

Profil instrumen tes kemampuan kognitif pada materi gerak parabola

Fathul Naufal Hamidi, Agus Danawan, Heni Rusnayati

Artikel ini telah dipresentasikan pada kegiatan Seminar Nasional Fisika (Sinafi X) & International Physics Conference (IPC)
Universitas Pendidikan Indonesia, Bandung, Indonesia
9 November 2024

Abstract

This study aims to evaluate the quality of a test instrument designed to measure high school students' cognitive abilities on the topic of Projectile Motion. This research employs a quantitative method with a quasi-experimental type and a one-shot case study design. The population of this study consists of 12th-grade students at SMA PGRI 1 Bandung, with a sample of 34 students from class XII-1 selected through purposive sampling. The instrument consists of 20 items analyzed using the Rasch Model to assess validity, reliability, and item difficulty. The analysis results show a raw variance of 28.7%, with an unexplained variance in the first contrast of 14.6%, indicating good unidimensionality. The validity test results show that all items met the fit criteria, with 4 items categorized as "Highly Appropriate" and 16 items as "Appropriate." The item difficulty distribution consists of 5 "Easy" items, 12 "Moderate" items, and 3 "Difficult" items. The item reliability score was 0.79 ("Moderate" category), while the person reliability score was 0.00 ("Low" category), indicating limitations in sample variation. The Cronbach's alpha (KR-20) score of 0.62 suggests that the instrument has reasonably good overall quality. This instrument can be used to map students' cognitive abilities in the topic of Projectile Motion.

Keywords: Projectile Motion · Rasch Model · Test Instrument · Cognitive Ability · Validity, Reliability, Item Difficulty Level

PENDAHULUAN

Instrumen tes kemampuan kognitif fisika sangat penting untuk mengukur pemahaman siswa dalam konsep fisika, seperti gerak parabola, secara akurat. Berdasarkan hasil PISA 2022, skor literasi sains siswa Indonesia turun menjadi 383, jauh di bawah rata-rata global sebesar 483 poin. Meskipun peringkat Indonesia naik dibandingkan tahun 2018, penurunan skor ini menunjukkan bahwa kemampuan sains siswa masih perlu ditingkatkan. Hal ini menegaskan pentingnya peningkatan dalam penilaian dan pembelajaran sains di sekolah. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan instrumen yang baik sehingga dapat diterapkan pada pembelajaran di kelas guna memberikan informasi yang baik terkait kemampuan peserta didik pada materi gerak parabola.

Nurhayati (2021) menyatakan bahwa instrumen yang teruji dapat memetakan kemampuan siswa secara akurat dan membantu guru menyesuaikan metode pengajaran. Hakim (2020) juga menemukan bahwa instrumen yang berfokus pada keterampilan berpikir tingkat tinggi, seperti gerak parabola, efektif dalam mengidentifikasi pemahaman siswa. Selain itu, Widodo dan Kadarwati (2022) menunjukkan bahwa penilaian kognitif yang terstruktur dalam sains

✉ Fathul Naufal Hamidi Agus Danawan Heni Rusnayati
nopal@upi.edu ad@upi.edu heni@upi.edu

Universitas Pendidikan Indonesia, Bandung, Indonesia

How to Cite: Hamidi, F. N., Danawan, A. & Rusnayati, H. (2024). Profil Instrumen Tes Kemampuan Kognitif pada Materi Gerak Parabola. *Prosiding Seminar Nasional Fisika & International Physics Conference*, 3(1), 186-198. <https://proceedings.upi.edu/index.php/sinafi/>

memberikan Gambar an yang lebih akurat terhadap profil siswa. Dengan demikian, pengembangan instrumen tes ini diharapkan mampu meningkatkan kualitas pembelajaran fisika di tingkat SMA. Dari pernyataan pada penelitian sebelumnya, maka dapat dikatakan bahwa instrumen tes yang valid dan reliabel sangat penting untuk memberikan Gambar an yang lebih detail kepada guru sehingga dapat meningkatkan kualitas mengajar serta lebih presisi tentang pada bagian mana saat mengajar yang perlu ditingkatkan.

Gerak Parabola adalah topik fundamental dalam fisika SMA yang memerlukan pemahaman mendalam tentang konsep matematika dan fisika secara bersamaan, menjadikannya landasan penting untuk topik-topik lebih kompleks seperti dinamika rotasi, energi, dan hukum Newton. Fokus pada gerak parabola didasarkan pada relevansinya dalam kehidupan nyata dan aplikasinya dalam berbagai disiplin ilmu. Menurut Popham (2008, hlm. 78), "tes harus mampu mengukur kemampuan kognitif dengan berbagai tingkat kesulitan dan jenis soal agar dapat mencerminkan pemahaman siswa yang sebenarnya." Selain melatih keterampilan analisis dan pemecahan masalah, gerak parabola juga memperkenalkan siswa pada konsep kinematika dan dinamika secara terintegrasi. Widodo dan Kadarwati (2022) menegaskan bahwa pemahaman yang kuat tentang gerak parabola menjadi dasar untuk mengkaji fenomena fisika lainnya. Oleh karena itu, meskipun penelitian ini terfokus pada gerak parabola, metode pengembangan instrumen yang dihasilkan diharapkan dapat diterapkan pada konsep fisika lain yang memiliki karakteristik serupa dalam hal kompleksitas dan kebutuhan evaluasi.

Dalam pengembangan instrumen evaluasi, penting untuk mempertimbangkan model pengukuran yang digunakan untuk memastikan kualitas soal yang dihasilkan. Model pengukuran klasik (Classical Test Theory/CTT) memiliki keterbatasan dalam memberikan informasi rinci pada tingkat butir soal, seperti ketidakmampuannya memisahkan karakteristik soal dan kemampuan siswa secara independen, yang dapat memengaruhi akurasi pengukuran (Hambleton & Jones, 1993). Model Rasch menawarkan solusi atas keterbatasan ini dengan menganalisis hubungan probabilistik antara kemampuan siswa dan tingkat kesulitan soal, memungkinkan peneliti untuk mengembangkan instrumen yang memiliki validitas dan reliabilitas tinggi (Boone, Staver, & Yale, 2014, hlm. 101). Rasch Model memastikan bahwa setiap butir soal memiliki kontribusi yang terukur dalam proses evaluasi, memberikan informasi lebih akurat dan tidak bias terhadap kelompok sampel tertentu. Selain itu, model ini mempermudah deteksi butir soal yang tidak sesuai atau terlalu sulit, meningkatkan kualitas instrumen secara keseluruhan dan menjadikan hasil pengukuran lebih konsisten serta dapat diandalkan dalam berbagai konteks evaluasi.

