

KEVALIDAN INSTRUMEN *AUTHENTIC ASSESSMENT* UNTUK MENGUKUR KEMAMPUAN BERPIKIR KRITIS PESERTA DIDIK PADA *PROBLEM BASED LEARNING (PBL)* MATERI HUKUM NEWTON TENTANG GERAK

Erna Wahyu Septianna^{1*}, Viyanti², Wayan Suana³

^{1,2,3} Universitas Lampung, Indonesia

Correspondence e-mail: wahyuernaseptianna@email.com^{1*}

Article History	ABSTRAK
<p>Accepted: February 24th 2025 Approved: July 28th 2025 Published: July 30th 2025</p>	<p>Penelitian ini mengembangkan instrumen <i>authentic assessment</i> berbasis <i>Problem Based Learning (PBL)</i> untuk mengukur kemampuan berpikir kritis peserta didik pada materi Hukum Newton tentang gerak. Metode <i>Research and Development (R&D)</i> digunakan dengan mengadaptasi model 4D Thiagarajan (1974) yang meliputi tahap: (1) <i>Define</i> (analisis kebutuhan di SMAN 1 Sukoharjo menunjukkan 62% guru belum memiliki instrumen penilaian berpikir kritis yang memadai), (2) <i>Design</i> (penyusunan kisi-kisi berdasarkan indikator Ennis, 1996), (3) <i>Develop</i> (validasi ahli dan uji empiris), serta (4) <i>Disseminate</i> (implementasi terbatas). Validasi oleh tiga ahli (2 dosen dan 1 guru) menghasilkan skor rata-rata 82,48% (<i>sangat valid</i>) dengan rincian: aspek konstruk (83,33%), substansi (78,02%), dan bahasa (86,10%). Uji coba pada 35 siswa menggunakan analisis model Rasch (<i>Ministeps 5.4.1</i>) mengonfirmasi 12 butir soal memenuhi kriteria validitas. Instrumen ini memiliki reliabilitas tinggi (<i>Cronbach's Alpha 0,81</i>) dan kepraktisan luar biasa (skor 96,16% berdasarkan penilaian guru). Keunikan penelitian terletak pada integrasi pendekatan PBL dengan rubrik penilaian autentik yang terstruktur sesuai tahapan pemecahan masalah, serta penggunaan analisis Rasch untuk validasi empiris. Temuan ini memberikan kontribusi praktis bagi guru fisika dalam menerapkan penilaian berbasis keterampilan abad 21 sekaligus mengisi celah literatur tentang instrumen spesifik untuk materi Hukum Newton.</p> <p>Kata Kunci: <i>Authentic Assessment</i>; kemampuan berpikir kritis; <i>problem based learning</i></p>
<p>DOI: doi.org/10.30822/magneton.v3i2.4513</p>	<h3 data-bbox="938 1514 1104 1541">ABSTRACT</h3> <p>This study developed an authentic assessment instrument based on Problem Based Learning (PBL) to measure students' critical thinking skills in Newton's laws of motion. The Research and Development (R&D) method was used by adapting Thiagarajan's 4D model (1974), which includes the following stages: (1) Define (analysis of needs at SMAN 1 Sukoharjo showed that 62% of teachers did not have adequate critical thinking assessment instruments), (2) Design (development of a rubric based on Ennis' indicators, 1996), (3) Develop (expert validation and empirical testing), and (4) Disseminate (limited implementation). Validation by three experts (2 lecturers and 1 teacher) resulted in an average score of 82.48% (highly valid), with details as follows: construct aspect (83.33%), substance (78.02%), and language (86.10%). A pilot test on 35 students using Rasch model analysis (Ministeps 5.4.1) confirmed that 12</p>



items met validity criteria. The instrument has high reliability (Cronbach's Alpha 0.81) and exceptional practicality (96.16% score based on teacher assessment). The uniqueness of this research lies in the integration of the PBL approach with an authentic assessment rubric structured according to problem-solving stages, as well as the use of Rasch analysis for empirical validation. These findings provide practical contributions for physics teachers in implementing 21st-century skills-based assessment while filling a gap in the literature on specific instruments for Newton's Law material.

Keywords: Authentic Assessment; critical thinking skills; problem-based learning

PENDAHULUAN

Era pendidikan abad ke-21 mengalami transformasi fundamental yang didorong oleh kemajuan teknologi digital sebagai katalisator perubahan. Integrasi teknologi dalam proses pembelajaran tidak hanya memodifikasi metode pengajaran tradisional, tetapi juga menciptakan paradigma baru dalam pengembangan kompetensi peserta didik. Dalam konteks ini, literasi digital dan keterampilan kompleks seperti analisis kritis, inovasi kreatif, serta kolaborasi lintas disiplin menjadi prasyarat utama untuk menghadapi tantangan global yang semakin dinamis (Redhana, 2024).

