

Implementasi Rasch Model dalam Pengembangan Instrumen Tes Diagnostik

Slamet Maulana, Ani Rusilowati*, Sunyoto Eko Nugroho, Endang Susilaningsih

Pendidikan IPA, Fakultas Pascasarjana, Universitas Negeri Semarang, Gedung A Kampus Pascasarjana Jl
Kelud Utara III, Semarang 50237, Indonesia

*Corresponding Author: rusilowati@mail.unnes.ac.id

Abstrak. Keberadaan kurikulum merdeka memberikan alasan untuk segera dilakukan penelitian baru. Dimana pada standar penilaian kurikulum merdeka, guru wajib untuk melakukan diagnostik kognitif awal peserta didik sebelum proses pembelajaran dimulai. Namun demikian, hasil observasi di lapangan menunjukkan masih terbatasnya instrumen tes diagnostik. Selain itu, IPA menjadi salah satu mata pelajaran yang perlu menjadi fokus. Kompetensi peserta didik dalam penguasaan konsep IPA masih terbilang rendah. Pengembangan instrumen tes diagnostik pada materi IPA menjadi salah satu kebutuhan dewasa ini. Rasch model menjadi salah satu model analisis yang dapat digunakan untuk mengetahui kelayakan suatu instrumen. Sehingga penelitian ini bertujuan untuk mengembangkan instrumen tes diagnostik pada materi IPA melalui implementasi Rasch Model. Penelitian ini merupakan penelitian dan pengembangan. Adapun model penelitian dan pengembangan yang digunakan adalah model ADDIE. Hasil implementasi Rasch Model menunjukkan bahwa instrumen tes diagnostik pada materi IPA yang telah dikembangkan adalah valid dari segi konten maupun konstruk, serta reliabel. Berdasarkan hasil penelitian serta pembahasan yang telah diuraikan, penelitian menghasilkan kesimpulan bahwa produk pendidikan berupa instrumen tes diagnostik pada materi IPA layak digunakan untuk digunakan pada khalayak umum untuk mendiagnosis pengetahuan awal peserta didik pada materi IPA.

Kata kunci: Rasch Model; Instrumen Tes Diagnostik; IPA; Kurikulum Merdeka.

Abstract. The existence of Merdeka Curriculum provides a reason for new research to be carried out immediately. Where in the Merdeka Curriculum's assessment standards, teachers are required to conduct initial cognitive diagnostics of students before the learning process begins. However, the results of observations in the field show that there are still limited diagnostic test instruments. In addition, science is one of the subjects that need to be focused. The competence of students in mastering science concepts is still relatively low. The development of diagnostic test instruments on science material is one of the needs today. The Rasch model is one of the analytical models that can be used to determine the feasibility of an instrument. So, this study aims to develop diagnostic test instruments on science material through the implementation of the Rasch Model. This research is research and development. The research and development model used is the ADDIE model. The results of the implementation of the Rasch Model show that the diagnostic test instruments on the science material that have been developed are valid in terms of content and construct, as well as reliable. Based on the results of the research and the discussion that has been described, the study resulted in the conclusion that educational products in the form of a diagnostic test instruments on science material are suitable for use in the general public to diagnose students' initial knowledge on science material.

Key words: Rasch Model; Diagnostic Test Instrument; Science; Merdeka Curriculum

How to Cite: Maulana, S., Rusilowati, A., Nugroho, S. E., & Susilaningsih, E. (2023). Implementasi Rasch Model dalam Pengembangan Instrumen Tes Diagnostik. *Prosiding Seminar Nasional Pascasarjana, 2023*, 748-757.

PENDAHULUAN

Pada perkembangan kurikulum terbaru saat ini (Kurikulum Merdeka), proses asesmen dilaksanakan dalam beberapa tahap, salah satunya adalah asesmen diagnostik. Asesmen diagnostik dilaksanakan sebagai bentuk penggalian informasi awal pengetahuan peserta didik.

Keberadaan kurikulum merdeka yang diterapkan memberikan alasan untuk segera dilakukan penelitian baru. Dimana pada standar penilaian kurikulum merdeka, guru wajib untuk melakukan diagnostik kognitif awal peserta didik sebelum proses pembelajaran dimulai. Namun demikian, hasil observasi di lapangan

menunjukkan masih terbatasnya instrument tes diagnostik.

Para guru menyatakan bahwa mereka masih *awam* terhadap pelaksanaan asesmen diagnostik sebelum pembelajaran. Hal ini menunjukkan adanya kebutuhan instrument tes diagnostik di sekolah.

IPA menjadi salah satu mata pelajaran yang perlu menjadi fokus. Kompetensi peserta didik dalam penguasaan konsep IPA masih terbilang rendah (Dewi, 2019; Suroso et al., 2021; Usmeldi & Amini, 2019). Hal ini tentu menjadi sebuah awal yang kurang baik dalam menumbuhkembangkan peserta didik.

Menindaklanjuti temuan tersebut, maka penelitian untuk mengembangkan instrumen tes

diagnostik masih sangat dibutuhkan. Dengan demikian, guru tidak lagi mengalami kesulitan dalam melakukan diagnostik awal peserta didik pada Kurikulum Merdeka di tahun 2024 kelak. Terlebih lagi, pelaksanaan asesmen diagnostik ini menjadi urgen dalam memetakan penguasaan konsep dasar peserta didik.

Pengembangan instrumen tes diagnostik dalam implementasinya pada Kurikulum Merdeka pada dewasa ini menjadi penting. Namun, apakah instrumen tes diagnostik yang dikembangkan tersebut dapat langsung digunakan untuk umum. Tentu instrumen tes diagnostik tersebut harus memiliki kekuatan yang semestinya. Sehingga ketika digunakan kelak, instrumen tes diagnostik tersebut memiliki kevalidan baik. Alhasil, hasil analisis yang diberikan oleh seperangkat tes diagnostik tersebut dapat dipercaya. Pada akhirnya, hasil evaluasi yang diberikan dari instrumen tes diagnostik yang dikembangkan dapat dengan jelas menggambarkan pengetahuan awal peserta didik.