Model Rasch telah banyak digunakan oleh peneliti dalam mengevaluasi kemampuan kognitif siswa, khususnya dalam bidang pendidikan sains dan fisika. Rahman (2021) menemukan bahwa model ini mampu mengidentifikasi butir soal yang tidak sesuai dengan tingkat kemampuan siswa, sehingga mempermudah proses revisi soal. Penelitian Ibrahim (2020) yang menerapkan Model Rasch pada soal gerak parabola menunjukkan bahwa model ini efektif dalam membedakan siswa berdasarkan tingkat kemampuan kognitif mereka. Mustafa dan Yusoff (2022) juga mengungkapkan bahwa Model Rasch mampu mendeteksi butir soal yang terlalu mudah atau sulit, sehingga membantu meningkatkan kualitas instrumen evaluasi fisika. Widjaja (2019) mendukung temuan ini dengan menunjukkan bahwa penggunaan Model Rasch meningkatkan reliabilitas soal, terutama dalam topik energi dan momentum.

Penelitian ini akan memanfaatkan Google Spreadsheet dan perangkat lunak Winsteps atau Ministep untuk menganalisis data menggunakan Model Rasch. Google Spreadsheet akan digunakan untuk mengelola, mengorganisasi, dan memproses data awal dari hasil uji coba soal. Aplikasi ini memungkinkan pengolahan data mentah, pembersihan, dan pengelompokan respons siswa berdasarkan hasil tes. Selanjutnya, data tersebut akan diimpor ke Winsteps, perangkat lunak analisis statistik yang dirancang khusus untuk Model Rasch, guna mengukur kemampuan siswa dan tingkat kesulitan setiap soal. Hasil analisis Winsteps yang mencakup statistik kecocokan, validitas, dan reliabilitas soal akan diekspor kembali ke Spreadsheet untuk diproses lebih lanjut, termasuk visualisasi data dan penyusunan laporan akhir.

METODE

Metode penelitian yang akan digunakan adalah penelitian kuantitatif. Jenis penelitian ini adalah *quasi-eksperimental* dengan desain penelitian *one-shot case study design*. Campbell, D. T., & Stanley, J. C. (1963) menjelaskan bahwa *One-shot case study design* adalah bentuk paling sederhana dari desain eksperimen, dimana peneliti hanya memberikan satu kali intervensi pada satu kelompok dan mengukur hasilnya tanpa adanya perbandingan dengan kelompok lain atau pretest.

Populasi dalam penelitian ini adalah seluruh siswa kelas XII SMA PGRI 1 Bandung. Sampel dalam penelitian ini adalah siswa kelas XII-1 SMA PGRI 1 Bandung. Teknik pengambilan sampel dalam penelitian ini adalah *purposive sampling*. Penelitian ini melibatkan 34 siswa dalam satu kelas tanpa adanya kelas kontrol. Kelas yang diambil dalam partisipan adalah siswa-siswi yang telah mempelajari materi gerak parabola. Penelitian dilakukan dengan 1 pertemuan, dimana peneliti akan memberikan soal kepada siswa untuk kemudian dijawab selama satu pertemuan, yang durasi satu pertemuan tersebut adalah 2 JP. Satu JP sama dengan 40 menit.

Sebelum instrumen tes kemampuan kognitif diberikan, instrumen terlebih dahulu divalidasi oleh dosen ahli materi sebagai bagian dari validitas logis dan validasi isi. Setelah instrumen direvisi sesuai arahan dari dosen ahli, instrumen kemudian siap untuk diterapkan. Instrumen tes berjumlah 20 butir soal dengan tingkat kesulitan berbeda-beda, mulai dari level terendah adalah level kognitif C2 (Memahami), level menengah C3 (Menerapkan), dan level tertinggi C4 (Menganalisis). Rincian sebaran butir soal dengan level kognitifnya dapat dilihat pada Tabel 1 berikut.

Tabel 1. Sebaran Level Kognitif Instrumen Tes

No	Tingkat Kognitif	Nomor Soal	Jumlah Soal
1	Memahami (C2)	1, 2, 8, 9, 10	5
2	Menerapkan (C3)	3, 4, 6, 7, 11, 12, 14, 15, 16, 18, 20	11
3	Menganalisis (C4)	5, 13, 17, 19	4

Analisis yang akan dilakukan ialah terkait Uji Validitas, Reliabilitas, dan Tingkat Kesukaran Soal dan dianalisis menggunakan Model Analisis Rasch Model dengan aplikasi *ministep* versi 5.8.3.0 dengan *output Tabel 23.0* dan *10.1*.

Uji Validitas

Uji validitas dilakukan menggunakan model analisis Rasch Model. Dalam model tersebut, pengujian validitas disebut dengan *Undimensionality* (Sumintono & Widhiarso, 2015), dimana hasil dari analisis tersebut dapat dilihat dari *raw variance explained by measures* menggunakan aplikasi *ministep* versi 5.8.3.0 yang diperoleh dari menu *output Tabel 23.0 : Item dimensionality*. Hasil yang didapat kemudian diinterpretasikan melalui Tabel berikut.

Tabel2. Interpretasi Raw Variance Explained by Measures Menggunakan Aplikasi Ministep

No	Raw Variance Explained By Measures	Interpretasi
1	$20\% \leq value \leq 40\%$	Terpenuhi
2	$40\% \leq value \leq 60\%$	Sesuai
3	$value \geq 60\%$	Istimewa

(Sumintono & Widhiarso, 2015)

Uji validitas berikutnya dilakukan untuk setiap butir soal dengan memakai permodelan Rasch menggunakan aplikasi *ministep* versi 5.8.3.0. Uji validitas ini digunakan untuk mengetahui kualitas setiap butir soal. Kelayakan setiap butir soal dapat dievaluasi berdasarkan kualitas item fit, yang dilihat dari nilai outfit mean square (MNSQ), outfit Z-standard (ZSTD), dan point measure correlation (Pt Measure Corr.). Nilai dari masing-masing kriteria ini ditampilkan pada Tabel berikut.