Association (2002) merumuskan kerangka kompetensi abad ke-21 dalam empat dimensi integral: (1) Penguasaan konten inti multidisiplin, (2) Kapasitas inkuiri dan inovasi, (3) Melek teknologi informasi, dan (4) Kecakapan hidup berbasis karier. Khusus pada dimensi kedua, framework 4C (*critical thinking, communication, collaboration, creativity*) menjadi fondasi pedagogis yang menggeser fokus pembelajaran dari hafalan menuju pengembangan kapasitas metakognitif. Implementasi kurikulum merdeka melalui pendekatan berbasis masalah (*problem-based approach*) menegaskan pentingnya kontekstualisasi pengetahuan, di mana peserta didik dilatih untuk mengkonstruksi pemahaman melalui eksplorasi masalah autentik.

Joyce & Weil (1972) menegaskan bahwa efektivitas pembelajaran sangat bergantung pada pemilihan model pedagogis yang tepat. *Problem Based Learning* (PBL) yang diadopsi dalam penelitian ini berfungsi sebagai katalis perkembangan kognitif melalui tiga mekanisme utama: (1) Stimulasi rasa ingin tahu intrinsik, (2) Pembelajaran mandiri berbasis inkuiri, dan (3) Penguatan kemampuan analisis melalui penyelesaian masalah kompleks (Astuti et al., 2018). Simbiosis antara PBL dan keterampilan berpikir kritis terlihat ketika peserta didik menerapkan indikator pemikiran kritis seperti analisis premis, evaluasi bukti, dan sintesis solusi dalam setiap tahapan PBL.

Dalam domain pendidikan sains khususnya fisika, PBL memainkan peran strategis. Walsh et al., (2007) mengemukakan bahwa hakikat pembelajaran fisika terletak pada kemampuan mentransfer konsep abstrak ke dalam solusi praktis masalah riil. Implementasi PBL dalam fisika Retug (2010); Christi et al. (2020); mengoptimalkan dua aspek kunci: (1) Desain tantangan pembelajaran yang sesuai dengan tingkat perkembangan kognitif, dan (2) Strukturisasi pengetahuan melalui scaffolding yang progresif. Depdiknas (2005) menambahkan bahwa keberhasilan pendekatan ini memerlukan desain pembelajaran yang partisipatif, melibatkan aktivitas higher-order thinking secara sistematis.

Indrapangastuti (2023) mengidentifikasi tiga keunggulan PBL dalam pengembangan berpikir kritis: (1) Eksplorasi kognitif tanpa batas, (2) Pembuktian ilmiah melalui investigasi, dan (3) Peningkatan signifikan dalam kemampuan pemecahan masalah. Efektivitas pendekatan ini semakin optimal ketika didukung oleh instrumen penilaian autentik yang komprehensif. Rosidin (2017) dan Wulandari et al. (2019) sepakat bahwa sistem penilaian modern harus mampu mengukur tiga aspek: (1) Proses kognitif, (2) Produk pemikiran, dan (3) Aplikasi pengetahuan dalam simulasi dunia nyata. Safitri & Harjono (2021) menekankan pentingnya alignment antara teknik penilaian dengan karakteristik kompetensi abad 21, di mana penilaian autentik berbasis kinerja menjadi pilihan ideal untuk mengukur dimensi kompleksitas berpikir kritis.

Berdasarkan analisis kebutuhan yang dilaksanakan di SMAN 1 Sukoharjo melalui angket analisis kebutuhan guru dengan aspek yang dianalisis yaitu aspek kurikulum dan capaian pembelajaran, instrumen penilaian yang digunakan, proses pembelajaran fisika dengan menggunakan model PBL dan penilaian yang digunakan dalam pembelajaran fisika, serta penilaian kemampuan berpikir kritis. Pada aspek kurikulum dan capaian pembelajaran mengungkapkan bahwa sebesar 62% guru fisika di SMA tersebut telah melakukan analisis kurikulum dan capaian pembelajaran sebelum pelaksanaan pembelajaran. Aspek instrumen penilaian yang digunakan mengungkapkan bahwa sebesar 62% guru fisika di SMAN 1 Sukoharjo telah melakukan penilaian dalam proses pembelajaran. Aspek proses pembelajaran fisika dengan menggunakan model PBL dan penilaian yang digunakan dalam pembelajaran fisika mengungkapkan sebesar 55%. Sedangkan pada aspek penilaian kemampuan berpikir kritis mengungkapkan sebesar 43%. Karena tidak adanya alat penilaian yang sederhana, berguna, dan sesuai selama proses pembelajaran, terbukti dari semua faktor tersebut bahwa guru fisika di sekolah menengah atas tersebut belum menggunakan penilaian yang obyektif, khususnya untuk mengukur kemampuan berpikir kritis siswa.

Wijaya et al. (2017) mengembangkan alat penilaian untuk berpikir kritis untuk menciptakan alat yang reliabel dan valid untuk mengevaluasi kemampuan berpikir kritis. Dengan menggunakan strategi pembelajaran inkuiri otentik, Sylvia et al. (2019) juga melakukan penelitian pengembangan berupa instrumen penilaian otentik. Sementara itu, tujuan dari pembuatan instrumen penilaian oleh Nurindanasari et al. (2020) adalah instrumen penilaian literasi, bukan instrumen penilaian berpikir kritis. Menurut penelitian, PBL dapat digunakan sebagai model pembelajaran alternatif bagi guru untuk membantu siswa mengembangkan kemampuan berpikir kritis dan lebih memahami apa yang mereka pelajari (Ismiyana et al., 2023).

