



Pengembangan dan Validasi Instrumen Keterampilan Berpikir Kritis Siswa SMA pada Materi Dinamika Rotasi: Pendekatan *Rasch* Model

Abdul Salam^{1*}, Lina Aviyanti¹, Nur Endah Susilowati¹, Silmi Fitriani^{1*}, Rahma Malizka¹

¹ Pendidikan Fisika, FPMIPA, Universitas Pendidikan Indonesia, Bandung, Indonesia.

DOI: <https://doi.org/10.29303/goescienceed.v7i3.2452>

Article Info:

Received : 20 Juni 2026
Revised : 27 Juni 2026
Accepted : 03 Juli 2026
Published : 07 Juli 2026

Correspondence:

Abdul Salam

Phone: +6283822454063

Abstract: Critical thinking is a core competency required in 21st-century education, but a valid and reliable instrument for assessing this skill within the topic of rotational dynamics remains limited. This study aims to develop and validate a critical thinking skills assessment instrument for senior high school students on the topic of rotational dynamics, employing the *Rasch* model for comprehensive validation. Instrument development followed the ADDIE model, involving 34 students during the implementation and evaluation phases. The instrument comprises 27 items constructed based on Facione's critical thinking aspects. Content validity analysis yielded a mean Aiken's V of 0.93, exceeding the critical threshold of 0.80. Unidimensionality testing revealed a raw variance explained by measures of 28.9% (>20%), with unexplained variance in contrasts 1–5 of the PCA residuals ranging from 5.8% to 9.9% (<15%), indicating that the instrument consistently measures a single construct. Twenty-six of the 27 items demonstrated fit to the *Rasch* model. Reliability analysis yielded an item reliability of 0.81 (good), a person reliability of 0.68 (fair), and a Cronbach's Alpha of 0.72 (good). Wright Map and Item Characteristic Curve (ICC) analyses further confirmed the alignment between item difficulty levels and student ability levels. The developed instrument is demonstrated to be valid and reliable for measuring critical thinking skills among senior high school students in the context of rotational dynamics.

Keywords: Critical Thinking Instrument; Critical Thinking Skills; *Rasch* Model; Rotational Dynamics.

Citation: Salam, A., Aviyanti, L., Susilowati, N. E., Fitriani, S., & Malizka, R. (2026). Pengembangan dan Validasi Instrumen Keterampilan Berpikir Kritis Siswa SMA pada Materi Dinamika Rotasi: Pendekatan *Rasch* Model. *Jurnal Pendidikan, Sains, Geologi, Dan Geofisika (GeoScienceEd Journal)*, 7(3), 3216–3226. <https://doi.org/10.29303/goescienceed.v7i3.2452>

Pendahuluan

Keterampilan berpikir kritis merupakan salah satu kompetensi inti yang menjadi tuntutan dalam pendidikan abad ke-21. *Organisation for Economic Co-operation and Development* (OECD) melalui kerangka *Programme for International Student Assessment* (PISA) secara konsisten menempatkan penalaran ilmiah dan pemikiran kritis sebagai salah satu indikator utama kemampuan siswa dalam penilaian internasional (OECD, 2023). Namun demikian, data PISA 2022 menunjukkan bahwa Indonesia berada pada peringkat ke-68 dari 81 negara peserta dalam domain sains, dengan rata-rata skor 383 poin, jauh di bawah rata-rata

OECD sebesar 485 poin. Kondisi ini mengindikasikan bahwa keterampilan berpikir kritis siswa Indonesia, khususnya dalam mata pelajaran sains perlu mendapat perhatian serius dan upaya peningkatan yang terencana.

Menurut Facione (1990) dalam laporan *Delphi Report*, berpikir kritis adalah penilaian yang bertujuan dan reflektif diri (*purposeful self-regulatory judgment*) yang menghasilkan interpretasi, analisis, evaluasi, dan inferensi, serta penjelasan dari pertimbangan bukti, konseptual, metodologis, atau kontekstual yang mendasari penilaian tersebut. Lebih lanjut, Facione (2023) menyempurnakan *framework* tersebut dengan mengidentifikasi enam aspek inti, yaitu *interpretation*,

analysis, inference, evaluation, explanation, dan self-regulation. Keenam aspek tersebut bersifat hierarkis dan saling berkaitan, sehingga pengukurannya menuntut instrumen yang dirancang secara cermat dan sistematis.

Dalam pembelajaran fisika, berpikir kritis memegang peranan yang sangat penting. Pembelajaran fisika menuntut siswa tidak hanya menguasai konsep dan prosedur perhitungan (Wafa & Jatmiko, 2022), tetapi juga mengembangkan kemampuan untuk menginterpretasikan fenomena alam, menganalisis hubungan sebab-akibat, mengevaluasi bukti empiris, dan menarik inferensi yang valid (Azizah, Handayanto, Wisodo, & Fernando, 2024; Nuryanti, Masykuri, & Susilowati, 2018; Susilawati, Samsudin, & Siahaan, 2020). Evaluasi keterampilan berpikir kritis siswa merupakan komponen penting dalam meningkatkan kualitas pengajaran. Tahap awal evaluasi adalah dengan melakukan pengukuran kemampuan siswa, dengan begitu pendidik dapat mengetahui sejauh mana siswa mampu menganalisis, mengevaluasi, dan memecahkan masalah kompleks secara kritis (Nurlailah & Hamdu, 2021). Dalam upaya mendiagnosis dan memantau perkembangan keterampilan berpikir kritis siswa secara sistematis, para pendidik umumnya menggunakan dua pendekatan, yaitu wawancara dan tes tertulis. Meskipun wawancara memungkinkan eksplorasi mendalam terhadap proses berpikir siswa secara individual, pendekatan ini kurang efisien untuk diterapkan pada jumlah siswa yang besar dalam konteks pembelajaran di kelas (Costan, Costan, Weißbach, & Kulgemeyer, 2026). Sebaliknya, tes tertulis dalam bentuk instrumen yang terstruktur memungkinkan pengukuran keterampilan berpikir kritis secara efisien, konsisten, dan dapat direplikasi pada skala yang lebih luas (Mesic et al., 2019). Namun demikian, efektivitas tes tertulis sangat bergantung pada kualitas instrumen yang digunakan. Instrumen yang tidak dirancang dan divalidasi secara memadai berisiko menghasilkan data yang tidak mencerminkan kemampuan siswa yang sesungguhnya, sehingga keputusan pedagogis yang diambil berdasarkan data tersebut pun menjadi tidak tepat sasaran (Retnawati, 2014). Pengembangan instrumen tes keterampilan berpikir kritis yang berkualitas tinggi menjadi prasyarat utama bagi terlaksananya penilaian yang akurat dan bermakna dalam pembelajaran fisika.