Tabel3. Kriteria Outfit MNSQ, ZSTD dan Pt Measure Corr

No	Kriteria Nilai Fit-Statistic	Nilai
1	Outfit Mean Square (MNSQ)	$0,50 < MNSQ < 1,50$
2	Outfit Z-Standard (ZSTD)	$-2,00 < ZSTD < 2,00$
3	Point Measure Correlation (Pt Measure Corr.)	$0,40 < Pt. Measure Corr. < 0,85$

(Sumintono & Widhiarso, 2015)

Setelah diperoleh kriteria untuk masing-masing item, data kemudian diinterpretasikan kembali untuk diklasifikasikan sesuai dengan Tabel berikut.

Tabel4. Interpretasi Kualitas Butir Soal dari Nilai Outfit MNSQ, ZSTD dan Pt Measure Corr

No	Kriteria	Interpretasi
1	Ketiga Kriteria Nilai Terpenuhi	Sangat Sesuai
2	Dua dari Tiga Kriteria Nilai Terpenuhi	Sesuai
3	Satu dari Tiga Kriteria Nilai Terpenuhi	Kurang Sesuai
4	Semua Kriteria Nilai Tidak Terpenuhi	Tidak Sesuai

(Sumintono & Widhiarso, 2014)

Uji Reliabilitas

Uji reliabilitas dalam penelitian ini dianalisis juga menggunakan pemodelan Rasch dengan bantuan *software ministep* versi 5.8.3.0 pada menu *output Tabel 3.1: Summary Statistic*. Data yang diperoleh dari menu *Summary Statistic* meliputi *person reliability* (p), *item reliability* (r), dan *cronbach alpha* (KR-20). *Person reliability* menunjukkan reliabilitas peserta, *item reliability* menggambarkan reliabilitas instrumen, dan *cronbach alpha* (KR-20) menunjukkan interaksi antara *person reliability* (p) dan *item reliability* (r). Interpretasi informasi mengenai *person reliability* (p), *item reliability* (r), dan *cronbach alpha* (KR-20) disajikan pada Tabel berikut.

Tabel 5. Kriteria serta Interpretasi Nilai dari Item dan Person Reliability

No	Statistik	Nilai Indeks	Interpretasi
1	Item dan Person Reliability	$r > 0,94$	Istimewa
		$0,90 < r \leq 0,94$	Sangat Baik
		$0,80 < r \leq 0,90$	Baik
		$0,67 < r \leq 0,80$	Cukup
		$r \leq 0,60$	Rendah
2	Cronbach Alpha (KR-20)	$KR - 20 \geq 0,80$	Sangat Tinggi
		$0,70 \leq KR - 20 < 0,80$	Tinggi
		$0,60 \leq KR - 20 < 0,70$	Baik
		$0,50 \leq KR - 20 < 0,60$	Sedang
		$KR - 20 < 0,50$	Rendah

(Sumintono & Widhiarso, 2015)

Uji Tingkat Kesukaran Butir Soal

Tingkat kesukaran butir soal dianalisis menggunakan analisis Rasch Model dengan bantuan aplikasi *ministep* versi 5.8.3.0. Tingkat kesukaran butir soal dapat dianalisis berdasarkan nilai *measure* (ME) dan standar deviasi (SD) dengan cara membandingkan nilai logit ME pada setiap item dengan nilai SD (Sumintono & Widhiarso, 2015). Interpretasi tingkat kesukaran butir soal dapat dilihat berdasarkan kriteria yang ditampilkan pada Tabel berikut.

Tabel 6. Interpretasi Kriteria Tingkat Kesukaran tiap Butir Soal

No	Kriteria Tingkat Kesukaran Soal	Interpretasi
1	$ME < -1SD$	Mudah
2	$-1SD < ME < +1SD$	Sedang
3	$ME > +1SD$	Sulit

(Sumintono & Widhiarso, 2015)

HASIL DAN PEMBAHASAN

Uji Validitas

Hasil dari analisis *Undimensionality* dapat dilihat pada Gambar 1 berikut.

```

TABLE 23.0 C:\Users\Nopal\Documents\Rasch2.PRN ZOU562WS.TXT Nov 05 2024 07:07
INPUT: 34 Person 20 Item REPORTED: 34 Person 20 Item 2 CATS MINISTEP 5.8.2.0
-----
Table of RAW RESIDUAL variance in Eigenvalue units
Eigenvalue  Observed  Expected
Total raw variance in observations = 22.2004 100.0% 100.0%
Raw variance explained by measures = 6.2004 27.9% 28.7%
Raw variance explained by persons = 1.1589 5.2% 5.4%
Raw Variance explained by items = 5.0415 22.7% 23.3%
Raw unexplained variance (total) = 16.0000 72.1% 100.0% 71.3%
Unexplnd variance in 1st contrast = 3.2492 14.6% 20.3%
Unexplnd variance in 2nd contrast = 2.1393 9.6% 13.4%
Unexplnd variance in 3rd contrast = 1.9897 9.0% 12.4%
Unexplnd variance in 4th contrast = 1.6817 7.6% 10.5%
Unexplnd variance in 5th contrast = 1.3367 6.0% 8.4%

Essential Unidimensionality (Rasch/Common variance) = 37.4%

RAW RESIDUAL VARIANCE SCREE PLOT
    
```

Gambar 1. Output Tabel Item Undimensionality

Gambar 1 menunjukkan nilai *raw variance explained by measures* yang diperoleh dari uji coba instrumen tes kemampuan kognitif dengan 20 butir soal adalah sebesar 28,7% dan presentase tersebut masuk dalam kategori interpretasi “Terpenuhi”. Dengan demikian, instrumen yang digunakan pada penelitian ini mampu mengukur satu variabel secara akurat tanpa adanya pengaruh dari variabel lain. Selain nilai *raw variance explained by measures*, *Unidimensionality* instrumen tes juga dapat dilihat dari nilai *unexplained variance in 1st contrast*. Sebuah instrumen dapat dikatakan unidimensionalitas yang baik jika *unexplained variance in 1st contrast* bernilai kurang dari 15% (Fisher, 2007). Dan seperti yang ditunjukkan pada Gambar 1, *unexplained variance in 1st contrast* yang didapat adalah 14,6%, yang artinya *unidimensionality* dari tes kemampuan kognitif yang dibuat masuk dalam kategori baik.