Penelitian ini menawarkan kebaruan melalui pengembangan instrumen *authentic assessment* berbasis *Problem Based Learning* (PBL) yang secara khusus dirancang untuk mengukur kemampuan berpikir kritis siswa pada materi Hukum Newton tentang gerak. Berbeda dengan penelitian sebelumnya yang umumnya berfokus pada instrumen penilaian konvensional, penelitian ini mengintegrasikan pendekatan PBL dengan teknik penilaian autentik, menghasilkan perangkat evaluasi yang tidak hanya valid dan reliabel (*Cronbach's Alpha* 0.81), tetapi juga sangat praktis (skor 96.16%) bagi guru. Instrumen ini juga dilengkapi dengan rubrik penilaian dan pedoman penskoran yang terstruktur sesuai tahapan PBL, menjadikannya solusi inovatif untuk pembelajaran fisika dalam konteks Kurikulum Merdeka.

Kebaruan lain terletak pada metodologi penelitian yang menggabungkan model pengembangan 4D Thiagarajan dengan analisis statistik menggunakan model Rasch. Pendekatan ini memungkinkan validasi instrumen yang lebih komprehensif, mencakup aspek konstruksi, substansi, dan bahasa, dengan hasil validasi ahli mencapai 82.48%. Selain itu, penelitian ini mengisi celah literatur dengan menyediakan instrumen spesifik untuk materi Hukum Newton yang sebelumnya belum tersedia, sekaligus menawarkan bukti empiris kuat melalui uji coba pada 35 siswa. Dengan demikian, penelitian ini tidak hanya memperkaya khazanah alat evaluasi pendidikan, tetapi juga memberikan kontribusi praktis bagi guru dalam menerapkan penilaian autentik berbasis PBL. Sebagai langkah untuk mengatasi keterbatasan instrumen penilaian autentik yang telah dikembangkan dalam penelitian sebelumnya dalam mengukur kemampuan berpikir kritis berbasis PBL, serta berdasarkan hasil analisis kebutuhan di sekolah, peneliti akan melakukan penelitian pengembangan dengan judul Pengembangan Instrumen *Authentic Assessment* berbasis *Problem Based Learning* (PBL) untuk Mengukur Kemampuan Berpikir Kritis Peserta didik pada Materi Hukum Newton tentang Gerak.

Penelitian ini bertujuan mengembangkan instrumen *Authentic Assessment* untuk mengukur kemampuan berpikir kritis siswa dalam model PBL pada materi Hukum Newton tentang Gerak, dengan mengevaluasi kelayakan konstruk, substansi, bahasa, validitas empiris, reliabilitas, dan kepraktisannya. Manfaat penelitian antara lain memberikan pengalaman berharga bagi peneliti dalam pengembangan instrumen, membantu guru memperoleh penilaian yang lebih obyektif, serta mendorong siswa untuk lebih aktif selama pembelajaran karena merasa dinilai secara langsung.

Penelitian ini diharapkan dapat memberikan kontribusi praktis bagi guru fisika melalui instrumen *authentic assessment* berbasis PBL yang valid (skor validasi 82.48%), reliabel ($\alpha=0.81$), dan praktis (skor 96.16%) untuk mengukur kemampuan berpikir kritis siswa pada materi Hukum Newton. Selain memperkaya khazanah alat evaluasi pendidikan, temuan ini diharapkan dapat memudahkan guru dalam menerapkan penilaian autentik sekaligus mendorong implementasi PBL yang lebih efektif di kelas, sehingga berdampak pada peningkatan kualitas pembelajaran sains di Indonesia.

METODE

Penelitian ini menggunakan metode penelitian pengembangan (*Research and Development/R&D*) sebagai kerangka dasar penyelidikan. Pendekatan ini secara khusus dipilih untuk menghasilkan suatu produk akademik berupa instrumen evaluasi pembelajaran sekaligus menguji validitas dan efektivitas penerapannya dalam setting pendidikan nyata. Hasil dari penelitian ini adalah sebuah instrumen penilaian untuk mengukur kemampuan berpikir kritis siswa berbasis PBL. Penelitian ini mengadopsi model pengembangan 4D menurut Thiagarajan (1974), dengan tahapan: (1) pendefinisian masalah, (2) perancangan produk, (3) pengembangan prototipe, dan (4) diseminasi terbatas.

Tahap *Define* (Pendefinisian)

Penelitian teoritis dan empiris digunakan untuk melaksanakan langkah pendefinisian. Para peneliti melakukan investigasi teoritis dengan menelaah literatur yang berkaitan dengan penelitian pengembangan yang mereka kumpulkan dari buku-buku dan majalah nasional dan internasional lainnya.

