5	S9	
10	456	313	-.16	.04	.88	-3.0	1.01	1.1	.46	.39	6.3	6.1	S10	
7	486	313	-.21	.04	.94	-1.5	.91	-1.0	.41	.38	6.3	6.1	S7	
22	489	313	-.22	.04	.92	-1.9	.88	-1.4	.42	.38	6.6	6.1	S22	
14	510	313	-.25	.04	1.07	1.6	1.12	1.3	.32	.37	7.9	8.3	S14	
12	534	313	-.29	.04	.94	-1.4	.91	-1.0	.39	.36	8.6	8.3	S12	
23	552	313	-.33	.04	1.08	1.8	1.35	3.1	.29	.35	7.6	8.4	S23	
15	558	313	-.34	.04	.89	-2.8	.89	-1.0	.41	.35	11.2	10.6	S15	
6	567	313	-.35	.04	.83	-4.1	1.13	1.2	.44	.35	10.9	10.7	S6	
21	612	313	-.43	.04	.89	-2.3	.85	-1.2	.39	.33	12.8	12.5	S21	
MEAN	378.3	313.0	.00	.05	1.00	-.3	1.06	.4			21.7	22.5		
S.D.	125.9	.2	.25	.01	.10	1.8	.18	1.5			18.4	19.2		

Gambar 2. Uji Tingkat Kesesuaian Butir Item

Analisis kesesuaian butir (*item fit*) dalam Rasch Model dilakukan untuk mengevaluasi apakah respons terhadap suatu item sesuai dengan prediksi model. Dalam penelitian ini, evaluasi kesesuaian butir diprioritaskan pada nilai OUTFIT Mean Square (MNSQ) dengan rentang penerimaan 0,5–1,5. Apabila kriteria tersebut tidak terpenuhi, maka analisis dilanjutkan menggunakan nilai OUTFIT ZSTD (-2 hingga +2) serta PT-Measure Correlation yang harus bernilai positif (Bond & Fox, 2015; Linacre, 2002). Dalam praktiknya, tidak semua kriteria harus terpenuhi secara simultan, namun interpretasi dilakukan secara bertahap untuk menentukan kelayakan item.

Berdasarkan analisis pada Gambar 2, sebanyak 24 dari 30 butir (80%) dinyatakan *fit*, sedangkan 6 (20%) butir (S27, S23, S17, S8, S11, dan S6)

tergolong *misfit* karena tidak memenuhi kriteria OUTFIT MNSQ dan/atau ZSTD (Bond & Fox, 2015; Linacre, 2002). Ketidaksesuaian ini disebabkan oleh faktor seperti redaksi yang ambigu, tingkat kesukaran yang tidak sesuai, serta pengecoh yang kurang berfungsi (Boone et al., 2014). Oleh karena itu, dilakukan revisi berupa penyederhanaan redaksi, penyesuaian tingkat kesukaran, perbaikan pengecoh, dan penguatan konteks soal, serta direkomendasikan untuk dilakukan kalibrasi ulang.

Selain itu, analisis *person fit* menunjukkan adanya responden yang tidak konsisten, yang dapat disebabkan oleh ketidaktelitian, menebak, atau kesalahan pemahaman (Boone, 2016). Untuk meningkatkan akurasi model, dilakukan *data cleaning* dengan menghapus responden *misfit* dan melakukan kalibrasi ulang agar diperoleh model yang lebih stabil (Sumintono & Widhiarso, 2015). Dengan demikian, meskipun sebagian besar item telah memenuhi kriteria kesesuaian, revisi dan kalibrasi ulang tetap diperlukan untuk memastikan validitas dan konsistensi instrumen dalam kerangka Rasch Model.

Tabel 6. Uji Kesesuaian *Person*

No. Entry	Person ID	Total Score	Person Measure	Infit MNSQ	Outfit MNSQ	ZSTD Outfit	Interpretasi
173	173L	87	1.22	1.10	3.73	1.7	Sangat misfit
70	70L	15	-0.60	1.26	2.33	2.3	Sangat misfit
161	161L	18	-0.52	1.28	2.16	2.4	Sangat misfit
154	154P	87	1.22	1.09	2.10	1.1	Misfit
180	180L	78	0.69	1.30	2.07	1.7	Misfit
198	198L	12	-0.69	1.08	2.08	1.7	Misfit