Hasil analisis menggunakan Rasch Model menunjukkan bahwa instrumen tes kemampuan kognitif fisika memiliki kualitas yang baik dalam mengukur pemahaman siswa pada materi gerak parabola. Nilai *raw variance explained by measures* sebesar 28,7% menunjukkan bahwa lebih dari seperempat varians dalam data dapat dijelaskan oleh model, yang mengindikasikan bahwa instrumen ini cukup mampu mengukur satu variabel utama tanpa terpengaruh faktor lain. Menurut Bond & Fox (2015), angka ini termasuk dalam kategori yang dapat diterima untuk instrumen pendidikan, karena menunjukkan bahwa sebagian besar varians dihasilkan oleh kemampuan siswa yang diukur, bukan oleh noise atau faktor eksternal. Selain itu, hasil *unexplained variance in 1st contrast* sebesar 14,6% juga memperkuat bukti unidimensionalitas instrumen, mengingat angka ini berada di bawah ambang batas 15% (Fisher, 2007). Nilai ini mengindikasikan bahwa butir-butir soal dalam tes memiliki konsistensi yang baik dan tidak menunjukkan adanya subdimensi tersembunyi yang dapat mengganggu interpretasi hasil. Dengan demikian, instrumen ini tidak hanya memenuhi kriteria validitas konstruksi, tetapi juga reliabilitas dalam mengukur kemampuan kognitif siswa secara akurat. Penggunaan instrumen ini dapat membantu guru dalam memetakan pemahaman siswa pada berbagai tingkat kesulitan soal, serta memberikan gambaran yang jelas mengenai area yang memerlukan peningkatan. Oleh karena itu, tes ini memiliki potensi besar untuk digunakan secara luas dalam evaluasi pembelajaran fisika di tingkat SMA, terutama dalam menilai konsep-konsep kompleks seperti gerak parabola yang memerlukan integrasi pemahaman matematika dan fisika.

Selanjutnya, uji validitas dilakukan untuk setiap butir soal guna menilai kualitas dari masing-masing butir tersebut. Hasil uji validitas butir soal diperoleh dari Tabel output 10.1: *Item Fit Order*. Kualitas setiap butir soal dapat dievaluasi melalui *item fit order*, dengan melihat nilai *outfit means square* (MNSQ), *outfit Z-standard* (ZSTD), dan *point measure correlation* (Pt Measure Corr.). Hasil uji validitas dari setiap butir soal kemampuan kognitif ditampilkan pada Gambar 2 berikut.

TABLE 10.1 C:\Users\Nopal\Documents\Rasch2.PRN ZOU562NS.TXT Nov 05 2024 07:07
 INPUT: 34 Person 20 Item REPORTED: 34 Person 20 Item 2 CATS MINISTEP 5.8.2.0

Person: REAL SEP.: .00 REL.: .00 ... Item: REAL SEP.: 1.43 REL.: .67

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEASURE CORR.	AL EXP.	EXACT OBS%	MATCH EXP%	Item
6	32	34	-1.03	.73	1.09	.34	1.93	1.23	A-.23	.13	94.1	94.1	S6
17	14	34	2.26	.36	1.20	1.84	1.23	1.84	B-.11	.26	55.9	63.3	S17
13	13	34	2.39	.36	1.15	1.31	1.19	1.33	C-.03	.25	58.8	64.9	S13
9	23	34	1.09	.38	1.15	1.01	1.13	.77	D-.00	.25	61.8	68.6	S9
4	22	34	1.23	.37	1.07	.59	1.13	.86	E-.10	.25	67.6	66.4	S4
7	32	34	-1.03	.73	1.00	.21	.98	.22	F-.12	.13	94.1	94.1	S7
11	33	34	-1.76	1.02	1.00	.32	.87	.25	G-.09	.09	97.1	97.0	S11
12	33	34	-1.76	1.02	1.00	.32	.87	.25	H-.09	.09	97.1	97.0	S12
15	32	34	-1.03	.73	.98	.17	.72	-.17	I-.23	.13	94.1	94.1	S15
8	32	34	-1.03	.73	.93	.09	.63	-.32	J-.32	.13	94.1	94.1	S8
10	32	34	-1.03	.73	.93	.09	.63	-.32	K-.32	.13	94.1	94.1	S10
20	31	34	-.58	.61	.93	.02	.71	-.36	L-.32	.15	91.2	91.1	S20
16	33	34	-1.76	1.02	.92	.22	.38	-.37	M-.36	.09	97.1	97.0	S16
18	21	34	1.36	.36	.91	-.78	.90	-.74	N-.42	.25	70.6	64.5	S18
14	24	34	.94	.39	.88	-.71	.82	-.90	O-.47	.24	76.5	71.1	S14
19	18	34	1.75	.36	.78	-2.68	.76	-2.62	P-.66	.26	85.3	61.1	S19
MEAN	28.1	34.0	-.60	.86	.99	.15	.93	.06			83.1	82.0	
P.SD	6.9	.0	1.77	.53	.11	.96	.34	1.03			14.8	14.6	

TABLE 10.3 C:\Users\Nopal\Documents\Rasch2.PRN ZOU562NS.TXT Nov 05 2024 07:07
 INPUT: 34 Person 20 Item REPORTED: 34 Person 20 Item 2 CATS MINISTEP 5.8.2.0

Item CATEGORY/OPTION/DISTRACTOR FREQUENCIES: MISFIT ORDER

Gambar 2. Hasil Uji Validitas Setiap Butir Soal

Berdasarkan Gambar 2 yang ditampilkan, maka didapat interpretasi tiap butir soal seperti pada Tabel 7 berikut.