Dalam beberapa tahun terakhir, pengembangan instrumen keterampilan berpikir kritis dalam pendidikan fisika menunjukkan tren yang mengarah pada pengembangan instrumen yang bersifat spesifik terhadap konteks materi dan didukung oleh analisis psikometrik yang lebih komprehensif menggunakan Model *Rasch*. Berbagai penelitian telah mengembangkan dan memvalidasi instrumen keterampilan berpikir kritis pada topik fisika seperti fluida dinamis oleh Afandi et al., (2025) dan Halim et al., (2026), usaha dan pesawat

sederhana oleh Noperi et al., (2025), pemanasan global oleh Aviyaniti et al., (2025), serta Gelombang mekanik oleh (Azizah et al., 2024). Perkembangan tersebut menunjukkan bahwa penelitian terkini tidak hanya fokus pada pengukuran hasil belajar (*achievement*), tetapi juga pada pengembangan instrumen diagnostik yang mampu memetakan profil keterampilan berpikir kritis secara detail sesuai konteks materi pembelajaran. Namun, kajian literatur menunjukkan bahwa pengembangan instrumen keterampilan berpikir kritis pada materi dinamika rotasi masih sangat terbatas dibandingkan topik fisika lainnya. Sejauh penelusuran penulis, hingga saat ini belum diterumkan instrumen keterampilan berpikir kritis pada materi dinamika rotasi yang dikembangkan secara komprehensif mencakup keenam aspek berpikir kritis Facione (2023) sekaligus divalidasi menggunakan pendekatan *Rasch* model. Ketidadaan instrumen yang memadai berdampak langsung pada perencanaan pembelajaran, karena tanpa alat ukur yang tepat, pendidik menghadapi kesulitan dalam mendiagnosis kemampuan siswa untuk merancang metode pembelajaran yang terarah (Retnawati, Djidu, Apino, & Anazifa, 2018). Kebutuhan ini menjadi semakin mendesak mengingat dinamika rotasi merupakan salah satu materi fisika dengan tingkat abstraksi dan kompleksitas representasi yang tinggi. Materi ini menuntut siswa mengintegrasikan konsep translasi dan rotasi secara simultan. Kompleksitas tersebut menjadikan dinamika rotasi sebagai materi yang rentan memunculkan miskonsepsi, sehingga proses berpikir kritis siswa pada materi ini berpotensi berbeda dibanding materi fisika lainnya. Oleh karena itu, pengembangan instrumen yang valid tidak hanya penting untuk diagnosis, tetapi juga untuk mendukung pendidik dalam merancang pembelajaran yang menghubungkan konsep akademis dengan konteks dunia nyata (Aviyaniti, Gani, Febriyanti, Ribie, & Nawas, 2024), sehingga meningkatkan relevansi pembelajaran bagi siswa. Oleh karena itu, penelitian ini mengisi celah penelitian tersebut melalui pengembangan instrumen keterampilan berpikir kritis berbasis kerangka Facione pada materi dinamika rotasi yang menggunakan pendekatan *Rasch* model. Kebaharuan penelitian ini terletak pada cakupan keenam aspek berpikir kritis secara utuh dan validasi psikometrik berbasis *Rasch* model belum pernah diterapkan pada materi dinamika rotasi.

Lebih jauh, meskipun berbagai instrumen telah dikembangkan, sebagian besar hanya dianalisis menggunakan *Classical Test Theory* (CTT) yang memiliki kelemahan yaitu parameter butir soal yang dihasilkan bersifat sampel-dependen sehingga karakteristik butir dan kemampuan siswa tidak dapat dipisahkan satu sama lain (Hambleton, Swaminathan, & Rogers, 1991). Keterbatasan ini dapat diatasi melalui *Rasch* model yang

menawarkan kerangka *psikometrik* yang lebih *robust*, dimana mempertimbangkan kesulitan butir soal dan kemampuan siswa, menghasilkan penilaian yang lebih akurat yang melampaui sekadar mengidentifikasi jawaban benar (Cascella, Giberti, & Bolondi, 2020). Model *Rasch* merumuskan kerangka kerja matematis yang menghubungkan siswa dan butir soal secara simultan pada skala interval yang sama (Sumintono, 2018). Model ini juga mengembangkan instrumen yang valid dan reliabel untuk mengukur kemampuan siswa pada berbagai tingkat kemampuan secara akurat pada beragam konteks pembelajaran (Yokhebed, Karmadi, & Nastiti, 2025). Sebagai model stokastik, analisis *Rasch* memungkinkan data mentah tes untuk ditransformasikan secara linier menjadi skor standar, sehingga memungkinkan deteksi tingkat keterampilan siswa dan kesulitan butir soal pada skala logit standar (Nurdini et al., 2020). Penggunaan model *Rasch* dalam mengukur keterampilan berpikir kritis membantu memastikan bahwa tes tersebut secara konsisten mengukur keterampilan berpikir kritis siswa di berbagai tingkat kemampuan, sehingga memastikan bahwa hasil penilaian benar-benar mencerminkan konstruk yang diukur (Kaur, Mantri, Nagabhushan, & Singh, 2024; John M Linacre, 2026).

Berdasarkan latar belakang tersebut, penelitian ini bertujuan mengembangkan dan memvalidasi instrumen keterampilan berpikir kritis pada materi dinamika rotasi menggunakan pendekatan *Rasch* model sehingga menghasilkan instrumen yang tidak hanya valid secara teoritis dan konseptual, tetapi memenuhi standar *psikometrik* yang ketat untuk pengukuran konstruk keterampilan berpikir kritis yang akurat dan konsisten.

Metode

Penelitian ini menerapkan metode penelitian dan pengembangan (*Research and Development*) dengan mengadopsi model ADDIE yang meliputi lima tahapan utama, yakni *analysis*, *design*, *development*, *implementation*, dan *evaluation* (Branch, 2009). Tahap analisis, dilakukan kajian literatur terkait instrumen keterampilan berpikir kritis dalam pembelajaran fisika serta telaah teoritis mengenai konsep berpikir kritis berdasarkan Facione (2023). Tahap desain, mencakup perancangan kerangka kerja dan kisi-kisi instrumen yang diselaraskan dengan tujuan pembelajaran fisika. Tahap pengembangan, meliputi penyusunan butir-butir soal berdasarkan setiap aspek berpikir kritis yang relevan dengan topik dinamika rotasi yang selanjutnya divalidasi isi oleh para ahli. Tahap implementasi, dilakukan melalui uji empiris kepada siswa. Tahap evaluasi mencakup analisis kualitas instrumen menggunakan *Rasch* model. Sampel penelitian ini berjumlah 34 siswa kelas XI pada salah satu sekolah di Bandung Barat. Data yang hilang atau respons yang tidak lengkap tidak ada dalam analisis ini,

karena semua siswa menyelesaikan instrumen sepenuhnya. Pengumpulan data dilakukan menggunakan instrumen keterampilan berpikir kritis. Instrumen berbentuk soal pilihan ganda yang terdiri dari 27 butir soal yang mewakili setiap aspek berpikir kritis seperti yang ditunjukkan pada Tabel 1. Setiap butir soal dinilai secara *dichotomous* yaitu nilai satu poin jika jawaban benar dan nol poin jika jawaban salah atau tidak dijawab. Total skor siswa merupakan jumlah jawaban benar dari 27 butir soal. Contoh butir soal instrumen keterampilan berpikir kritis (aspek *inference*) ditunjukkan pada Gambar 1.

Tabel 1. Aspek Keterampilan Berpikir Kritis dan Distribusi Butir Soal

Aspek	Sub Aspek	Butir Soal
Interpretation	Categorize	1, 7, 13, 19
	Decode significance	
	Clarify meaning	
Analysis	Examine ideas	2, 8, 14, 20
	Identify arguments	
	Identify reasons and claims	
Inference	Query evidence	3, 9, 15, 21
	Conjecture alternatives	
	Draw logically valid or justified conclusions	
Evaluation	Assess credibility of claims	4, 10, 16, 22, 25
	Assess quality of arguments that were made using inductive or deductive reasoning	
	State results	
Explanation	Justify procedures	5, 11, 17, 23, 26
	Present arguments	
Self-Regulation	Self-monitor	6, 12, 18, 24, 27
	Self-correct	

(Facione, 2023)

Instrumen melewati proses validasi isi oleh lima orang ahli yang kompeten di bidang pendidikan fisika. Lima orang ahli tersebut terdiri dari tiga Dosen Pendidikan Fisika dan dua Guru Fisika. Para ahli meninjau instrumen secara independen dan memberikan skala penilaian pada setiap butir soal menggunakan skala lima poin untuk setiap kriteria (5 = sangat sesuai, 4 = sesuai, 3 = cukup sesuai, 2 = tidak sesuai, dan 1 = sangat tidak sesuai). Kriteria penilaian ahli diadaptasi dari penelitian Terwee et al., (2007) yang terdiri dari aspek materi, konstruksi, dan interpretasi dengan lima skala penilaian. Validitas isi dianalisis menggunakan Aiken's V dengan batas nilai kritis 0,80 pada taraf signifikansi 5% (Aiken, 1985).