Rasch model merupakan salah satu model analisis yang dapat digunakan untuk mengetahui kelayakan suatu instrumen. Rasch Model dapat digunakan dalam menentukan indeks reliabilitas hingga analisis item ke masing-masing tingkat, reliabilitas responden, serta dimensionalitas dan mendeteksi adanya bias dari item pada instrumen (Suryani, 2018). Rasch Model juga mampu menentukan kemampuan siswa, dimana kemungkinan seorang siswa menjawab suatu item dengan benar akan meningkat secara monoton dengan kemampuan yang dimiliki siswa tersebut (Islam et al., 2020; Xiao et al., 2018).

Rasch Model telah digunakan dalam analisis berbagai instrumen penilaian dalam memberikan informasi rinci tentang beberapa aspek kualitas dari tes (Ling et al., 2014). Rasch Model tidak hanya digunakan untuk menganalisis item instrumen, tetapi juga dapat menganalisis peserta didik (Matore et al., 2018; Zamri & Nordin, 2015).

Analisis Rasch Model mewakili pendekatan dinamis untuk mencapai validitas konstruk (Semakula et al., 2017). Analisis Rasch Model mendukung dalam pengembangan dan memvalidasi pengukuran invarian dalam menyediakan bukti empiris dan wawasan untuk sifat psikometrik penting dari tes dan kuesioner (Sjaastad, 2014). Rasch Model juga mampu menjelaskan perbedaan dalam kesulitan tingkatan yang berbeda (Fulmer et al., 2016). Rasch Model juga mampu menentukan kemampuan siswa, dimana kemungkinan seorang siswa menjawab

suatu item dengan benar akan meningkat secara monoton dengan kemampuan yang dimiliki siswa tersebut (Xiao et al., 2018).

Berdasarkan uraian-uraian di atas, maka penelitian ini bertujuan untuk mengembangkan instrumen tes diagnostik pada materi IPA melalui implementasi Rasch Model.

METODE

Penelitian ini merupakan penelitian dan pengembangan. Adapun model penelitian dan pengembangan yang digunakan adalah model ADDIE. Penelitian ini terdiri dari lima tahapan yaitu *Analyze, Design, Develop, Implement, dan Evaluation*.

Pada tahapan *analyze*, penelitian diawali dengan melakukan studi literatur dan observasi ke lapangan. Hasil studi literatur menunjukkan bahwa instrumen tes diagnostik masih menjadi salah satu masalah pendidikan, khususnya pada mata pelajaran IPA di dalam Kurikulum Merdeka. Namun, instrumen tes diagnostik tersebut juga memerlukan validasi terlebih dahulu sebelum diimplementasikan secara umum. Hal ini juga didukung dari hasil observasi bahwa sejumlah guru telah memiliki seperangkat tes diagnostik, namun belum mengetahui kevalidan perangkat tes tersebut. Sehingga pada tahapan ini diperoleh tujuan untuk mengembangkan instrumen tes diagnostik pada materi IPA yang valid.

Pada tahapan *design, blueprint* instrumen tes diagnostik pada materi IPA mulai dikembangkan. Instrumen tes diagnostik pada materi IPA yang akan dikembangkan mengacu pada kesesuaian Capaian Pembelajaran dan Analisis Tujuan Pembelajaran pada Kurikulum Merdeka. Instrumen tes diagnostik pada materi IPA yang akan dikembangkan nantinya akan divalidasi secara empirik melalui Rasch Model.

Pada tahapan *develop, blueprint* instrumen tes diagnostik pada materi IPA yang telah selesai dibuat mulai dikembangkan dan direalisasikan. Instrumen validasi teoritik yang telah dikembangkan ini kemudian divalidasi terlebih dahulu dengan menghitung nilai CVR dan CVI.

Pada tahapan *implement*, instrumen tes diagnostik pada materi IPA yang telah dinyatakan valid melalui skor CVR dan CVI diimplementasikan di lapangan. Instrumen tes diagnostik pada materi IPA tersebut kemudian digunakan oleh sejumlah peserta didik. Hasil pengisian kemudian dikumpulkan dan dianalisis melalui Rasch model. Melalui implementasi Rasch model, validasi secara empirik dilakukan

untuk mengetahui kevalidan instrumen tes diagnostik pada materi IPA yang telah dikembangkan.	12	1
	13	1
	14	1
	15	0,75
	16	1
	17	0,75
	18	0,75
	19	1
	20	1

HASIL DAN PEMBAHASAN

Penelitian ini telah menghasilkan instrumen tes diagnostik materi IPA yang dikembangkan dengan 20 butir pertanyaan di dalamnya.

CVR and CVI

CVR dihitung untuk mengetahui validitas setiap butir pernyataan dari setiap instrumen yang telah dikembangkan. Sedangkan nilai CVI digunakan untuk mengetahui validitas suatu instrumen secara utuh satu kesatuan. Instrumen tes diagnostik materi IPA menggunakan delapan validator untuk memberikan penilaian apakah setiap butir soal di set tes tersebut valid atau tidak.

Hasil validasi terhadap instrumen tes diagnostik materi IPA memberikan nilai CVR yang beragam. Sebanyak 7 butir soal memiliki nilai CVR sebesar 0,75 dan sisanya sebanyak 13 butir soal memiliki nilai CVR sebesar 1. Nilai tersebut menginformasikan bahwa setiap butir soal pada set instrumen tes diagnostik materi IPA dinyatakan valid. Adapun nilai CVI untuk set instrumen tes diagnostik materi IPA bernilai akhir 0,9125. Nilai tersebut juga menginformasikan bahwa set instrumen tes diagnostik materi IPA dinyatakan valid dengan kriteria sangat baik. Sehingga instrumen tes diagnostik materi IPA dapat dinyatakan valid baik setiap butirnya maupun secara keseluruhan instrumen.