Tabel 7. Data Kualitas tiap Butir Soal

Butir Soal	Nilai Outfit		Pt. Measure corr	Jumlah Kriteria Terpenuhi	Interpretasi
	MNSQ	ZTSD			
S1	1,23	0,72	0,51	3 kriteria	Sangat sesuai
S2	0,71	-0,15	0,27	2 kriteria	Sesuai
S3	0,77	-0,41	0,36	2 kriteria	Sesuai
S4	1,13	0,86	0,25	2 kriteria	Sesuai
S5	0,90	-0,76	0,44	3 kriteria	Sangat Sesuai
S6	1,43	1,23	0,23	2 kriteria	Sesuai
S8	0,63	-0,32	0,32	2 kriteria	Sesuai
S9	1,13	0,77	0,00	2 kriteria	Sesuai
S10	0,63	-0,32	0,32	2 kriteria	Sesuai
S11	0,87	0,25	0,09	2 kriteria	Sesuai
S12	0,87	0,25	0,09	2 kriteria	Sesuai
S13	1,19	1,33	0,03	2 kriteria	Sesuai
S14	0,76	-0,47	0,41	3 kriteria	Sangat Sesuai
S15	0,72	-0,17	0,23	2 kriteria	Sesuai
S16	0,38	-0,37	0,46	2 kriteria	Sesuai
S17	1,23	1,84	0,11	2 kriteria	Sesuai
S18	0,90	-0,74	0,42	3 kriteria	Sangat Sesuai
S19	0,76	-2,62	0,66	2 kriteria	Sesuai
S20	0,71	-0,36	0,32	2 kriteria	Sesuai

Berdasarkan data pada tabel, maka diketahui hasil dari uji validitas tiap butir diperoleh 4 soal dengan interpretasi “Sangat Sesuai” dan 16 soal “Sesuai”. Berdasarkan hasil tersebut, peneliti memutuskan untuk menggunakan semua butir soal yang didapat. Dengan demikian, jumlah soal yang memenuhi kriteria *item-fit* dan layak digunakan pada penelitian ini adalah 20 butir soal.

Hasil analisis uji validitas menunjukkan bahwa dari 20 butir soal yang diuji, 4 butir soal mendapat interpretasi “Sangat Sesuai” dan 16 butir soal berada dalam kategori “Sesuai”. Interpretasi ini mencerminkan bahwa semua butir soal memiliki tingkat kesesuaian yang baik dengan indikator yang diukur. Keberadaan 4 butir soal dengan interpretasi sangat sesuai menunjukkan bahwa soal tersebut memiliki daya ukur yang sangat kuat dan mampu memberikan informasi yang jelas terkait dengan kemampuan kognitif siswa. Sementara itu, 16 butir soal yang dikategorikan sesuai tetap dianggap valid dan dapat memberikan kontribusi dalam pengukuran, meskipun tidak sekuat 4 butir lainnya. Keputusan untuk mempertahankan semua butir soal didasarkan pada pertimbangan bahwa seluruh instrumen masih berada dalam rentang yang dapat diterima dalam model Rasch, yang menunjukkan bahwa tidak ada soal yang memberikan dampak negatif terhadap keseluruhan reliabilitas tes.

Selain itu, instrumen ini dinilai memiliki *item-fit* yang memenuhi kriteria, yang berarti setiap butir soal memberikan kontribusi yang signifikan dalam mengukur variabel yang diinginkan. Dalam konteks pengembangan instrumen evaluasi, mempertahankan semua butir soal memberikan keuntungan dalam menciptakan variasi tingkat kesulitan yang lebih luas. Hal ini penting untuk mengakomodasi perbedaan kemampuan siswa, dari yang memiliki pemahaman dasar hingga yang mampu menerapkan konsep pada level yang lebih tinggi. Soal dengan interpretasi “Sesuai” memiliki peran penting dalam menciptakan distribusi skor yang lebih merata, yang pada akhirnya dapat meningkatkan reliabilitas instrumen secara keseluruhan.

Lebih lanjut, validitas butir soal yang tinggi menunjukkan bahwa setiap soal benar-benar mengukur aspek yang telah dirancang dalam kisi-kisi tes, tanpa adanya bias atau kesalahan konseptual. Hal ini sejalan dengan prinsip *construct validity* yang menegaskan bahwa instrumen harus secara akurat mencerminkan konstruk teoretis yang hendak diukur. Dengan demikian, instrumen ini tidak hanya memiliki validitas secara statistik tetapi juga memiliki dasar teoritis yang kuat. Validitas yang baik juga meningkatkan ketepatan interpretasi hasil tes, memungkinkan guru dan peneliti untuk menggunakan hasil tersebut dalam mengevaluasi pembelajaran dan merancang strategi pengajaran yang lebih efektif.

Dengan demikian, keputusan untuk menggunakan semua butir soal dalam penelitian ini merupakan langkah yang tepat dalam menjaga keseimbangan antara validitas dan reliabilitas instrumen. Instrumen ini tidak hanya mampu memberikan gambaran yang komprehensif tentang kemampuan kognitif siswa, tetapi juga dapat diandalkan untuk digunakan dalam berbagai konteks evaluasi pendidikan, khususnya dalam pembelajaran fisika di tingkat SMA.

Uji Reliabilitas

Reliabilitas instrumen merupakan aspek krusial dalam pengembangan tes, karena menunjukkan sejauh mana instrumen tersebut konsisten dalam mengukur apa yang seharusnya diukur. Instrumen yang reliabel akan memberikan hasil yang stabil dan dapat diandalkan meskipun digunakan dalam kondisi atau waktu yang berbeda. Dalam konteks evaluasi kemampuan

kognitif fisika, reliabilitas yang tinggi memastikan bahwa setiap butir soal mampu memberikan gambaran yang akurat mengenai tingkat pemahaman siswa, tanpa dipengaruhi oleh faktor eksternal yang tidak relevan. Berikut adalah hasil analisis reliabilitas yang diperoleh dari uji coba instrumen ini.

Reliabilitas dalam penelitian ini dianalisis menggunakan model Rasch dengan bantuan software Ministep versi 5.8.3.0 melalui fitur output pada menu *Summary Statistic*. Data yang diperoleh dari menu tersebut mencakup *person reliability* (p), *item reliability* (r), dan *Cronbach's alpha* (KR-20). *Person reliability* merepresentasikan tingkat konsistensi dari peserta, sementara *item reliability* menggambarkan konsistensi butir soal dalam instrumen yang digunakan. Sementara itu, *Cronbach's alpha* (KR-20) menunjukkan tingkat keterkaitan antara *person reliability* dan *item reliability*, memberikan gambaran keseluruhan mengenai kualitas instrumen dalam mengukur kemampuan siswa. Data yang didapat untuk *person reliability*, *item reliability*, dan *Cronbach's alpha* (KR-20) disajikan pada Gambar berikut.