Tahap *Design* (Perancangan)

Berikut ini adalah hasil yang diperoleh peneliti pada tahap ini setelah merumuskan tujuan penelitian dan langkah-langkah desain produk: 1) Dengan menggunakan paradigma pengembangan 4D dari (Thiagarajan, 1974), penelitian ini berusaha untuk menciptakan produk instrumen penilaian autentik yang valid, reliabel, dan bermanfaat untuk mengukur kemampuan berpikir kritis siswa dalam pembelajaran fisika berbasis PBL. 2) Kisi-kisi instrumen penilaian dikembangkan dengan menggunakan modul pembelajaran fisika kurikulum mandiri pada materi Hukum Newton tentang Gerak kelas XI dan indikator kemampuan berpikir kritis berdasarkan (Robert, 1996).

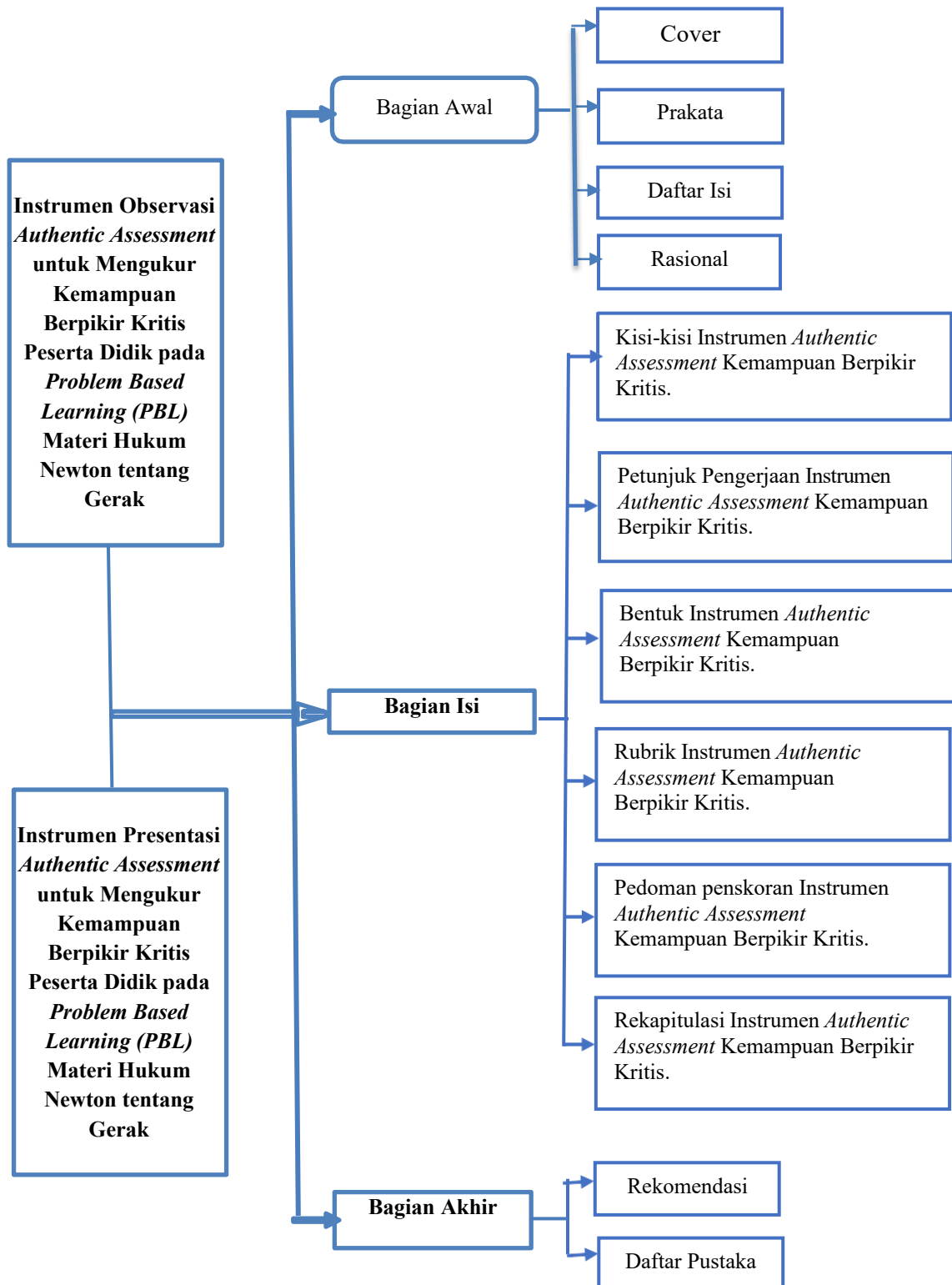
Tahap *Develop* (Pengembangan)

Pada tahap ini dilakukan pengembangan produk awal berupa instrumen penilaian kemampuan berpikir kritis pada pembelajaran fisika berbasis PBL, dengan tahapan pengembangan produk yaitu penyusunan spesifikasi instrumen penilaian, penulisan instrumen, menentukan rubrik penilaian, menentukan pedoman penskoran, uji validasi ahli, revisi hasil uji coba, uji coba lapangan, menganalisis validitas dan reliabilitas produk. Kemudian revisi produk

Tahap *Disseminate* (Penyebaran)

Tahapan selanjutnya, yaitu penyebarluasan terbatas instrumen penilaian yang sudah valid dan reliabel kepada guru fisika SMAN 1 Sukoharjo untuk mengetahui kepraktisan produk yang dikembangkan. Uji kepraktisan instrumen penilaian kemampuan berpikir kritis dilakukan pada tiga aspek, yaitu kemudahan penggunaan, kemenarikan sajian, dan kebermanfaatannya. Uji kepraktisan ini dilakukan kepada dua orang guru fisika sebagai ahli praktisi.