Tabel 1. Nilai CVR Setiap Butir Soal pada Instrumen Tes Diagnostik

Nomor Butir Soal	Nilai CVR
1	0,75
2	1
3	0,75
4	0,75
5	1
6	0,75
7	1
8	1
9	1
10	1
11	1

Validasi teoritik menggunakan perhitungan nilai *Content Validity Ratio* (CVR) dan *Content Validity Index* (CVI). Hasil perhitungan nilai CVR untuk setiap butir soal memberikan nilai memenuhi batas minimum. Nilai tersebut menjelaskan bahwa setiap butir soal yang telah dikembangkan telah dinyatakan valid dan dapat digunakan. Hal ini sejalan dengan penelitian terdahulu mengenai pengembangan suatu instrumen yang menyatakan bahwa validitas konten (melalui perhitungan CVR) sangat penting dan krusial (Alizadeh-Siuki et al., 2020; Waltz et al., 2016).

Setiap butir soal yang dinyatakan valid belum tentu menjadi valid jika telah disatukan menjadi suatu instrumen tes. Maka setelah setiap butir soal dinyatakan valid berdasarkan kriteria minimum nilai CVR, maka dilakukan perhitungan nilai *Content Validity Index* (CVI) terhadap instrumen tes yang telah dikembangkan dan tersusun dari 20 soal yang telah dikembangkan sebelumnya. Hasil perhitungan nilai CVI memberikan nilai yang juga melebihi batas minimum. Hal ini menjelaskan bahwa seluruh butir soal yang telah dikembangkan jika ditinjau secara keseluruhan sebagai suatu instrumen utuh dinyatakan valid dan dapat digunakan. Konten validitas memastikan bahwa butir item maupun instrumen tes yang telah dirancang dan dikembangkan benar dapat mengukur (Al-Taweel & Awad, 2020).

Uraian-uraian diatas jelas menginformasikan bahwa instrumen tes diagnostik pada materi IPA dapat dinyatakan valid baik setiap butirnya maupun secara keseluruhan instrumen. Selain itu, instrumen

Rasch Model Analysis

Setelah instrumen tes diagnostik pada materi IPA dinyatakan valid melalui nilai CVR dan CVI, instrumen tes diagnostik pada materi IPA kemudian diimplementasikan ke lapangan. Instrumen validasi teoritik tersebut diimplementasikan untuk menilai validitas dari instrumen tes diagnostik yang telah dikembangkan juga. Hasil implementasi di

lapangan kemudian dianalisis menggunakan Rasch Model untuk mengetahui kualitas instrumen yang telah dikembangkan.

Hasil analisis Rasch Model menunjukkan bahwa instrumen tes diagnostik materi IPA memiliki nilai *Raw Variance* sebesar 27,6%. Hal

ini menginformasikan bahwa instrumen tes diagnostik materi IPA memiliki validitas konstruk yang baik. Hal ini menginformasikan bahwa instrumen tes diagnostik materi IPA benar-benar mengukur apa yang seharusnya diukur.

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units			
	Eigenvalue	Observed	Expected
Total raw variance in observations =	23.2328	100.0%	100.0%
Raw variance explained by measures =	6.2328	26.8%	27.6%
Raw variance explained by persons =	2.8292	12.2%	12.5%
Raw Variance explained by items =	3.4035	14.6%	15.1%
Raw unexplained variance (total) =	17.0000	73.2%	100.0%
Unexplned variance in 1st contrast =	2.1704	9.3%	12.7%
Unexplned variance in 2nd contrast =	1.9423	8.4%	11.5%
Unexplned variance in 3rd contrast =	0.6884	3.0%	4.1%
Unexplned variance in 4th contrast =	0.6797	2.9%	3.9%
Unexplned variance in 5th contrast =	0.4458	1.9%	2.6%

Gambar 1. Nilai *Raw Variance* Instrumen Tes Diagnostik.

Hasil analisis Rasch Model lainnya juga menunjukkan bahwa instrumen tes diagnostik materi IPA memiliki reliabilitas yang cukup bagus. Hal ini ditunjukkan dari nilai item reliability

sebesar 0,75. Sehingga dapat dimaknai bahwa reliabilitas setiap butir soal di dalam instrumen tes diagnostik materi IPA memiliki reliabilitas yang cukup bagus.

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MEAN	60.7	73.0	.00	.37	1.00	.09	.96	.01
SEM	1.4	.0	.18	.02	.02	.15	.06	.20
P.SD	6.1	.0	.76	.09	.09	.64	.28	.88
S.SD	6.2	.0	.78	.10	.09	.66	.29	.90
MAX.	70.0	73.0	1.25	.60	1.21	1.96	1.80	2.12
MIN.	48.0	73.0	-1.51	.27	.86	-.97	.48	-1.25
REAL RMSE	.38	TRUE SD	.66	SEPARATION	1.71	Item	RELIABILITY	.75
MODEL RMSE	.38	TRUE SD	.66	SEPARATION	1.74	Item	RELIABILITY	.75
S.E. OF Item	MEAN = .18							

Gambar 2. Nilai Reliabilitas Instrumen Tes Diagnostik.

Hasil analisis Rasch Model kemudian dilanjutkan untuk mengetahui validitas setiap butir soal pada instrumen tes diagnostik materi IPA. Suatu butir soal dinyatakan valid jika memenuhi satu dari tiga syarat. Adapun ketiga syarat tersebut adalah nilai *outfit means-square* antara 0,5 dan 1,5; nilai *outfit z-standard* antara -