SUMMARY OF 34 MEASURED Person								
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	16.5	20.0	1.87	.72	.99	-.02	.93	.05
SEM	.2	.0	.10	.01	.07	.20	.14	.16
P.SD	1.1	.0	.55	.06	.41	1.13	.80	.90
S.SD	1.1	.0	.56	.06	.41	1.14	.81	.92
MAX.	18.0	20.0	2.69	.83	2.03	2.38	3.04	2.31
MIN.	14.0	20.0	.72	.64	.51	-1.69	.27	-1.30
REAL RMSE	.78	TRUE SD	.00	SEPARATION	.00	Person RELIABILITY	.00	
MODEL RMSE	.72	TRUE SD	.00	SEPARATION	.00	Person RELIABILITY	.00	
S.E. OF Person MEAN = .10								
Person RAW SCORE-TO-MEASURE CORRELATION = 1.00								
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .62 SEM = 1.09								
STANDARDIZED (50 ITEM) RELIABILITY = .00								
SUMMARY OF 16 MEASURED Item								
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	26.6	34.0	.00	.62	.99	.15	.93	.06
SEM	1.8	.0	.38	.06	.03	.25	.09	.26
P.SD	7.0	.0	1.46	.25	.11	.96	.34	1.03
S.SD	7.2	.0	1.51	.26	.11	1.00	.35	1.06
MAX.	33.0	34.0	2.39	1.02	1.20	1.84	1.93	1.84
MIN.	13.0	34.0	-1.76	.36	.78	-2.68	.38	-2.62
REAL RMSE	.67	TRUE SD	1.30	SEPARATION	1.93	Item RELIABILITY	.79	
MODEL RMSE	.67	TRUE SD	1.30	SEPARATION	1.95	Item RELIABILITY	.79	
S.E. OF Item MEAN = .38								
MAXIMUM EXTREME SCORE: 4 Item 20.0%								

Gambar 3. Hasil uji reliabilitas butir soal

Berdasarkan Gambar 3, terlihat bahwa nilai reliabilitas peserta (*person reliability*) adalah 0,00, yang diinterpretasikan sebagai "rendah". Hal ini mengindikasikan bahwa jumlah sampel dan butir soal pada instrumen tes ini mungkin masih kurang optimal. Untuk mencapai nilai reliabilitas peserta yang tinggi, diperlukan sampel dengan variasi kemampuan yang luas serta jumlah butir soal yang lebih banyak (Sumintono & Widhiarso, 2015). Sementara itu, reliabilitas butir soal (*item reliability*) mencapai 0,79, yang diinterpretasikan sebagai "cukup". Ini menunjukkan bahwa instrumen tes kemampuan kognitif yang digunakan dalam penelitian ini memiliki konsistensi yang baik dalam bobot soal dan pilihan jawabannya. Selain itu, nilai *cronbach alpha* (KR-20) adalah 0,62, yang dikategorikan sebagai "Baik", menandakan bahwa

instrumen ini memiliki kualitas yang cukup. Dengan demikian, instrumen tes ini mampu mengidentifikasi hubungan antara siswa (reliabilitas peserta) dan butir soal (reliabilitas butir soal) (Zakwandi & Istiyono, 2022).

Rendahnya reliabilitas peserta menunjukkan adanya keterbatasan dalam variasi kemampuan siswa yang mengikuti uji coba, yang dapat berdampak pada kurang optimalnya hasil pengukuran. Hal ini mengindikasikan bahwa instrumen tes mungkin belum cukup sensitif untuk membedakan siswa dengan kemampuan yang beragam. Keterbatasan ini bisa disebabkan oleh jumlah sampel yang terlalu kecil atau homogen, baik dari segi kemampuan akademik maupun latar belakang pendidikan. Dalam konteks ini, strategi untuk meningkatkan reliabilitas peserta meliputi pengambilan sampel yang lebih luas dan representatif, mencakup siswa dari berbagai tingkat kelas, sekolah, atau bahkan daerah yang berbeda. Dengan melibatkan lebih banyak siswa, instrumen akan diuji dalam kondisi yang lebih variatif, sehingga hasil pengukuran menjadi lebih akurat dan dapat diandalkan. Selain itu, penambahan butir soal yang mencakup tingkat kesulitan berjenjang—dari level mudah hingga sulit—akan memperkaya instrumen dan meningkatkan peluang untuk mengidentifikasi perbedaan kemampuan siswa secara lebih detail.

Di sisi lain, nilai reliabilitas butir soal yang tergolong cukup mencerminkan bahwa kualitas soal dalam instrumen ini sudah berada pada tingkat yang memadai. Hal ini menunjukkan bahwa meskipun terdapat kendala dalam reliabilitas peserta, struktur dan desain butir soal telah memenuhi standar untuk mengukur kemampuan kognitif siswa dengan konsistensi yang baik. Konsistensi ini penting dalam memastikan bahwa setiap butir soal memberikan kontribusi yang signifikan dalam mengukur aspek kognitif yang relevan dengan topik yang diujikan, dalam hal ini gerak parabola. Namun, meskipun reliabilitas butir soal sudah cukup, tetap diperlukan upaya untuk meningkatkan kualitas instrumen secara keseluruhan. Salah satu cara yang dapat dilakukan adalah dengan melakukan analisis lebih mendalam terhadap butir soal yang memiliki tingkat kesulitan ekstrem—baik yang terlalu mudah maupun terlalu sulit—dan menyesuaikannya agar lebih sesuai dengan kemampuan rata-rata siswa. Pendekatan ini tidak hanya akan meningkatkan reliabilitas butir soal, tetapi juga akan berkontribusi pada peningkatan reliabilitas peserta secara tidak langsung, karena siswa dihadapkan pada soal yang sesuai dengan rentang kemampuan mereka. Dengan demikian, instrumen ini akan lebih efektif dalam mengevaluasi dan memetakan kemampuan kognitif siswa, sekaligus memberikan wawasan yang lebih komprehensif bagi guru dalam merancang strategi pembelajaran yang tepat.

Uji Tingkat Kesukaran Butir Soal

Hasil uji tingkat kesukaran butir soal dapat dilihat dari Measure (ME) dan Standar Deviasi dari tiap soal. Dan data ini dapat dilihat dari Gambar 4. Berdasarkan Gambar 4 dan interpretasi data tingkat kesukaran butir soal yang disampaikan pada bab metode penelitian, maka didapatkan interpretasi tiap butir soal seperti pada Tabel 8.