Sebuah helikopter penyelamat yang memiliki empat baling-baling dengan panjang L dan massa M . Baling-baling berbentuk batang tipis yang berputar pada sumbu rotasi yang tegak lurus terhadap panjangnya. Helikopter bergerak dengan kecepatan sudut ω .

Gambar 13. Dua helikopter dengan panjang baling-baling berbeda

Jika panjang setiap baling-baling helikopter diperbesar menjadi dua kali lipat, sementara massa tiap baling dan kecepatan sudut tetap, maka perubahan energi kinetik rotasi total sistem adalah...

Alur penalaran yang diharapkan siswa:

Untuk batang tipis yang berputar tegak lurus terhadap panjangnya: $I_{\text{batang}} = \frac{1}{3}ML^2$

Jika panjang menjadi $2L$: $I'_{\text{batang}} = \frac{1}{3}M(2L)^2 = 4(\frac{1}{3}ML^2) = 4I_{\text{batang}}$

Energi kinetik rotasi: $E_k = \frac{1}{2}I\omega^2$ (saat tetap)

Total momen inersia (4 baling): $I_{\text{total}} = 4I_{\text{batang}} \Rightarrow I'_{\text{total}} = 4I_{\text{total}}$

Maka, $E'_k = \frac{1}{2}I'_{\text{total}}\omega^2 = 4E_k$

Energi kinetik rotasi menjadi empat kali lipat.

A. energi kinetik rotasi menjadi satu per empatnya karena memiliki hubungan akar kuadrat dengan panjang baling-baling
 B. energi kinetik rotasi menjadi empat kali lipat karena memiliki hubungan kuadrat dengan panjang baling-baling
 C. energi kinetik rotasi menjadi setengahnya karena berbanding terbalik dengan panjang baling-baling
 D. energi kinetik rotasi menjadi dua kali lipat karena berbanding lurus dengan panjang baling-baling
 E. energi kinetik rotasi tetap karena panjang baling-baling tidak memengaruhi

Query Evidence (Memeriksa bukti)
 Siswa mengidentifikasi informasi penting yang menjadi bukti/data: Panjang baling-baling (L)
 Massa baling-baling (M)
 Kecepatan sudut (ω)

Draw Logically Valid or Justified Conclusion (Menarik kesimpulan logis dan beraturan)
 Siswa menarik kesimpulan jika panjang baling-baling diperbesar terhadap energi kinetik rotasi

Conjecture Alternatives (Mempertimbangkan berbagai alternatif)
 Siswa mengevaluasi beberapa pilihan jawaban, menghubungkan dengan konsep fisika kemudian memilih jawaban yang dianggap benar

Gambar 1. Contoh Butir Soal Instrumen Keterampilan Berpikir Kritis aspek *Inference*

Analisis data pada penelitian ini menggunakan pendekatan *Rasch* model untuk mengukur kualitas instrumen keterampilan berpikir kritis yang dikembangkan. Validitas konstruk diukur menggunakan uji unidimensionalitas jika nilai raw variance explained by measures $> 20\%$ dan nilai unexplained variance in contrast 1-5 of PCA residuals $< 15\%$. Jika nilai memenuhi kriteria tersebut, maka instrumen mengukur satu konstruk yaitu berpikir kritis (Sumintono & Widhiarso, 2015).

Tabel 2. Kriteria *Infit* dan *Outfit Mean Square (MNSQ)*

Nilai Mean Square (MNSQ)	Interpretasi
$> 2,0$	Menurunkan kualitas pengukuran
$1,5 < MNSQ \leq 2,0$	Tidak produktif untuk pengukuran, tapi tidak menurunkan kualitas
$0,5 \leq MNSQ \leq 1,5$	Produktif untuk pengukuran
$< 0,5$	Kurang produktif untuk pengukuran, tapi tidak menurunkan keakuratan

Kecocokan item (*item fit*) dievaluasi menggunakan dua statistik utama, yaitu *infit* dan *outfit*. *Infit* (*inlier-sensitive fit*) merupakan statistik yang sensitif terhadap pola respons pada butir soal yang tingkat kesulitannya sesuai atau mendekati kemampuan siswa (John Michael Linacre, 2002). Sedangkan *Outfit* (*outlier-sensitive fit*) merupakan statistik yang sensitif terhadap pola respons pada item yang tingkat kesulitannya jauh dari kemampuan siswa. Kecocokan item melalui *infit* dan *outfit* dinyatakan dalam *mean square* (MNSQ) dan *z-*

standardized (ZSTD). Kriteria kecocokan item dianalisis pada Tabel 2 dan Tabel 3 berdasarkan (John Michael Linacre, 2002).

Tabel 3. Kriteria *Infit* dan *Outfit Z-Standardizer (ZSTD)*

Nilai Standardizer (ZSTD)	Z-	Interpretasi
$\geq 3,0$		Data tidak dapat diprediksi, tapi pada sampel besar ($N > 500$) ketidaksesuaian tidak terlalu besar
$2,0 < ZSTD \leq 2,9$		Data sulit diprediksi
$-1,9 \leq ZSTD \leq 1,9$		Data dapat di prediksi
$\leq -2,0$		Data terlalu mudah diprediksi, indikasi terdapat dimensi lain

Reliabilitas instrumen dianalisis berdasarkan output *summary statistics* yang meliputi *item reliability*, *person reliability*, dan *cronbach's alpha* berdasarkan kriteria Sumintono & Widhiarso (2015). Selanjutnya dilakukan analisis tingkat kesulitan butir soal berdasarkan output *person measures*, Wright maps, profil keterampilan berpikir kritis siswa dan *item characteristic curve*.

Hasil dan Diskusi

Hasil analisis validitas isi menggunakan formula Aiken's V menunjukkan bahwa instrumen keterampilan berpikir kritis yang dikembangkan telah memenuhi kriteria validitas isi yang ditetapkan. Nilai Aiken's V pada ketiga aspek penilaian, yaitu aspek materi (0,93), aspek konstruksi (0,92), dan aspek interpretasi (0,93), seluruhnya melebihi nilai kritis 0,80 pada taraf signifikansi 5%. Temuan ini sejalan dengan penelitian Bhakti et al., (2024) yang menegaskan bahwa validitas isi merupakan syarat fundamental dalam pengembangan instrumen, di mana kesepakatan para ahli yang tinggi terhadap kesesuaian butir soal dengan konstruk yang diukur menjadi indikator utama kualitas instrumen.