2,0 dan 2,0; serta nilai *point measure correlation* antara 0,4 dan 0,85. Hasil analisis Rasch Model menunjukkan bahwa seluruh butir soal pada instrumen tes diagnostik materi IPA dinyatakan valid karena memenuhi setidaknya satu dari tiga syarat.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S. E.	INFIT		OUTFIT		PTMEASUR-AL		EXACT OBS%	MATCH EXP%	Item	
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.				
10	62	73	.00	.34	1.18	.87	1.80	2.12	A	.07	.29	83.3	83.3	S10
11	59	73	.33	.32	1.12	.76	1.28	1.09	B	.20	.32	78.8	78.9	S11
20	48	73	1.25	.27	1.21	1.96	1.27	1.86	C	.24	.41	60.6	67.5	S20
13	63	73	-.12	.36	.98	-.02	1.16	.56	D	.27	.28	84.8	84.9	S13
9	63	73	-.12	.36	1.11	.53	1.12	.46	E	.19	.28	84.8	84.9	S9
1	55	73	.70	.29	.97	-.17	1.06	.36	F	.37	.36	75.8	73.7	S1
3	63	73	-.12	.36	1.04	.24	.97	.04	G	.26	.28	84.8	84.9	S3
19	55	73	.70	.29	1.03	.29	1.04	.26	H	.33	.36	69.7	73.7	S19
16	65	73	-.40	.39	1.02	.16	.86	-.23	I	.25	.25	87.9	87.9	S16
4	67	73	-.74	.44	.99	.09	.80	-.28	J	.24	.22	90.9	90.9	S4
5	67	73	-.74	.44	.99	.09	.83	-.20	j	.24	.22	90.9	90.9	S5
8	68	73	-.95	.48	.99	.09	.96	.10	i	.21	.20	92.4	92.4	S8
12	58	73	.43	.31	.99	-.01	.93	-.21	h	.34	.33	80.3	77.5	S12
6	70	73	-1.51	.60	.98	.14	.65	-.29	g	.21	.15	95.5	95.5	S6
2	59	73	.33	.32	.97	-.12	.82	-.67	f	.37	.32	78.8	78.9	S2
7	70	73	-1.51	.60	.92	.02	.48	-.62	e	.27	.15	95.5	95.5	S7
18	49	73	1.18	.27	.90	-.97	.89	-.78	d	.48	.40	75.8	68.1	S18
17	57	73	.52	.30	.89	-.73	.78	-.96	c	.44	.34	78.8	76.2	S17
15	58	73	.43	.31	.88	-.73	.71	-1.25	b	.45	.33	77.3	77.5	S15
14	59	73	.33	.32	.86	-.79	.71	-1.16	a	.45	.32	81.8	78.9	S14
MEAN	60.7	73.0	.00	.37	1.00	.09	.96	.01				82.4	82.1	
P. SD	6.1	.0	.76	.09	.09	.64	.28	.88				8.5	8.2	

Gambar 3. Validitas Setiap Butir Soal Instrumen Tes Diagnostik.

Hasil analisis Rasch Model selanjutnya adalah menganalisis tingkat kesukaran setiap butir soal pada instrumen tes diagnostik materi IPA. Hasil analisis menunjukkan nilai Mean sebesar 0,00 dan S.D. sebesar 0,76. Hal ini menginformasikan bahwa butir soal yang memiliki nilai Measure lebih dari Mean + S.D. termasuk dalam kategori butir soal sukar, antara

Mean – S.D. hingga Mean + S.D. termasuk dalam kategori butir soal sedang, dan lebih kecil dari Mean – S.D. termasuk dalam kategori butir soal mudah. Hasil analisis menunjukkan butir soal nomor 20 dan 18 adalah butir soal yang sukar, butir soal nomor 6 hingga 8 adalah butir soal yang mudah, sedangkan sisanya memiliki tingkat kesulitan yang sedang.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S. E.	INFIT		OUTFIT		PTMEASUR-AL		EXACT OBS%	MATCH EXP%	Item	
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.				
20	48	73	1.25	.27	1.21	1.96	1.27	1.86	.24	.41	60.6	67.5	S20	
18	49	73	1.18	.27	.90	-.97	.89	-.78	.48	.40	75.8	68.1	S18	
1	55	73	.70	.29	.97	-.17	1.06	.36	.37	.36	75.8	73.7	S1	
19	55	73	.70	.29	1.03	.29	1.04	.26	.33	.36	69.7	73.7	S19	
17	57	73	.52	.30	.89	-.73	.78	-.96	.44	.34	78.8	76.2	S17	
12	58	73	.43	.31	.99	-.01	.93	-.21	.34	.33	80.3	77.5	S12	
15	58	73	.43	.31	.88	-.73	.71	-1.25	.45	.33	77.3	77.5	S15	
2	59	73	.33	.32	.97	-.12	.82	-.67	.37	.32	78.8	78.9	S2	
11	59	73	.33	.32	1.12	.76	1.28	1.09	.20	.32	78.8	78.9	S11	
14	59	73	.33	.32	.86	-.79	.71	-1.16	.45	.32	81.8	78.9	S14	
10	62	73	.00	.34	1.18	.87	1.80	2.12	.07	.29	83.3	83.3	S10	
3	63	73	-.12	.36	1.04	.24	.97	.04	.26	.28	84.8	84.9	S3	
9	63	73	-.12	.36	1.11	.53	1.12	.46	.19	.28	84.8	84.9	S9	
13	63	73	-.12	.36	.98	-.02	1.16	.56	.27	.28	84.8	84.9	S13	
16	65	73	-.40	.39	1.02	.16	.86	-.23	.25	.25	87.9	87.9	S16	
4	67	73	-.74	.44	.99	.09	.80	-.28	.24	.22	90.9	90.9	S4	
5	67	73	-.74	.44	.99	.09	.83	-.20	.24	.22	90.9	90.9	S5	
8	68	73	-.95	.48	.99	.09	.96	.10	.21	.20	92.4	92.4	S8	
6	70	73	-1.51	.60	.98	.14	.65	-.29	.21	.15	95.5	95.5	S6	
7	70	73	-1.51	.60	.92	.02	.48	-.62	.27	.15	95.5	95.5	S7	
MEAN	60.7	73.0	.00	.37	1.00	.09	.96	.01				82.4	82.1	
P. SD	6.1	.0	.76	.09	.09	.64	.28	.88				8.5	8.2	

Gambar 4. Tingkat Kesukaran Setiap Butir Soal Instrumen Tes Diagnostik.

Analisis selanjutnya yang dilakukan terhadap instrumen tes diagnostik materi IPA adalah analisis bias soal. Adapun bias yang dilakukan berfokus pada jenis kelamin peserta tes saja. Hasil analisis Rasch Model menunjukkan

bahwa nilai DIF tidak ada yang kurang dari 5% (0,05). Hal ini menunjukkan bahwa tidak ada satu butir soal pun pada instrumen tes diagnostik materi IPA yang bias terhadap jenis kelamin.