TABLE 10.1 C:\Users\Nopal\Documents\Rasch2.PRN ZOU5G2WS.TXT Nov 05 2024 07:07
 INPUT: 34 Person 20 Item REPORTED: 34 Person 20 Item 2 CATS MINISTEP 5.8.2.0
 Person: REAL SEP.: .00 REL.: .00 ... Item: REAL SEP.: 1.43 REL.: .67

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S E.	INFIT MNSQ ZSTD	OUTFIT MNSQ ZSTD	PTMEASUR-AL CORR.	EXP.	EXACT OBS%	MATCH EXP%	Item
6	32	34	-1.03	.73	1.09 .34	1.93 1.23	A-.23 .13	94.1	94.1	S6	
17	14	34	2.26	.36	1.20 1.84	1.23 1.84	B-.11 .26	55.9	63.3	S17	
13	13	34	2.39	.36	1.15 1.31	1.19 1.33	C-.03 .25	58.8	64.9	S13	
9	23	34	1.09	.38	1.15 1.01	1.13 .77	D-.00 .25	61.8	68.6	S9	
4	22	34	1.23	.37	1.07 .59	1.13 .86	E-.10 .25	67.6	66.4	S4	
7	32	34	-1.03	.73	1.00 .21	.98 .22	F-.12 .13	94.1	94.1	S7	
11	33	34	-1.76	.02	1.00 .32	.87 .25	G-.09 .09	97.1	97.0	S11	
12	33	34	-1.76	.02	1.00 .32	.87 .25	H-.09 .09	97.1	97.0	S12	
15	32	34	-1.03	.73	.98 .17	.72 -.17	h-.23 .13	94.1	94.1	S15	
8	32	34	-1.03	.73	.93 .09	.63 -.32	g-.32 .13	94.1	94.1	S8	
10	32	34	-1.03	.73	.93 .09	.63 -.32	f-.32 .13	94.1	94.1	S10	
20	31	34	-.58	.61	.93 .02	.71 -.36	e-.32 .15	91.2	91.1	S20	
16	33	34	-1.76	.02	.92 .22	.38 -.37	d-.36 .09	97.1	97.0	S16	
18	21	34	1.36	.36	.91 -.78	.90 -.74	c-.42 .25	70.6	64.5	S18	
14	24	34	.94	.39	.88 -.71	.82 -.90	b-.47 .24	76.5	71.1	S14	
19	18	34	1.75	.36	.78 -2.68	.76 -2.62	a-.66 .26	85.3	61.1	S19	
MEAN	28.1	34.0	-.60	.86	.99 .15	.93 .06		83.1	82.0		
P.SD	6.9	.0	1.77	.53	.11 .96	.34 1.03		14.8	14.6		

Gambar 4. Data Measure dan Standar Deviasi Butir Soal

Tabel 8. Data tingkat kesukaran tiap butir soal

Nomor Butir Soal	Measure (ME)	Standar Deviasi (SD)	Kriteria	Interpretasi
S1	-1,78	1,77	- 1,71 < - 1,77	Mudah
S2	-2,09	1,77	- 2,09 < - 1,77	Mudah
S3	-2,09	1,77	- 2,09 < - 1,77	Mudah
S4	1,23	1,77	- 1,77 ≤ 1,23 ≤ 1,77	Sedang
S5	-1,81	1,77	- 1,81 < - 1,77	Mudah
S6	-1,03	1,77	- 1,77 ≤ - 1,03 ≤ 1,77	Sedang
S7	-1,03	1,77	- 1,77 ≤ - 1,03 ≤ 1,77	Sedang
S8	-1,03	1,77	- 1,77 ≤ - 1,03 ≤ 1,77	Sedang
S9	1,09	1,77	- 1,77 ≤ 1,09 ≤ 1,77	Sedang
S10	-1,03	1,77	- 1,77 ≤ - 1,03 ≤ 1,77	Sedang
S11	-1,76	1,77	- 1,77 ≤ - 1,76 ≤ 1,77	Sedang
S12	-1,76	1,77	- 1,77 ≤ - 1,76 ≤ 1,77	Sedang
S14	0,94	1,77	- 1,77 ≤ 0,94 ≤ 1,77	Sedang
S15	-1,03	1,77	- 1,77 ≤ - 1,03 ≤ 1,77	Sedang
S16	-1,78	1,77	- 1,78 < - 1,77	Mudah
S17	2,26	1,77	2,26 > 1,77	Sulit
S18	1,36	1,77	- 1,77 ≤ 1,36 ≤ 1,77	Sedang
S19	2,39	1,77	2,39 > 1,77	Sulit
S20	-0,58	1,77	- 1,77 ≤ - 0,58 ≤ 1,77	Sedang

Berdasarkan Tabel 8 di atas, maka diketahui lah dari total sebaran 20 soal, 5 soal masuk kedalam kategori “Mudah”, 12 soal masuk dalam kategori “Sedang”, dan 3 soal masuk dalam kategori “Sulit”. Dengan distribusi ini, instrumen tes memiliki keseimbangan antara soal yang lebih mudah dan lebih sulit, yang penting dalam memastikan bahwa tes dapat mengukur berbagai level kemampuan siswa secara efektif.

Berdasarkan distribusi soal yang ada, keseimbangan antara soal mudah, sedang, dan sulit memberikan Gambaran yang cukup baik mengenai efektivitas instrumen tes dalam mengukur

kemampuan kognitif siswa. Dengan adanya variasi tingkat kesulitan soal, instrumen ini memungkinkan penilaian yang lebih komprehensif terhadap pemahaman siswa. Soal yang lebih mudah memberikan dasar bagi siswa dengan pemahaman yang lebih rendah untuk berhasil, sementara soal yang lebih sulit dapat mengungkapkan tingkat pemahaman yang lebih mendalam pada siswa yang lebih maju. Hal ini juga membantu memastikan bahwa tes dapat membedakan antara tingkat kemampuan siswa yang beragam.

Selain itu, distribusi soal yang demikian penting untuk menjaga validitas tes. Dengan adanya variasi dalam tingkat kesulitan soal, tes tidak hanya berfokus pada kemampuan dasar, tetapi juga mengukur keterampilan analitis dan aplikatif siswa. Pembagian ini memudahkan pendidik untuk menganalisis hasil tes secara lebih menyeluruh, dan memberikan informasi yang lebih mendalam tentang area mana yang perlu diperkuat dalam pembelajaran. Oleh karena itu, instrumen tes dengan distribusi soal yang proporsional ini diharapkan dapat memberikan hasil yang lebih akurat dalam mengevaluasi kemampuan kognitif siswa secara keseluruhan.