Tabel 4. Interpretasi uji unidimensionalitas

Unidimensionalitas	Nilai	Interpretasi
Raw variance explained by measures	28,9%	Terpenuhi
Unexplained variance in contrast 1-5 of PCA residuals	5,8 - 9,9%	Baik

Validitas konstruk instrumen diukur melalui uji unidimensionalitas dengan analisis *Principal Component Analysis* (PCA) terhadap residual *Rasch* dalam program Winsteps 4.5.0. Kriteria *unidimensionality* dipenuhi jika: (1) variansi ter jelaskan oleh *measures* $> 20\%$, dan (2) variansi *unexplained* dalam *contract residual* $< 15\%$ (Sumintono & Widhiarso, 2015). Hasil analisis pada

Tabel 4 menunjukkan bahwa nilai *raw variance explained by measures* sebesar 28,9%, nilai ini telah memenuhi kriteria minimum yang ditetapkan yaitu $> 20\%$. Selain itu, nilai *unexplained variance in contrast 1-5 of PCA residuals* berada pada rentang 5,8 - 9,9%, yang berada di bawah batas maksimum yang diperbolehkan yaitu $< 15\%$. Dengan terpenuhinya kedua kriteria tersebut, dapat disimpulkan bahwa instrumen keterampilan berpikir kritis yang dikembangkan bersifat unidimensional (Noperi et al., 2025), artinya seluruh butir soal dalam instrumen mengukur satu konstruk

yang sama yaitu keterampilan berpikir kritis (Sumintono & Widhiarso, 2015; Susilowati, Al-akbari, Arifiyanti, Saputra, & Fratiwi, 2025).

Kecocokan butir soal (*item fit*) instrumen keterampilan berpikir kritis dianalisis menggunakan statistik *infit* dan *outfit* yang dinyatakan dalam bentuk *mean-square* (MNSQ) dan *z-standardized* (ZSTD). Kriteria kecocokan butir soal yang digunakan mengacu pada Linacre (2002), yaitu nilai MNSQ berada pada rentang 0,5-1,5 dan nilai ZSTD berada pada rentang -2 hingga +2. Hasil uji kecocokan butir soal dapat dilihat pada Tabel 5.

Tabel 5. Hasil Uji Kecocokan Butir Soal

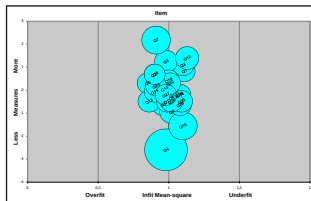
Soal	Butir	Infit MNSQ	ZSTD	Outfit MNSQ	ZSTD	Kriteria
1		1,11	0,91	1,13	0,92	Fit
2		1,10	0,69	1,11	0,66	Fit
3		0,98	-0,04	0,98	-0,02	Fit
4		0,98	0,17	0,88	0,09	Fit
5		0,96	-0,19	0,90	-0,45	Fit
6		1,02	0,20	1,02	0,15	Fit
7		0,91	-0,17	0,72	-0,61	Fit
8		0,85	-1,68	0,83	-1,75	Fit
9		1,02	0,15	1,02	0,16	Fit
10		1,10	0,39	1,71	1,56	Tidak Fit
11		1,03	0,33	1,04	0,30	Fit
12		1,13	0,74	1,21	0,92	Fit
13		0,86	-0,95	0,86	-0,77	Fit
14		0,97	-0,29	1,00	0,05	Fit
15		0,90	-1,01	0,86	-1,13	Fit
16		1,07	0,66	1,04	0,31	Fit
17		0,99	-0,01	0,93	-0,31	Fit
18		1,08	0,70	1,09	0,66	Fit
19		1,00	0,04	0,99	-0,11	Fit
20		1,08	0,53	1,15	0,75	Fit
21		0,90	-0,95	0,88	-1,02	Fit
22		1,01	0,16	1,03	0,34	Fit
23		0,91	-1,02	0,90	-1,01	Fit
24		0,90	-0,88	0,89	-0,87	Fit
25		1,09	0,65	1,20	1,10	Fit
26		1,02	0,19	0,98	-0,05	Fit
27		0,98	-0,14	0,95	-0,30	Fit

Hasil analisis menunjukkan bahwa dari 27 butir soal yang dianalisis, sebanyak 26 butir soal dinyatakan fit terhadap model *Rasch* karena memiliki nilai *infit* MNSQ dan *outfit* MNSQ yang berada dalam rentang 0,5-1,5 serta nilai ZSTD yang berada dalam rentang -2 hingga +2. Satu butir soal, yaitu butir soal nomor 10, dinyatakan tidak fit karena nilai *outfit* MNSQ sebesar 1,71 melebihi batas atas yang diperbolehkan (John Michael Linacre, 2022). Nilai *outfit* MNSQ yang melebihi

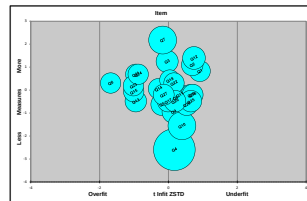
1,5 mengindikasikan adanya *underfit*, yang berarti terdapat respons yang tidak terduga pada item tersebut, kemungkinan disebabkan oleh *lucky guess* atau kecerobohan (*careless mistake*) peserta dalam menjawab butir soal yang tingkat kesulitannya jauh dari kemampuan mereka (John Michael Linacre, 2002).

Hasil analisis secara visual disajikan dalam bentuk *bubble chart* pada Gambar 2 hingga Gambar 5. Sumbu X pada masing-masing grafik menunjukkan nilai

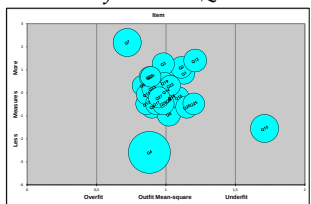
statistik kecocokan butir soal, sumbu Y menunjukkan tingkat kesulitan item dalam skala *logit*, dan ukuran gelembung mencerminkan besarnya *standard error* estimasi tingkat kesulitan item (John M Linacre, 2026). Gelembung yang lebih besar mengindikasikan estimasi tingkat kesulitan yang kurang presisi, dapat terjadi pada item dengan tingkat kesulitan yang sangat sulit atau sangat mudah.



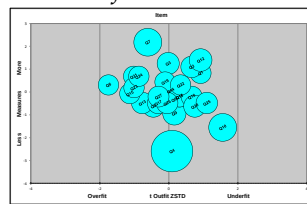
Gambar 2. Visualisasi Infit MNSQ



Gambar 3. Visualisasi Infit ZSTD



Gambar 4. Visualisasi Outfit MNSQ



Gambar 5. Visualisasi Outfit ZSTD

Gambar 2 terkait *infit* MNSQ menunjukkan bahwa seluruh 27 butir soal berada dalam zona fit (0,5-1,5). Butir soal Q7 dengan *measure* tertinggi menunjukkan butir soal paling sulit dengan MNSQ

optimal, sedangkan Q4 menunjukkan butir soal paling mudah dengan gelembung terbesar mengindikasikan *standart error* lebih besar. Gambar 3 *infit* ZSTD menampilkan semua butir soal dalam rentang -2 hingga +2, mengkonfirmasi fit yang konsisten pada semua tingkat kesulitan. Gambar 4 *outfit* MNSQ menunjukkan pola yang serupa dengan Gambar 2, namun butir soal Q10 terlihat jelas melampaui batas atas (1,71>1,5) mengindikasikan butir soal tersebut *underfit*. Gambar 5 *outfit* ZSTD mengkonfirmasi bahwa butir soal Q10 tidak memenuhi kriteria fit, sementara butir soal lainnya konsisten dalam zona fit. Meskipun demikian, butir soal Q10 yang dinyatakan tidak fit tersebut tidak serta-merta dieliminasi dari instrumen. Mengacu pada pertimbangan konten dan keterwakilan aspek berpikir kritis yang diukur, butir soal tersebut diputuskan untuk direvisi guna memperbaiki kualitas redaksi maupun kualitas pengecohnya.