Person CLASSES	SUMMARY DIF			BETWEEN-CLASS/GROUP		Item	
	CHI-SQUARED	D. F.	PROB.	UNWTD	MNSQ	ZSTD	Number Name
10	6.8781	9	.6498	1.2128	.58		1 S1
10	7.8031	9	.5541	1.4314	.97		2 S2
10	7.8936	9	.5448	1.6048	1.24		3 S3
10	3.2946	9	.9514	.7686	-.38		4 S4
10	5.0514	9	.8298	1.1800	.52		5 S5
10	4.4031	9	.8829	1.0104	.18		6 S6
10	3.3000	9	.9512	.7288	-.48		7 S7
10	4.3471	9	.8871	1.0124	.18		8 S8
10	4.0469	9	.9083	.8005	-.30		9 S9
10	4.8800	9	.8446	.9214	-.01		10 S10
10	5.2471	9	.8122	1.1067	.38		11 S11
10	5.6151	9	.7777	1.0515	.26		12 S12
10	2.5725	9	.9789	.4868	-1.20		13 S13
10	4.3188	9	.8892	.8577	-.16		14 S14
10	4.5699	9	.8701	.9108	-.04		15 S15
10	3.9019	9	.9177	.8121	-.27		16 S16
10	4.8014	9	.8512	.8606	-.15		17 S17
10	2.9694	9	.9655	.4115	-1.47		18 S18
10	6.7623	9	.6618	1.1889	.54		19 S19
10	6.5883	9	.6799	1.0320	.22		20 S20

Gambar 5. Nilai *DIF* Setiap Butir Soal pada Instrumen Tes Diagnostik.

Hasil analisis Rasch Model juga dilanjutkan untuk mengetahui validitas dari responden. Sama halnya dengan validitas item, responden dinyatakan valid jika memenuhi satu dari tiga syarat. Adapun ketiga syarat tersebut adalah nilai *outfit means-square* antara 0,5 dan 1,5; nilai *outfit z-standard* antara -2,0 dan 2,0; serta nilai *point measure correlation* antara 0,4 dan 0,85. Hasil analisis Rasch Model menunjukkan bahwa setiap responden dinyatakan valid karena memenuhi setidaknya satu dari tiga syarat valid. Hal ini menginformasikan bahwa selama pengisian instrumen, setiap responden tidak terindikasi melakukan kecurangan, kecerobohan, atau keberuntungan dalam mengisi instrumen tes diagnostik pada materi IPA.

Hasil analisis Rasch Model selanjutnya adalah menganalisis tingkat kemampuan setiap peserta tes dalam mengerjakan soal pada instrumen tes diagnostik materi IPA. Hasil analisis menunjukkan nilai Mean sebesar 2,07 dan S.D. sebesar 1,09. Hal ini menginformasikan bahwa peserta tes yang memiliki nilai Measure lebih dari Mean + S.D. termasuk dalam kategori peserta tes berkemampuan tinggi, antara Mean – S.D. hingga Mean + S.D. termasuk dalam kategori peserta tes berkemampuan sedang, dan lebih kecil dari Mean – S.D. termasuk dalam kategori peserta tes berkemampuan rendah. Hasil analisis menunjukkan dari 83 peserta tes, sebanyak 17 peserta tes adalah peserta tes berkemampuan tinggi, sebanyak 14 peserta tes adalah peserta tes berkemampuan rendah, sedangkan sisanya sebanyak 52 peserta tes adalah peserta tes berkemampuan sedang.

Model Rasch digunakan karena model ini mampu mendefinisikan konstruk dari item yang

valid dan memberikan definisi yang jelas tentang konstruksi terukur yang konsisten dengan harapan teoritis (Mohamad et al., 2015). Adapun soal yang dianalisis menggunakan model pengukuran Rasch Model dan dinyatakan fit, maka hasil itu sudah cukup memberikan informasi bahwa soal yang dikembangkan valid dan dapat mengukur apa yang seharusnya diukur (Boone, 2016; Planinic et al., 2019).

Hasil analisis Rasch Model menunjukkan bahwa set instrumen tes diagnostik materi IPA memiliki nilai *Raw Variance* sebesar 27,6%. Hal ini menginformasikan bahwa set instrumen tes diagnostik materi IPA memiliki validitas konstruk yang baik. Hasil analisis Rasch Model juga menunjukkan bahwa instrumen validasi teoritik aspek materi IPA memiliki nilai *Raw Variance* sebesar 21,0%. Hal ini menginformasikan bahwa instrumen instrumen validasi teoritik aspek materi IPA telah memenuhi kriteria validasi konstruk. Selain itu, hasil analisis Rasch Model menunjukkan bahwa instrumen validasi teoritik aspek penilaian memiliki nilai *Raw Variance* sebesar 33,5%. Hal ini menginformasikan bahwa instrumen instrumen validasi teoritik aspek penilaian telah memenuhi kriteria validasi konstruk.

Raw variance menjadi sifat laten dari suatu instrumen yang diukur (Al Zoubi et al., 2018; Nielsen & Dammeyer, 2019) dalam memverifikasi konstruksi pengukuran dari instrumen tersebut (Chang et al., 2020; Lo et al., 2015). Validitas tidak hanya sekedar menginformasikan mengenai validitas isi tes, tapi juga menginformasikan konsekuensi dari penggunaan skor tes tersebut (Mokshein et al., 2019). Dengan menggunakan model pengukuran

Rasch, hasil analisis dapat menetapkan kesesuaian item dengan konstruksi yang diidentifikasi (Md Yunos et al., 2017) terhadap instrumen tes secara umum (R. M. Yasin et al., 2015; S. N. T. M. Yasin et al., 2018).