SIMPULAN

Berdasarkan hasil analisis, instrumen tes kemampuan kognitif siswa SMA pada materi Gerak Parabola menunjukkan kualitas yang memadai untuk mengukur tingkat pemahaman siswa. Dari segi unidimensionalitas, nilai *raw variance explained by measures* sebesar 28,7% menunjukkan bahwa instrumen ini mampu mengukur satu variabel dengan baik tanpa adanya pengaruh dari variabel lain, ditunjang oleh nilai *unexplained variance in 1st contrast* sebesar 14,6% yang masuk dalam kategori baik. Uji validitas menunjukkan bahwa seluruh 20 butir soal memenuhi kriteria item-fit, dengan 4 butir soal tergolong "Sangat Sesuai" dan 16 butir "Sesuai", sehingga layak digunakan. Distribusi tingkat kesukaran soal juga seimbang, dengan 5 soal "Mudah", 12 soal "Sedang", dan 3 soal "Sulit", yang memungkinkan instrumen ini untuk mencakup berbagai tingkat kemampuan siswa. Dari segi reliabilitas, *item reliability* bernilai 0,79 yang cukup baik, sedangkan *person reliability* hanya mencapai 0,00, menunjukkan adanya kemungkinan keterbatasan pada sampel atau butir soal dalam instrumen ini. Meskipun demikian, nilai *Cronbach alpha* (KR-20) sebesar 0,62 masih termasuk kategori "Baik", yang mendukung kualitas instrumen sebagai alat ukur yang cukup memadai dalam menilai kemampuan kognitif siswa. Dengan demikian, instrumen ini dapat digunakan oleh guru di sekolah sebagai instrumen tes kemampuan kognitif.

DAFTAR PUSTAKA

- Fadilah, F., & Sari, D. (2021). Implementasi model Rasch dalam pengembangan instrumen tes untuk mengukur kemampuan kognitif siswa SMA. *Jurnal Pendidikan dan Pembelajaran*, 29(1), 85-92.
- Haryanto, H., & Rinaldi, Y. (2019). Penggunaan model Rasch dalam menganalisis instrumen penilaian kemampuan kognitif siswa pada materi fisika. *Jurnal Ilmiah Pendidikan Fisika*, 7(1), 34-42.
- Kusnadi, N., & Hidayati, N. (2020). Penggunaan analisis Rasch dalam evaluasi kualitas soal kemampuan kognitif siswa pada materi biologi. *Jurnal Biologi dan Pembelajaran*, 11(3), 175-184.
- Nurdin, A., & Safitri, E. (2020). Evaluasi kualitas soal menggunakan model Rasch: studi pada tes kemampuan kognitif siswa dalam pembelajaran matematika. *Jurnal Pendidikan Matematika*, 8(1), 55-64.

- Setiawan, A., & Dewi, R. (2022). Pengukuran kemampuan kognitif siswa dengan analisis Rasch pada instrumen tes fisika. *Jurnal Sains dan Pendidikan*, 12(2), 220-229.
- Yulianti, Y., & Amin, M. (2018). Analisis kemampuan kognitif siswa pada pembelajaran fisika menggunakan pendekatan Rasch. *Jurnal Penelitian Pendidikan*, 25(3), 147-158.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. Trim Komunikata.
- Popham, W. J. (2008). *Classroom Assessment: What Teachers Need to Know* (hlm. 78). Pearson Education.
- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Advances in Science Education*. Springer. (hlm. 101).
- Rahman, M., Salam, M. D., Rahim, M., & Johari, M. (2021). Pengembangan instrumen evaluasi untuk pendidikan sains. *Journal of Educational Measurement*, 58(2), 143-157.
- Ibrahim, M. R., Farooq, M. S., & Malik, M. (2020). Evaluasi kognitif siswa SMA dengan pendekatan Rasch. *Journal of Assessment in Education*, 37(1), 45-59.
- Mustafa, M., & Yusoff, M. S. B. (2022). Analisis keterampilan kognitif dengan pendekatan model Rasch. *International Journal of Science Education*, 44(4), 214-230.
- Widjaja, Y. M., Sutrisno, R., & Nur, H. (2019). Evaluasi kemampuan kognitif siswa pada topik gerak parabola menggunakan model Rasch. *Jurnal Pendidikan Fisika Indonesia*, 15(3), 109-120.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Model Rasch untuk Penelitian Ilmu-Ilmu Sosial*. Trim Komunikata.
- Sumintono, B., & Widhiarso, W. (2014). *Pengantar Model Rasch untuk Penilaian Pendidikan*. Trim Komunikata Press.
- Fischer, G. H. (1995). *Derivations of Rasch models*. Springer-Verlag.
- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. MESA Press.
- Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge.
- Engelhard, G. (2013). *Invariant measurement: Using Rasch models in the social, behavioral, and health sciences*. Routledge.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Psychology Press.
- Sumardi, S., & Prasetyo, A. (2019). Penerapan model Rasch dalam evaluasi pembelajaran fisika berbasis kognitif. *Jurnal Inovasi Pendidikan Fisika*, 8(2), 112-120.
- Ahmad, F., & Rahmawati, I. (2021). Validitas dan reliabilitas instrumen evaluasi fisika menggunakan analisis Rasch. *Jurnal Fisika dan Aplikasinya*, 13(1), 45-56.
- Rusman, T. R., & Ardiansyah, A. (2020). Penggunaan model Rasch dalam mengukur kesesuaian butir soal pada tes fisika. *Jurnal Pendidikan Fisika Indonesia*, 16(2), 120-132.
- Zakwandi, R., & Istiyono, E. (2022). Evaluating Student Computational Thinking Skills in Physics Experimental Class. *2022 13th International Conference on E-Education, E-Business, E-Management, and E-Learning (IC4E)*, 104-109.
- Zulkifli, Z., & Ramdani, I. (2022). Analisis unidimensionalitas pada soal kemampuan kognitif siswa menggunakan model Rasch. *Jurnal Pendidikan dan Teknologi*, 19(4), 231-245.