Kuesioner analisis kebutuhan yang diberikan secara langsung kepada narasumber, yaitu guru fisika SMAN 1 Sukoharjo, merupakan salah satu alat penelitian yang digunakan dalam penelitian pengembangan ini. Angket ini berfungsi sebagai panduan bagi peneliti untuk mengumpulkan data yang berkaitan dengan penelitian. Selain itu, lembar uji validasi ahli digunakan untuk memastikan tingkat validitas produk. Lembar uji ini dapat memberikan informasi mengenai layak atau tidaknya produk yang dikembangkan, serta dapat dijadikan panduan oleh guru untuk mengukur kemampuan berpikir kreatif siswa selama proses pembelajaran. Setelah itu, angket kepraktisan digunakan untuk mengetahui bagaimana pendapat dosen fisika tentang kegunaan, daya tarik penyajian, dan kemudahan penggunaan instrumen *authentic assessment* yang mengukur kemampuan berpikir kritis peserta didik.



Gambar 1. Rancangan Instrumen *Authentic assessment* Berbasis *Problem Based Learning* Kemampuan Berpikir Kritis Peserta Didik

Validasi ahli melibatkan dua dosen dan satu guru yang mengevaluasi aspek bahasa, konstruksi, serta substansi. Data kuantitatif dalam proses ini dikumpulkan menggunakan skala Likert dengan rentang nilai 1 hingga 4. Skor yang diperoleh selanjutnya dianalisis melalui proses perhitungan sebagai berikut:

$$P = \frac{\text{Jumlah skor yang diperoleh}}{\text{Jumlah skor tertinggi}} \times 100\% \quad (1)$$

Selanjutnya, nilai rata-rata validasi instrumen dikelompokkan berdasarkan persyaratan kelayakan Tabel 1 (Sugiyono, 2015).

Tabel 1. Kriteria hasil persentase kelayakan validasi instrumen

No	Presentasi	Kriteria
1	80,1% - 100%	Sangat Valid
2	60,1% - 80%	Valid
3	40,1% - 60%	Cukup Valid
4	<40%	Tidak Valid

Uji validitas empiris penelitian ini menggunakan model *Rasch* dari Linacre (2006) dengan perangkat lunak Ministep 5.4.1. Menurut Boone, Staver & Yale (2014), model *Rasch* dapat secara simultan mengamati interaksi antara item dan responden. Model *Rasch* dipilih karena dapat memberikan informasi mendalam tentang kualitas item soal dan konsistensi respon siswa, sehingga validitas dan reliabilitas instrumen lebih terukur. Kriteria yang digunakan untuk menilai kesesuaian atau ketepatan antara responden dan item pertanyaan antara lain sebagai berikut:

1. Nilai *outfit mean square* (MNSQ) yang diterima: $0,5 < \text{MNSQ} < 1,5$.
2. Nilai *outfit Z-standars* (ZSTD) yang diterima: $-2,0 < \text{ZSTD} < +2,0$.
3. Nilai *outfit Point Measure Correlation* (Pt Mean Corr) yang diterima: $0,4 < \text{Pt Measure Corr} < 0,85$.

Pengujian reliabilitas produk ini dilakukan menggunakan model *Rasch* dengan bantuan perangkat lunak Ministep 5.4.1. Pada penelitian ini terdapat dua analisis reliabilitas, yaitu: *item reliability* dan *person reliability*. Rumus *Cronbach's alpha* harus digunakan untuk menilai reliabilitas dengan model *Rasch*. Ketergantungan interaksi orang-item dan item secara keseluruhan diukur dengan nilai *Cronbach's alpha*. Kriteria nilai *Cronbach's alpha* tercantum dalam Tabel 2, dan *item reliability* dan *person reliability* tercantum dalam Tabel 3 (Sumintono, B., dan Widhiarso, 2015).