Selain itu dalam analisis Rasch Model, reliabilitas memberikan informasi yang lebih rinci antara properti item pada instrumen dan individu yang menanggapi item tersebut (Bodzin et al., 2020). Reliabilitas juga menginformasikan mengenai konsistensi baik konsistensi item maupun konsistensi person (Koçak, 2020). Hasil analisis Rasch Model menunjukkan bahwa set instrumen tes diagnostik materi IPA memiliki reliabilitas yang cukup bagus. Hal ini ditunjukkan dari nilai item reliability sebesar 0,75. Hasil analisis Rasch Model juga menunjukkan bahwa instrumen validasi teoritik aspek materi IPA memiliki reliabilitas yang cukup bagus. Hal ini ditunjukkan dari nilai item reliability sebesar 0,69. Hasil analisis Rasch Model lainnya juga menunjukkan bahwa instrumen validasi teoritik aspek penilaian memiliki reliabilitas yang cukup bagus. Hal ini ditunjukkan dari nilai item reliability sebesar 0,72.

Hasil analisis Rasch Model menunjukkan bahwa seluruh butir soal pada set instrumen tes diagnostik materi IPA dinyatakan valid. Adapun untuk instrumen validasi teoritik, hasil analisis Rasch Model menunjukkan bahwa seluruh butir pernyataan pada instrumen validasi teoritik aspek penilaian dinyatakan valid. Namun hasil analisis Rasch Model pada instrumen validasi teoritik pada aspek materi IPA menunjukkan bahwa 19 dari 20 butir pernyataan pada instrumen validasi teoritik aspek materi IPA dinyatakan valid, dimana butir pernyataan nomor 14 tidak valid karena tidak ada sama sekali syarat yang memenuhi kriteria valid.

Selain memberikan analisis terhadap kualitas soal yang telah dikembangkan, penerapan Rasch Model juga memberikan informasi tentang kualitas siswa yang melakukan tes. Hal ini menginformasikan pendidik dalam meningkatkan kualitas siswa mereka (Zamri & Nordin, 2015). Berlandaskan pada informasi yang lebih mendalam mengenai kualitas peserta didiknya, pendidik dapat terbantu dalam mengevaluasi proses pembelajaran di kelas (Mursidi & Soeharto, 2016; Rahmani, 2018; Suranata et al., 2018).

Hasil analisis menunjukkan butir soal nomor 20 dan 18 pada set instrumen tes diagnostik materi IPA adalah butir soal yang sukar, butir soal nomor 6 hingga 8 pada set instrumen tes

diagnostik materi IPA adalah butir soal yang mudah, sedangkan sisanya sebanyak 15 butir soal pada set instrumen tes diagnostik materi IPA memiliki tingkat kesulitan yang sedang. Hasil analisis Rasch Model juga menunjukkan bahwa nilai DIF tidak ada yang kurang dari 5% (0,05). Hal ini menunjukkan bahwa tidak ada satu butir soal pun pada set instrumen tes diagnostik materi IPA yang bias terhadap jenis kelamin.

Tingkat kesukaran soal yang bervariasi menjadi salah satu kriteria lain dari suatu alat ukur yang baik (Barus et al., 2019; Hamdu et al., 2020). Adapun tingkat kesukaran butir item tidaklah harus berurutan dalam suatu instrumen tes (Isnani et al., 2019). Selain tingkat kesukaran, probabilitas terjadinya bias pada soal menjadi hal yang patut dianalisis. Nilai DIF mengecek kemungkinan item yang bias yang disebabkan oleh respon dari group yang berbeda (Rahmani, 2018; Veas et al., 2016; Xu et al., 2020). Soal yang bias timbul karena adanya dua kelompok berbeda dengan kemampuan yang sama, namun terdapat perbedaan yang signifikan saat menjawab butir soal (Alwi, 2017).

Hasil analisis Rasch Model lainnya juga menunjukkan bahwa setiap responden dinyatakan valid karena memenuhi setidaknya satu dari tiga syarat valid. Hal ini menginformasikan bahwa selama pengisian instrumen, baik set instrumen tes diagnostik materi IPA maupun instrumen validasi teoritik bersifat valid. Hal ini dimaknai bahwa setiap responden tidak terindikasi melakukan kecurangan, kecerobohan, atau keberuntungan dalam mengisi instrumen validasi teoritik aspek penilaian.

Tampak jelas bahwa analisis Rasch Model mampu menjelaskan tidak hanya kualitas butir soal yang dikembangkan, namun juga kualitas siswa peserta tes (Maulana et al., 2020; Supriyati et al., 2021). Rasch Model memberikan informasi mengenai konsistensi peserta tes dalam menjawab soal hingga adanya kemungkinan peserta tes yang melakukan kecerobohan dalam menjawab, menebak, atau pun mencontek (Isnani et al., 2019; Susongko, 2016). Hal ini tentunya sangat membantu guru dalam memberikan evaluasi yang sesuai dengan kemampuan siswa peserta tes (Mursidi & Soeharto, 2016; Suranata et al., 2018).

Interpretasi menggunakan model pengukuran Rasch memberikan informasi lebih rinci mengenai kualitas item dan person, serta kesesuaiannya dalam mengukur apa yang seharusnya diukur (Bichi et al., 2019; Tornabene et al., 2018). Hal ini juga memberikan kemudahan

bagi para guru untuk memeriksa/mengevaluasi dan menentukan pemberian *feedback* yang tepat terhadap setiap siswanya (Börkan, 2017). Sehingga evaluasi siswa peserta tes oleh guru menjadi lebih mudah dan tepat (Erman Aslanoglu et al., 2020).

SIMPULAN

Berdasarkan hasil penelitian serta pembahasan yang telah diuraikan, penelitian menghasilkan kesimpulan bahwa produk pendidikan berupa instrumen tes diagnostik materi IPA dinyatakan layak dan dapat digunakan untuk implementasi pada khalayak umum. Hal ini dibuktikan melalui hasil pengujian secara teoritis (CVR dan CVI) dan empiris (Rasch Model) yang menyatakan bahwa instrumen tes diagnostik pada materi IPA valid dari segi konten maupun konstruk, serta reliabel. Sehingga instrumen tes diagnostik pada materi IPA yang telah dikembangkan benar-benar dapat menginformasikan pengetahuan awal peserta didik pada materi IPA.

UCAPAN TERIMA KASIH

Ucapan terima kasih ditujukan kepada program studi Pendidikan IPA, Fakultas Pascasarjana, Universitas Negeri Semarang yang telah mendukung dalam terselesaikannya penelitian ini.