Berdasarkan analisis *Person Statistics: Misfit Order*, terdapat enam responden yang teridentifikasi misfit, yaitu 173L, 70L, 161L, 154P, 180L, dan 198L. Hal ini ditunjukkan oleh nilai OUTFIT dan INFIT MNSQ yang berada di luar rentang ideal serta nilai ZSTD yang melebihi batas +2, sehingga menunjukkan pola respons yang tidak sesuai dengan prediksi Model Rasch. Dalam analisis Rasch, statistik *fit* digunakan untuk mengevaluasi kesesuaian pola respons individu terhadap model probabilistik, dengan kriteria MNSQ berada pada rentang 0,5–1,5 dan ZSTD antara –2 hingga +2 (Bond & Fox, 2015; Linacre, 2002). Ketidaksesuaian respons ini dapat disebabkan oleh berbagai faktor, seperti ketidaktelitian, kecenderungan menebak, kelelahan, atau kesalahan dalam memahami soal, yang berpotensi meningkatkan *measurement error* dan menurunkan akurasi estimasi kemampuan (Boone et al., 2014; Karabatsos, 2003).

Temuan ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa Keberadaan responden *misfit* dalam jumlah kecil masih dapat ditoleransi selama tidak mengganggu validitas konstruk secara keseluruhan (Sumintono & Widhiarso, 2015), Namun, responden dengan tingkat misfit yang tinggi tidak hanya diidentifikasi, tetapi juga telah ditindaklanjuti melalui proses *data cleaning* berupa eliminasi selektif dan analisis ulang (*recalibration*) untuk meningkatkan akurasi model pengukuran (Boone, 2016). Ketidaksesuaian respons juga dapat dipengaruhi oleh kompleksitas konteks pencemaran lingkungan yang

DOI: <https://doi.org/10.26877/jp3.v12i1.531>

memungkinkan interpretasi berbeda antarsiswa, sehingga revisi kejelasan konteks dan relevansi skenario diperlukan untuk meningkatkan akurasi dan konsistensi pengukuran kemampuan *design thinking* (KLHK, 2020; Tchobanoglous et al., 2014).

4. Analisis *Reliability* Instrumen

Reliabilitas instrumen dalam *Rasch Model* dievaluasi melalui *person reliability*, *item reliability*, dan *Cronbach's Alpha* sebagai indikator konsistensi pengukuran pada tingkat responden maupun butir soal. Hasil analisis menunjukkan nilai *Cronbach's Alpha* sebesar 0,86 (sangat tinggi), *person reliability* sebesar 0,78 (memadai dalam membedakan tingkat kemampuan responden), dan *item reliability* sebesar 0,97 (stabilitas hierarki kesukaran item sangat baik). Kombinasi ketiga indikator tersebut menegaskan bahwa instrumen memiliki konsistensi pengukuran yang kuat, dengan nilai reliabilitas yang lebih tinggi dibandingkan pendekatan *Classical Test Theory* (Królikowska et al., 2023), sehingga instrumen dinyatakan reliabel untuk mengukur kemampuan *design thinking* siswa. Adapun untuk nilai reliabilitas terlihat dari hasil *Summary Statistics* yang disajikan pada Gambar 3 dan Gambar 4.

SUMMARY OF 313 MEASURED (EXTREME AND NON-EXTREME) PERSON

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	36.3	30.0	-.13	.17				
S.D.	18.6	.1	.50	.16				
MAX.	90.0	30.0	2.00	1.06				
MIN.	3.0	29.0	-1.20	.13	.66	-2.4	.50	-2.3
REAL RMSE	.24	TRUE SD	.45	SEPARATION	1.88	PERSON RELIABILITY		.78
MODEL RMSE	.23	TRUE SD	.45	SEPARATION	1.91	PERSON RELIABILITY		.78
S.E. OF PERSON MEAN = .03								

PERSON RAW SCORE-TO-MEASURE CORRELATION = .95
 CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .86

Gambar 3. Uji Reliabilitas *Instrument Person*

SUMMARY OF 30 MEASURED (NON-EXTREME) ITEM

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	378.3	313.0	.00	.05	1.00	-.3	1.06	.4
S.D.	125.9	.2	.25	.01	.10	1.8	.18	1.5
MAX.	612.0	313.0	.59	.06	1.19	3.1	1.63	3.7
MIN.	138.0	312.0	-.43	.04	.83	-4.1	.83	-1.7
REAL RMSE	.05	TRUE SD	.25	SEPARATION	5.33	ITEM RELIABILITY		.97
MODEL RMSE	.05	TRUE SD	.25	SEPARATION	5.45	ITEM RELIABILITY		.97
S.E. OF ITEM MEAN = .05								