Tabel 2. Kriteria *alpha cronbach*

No	Nilai	Kriteria
1	> 0,8	Bagus sekali
2	0,7 – 0,8	Bagus
3	0,6 – 0,7	Cukup
4	0,5 – 0,6	Jelek
5	< 0,5	Buruk

Tabel 3. Kriteria *item reliability* dan *person reliability*

No	Nilai	Kriteria
1	> 0,94	Istimewa
2	0,91 - 0,94	Bagus sekali
3	0,81 - 0,90	Bagus
4	0,67 - 0,80	Cukup
5	< 0,67	Lemah

Tabel 4. Kriteria kepraktisan perangkat pembelajaran

No	Nilai	Kriteria
1	$P_n > 80$	Sangat Praktis
2	$60 < P_n \leq 81$	Praktis
3	$40 < P_n \leq 60$	Cukup Praktis
4	$20 < P_n \leq 40$	Kurang Praktis
5	$P_n \leq 20$	Tidak Praktis

Guru-guru fisika diberikan kuesioner sebagai bagian dari uji kepraktisan untuk memperoleh tanggapan mereka terhadap kemudahan penggunaan, kejelasan, serta

kebermanfaatan perangkat evaluasi yang dikembangkan. Data hasil isian kuesioner kemudian dianalisis menggunakan perhitungan kepraktisan yang ditunjukkan dalam Persamaan (2).

$$P_n = \frac{\sum n}{\sum n_{maks}} \times 100 \quad (2)$$

Persamaan ini menghitung persentase nilai kepraktisan berdasarkan total skor aktual yang diperoleh dari responden dibandingkan dengan total skor maksimum yang mungkin dicapai. Nilai hasil perhitungan ini (P_n) digunakan untuk menginterpretasikan tingkat kepraktisan perangkat berdasarkan kriteria yang telah ditetapkan. Tabel 4 menjelaskan kategori kepraktisan berdasarkan rentang nilai P_n , yang mengacu pada klasifikasi dari (Riduwan, 2012). Semakin tinggi nilai P_n , semakin praktis perangkat evaluasi yang dikembangkan menurut pendapat para guru sebagai praktisi.

HASIL DAN PEMBAHASAN

Instrumen penilaian yang dibuat melalui penelitian pengembangan ini untuk mengukur kemampuan berpikir kritis siswa mengenai materi hukum Newton yang berkaitan dengan gerak berbasis PBL. Alat evaluasi yang dibuat merupakan instrumen penilaian murni yang berfokus pada kemampuan berpikir kritis dalam PBL dan mencakup alat presentasi dan non-tes. Kisi-kisi instrumen, petunjuk penggunaan, bentuk instrumen, rubrik instrumen, dan rangkuman nilai akhir instrumen merupakan bagian dari alat evaluasi yang peneliti buat. Produk instrumen penilaian yang dihasilkan dianggap praktis, valid, dan dapat diandalkan.

Validitas

Produk instrumen penilaian untuk mengukur kemampuan berpikir kritis siswa pada pembelajaran fisika berbasis *Problem Based Learning* (PBL) dinyatakan valid oleh dua dosen ahli dan satu guru fisika berdasarkan tiga aspek yang diujikan, yaitu konstruk, substansi, dan bahasa. Proses validasi dilakukan untuk menilai sejauh mana instrumen telah memenuhi kriteria kelayakan dari segi isi, struktur, dan keterbacaan, sehingga dapat digunakan secara efektif dalam konteks pembelajaran. Penilaian dilakukan dengan menggunakan lembar validasi yang mencantumkan indikator pada masing-masing aspek, yang kemudian diberi skor oleh para validator.

Tabel 5. Hasil Validasi Instrumen

No	Aspek	Ahli			Skor Maks	Presentase Penilaian	Kategori
		1	2	3			
1.	Konstruk	26	26	28	32	83,33%	Sangat Valid
2.	Substansi	34	34	35	44	78,02%	Valid
3.	Bahasa	10	10	11	12	86,10%	Sangat Valid
Rata-rata persentase penilaian						82,48%	Sangat Valid

Tabel 5 menunjukkan bahwa berdasarkan penilaian tiga validator, instrumen memperoleh rata-rata skor validitas sebesar 82,48%, yang termasuk dalam kategori sangat valid. Aspek bahasa memperoleh skor tertinggi (86,10%), menunjukkan bahwa instrumen mudah dipahami dan tidak mengandung ambiguitas. Aspek konstruk memperoleh nilai 83,33%, yang menandakan bahwa penyusunan butir soal telah sesuai dengan indikator kemampuan berpikir kritis. Sementara itu, aspek substansi memperoleh nilai terendah (78,02%) namun tetap berada dalam kategori valid, yang menunjukkan bahwa isi materi pada instrumen relevan dengan tujuan pembelajaran meskipun memerlukan sedikit perbaikan. Hasil ini mengindikasikan bahwa instrumen layak digunakan dalam proses pembelajaran untuk mengukur kemampuan berpikir kritis siswa, dengan rekomendasi untuk melakukan revisi minor pada aspek substansi agar kualitas instrumen semakin optimal.

Peneliti menggunakan instrumen penilaian esai sebanyak 12 pertanyaan untuk mengukur kemampuan berpikir kritis siswa sebagai bagian dari analisis validitas. Tingkat kecocokan butir soal, atau validitas, digunakan untuk menentukan apakah butir soal beroperasi secara normal atau tidak.

Tabel 6. Analisis *item fit* pada instrumen penilaian berpikir kritis peserta didik.