REFERENSI

- Al-Taweel, D., & Awad, A. (2020). Development and validation of medication assessment tools to evaluate prescribing adherence to evidence-based guidelines for secondary prevention of coronary heart disease in post-acute coronary syndromes patients in Kuwait. *PLoS ONE*, *15*(11 November). <https://doi.org/10.1371/journal.pone.0241633>
- Al Zoubi, F., Mayo, N., Rochette, A., & Thomas, A. (2018). Applying modern measurement approaches to constructs relevant to evidence-based practice among Canadian physical and occupational therapists. *Implementation Science*, *13*(1). <https://doi.org/10.1186/s13012-018-0844-4>
- Alizadeh-Siuki, H., Tehrani, H., Gholian-Aval, M., Ebrahimpour, H., Jafari, A., & Vahedian-Shahroodi, M. (2020). Psychometric properties of a questionnaire on brucellosis prevention behaviors based on the precede model among rural farmers and their family members. *Risk Management and Healthcare Policy*, *13*. <https://doi.org/10.2147/RMHP.S247356>
- Alwi, I. (2017). Sensitivity of Mantel Haenszel Model and Rasch Model as Viewed from Sample Size. *Jurnal Evaluasi Pendidikan*, *2*(1). <https://doi.org/10.21009/jep.021.02>
- Barus, G., Sinaga, J. D., & Moron, Y. D. (2019). Testing the quality of film-based creative and innovative character education test with rasch model. *Jurnal Konseling Dan Pendidikan*, *7*(2). <https://doi.org/10.29210/136000>
- Bichi, A. A., Talib, R., Atan, N. A., Ibrahim, H., & Yusof, S. M. (2019). Validation of a developed university placement test using classical test theory and Rasch measurement approach. *International Journal of Advanced and Applied Sciences*, *6*(6), 22–29. <https://doi.org/10.21833/ijaas.2019.06.004>
- Bodzin, A., Hammond, T., Fu, Q., & Farina, W. (2020). Development of Instruments to Assess Students' Spatial Learning Attitudes (SLA) and Interest in Science, Technology and Geospatial Technology (STEM-GEO). *International Journal of Educational Methodology*, *6*(1), 67–81. <https://doi.org/10.12973/ijem.6.1.67>
- Boone, W. J. (2016). Rasch Analysis for Instrument Development: Why, When, and How? *CBE Life Sciences Education*, *15*(4), 1–7. <https://doi.org/10.1187/cbe.16-04-0148>
- Börkan, B. (2017). Exploring Variability Sources in Student Evaluation of Teaching via Many-Facet Rasch Model. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*. <https://doi.org/10.21031/epod.298462>
- Chang, J., Li, Y., Song, H., Yong, L., Luo, L., Zhang, Z., & Song, N. (2020). Assessment of Validity of Children's Movement Skill Quotient (CMSQ) Based on the Physical Education Classroom Environment. *BioMed Research International*, *2020*. <https://doi.org/10.1155/2020/8938763>
- Dewi, A. K. (2019). Improving Students Learning Outcomes Through Mind Map in Human Reproductive System Topic in Natural Science Learning. *International Journal for Educational and Vocational Studies*, *1*(7). <https://doi.org/10.29103/ijevs.v1i7.1675>
- Erman Aslanoglu, A., Karakaya, I., & Sata, M. (2020). Evaluation of university students' rating behaviors in self and peer rating process via many facet rasch model.

- Eurasian Journal of Educational Research*, 2020(89).
<https://doi.org/10.14689/ejer.2020.89.2>
- Fulmer, G. W., Chu, H., Treagust, D. F., & Neumann, K. (2016). Is it harder to know or to reason? Analyzing two-tier science assessment items using the Rasch measurement model. *Asia-Pacific Science Education*, 1(2015), 1–16.
<https://doi.org/10.1186/s41029-015-0005-x>
- Hamdu, G., Fuadi, F. N., Yulianto, A., & Akhirani, Y. S. (2020). Items Quality Analysis Using Rasch Model To Measure Elementary School Students' Critical Thinking Skill On Stem Learning. *JPI (Jurnal Pendidikan Indonesia)*, 9(1).
<https://doi.org/10.23887/jpi-undiksha.v9i1.20884>
- Islam, A. Y. M. A., Gu, X., Crook, C., & Spector, J. M. (2020). Assessment of ICT in Tertiary Education Applying Structural Equation Modeling and Rasch Model. *SAGE Open*, 10(4).
<https://doi.org/10.1177/2158244020975409>
- Isnani, I., Utami, W. B., Susongko, P., & Lestiani, H. T. (2019). Estimation of college students' ability on real analysis course using Rasch model. *Research and Evaluation in Education*, 5(2).
<https://doi.org/10.21831/reid.v5i2.20924>
- Koçak, D. (2020). Investigation of rater tendencies and reliability in different assessment methods with many facet rasch model. *International Electronic Journal of Elementary Education*, 12(4).
<https://doi.org/10.26822/iejee.2020459464>
- Ling, M., Pang, V., & Salwana, B. (2014). Rasch Analysis: Psychometric Properties of Malaysian Secondary School Students' Leadership Scale. *Man In India*, 96(1–2), 1–8.
- Lo, C., Liang, W., Hang, L., Wu, T., Chang, Y., & Chang, C. (2015). A Psychometric Assessment of the St. George's Respiratory Questionnaire in Patients with COPD using Rasch Model Analysis. *Health and Quality of Life Outcomes*, 13(131), 1–16.
<https://doi.org/10.1186/s12955-015-0320-7>
- Matore, M. E. E. M., Maat, S. M., Affandi, H. M., Mohamad, S., & Khairani, A. Z. (2018). Assessment of psychometric properties for Raven Advanced Progressive Matrices in measuring intellectual quotient (IQ) using Rasch model. *Asian Journal of Scientific Research*, 11(3).
<https://doi.org/10.3923/ajsr.2018.393.400>
- Maulana, S., Supriyati, Y., & Astra, I. M. (2020). Development of Physics Problems for The Final Assessment in High School Level with Rasch Model Analysis. *Unnes Science Education Journal*, 9(1), 22–29.
- Md Yunos, J., Ibrahim Mukhtar, M., Alias, M., Lee, M. F., Tee, T. K., Rubani, S. N. K., Hamid, H., Yunus, F. A. N., Sulaiman, J., & Sumarwati, S. (2017). Validity of vocational pedagogy constructs using the rasch measurement model. *Journal of Technical Education and Training*, 9(2).
- Mohamad, M. M., Lisa, N., Sern, L. C., & Mohd, K. (2015). Measuring the Validity and Reliability of Research Instruments. *Procedia - Social and Behavioral Sciences*, 204(1), 164–171.
<https://doi.org/10.1016/j.sbspro.2015.08.129>
- Mokshein, S. E., Ishak, H., & Ahmad, H. (2019). The use of rasch measurement model in English testing. *Cakrawala Pendidikan*, 38(1).
<https://doi.org/10.21831/cp.v38i1.22750>
- Mursidi, A., & Soeharto. (2016). An Introduction: Evaluation of Quality Assurance for Higher Educational Institutions using Rasch Model. *Journal of Education, Teaching and Learning*, 1(1), 1–6.
- Nielsen, T., & Dammeyer, J. (2019). Measuring higher education students' perceived stress: An IRT-based construct validity study of the PSS-10. *Studies in Educational Evaluation*, 63.
<https://doi.org/10.1016/j.stueduc.2019.06.007>
- Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research: Why measurement matters. *Physical Review Physics Education Research*, 15(2).
<https://doi.org/10.1103/PhysRevPhysEducRes.15.020111>
- Rahmani, B. D. (2018). Differential Item Functional Analysis on Pedagogic and Content Knowledge (PCK) Questionnaire for Indonesian Teachers using RASCH Model. *Journal of Physics: Conference Series*.
- Semakula, D., Nsangi, A., Oxman, A. D., Sewankambo, N. K., Guttersrud, O., & Austvoll-Dahlgren, A. (2017). Measuring ability to assess claims about treatment effects in English and Luganda: evaluation of multiple-choice questions from the “Claim Evaluation Tools”