Gambar 4. Uji Reliabilitas *Instrument Item*

Temuan ini sejalan dengan penelitian validasi instrumen sebelumnya yang melaporkan nilai Cronbach's Alpha pada rentang 0,70–0,93 sebagai indikator reliabilitas tinggi (ŞahİN & Yağci, 2017), serta menunjukkan kualitas psikometrik

DOI: <https://doi.org/10.26877/jp3.v12i1.531>

yang sebanding dengan instrumen internasional berbasis Rasch. Tingginya reliabilitas instrumen memberikan implikasi penting dalam pengukuran kemampuan *design thinking*, karena memungkinkan evaluasi kemampuan siswa secara konsisten dalam menganalisis dan merumuskan solusi terhadap permasalahan lingkungan nyata, seperti pencemaran sungai di wilayah perkotaan (Cai & Yang, 2023; Kementerian Lingkungan Hidup dan Kehutanan, 2020; Jambeck et al., 2015).

KESIMPULAN DAN SARAN

Berdasarkan analisis *Rasch Model*, instrumen kemampuan *design thinking* siswa SMA Negeri 15 Semarang dinyatakan valid dan reliabel, ditunjukkan oleh terpenuhinya kriteria unidimensionalitas (*raw variance explained by measures* 23,5% dan *unexplained variance in 1st contrast* 4,8%) serta reliabilitas yang sangat baik (Cronbach Alpha 0,86; *person reliability* 0,78; *item reliability* 0,97). Sebanyak 80% butir soal memenuhi kriteria *item fit*, sementara beberapa butir memerlukan revisi. Dominasi butir mudah menunjukkan instrumen lebih sensitif pada kemampuan rendah hingga sedang dan berpotensi menimbulkan *ceiling effect*. Secara umum, instrumen layak digunakan untuk mengukur kemampuan *design thinking* siswa pada materi pencemaran lingkungan. Disarankan dilakukan revisi butir *misfit* dan penambahan soal berkesukaran tinggi guna meningkatkan akurasi pengukuran. Penelitian selanjutnya dapat memperluas penerapan instrumen pada materi biologi lain maupun sampel yang lebih beragam, serta mengintegrasikan konteks lingkungan lokal dan asesmen autentik berbasis proyek untuk memperoleh pengukuran kemampuan *design thinking* yang lebih komprehensif.

DAFTAR PUSTAKA

- Aryadoust, V., Tan, H. A. H., & Ng, L. Y. (2019). A Scientometric Review of Rasch Measurement: The Rise and Progress of a Specialty. *Frontiers in Psychology, 10*(2197). <https://doi.org/10.3389/fpsyg.2019.02197>
- Azwar, S. (2017). *Reliabilitas dan validitas* (Edisi 4). Pustaka Pelajar.
- Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch Model*. Routledge.
- Boone, W. J. (2016). Rasch Analysis for Instrument Development: Why , When , and How? *CBE-Life Sciences Education, 1*–7. <https://doi.org/10.1187/cbe.16-04-0148>
- Boone, W. J., & Noltemeyer, A. (2017). Rasch analysis: A primer for school psychology researchers and practitioners. *Cogent Education, 4*(1). <https://doi.org/10.1080/2331186X.2017.1416898>
- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences* (1st ed.). Springer Dordrecht. <https://doi.org/https://doi.org/10.1007/978-94-007-6857-4>
- Cai, Y., & Yang, Y. (2023). The development and validation of the scale of design thinking for teaching (SDTT). *Thinking Skills and Creativity, 48*. <https://doi.org/doi.org/10.1016/j.tsc.2023.101255>
- Carroll, M., Goldman, S., Britos, L., Koh, J., Royalty, A., & Hornstein, M. (2010).