Tingkat kesulitan butir soal instrumen keterampilan berpikir kritis dianalisis berdasarkan nilai *item measure* dalam skala *logit* yang dihasilkan oleh model *Rasch*. Klasifikasi tingkat kesulitan item menggunakan kriteria standar deviasi (SD), di mana item dinyatakan sulit apabila nilai *measure* >+1SD, item dinyatakan mudah apabila nilai *measure* <-1SD, dan item dinyatakan sedang apabila nilai *measure* berada di antara -1SD hingga +1SD. Berdasarkan hasil analisis, diperoleh nilai standar deviasi sebesar 0,93, sehingga batas kategori sulit adalah *measure* > +0,93 dan batas kategori mudah adalah *measure* < -0,93.

Tabel 7. Distribusi Tingkat Kesulitan Butir Soal

Butir Soal	Item Measures	Standar Error	Interpretasi
1	0,83	0,37	Sedang
2	1,1	0,38	Sulit
3	1,25	0,39	Sulit
4	-2,59	0,73	Mudah
5	-0,62	0,39	Sedang
6	-0,34	0,37	Sedang
7	2,18	0,49	Sulit
8	0,31	0,36	Sedang
9	-0,94	0,41	Mudah
10	-1,55	0,49	Mudah
11	-0,2	0,37	Sedang
12	1,4	0,40	Sulit
13	-0,48	0,38	Sedang
14	0,06	0,36	Sedang
15	-0,07	0,36	Sedang
16	-0,2	0,37	Sedang
17	-0,48	0,38	Sedang
18	-0,2	0,37	Sedang
19	0,44	0,36	Sedang
20	-0,62	0,39	Sedang
21	0,69	0,36	Sedang
22	0,31	0,36	Sedang

Butir Soal	Item Measures	Standar Error	Interpretasi
23	0,19	0,36	Sedang
24	0,69	0,36	Sedang
25	-0,48	0,38	Sedang
26	-0,48	0,38	Sedang
27	-0,2	0,37	Sedang

Hasil analisis menunjukkan bahwa dari 27 butir soal yang dianalisis pada Tabel 6, sebanyak 4 butir soal termasuk dalam kategori sulit, yaitu butir soal nomor 2, 3, 7, dan 12. Butir soal nomor 7 merupakan butir soal dengan tingkat kesulitan tertinggi (*measure* = 2,18 logit), yang berarti hanya siswa dengan kemampuan sangat tinggi yang dapat menjawab butir soal tersebut dengan benar. Sebanyak 3 butir soal termasuk dalam kategori mudah, yaitu butir soal nomor 4, 9, dan 10. Butir soal nomor 4 merupakan butir soal termudah dalam instrumen (*measure* = -2,59 logit), yang berarti hampir seluruh siswa dapat menjawab butir soal tersebut dengan benar. Sementara itu, sebanyak 20 butir soal termasuk dalam kategori sedang dengan nilai *measure* berada pada rentang -0,62 hingga +0,83 logit. Secara keseluruhan, distribusi tingkat kesulitan butir soal instrumen keterampilan berpikir kritis didominasi oleh butir soal berkategori sedang (74,1%), diikuti kategori sulit (14,8%), dan butir soal berkategori mudah (11,1%). Instrumen yang memiliki rentang tingkat kesulitan yang bervariasi mampu mengukur kemampuan siswa pada berbagai level secara lebih komprehensif, mulai dari peserta dengan kemampuan rendah hingga peserta dengan kemampuan sangat tinggi (Agustini, Arafah, Sanusi, Faisal, & Fitria, 2026).

Hasil analisis menunjukkan bahwa nilai *item reliability* sebesar 0,81 yang termasuk dalam kategori bagus. Nilai ini mengindikasikan bahwa hierarki tingkat kesulitan butir soal dalam instrumen konsisten apabila instrumen diberikan kepada kelompok siswa yang berbeda dengan karakteristik serupa (Aviyanti et al., 2025). Nilai *person reliability* sebesar 0,68 termasuk dalam kategori cukup, yang mengindikasikan bahwa kemampuan peserta dapat diestimasi dengan tingkat konsistensi yang memadai, meskipun masih terdapat ruang untuk peningkatan, misalnya dengan menambah jumlah butir soal atau memperbesar ukuran sampel penelitian yang digunakan. Adapun nilai *Cronbach Alpha* sebesar 0,72 termasuk dalam kategori bagus, yang menunjukkan bahwa secara keseluruhan interaksi antara *person* dan *item* dalam instrumen ini memiliki konsistensi internal yang baik (Sumintono & Widhiarso, 2015). Berdasarkan ketiga nilai reliabilitas tersebut, dapat disimpulkan bahwa instrumen keterampilan berpikir kritis yang dikembangkan memiliki tingkat reliabilitas yang memadai untuk digunakan sebagai alat ukur dalam penelitian.

Tabel 6. Hasil Uji Reliabilitas

Summary Statistics	Nilai	Kriteria
Item Reliability	0,81	Bagus
Person Reliability	0,68	Cukup
Cronbach Alpha	0,72	Bagus

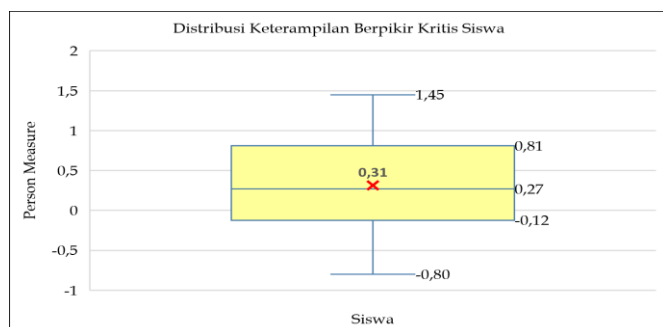
Hasil analisis reliabilitas instrumen keterampilan berpikir kritis pada Tabel 5 menunjukkan nilai yang memadai pada ketiga *summary statistics* yang digunakan. Nilai *item reliability* sebesar 0,81 yang termasuk kategori bagus mengindikasikan bahwa hierarki tingkat kesulitan butir soal dalam instrumen bersifat stabil apabila instrumen diberikan kepada kelompok siswa yang berbeda dengan karakteristik serupa (Aviyanti et al., 2025; John Michael Linacre, 2022). Sejalan dengan Bond & Fox (2015) yang menyatakan bahwa nilai *item reliability* yang tinggi mencerminkan kualitas butir soal yang konsisten dalam mendefinisikan konstruk yang diukur, sehingga urutan tingkat kesulitan butir soal tidak akan berubah secara signifikan meskipun diterapkan pada sampel yang berbeda (Salam, Aviyanti, & Sirnawati, 2026).

Nilai *person reliability* sebesar 0,68 yang termasuk kategori cukup mengindikasikan bahwa keterampilan berpikir kritis siswa dapat diestimasi dengan tingkat konsistensi yang memadai. Adapun nilai *Cronbach Alpha* sebesar 0,72 yang termasuk kategori bagus menunjukkan bahwa secara keseluruhan instrumen memiliki konsistensi internal yang baik (John Michael Linacre, 2022). Hasil reliabilitas yang baik ini mengkonfirmasi temuan Negoro et al., (2020) bahwa tes yang konsisten adalah tes yang secara akurat menangkap setiap aspek keterampilan berpikir kritis secara terfokus untuk meminimalkan bias.