- database using Rasch modelling. *Informed Health Choices*, 1–17.
- Sjaastad, J. (2014). Enhancing measurement in science and mathematics education research through Rasch analysis: Rationale and properties. *NORDINA*, 10(2), 212–230.
- Supriyati, Y., Susanti, D., & Maulana, S. (2021). Computer-Based Application for High School Physics Exams using IRT Model 1P. *AIP Conference Proceedings*, 2320(March). <https://doi.org/10.1063/5.0037567>
- Suranata, K., Rangka, I. B., Ifdil, I., Ardi, A., Susiani, K., Prasetyaningtyas, W. E., Daharnis, D., Alizamar, A., Erlinda, L., & Rahim, R. (2018). Diagnosis of students zone proximal development on math design instruction: A Rasch analysis. *Journal of Physics: Conference Series*.
- Suroso, J., Indrawati, Sutarto, & Mudakir, I. (2021). Profile of high school students science literacy in east java. *Journal of Physics: Conference Series*, 1832(1). <https://doi.org/10.1088/1742-6596/1832/1/012040>
- Suryani, Y. E. (2018). Aplikasi Rasch Model dalam Mengevaluasi Intelligenz Structure Test. *Psikohumaniora: Jurnal Penelitian Psikologi*, 3(1), 73–100.
- Susongko, P. (2016). Validation of Science Achievement Test with the Rasch Model. *Jurnal Pendidikan IPA Indonesia*, 5(2), 268–277. <https://doi.org/10.15294/jpii.v5i2.7690>
- Tornabene, R. E., Lavington, E., & Nehm, R. H. (2018). Testing validity inferences for Genetic Drift Inventory scores using Rasch modeling and item order analyses. *Evolution: Education and Outreach*, 11(1). <https://doi.org/10.1186/s12052-018-0082-x>
- Usmeldi, U., & Amini, R. (2019). The effect of integrated learning model to the students competency on the natural science. *Journal of Physics: Conference Series*, 1157(2). <https://doi.org/10.1088/1742-6596/1157/2/022022>
- Veas, A., Castejón, J. L., Gilar, R., & Miñano, P. (2016). Validation of the Spanish adaptation of the School Attitude Assessment Survey-Revised using multidimensional Rasch analysis. *Anales de Psicología*, 33(1). <https://doi.org/10.6018/analesps.33.1.235271>
- Waltz, C. F., Strickland, O. L., & Lenz, E. R. (2016). Measurement in Nursing and Health Research. In *Measurement in Nursing and Health Research*. <https://doi.org/10.1891/9780826170620>
- Xiao, Y., Han, J., Koenig, K., Xiong, J., & Bao, L. (2018). Multilevel Rasch modeling of two-tier multiple choice test: A case study using Lawson 's classroom test of scientific reasoning. *Physical Review Physics Education Research*, 14(2), 20104. <https://doi.org/10.1103/PhysRevPhysEducRes.14.020104>
- Xu, R. H., Wong, E. L. yi, Lu, S. Y. jun, Zhou, L. M., Chang, J. H., & Wang, D. (2020). Validation of the Toronto Empathy Questionnaire (TEQ) Among Medical Students in China: Analyses Using Three Psychometric Methods. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.00810>
- Yasin, R. M., Yunus, F. A. N., Rus, R. C., Ahmad, A., & Rahim, M. B. (2015). Validity and Reliability Learning Transfer Item Using Rasch Measurement Model. *Procedia - Social and Behavioral Sciences*, 204. <https://doi.org/10.1016/j.sbspro.2015.08.143>
- Yasin, S. N. T. M., Yunus, M. F. M., & Ismail, I. (2018). The use of rasch measurement model for the validity and reliability. *Journal of Counseling and Educational Technology*, 1(2). <https://doi.org/10.32698/0111>
- Zamri, A., & Nordin. (2015). Modeling a Multiple Choice Mathematics Test with the Rasch Model. *Indian Journal of Science and Technology*, 8(12), 1–6. <https://doi.org/10.17485/ijst/2015/v8i12/70650>