MEASURE	OUTFIT		PT-MEASURE	Item
	MNSQ	ZSTD	CORR	
2,56	1,02	0,18	0.53	BK11
0,59	0,95	-0,11	0.66	BK8
0,34	0,57	-1,86	0.69	BK1
0,34	1,18	0,72	0,51	BK4
0,22	1,09	0,43	0,51	BK6
0,22	1,07	0,34	0,60	BK7
0,10	0,68	-1,38	0,61	BK9
-0,02	1,09	0,45	0,57	BK10
-0,14	1,01	0,13	0,56	BK12
-0,82	1,06	0,38	0,54	BK3
-1,05	1,09	0,54	0,75	BK2
-2,34	0,97	-0,01	0,58	BK5

Mengacu pada Tabel 6, sebanyak 12 butir soal dalam instrumen penilaian kemampuan berpikir kritis telah memenuhi kriteria validitas sesuai dengan parameter yang ditetapkan oleh Boone, Stever, dan Yale (2014), yaitu: (1) nilai *Outfit Mean Square* (MNSQ) berada dalam rentang $0,5 < \text{MNSQ} < 1,5$; (2) nilai *Outfit Z-standard* (ZSTD) berada dalam kisaran $-2,0 < \text{ZSTD} < +2,0$; serta (3) nilai *Point Measure Correlation* (*Pt Measure Corr*) berkisar antara $0,4 < \text{Pt Measure Corr} < 0,85$. Berdasarkan hasil analisis tabel tersebut, seluruh butir soal dinyatakan valid dan layak digunakan.

Dua belas butir soal pada alat penilaian kemampuan berpikir kritis, seperti yang ditunjukkan pada Tabel 6, memenuhi persyaratan untuk parameter yang valid seperti yang dinyatakan oleh Boone, Stever, dan Yale (2014). Butir-butir soal tersebut adalah sebagai berikut: Nilai-nilai dari (1) *outfit mean square* (MNSQ) dan (2) *outfit Z-standard* (ZSTD) diperoleh sebagai berikut: (1) $0.5 < \text{MNSQ} < 1.5$; (2) $-2.0 < \text{ZSTD} < +2.0$; dan (3) $0.4 < \text{Pt Measure Corr} < 0.85$. Berdasarkan tabel temuan analisis data di atas, setiap item dianggap valis dan dapat digunakan.

Reliabilitas

Pada analisis reliabilitas instrumen penilaian kemampuan berpikir kritis peserta didik diperoleh nilai alpha Cronbach adalah 0,81 sehingga termasuk dalam kategori sangat baik (Sumintono dan Widhiarso, 2015). Nilai alpha Cronbach dalam reliabilitas merupakan nilai interaksi antara person reliability dan item reliability secara keseluruhan. Nilai person reliability dan item reliability yang dipakai adalah REAL RMSE karena nilai ini merupakan kondisi terburuk reliabilitas batas bawah berdasarkan instrumen yang dipakai (Sumintono dan Widhiarso, 2015). Adapun person reliability untuk instrumen penilaian kemampuan berpikir kritis dapat dilihat pada Tabel 7.

Tabel 7. Analisis *person reliability* instrumen *authentic assessment* kemampuan berpikir kritis.

TOTAL SCORE	CNT	MEASURE	MODEL S. E.	INFIT		OUTFIT		
				MNSQ	ZSTD	MNSQ	ZSTD	
MEAN	36,7	12.0	.23	.60	1.00	-.00	.91	-.08
SEM	.7	.0	.26	.01	.07	.20	.08	.18
P.SD	4.2	.0	1.51	.04	.43	1.14	.40	1.03
S.SD	4.3	.0	1.54	.04	.44	1.16	.41	1.04
MAX.	42.0	12.0	2.09	.84	1.91	2.37	1.82	1.64
MIN.	26.0	12.0	-3.92	.58	.30	-2.21	.19	-2.26
REAL RMSE	.66	TRUE SD	1.36	SEPARATION	2.07	Person		
RELIABILITY	.81							
MODEL RMSE	.60	TRUE SD	1.39	SEPARATION	2.30	Person		
RELIABILITY	.84							
S. E. OF Person MEAN	= .26							

Pada Tabel 7 diketahui INFIT MNSQ dan OUTFIT MNSQ untuk tabel person nilai rata-ratanya secara berurutan adalah 1,00 dan 0,91 yang berarti nilainya semakin mendekati dan

sudah pada nilai 1,00 sehingga dikategorikan semakin baik karena nilai idealnya ialah 1,00. INFIT ZSTD dan OUTFIT ZSTD, nilai rata-rata tabel *person* secara berurutan sebesar 0,00 dan 0,08 yang berarti nilainya mendekati dan sudah pada nilainya 0,00 maka dikategorikan kualitas semakin baik. Hal ini dikarenakan nilai idealnya sebesar 0,00. Sedangkan nilai *person reliability* sebesar 0,81 yang menunjukkan bahwa konsistensi jawaban dari responden bagus. Adapun item *reliability* untuk instrumen *authentic assessment* kemampuan berpikir kritis dapat dilihat pada Tabel 8.