Keterampilan berpikir kritis siswa dianalisis berdasarkan nilai *person measure* dalam skala logit yang dihasilkan oleh model *Rasch*. Nilai *person measure* menunjukkan estimasi kemampuan setiap siswa dalam skala logit, di mana nilai yang lebih tinggi mengindikasikan keterampilan berpikir kritis yang lebih baik. Gambar 6 menunjukkan grafik *box plot* distribusi keterampilan berpikir kritis siswa berdasarkan nilai *person measure*.

Berdasarkan Gambar 6 menunjukkan bahwa keterampilan berpikir kritis siswa tersebar pada rentang -0,80 hingga +1,45 logit dengan rata-rata *person measure* sebesar 0,31 logit dan standar deviasi sebesar 0,57. Nilai

rata-rata yang berada di atas nol (0,31 logit) mengindikasikan bahwa secara umum keterampilan berpikir kritis siswa berada sedikit di atas rata-rata tingkat kesulitan butir soal, artinya sebagian besar siswa mampu menjawab butir soal dengan tingkat kesulitan rata-rata dengan benar.



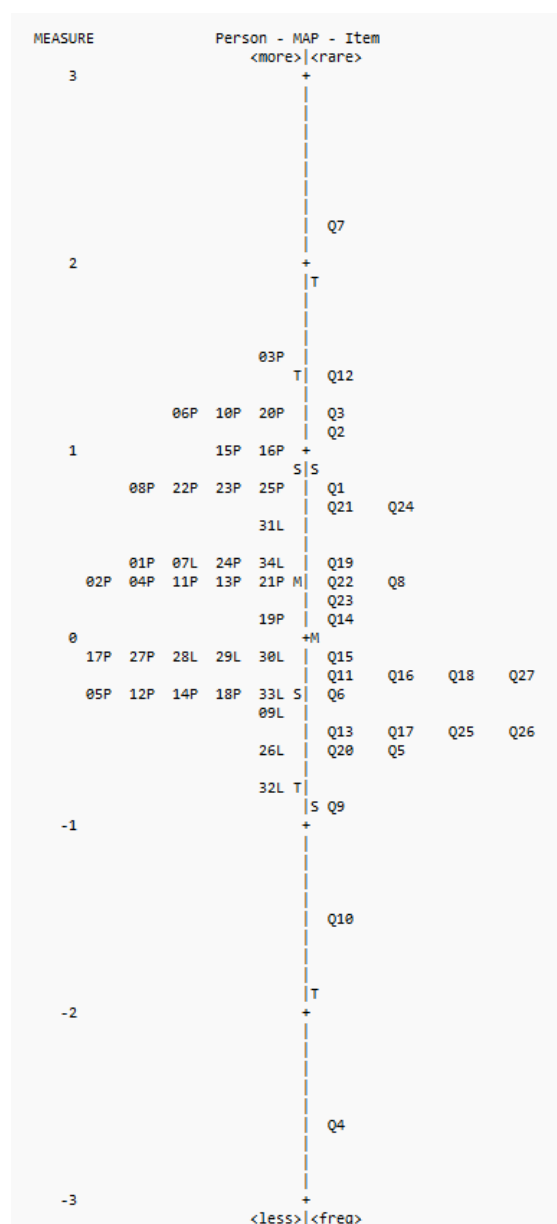
Gambar 6. Grafik Box Plot Keterampilan Berpikir Kritis Siswa

Siswa dengan keterampilan berpikir kritis tertinggi dengan nilai *measure* sebesar +1,45 logit, ini menunjukkan bahwa mampu menjawab dengan benar butir soal yang termasuk kategori sulit. Sebaliknya, siswa dengan keterampilan berpikir kritis terendah dengan nilai *measure* sebesar -0,80 logit, yang mengindikasikan bahwa siswa tersebut mengalami kesulitan dalam menjawab butir soal dengan tingkat kesulitan rata-rata. Distribusi keterampilan siswa yang relatif sempit (rentang -0,80 hingga +1,45 logit) dengan standar deviasi 0,57 menunjukkan bahwa keterampilan berpikir kritis antar siswa dalam kelompok ini cukup homogen (Mardhiya & Sari, 2025). Informasi mengenai profil kemampuan ini dapat dimanfaatkan sebagai dasar dalam merancang strategi pembelajaran yang lebih sesuai dengan karakteristik siswa, termasuk dalam menentukan bentuk dukungan atau diferensiasi pembelajaran yang diperlukan.

Distribusi tingkat kesulitan butir soal dan kemampuan siswa divisualisasikan melalui *Wright Map* (Person-maps-Item) sebagaimana disajikan pada Gambar 7. Peta ini menggambarkan sebaran kemampuan siswa (sisi kiri) dan butir soal (sisi kanan) pada skala *logit* yang sama, sehingga dapat diketahui kesesuaian antara tingkat kesulitan butir soal dengan kemampuan siswa (Sumintono & Widhiarso, 2015).

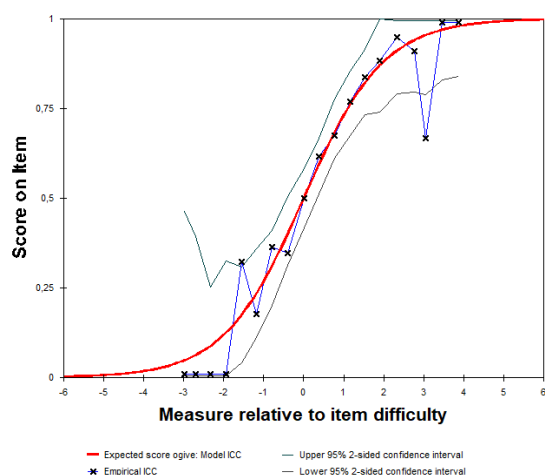
Berdasarkan Gambar 7, butir soal dalam instrumen keterampilan berpikir kritis tersebar pada rentang logit -2,59 hingga +2,18. Butir soal Q7 merupakan soal dengan tingkat kesulitan tertinggi yang berada pada logit +2,18, artinya butir soal ini hanya dapat dijawab dengan benar oleh siswa dengan kemampuan sangat tinggi. Butir soal Q12, Q3, dan Q2 berada pada kisaran logit +1,25 hingga +1,40, termasuk dalam kategori sulit. Sebagian besar item mengelompok

pada rentang logit -0,62 hingga +0,69, yang mencerminkan tingkat kesulitan sedang. Butir soal Q9 berada pada logit -0,94, butir soal Q10 pada logit -1,55, dan butir soal Q4 merupakan soal termudah dengan logit -2,59, artinya hampir seluruh siswa dapat menjawab butir soal ini dengan benar. Kemampuan siswa tersebar pada rentang logit yang relatif sempit, yaitu -0,62 hingga +1,25. Sebagian besar siswa mengelompok di sekitar rata-rata kemampuan (logit 0,00), yang ditandai dengan tanda M pada peta. Terdapat beberapa siswa dengan kemampuan di atas rata-rata seperti 03P, 06P, 10P, dan 20P yang berada pada kisaran logit +1,25 hingga +1,55, serta siswa 08P, 22P, 23P, dan 25P yang berada sedikit di atas rata-rata. Sementara itu, peserta dengan kemampuan di bawah rata-rata seperti siswa 32L berada pada logit -0,75.