DOI: <https://doi.org/10.26877/jp3.v12i1.531>

- Destination , Imagination and the Fires Within : Design Thinking in a Middle School Classroom. *JADE*, 29(1).
- Dosi, C., Rosati, F., & Vignoli, M. (2018). Measuring Design Thinking Mindset. *International Design Conference*, 1991–2002. <https://doi.org/doi.org/10.21278/idc.2018.0493>
- Fadhilah, R., Setiawati, S., & Ahman, A. (2024). Skala Adiksi Media Sosial: Analisis Validitas dan Reliabilitas Menggunakan Rasch Model. *Jurnal EDUCATIO (Jurnal Pendidikan Indonesia)*, 10(2), 158–168. <https://doi.org/https://doi.org/10.29210/1202424135>
- Forum, W. E. (2023). *Future of Jobs Report*. weforum.
- Gunawan, F. M., Suryana, D., & Rahmanadia, H. (2023). Analysis of Validity and Reliability of Self-Acceptance Scale Using Rasch Model. *Jurnal Pendidikan*, 24(2), 71–80. <https://doi.org/10.33830/jp.v24i2.4604.2023>
- Indonesia, K. L. H. dan K. R. (2020). *Status Lingkungan Hidup Indonesia*.
- Inhelder, B., & Piaget, J. (1958). *The Growth of Logical Thinking from Childhood to Adolescence*. Basic Books, Inc.
- Jambeck, J. R., Geyer, R., Wilcox, C., Siegler, T. R., Perryman, M., Andrady, A., Narayan, R., & Law, K. L. (2015). *Plastic Waste Inputs from Land Into the Ocean* (Issue 6223).
- Karyanto, Roesminingsih, M. V., & Riyanto, Y. (2024). Soft Skills-Based School Evaluation in Indonesian Secondary Schools. *International Journal of Innovative Research in Education*, 11(2), 141–149. <https://doi.org/doi.org/10.18844/ijire.v11i2.9696>
- Kementerian Lingkungan Hidup dan Kehutanan. (2020). *Status lingkungan hidup Indonesia*.
- Królikowska, A., Reichert, P., Karlsson, J., Mouton, C., Becker, R., & Prill, R. (2023). Improving the Reliability of Measurements in Orthopaedics and Sports Medicine. *Knee Surgery, Sports Traumatology, Arthroscopy*, 31(12), 5277–5285. <https://doi.org/10.1007/s00167-023-07635-1>
- Mardapi, D. (2017). *Pengukuran Penilaian dan Evaluasi Pendidikan* (2nd ed.). Parama Publishing.
- Maulana, S., Rusilowati, A., Nugroho, S. E., & Susilaningsih, E. (2021). Implementasi Rasch Model dalam Pengembangan Instrumen Tes Diagnostik. *Prosiding Seminar Nasional Pascasarjana Universitas Negeri Semarang*, 748–757. <http://pps.unnes.ac.id/pps2/prodi/prosiding-pascasarjana-unnes>
- Muntazhimah, Putri, S., & Khusna, H. (2020). Rasch Model untuk Memvalidasi Instrumen Resiliensi Matematis Mahasiswa Calon Guru Matematika. *Jkpm: Jurnal Kajian Pendidikan Matematika*, 2682(1), 65–74. <http://journal.lppmunindra.ac.id/index.php/jkpm/>
- OECD. (2019). *OECD future of education and skills 2030*. OECD Publishing.
- Patria, A. S., Sutiadiningsih, A., & Kristiana, N. (2024). Instrument Development of Design Thinking Test for Graphic Design Students. *IJORER: International Journal of Recent Educational Research*, 5(4), 796–807. <https://doi.org/doi.org/10.46245/ijorer.v5i4.630>

DOI: <https://doi.org/10.26877/jp3.v12i1.531>

- Razzouk, R., & Shute, V. (2012). What is design thinking and why is it important? *Review of Educational Research*, 82(3), 330–348.
- Şahin, C., & Yağci, M. (2017). Sosyal Medya Bağımlılığı Ölçeği- Yetişkin Formu: Geçerlilik ve Güvenirlik Çalışması. *Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi (KEFAD)*, 18(1), 523–538.
- Sasongko, P., Ratu, I. M., & Hayati, M. N. (2021). Tes Kemampuan Berfikir Tingkat Tinggi untuk Mengukur Literasi Lingkungan Hidup Bagi Siswa Sekolah Menengah Pertama dengan Aplikasi Model Rasch. *Pancasakti Science Education Journal*, 6(2), 93–101. <https://doi.org/10.24905/psej.v6i2.127>
- Semarang, B. P. S. K. (2022). *Profil Lingkungan Hidup Kota Semarang 2022*. BPS Kota Semarang.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan RASCH pada Assessment Pendidikan*. Penerbit Trim Komunikata.
- Tchobanoglous, G., Stensel, H. D., Tsuchihashi, R., & Burton, F. L. (2014). *Wastewater Engineering: Treatment and Resource Recovery* (5th ed.). Metcalf & Eddy.