Gambar 7. Peta Wright Maps

Karakteristik butir soal instrumen keterampilan berpikir kritis secara keseluruhan divisualisasikan melalui *Item Characteristic Curve* (ICC). ICC menggambarkan hubungan antara kemampuan peserta (*measure relative to item difficulty*) pada sumbu X dengan probabilitas menjawab benar (*score on item*) pada sumbu Y. Terdapat tiga komponen utama dalam grafik ICC, yaitu kurva merah yang merepresentasikan *expected score ogive* berdasarkan model *Rasch*, tanda silang (x) yang merepresentasikan nilai *empirical ICC* atau data respons aktual peserta, serta dua garis abu-abu yang merepresentasikan batas atas dan batas bawah *confidence interval* 95% (John M Linacre, 2026).



Gambar 8. Grafik *Item Characteristic Curve*

Secara umum, titik-titik *empirical ICC* (x) pada Gambar 8 mengikuti pola kurva model dengan cukup baik, terutama pada rentang kemampuan -2 hingga +2 logit di mana sebagian besar siswa berada. Hal ini mengindikasikan bahwa data empiris relatif konsisten dengan prediksi model *Rasch*. Namun demikian, terdapat beberapa titik empiris yang menyimpang dari kurva model, khususnya pada rentang kemampuan ekstrem yaitu di atas +3 logit. Penyimpangan pada rentang ekstrem ini merupakan hal yang wajar mengingat jumlah siswa pada rentang kemampuan tersebut relatif sedikit sehingga estimasi empirisnya kurang stabil. Sebagian besar titik empiris masih berada di dalam atau mendekati batas *confidence interval* 95%, yang mengonfirmasi bahwa kecocokan antara data empiris dan model *Rasch* secara keseluruhan dapat diterima (John M Linacre, 2026). Berdasarkan hasil analisis ICC tersebut, dapat disimpulkan bahwa instrumen keterampilan berpikir kritis yang dikembangkan memiliki karakteristik butir soal yang sesuai dengan model *Rasch*, di mana peningkatan kemampuan peserta diikuti oleh peningkatan probabilitas menjawab benar secara konsisten (Afandi et al., 2025).

Kesimpulan

Penelitian ini berhasil mengembangkan dan memvalidasi instrumen keterampilan berpikir kritis siswa SMA pada materi dinamika rotasi yang terdiri dari 27 butir soal pilihan ganda berbasis enam aspek berpikir kritis Facione (2023) menggunakan pendekatan *Rasch* model. Hasil validasi menunjukkan bahwa instrumen memiliki validitas isi yang sangat baik dengan nilai rata-rata Aiken's V sebesar 0,93, melampaui nilai batas kritis 0,80. Uji *unidimensionalitas* menunjukkan bahwa variansi yang dijelaskan oleh *measures* sebesar 28,9% melampaui batas minimum 20% yang mengindikasikan bahwa konstruk keterampilan berpikir kritis dijelaskan dengan baik dan *unexplained variance in contrast 1-5 of PCA residuals* berkisar 5,8-9,9%. Sebanyak 26 butir soal (96,3%) dinyatakan fit terhadap model *Rasch*, sedangkan satu butir soal tidak fit (3,7%) akibat nilai *outfit* MNSQ yang melampaui batas yang diperbolehkan. Reliabilitas instrumen secara keseluruhan tergolong baik, dengan nilai *item reliability* sebesar 0,81 (baik), *person reliability* sebesar 0,68 (cukup), dan *cronbach alpha* sebesar 0,72 (baik), menunjukkan bahwa instrumen memiliki konsistensi internal yang kuat dan presisi pengukuran individu yang dapat ditingkatkan. Distribusi tingkat kesulitan butir soal didominasi oleh kategori sedang (74,1%), diikuti sulit (14,8%) dan mudah (11,1%), yang mencerminkan kemampuan instrumen dalam mengukur keterampilan berpikir kritis siswa pada berbagai level kemampuan secara komprehensif. Temuan-temuan tersebut dikonfirmasi lebih lanjut melalui analisis *Wright Map* dan *Item Characteristic Curve* (ICC). Hasilnya menunjukkan kesesuaian yang baik antara sebaran kemampuan siswa dan tingkat kesulitan butir soal pada skala logit yang sama serta peningkatan kemampuan siswa diikuti oleh peningkatan probabilitas menjawab benar secara konsisten. Bagi penelitian selanjutnya, disarankan agar dilakukan analisis *Differential Item Functioning* (DIF) untuk memastikan instrumen tidak bias pada kelompok siswa (gender, sekolah asal, atau asal daerah), sampel yang digunakan bisa lebih besar untuk meningkatkan presisi dan stabilitas estimasi parameter butir soal, dan diintegrasikan lebih lanjut ke dalam platform penilaian adaptif seperti *computerized adaptive testing*.

Ucapan Terimakasih

Penulis mengucapkan terima kasih kepada seluruh pihak yang terlibat dan berpartisipasi dalam penelitian ini. Kontribusi, dukungan, dan kerja sama yang diberikan sangat membantu dalam proses pengembangan dan validasi instrumen yang dilakukan.

Referensi

Afandi, A., Aviyanti, L., Saepuzaman, D., & Setiawan, A. (2025). *Assesing Critical Thinking Skills in*

- Dynamic Fluids at Senior High School: *Rasch* Model Analysis. *Jurnal Ilmiah Pendidikan Fisika Al-Biruni*, 14(2), 87-100. <https://doi.org/10.24042/jipfalbiruni.v14i2.27836>
- Agustini, S., Arafah, K., Sanusi, D. K., Faisal, R., & Fitria, N. (2026). Analysis of the difficulty level of physics cognitive test items using the *Rasch* approach: Implementation of analysis with R Software. *Priviet Social Sciences Journal*, 6(4), 433-444. <https://doi.org/10.55942/pssj.v6i4.1036>
- Aiken, L. R. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, 45(1), 131-142. <https://doi.org/10.1177/0013164485451012>
- Aviyanti, L., Fratiwi, N. J., Nurdini, N., Salam, A., & Nawas, A. (2025). Can Multiple-Choice Items Measure Critical Thinking in Socio-Scientific Environmental Issues? Evidence from a Global Warming Assessment of Grade 10 Students Using *Rasch* Analysis. *Jurnal Ilmiah Pendidikan Fisika Al-Biruni*, 14(2), 33-53. <https://doi.org/10.24042/jipfalbiruni.v14i2.28194>
- Aviyanti, L., Gani, A. W., Febriyanti, T., Ribie, S., & Nawas, A. (2024). Promoting higher-order thinking skills and learning motivation through the teams games tournaments learning model in physics education: A *Rasch* model analysis Meningkatkan higher-order thinking skills dan motivasi belajar melalui model pembelajaran teams. *Indonesian Journal of Science and Mathematics Education (IJSME)*, 07(November), 560-576. <https://doi.org/10.24042/ijsme.v5i1.23648>
- Azizah, F., Handayanto, S. K., Wisodo, H., & Fernando, T. J. (2024). Development and Validation of Critical Thinking Skills Instruments on Mechanical Waves for Senior High School Student. *Kasuari: Physics Education Journal (KPEJ)*, 7(1), 195-205. <https://doi.org/10.37891/kpej.v7i1.624>
- Bhakti, Y. B., Arthur, R., & Supriyati, Y. (2024). The Content Validity: Multiple Choices Complex Physics Instrument to Measure Critical Thinking Skills. *JIPF (Jurnal Ilmu Pendidikan Fisika)*, 9(1), 56-64. <https://doi.org/10.26737/jipf.v9i1.4765>
- Bond, T. G., & Fox, C. M. (2015). Applying the *Rasch* Model, Fundamental Measurement in the Human Sciences (3rd editio). New York: Routledge.
- Branch, R. M. (2009). Instructional Design: The ADDIE Approach. New York: Springer Science business.
- Cascella, C., Giberti, C., & Bolondi, G. (2020). An analysis of Differential Item Functioning on INVALSI tests, designed to explore gender gap in mathematical tasks. *Studies in Educational Evaluation*, 64. <https://doi.org/10.1016/j.stueduc.2019.100819>
- Costan, M., Costan, K., Weißbach, A., & Kulgemeyer, C. (2026). Exploring physics teachers' views on physics education research: a case of science scepticism? *International Journal of Science Education*, 1-25. <https://doi.org/10.1080/09500693.2026.2629345>
- Facione, P. A. (1990). Critical Thinking A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction. Research Findings and Recommendations. Retrieved from <https://eric.ed.gov/?q=Facione&ft=on&pg=2&id=ED315423>
- Facione, P. A. (2023). Critical Thinking : What It Is and Why It Counts 2023 Update. In *Insight assessment* (pp. 1-28).
- Halim, M. G., Saepuzaman, D., Aviyanti, L., Utama, J. A., & Nawas, A. (2026). Measuring Critical Thinking in Physics: A *Rasch* Analysis of Instrument Quality and Gender Equivalence. *Jurnal Pendidikan MIPA*, 27(1), 452-475. <https://doi.org/10.23960/jpmipa.v27i1.pp452-475>
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. California: Sage Publication.
- Kaur, R., Mantri, A., Nagabhushan, P., & Singh, G. (2024). *Rasch* Computing Analysis of Two Tier Concept Inventory to Assess Engineering Students' Conceptual Knowledge. *SN Computer Science*, 5(5). <https://doi.org/10.1007/s42979-024-02955-6>
- Linacre, John M. (2026). A User's Guide to Winsteps: *Rasch*-Model Computer Program. Chicago: MESA Press.
- Linacre, John Michael. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions*, 16(2), 878. <https://doi.org/https://www.Rasch.org/rmt/rmt162f.htm>
- Linacre, John Michael. (2022). R Statistics : survey and review of packages for the estimation of *Rasch* models. *International Journal of Medical Education*, 13, 171-175. <https://doi.org/10.5116/ijme.629d.d88f>
- Mardhiya, J., & Sari, D. P. (2025). Measuring and Profiling Studnets' Critical Thinking Skills on Chemical Equilibrium using The *Rasch* Model. *UNESA Journal of Chemical Education*, 14(3), 191-201. <https://doi.org/10.26740/ujced.v14n3.p191-201>
- Mesic, V., Neumann, K., Aviani, I., Hasovi, E., Boone, W. J., Grubelnik, V., ... Repnik, R. (2019). Measuring students' conceptual understanding of wave

- optics: A *Rasch* modeling approach. *Physical Review Physics Education Research*, 010115, 1–20. <https://doi.org/10.1103/PhysRevPhysEducRes.1.5.010115>
- Negoro, R. A., Rusilowati, A., Aji, M. P., & Jaafar, R. (2020). Critical thinking in physics: Momentum critical thinking test for pre-service teacher. *Jurnal Ilmiah Pendidikan Fisika Al-Biruni*, 9(1), 73. <https://doi.org/10.24042/jipfalbiruni.v9i1.4834>
- Noperi, H., Diani, R., Irwandani, I., Suryani, Y., & Sodikin, S. (2025). Psychometric analysis of the critical thinking skills instrument in physics learning using the *Rasch* model. *Indonesian Journal of Science and Mathematics Education (IJSME)*, 8(3), 627–641. <https://doi.org/10.24042/ijsme.v8i3.28770>
- Nurdini, N., Suhandi, A., Ramalis, T., Samsudin, A., Fratiwi, N. J., & Costu, B. (2020). Developing Multitier Instrument of Fluids Concepts (MIFO) to Measure Student's Conception: A *Rasch* Analysis Approach. *Journal of Advanced Research in Dynamical and Control Systems*, 12(6), 3069–3083. <https://doi.org/10.5373/JARDCS/V12I6/S20201273>
- Nurlailah, S., & Hamdu, G. (2021). Implementasi Assessment Sikap Berpikir Kritis Berbasis Education for Sustainable Development (ESD) di Sekolah Dasar. *Jurnal IDEAS Pendidikan, Sosial, Dan Budaya*, 7(3), 309–316. <https://doi.org/10.32884/ideas.v7i3.390>
- Nuryanti, S., Masykuri, M., & Susilowati, E. (2018). Analisis Iteman dan model *Rasch* pada pengembangan instrumen kemampuan berpikir kritis peserta didik sekolah menengah kejuruan. *Jurnal Inovasi Pendidikan IPA*, 4(2), 224–233. <https://doi.org/10.21831/jipi.v4i2.21442>
- OECD. (2023). *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*. Paris: OECD Publishing. <https://doi.org/https://doi.org/10.1787/53f23881-en>
- Retnawati, H. (2014). *Teori Respons Butir dan Penerapannya*. Parama Nuha.
- Retnawati, H., Djidu, H., Apino, E., & Anazifa, R. D. (2018). Teachers' Knowledge About Higher-Order Thinking Skills and Its Learning Strategy. *Problems of Education in the 21st Century*, 76(2), 215–230. <https://doi.org/10.33225/pec/18.76.215>
- Salam, A., Aviyanti, L., & Sirnawati, R. (2026). Assessing students' critical thinking skills at senior high schools in Bandung City. *Gravity: Jurnal Ilmiah Penelitian Dan Pembelajaran Fisika*, 12(1), 37–52. <https://doi.org/10.62870/gravity.v12i1.38484>
- Sumintono, B. (2018). *Rasch Model Measurements as Tools in Assesment for Learning*. 173(Icei 2017), 38–42. <https://doi.org/10.2991/icei-17.2018.11>
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan Rasch: pada assesment pendidikan*. Cimahi: Tria Komunika.
- Susilawati, E., Samsudin, A., & Siahaan, P. (2020). Analisis Tingkat Keterampilan Berpikir Kritis Siswa SMA. *Jurnal Pendidikan Fisika Dan Teknologi (JPFT)*, 6(1), 11–16. <https://doi.org/http://dx.doi.org/10.29303/jpft.v6i1.1453>
- Susilowati, N. E., Al-akbari, S., Arifiyanti, F., Saputra, M. R. D., & Fratiwi, N. J. (2025). Measuring students' creative thinking skills with SCI-Fi (Science Creative Instrument in Fluids): How Suitable Is It? *Tarbiyah: Jurnal Ilmiah Kependidikan*, 14(2), 247–263. <https://doi.org/http://dx.doi.org/10.16085/tarbiyah.v12i1>
- Terwee, C. B., Bot, S. D. M., Boer, M. R. De, Windt, A. W. M. Van Der, Knol, D. L., & Dekker, J. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. 60, 34–42. <https://doi.org/10.1016/j.jclinepi.2006.03.012>
- Wafa, Z., & Jatmiko, B. (2022). Learning Physics with a Free Discovery Model to Improve Critical Thinking Skills of High School Students. *PrismaSains: Jurnal Pengkajian Ilmu Dan Pembelajaran Matematika Dan IPA IKIP Mataram*, 10(3), 637–646. <https://doi.org/https://doi.org/10.33394/j-ps.v10i3.5375>
- Yokhebed, Y., Karmadi, R. M. D., & Nastiti, L. R. (2025). Validity and reliability analysis of a socioscientific issues-based critical thinking self-assessment instrument using the *Rasch* model. *JPBI (Jurnal Pendidikan Biologi Indonesia)*, 11(1), 73–82. <https://doi.org/10.22219/jpbi.v11i1.38902>