Karena nilai idealnya adalah 1.00, INFIT MNSQ dan OUTFIT MNSQ pada Tabel 7 untuk tabel *person* memiliki nilai rata-rata 1.00 dan 0.91. Hal ini menunjukkan bahwa nilai tersebut mendekati dan telah mencapai 1.00, sehingga diklasifikasikan sebagai peningkatan. Nilai rata-rata tabel *person* untuk INFIT ZSTD dan OUTFIT ZSTD masing-masing adalah 0.00 dan 0.08. Hal ini menunjukkan bahwa nilai tersebut semakin mendekati dan telah mencapai 0.00, yang menunjukkan bahwa kualitasnya semakin tinggi. Hal ini disebabkan oleh fakta bahwa 0.00 adalah nilai yang ideal. Namun, nilai *person reliability* adalah 0,81, yang mengindikasikan bahwa tanggapan responden secara umum konsisten. Tabel 8 berikut menampilkan reliabilitas butir soal untuk instrumen *authentic assessment* keterampilan berpikir kritis.

Tabel 8. Analisis *Item Reliability* instrumen *authentic assessment* kemampuan berpikir kritis.

TOTAL SCORE	CNT MEASURE		MODEL S. E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MEAN 108.8	37.0	.00	.35	1.00	.04	.98	-.01
SEM 2.7	.0	.33	.00	.06	.28	.05	.23
P.SD 8.9	.0	1.10	.01	.20	.92	.17	.76
S.SD 9.3	.0	1.15	.01	.21	.96	.18	.79
MAX. 127.0	37.0	2.56	.39	1.19	0.94	1.18	0.72
MIN. 88.0	37.0	-	.34	.57	-1.96	.57	-1.86
		2.34					
REAL RMSE RELIABILITY	.37	TRUE SD .89	1.04	SEPARATION 2.83	Item		
MODEL RMSE RELIABILITY	.35	TRUE SD .90	1.04	SEPARATION 2.96	Item		
S. E. OF Item Mean	=	.33					

Nilai rata-rata INFIT MNSQ dan OUTFIT MNSQ masing-masing adalah 1.00 dan 0.98, seperti yang ditunjukkan pada Tabel 8. Hal ini mengindikasikan bahwa nilai tersebut membaik karena sudah mendekati nilai optimum yaitu 1,00. Kualitas butir soal meningkat karena nilai rata-rata INFIT ZSTD dan OUTFIT ZSTD sudah mendekati nilai ideal 0.0, yaitu masing-masing 0.04 dan -0.1. Butir-butir soal alat penilaian berpikir kritis dapat mengukur apa yang perlu diukur, seperti yang ditunjukkan oleh nilai *item reliability* butir soal sebesar 0.89, yang menunjukkan bahwa butir-butir soal memiliki kualitas yang baik.

Karena instrumen penilaian yang dikembangkan sudah memenuhi aspek kelayakan instrumen, maka dapat dinyatakan bahwa instrumen *authentic assessment* untuk mengukur kemampuan berpikir kritis peserta didik pada pembelajaran fisika berbasis PBL sudah valid dan reliabel untuk digunakan.

Kepraktisan

Uji kepraktisan instrumen penilaian kemampuan berpikir kritis dilakukan pada tiga aspek, yaitu kemudahan penggunaan, kemenarikan sajian, dan kebermanfaatannya. Uji kepraktisan ini dilakukan kepada dua orang guru fisika sebagai ahli praktisi. Adapun kriteria guru yang menjadi subjek uji kepraktisan instrumen penilaian kemampuan berpikir kritis yang peneliti tetapkan adalah guru dengan jenjang pendidikan guru serta latar belakang pendidikan fisika dan sekolah tempat praktisi mengajar sudah menerapkan Kurikulum Merdeka.

Hasil rata-rata terkait penilaian kepraktisan instrumen *authentic assessment* yang telah dinilai oleh praktisi disajikan dalam Tabel 9. Berdasarkan Tabel 5 data uji kepraktisan instrumen penilaian kemampuan berpikir kritis didapatkan bahwa pada aspek kemudahan penggunaan diperoleh skor rata-rata yaitu 93,75. Selanjutnya pada aspek kemenarikan sajian diperoleh skor

rata-rata yaitu 75 serta skor rata-rata untuk aspek kebermanfaatan yaitu sebesar 98,21. Dari ketiga skor rata-rata untuk ketiga aspek tersebut diperoleh skor perolehan rata-rata instrumen penilaian yaitu sebesar 88,98 dengan kriteria sangat tinggi. Kepraktisan merupakan salah satu aspek yang menentukan kualitas suatu alat evaluasi, yang dinilai berdasarkan kemudahan penggunaannya oleh guru dan siswa dalam proses pembelajaran.

Tabel 9. Hasil Skor Rata-rata Penilaian Kepraktisan Instrumen *Authentic Assessment* Kemampuan Berpikir Kritis.

No	Aspek	Praktisi (%)		Skor rata-rata (%)
		1	2	
1.	Kemudahan penggunaan	96,88	93,75	95,31
2.	Kemenarikkan sajian	93,75	93,75	93,75
3.	kemanfaatan	100	92,85	96,42
Skor yang diperoleh				95,16

Tabel 9 menampilkan temuan rata-rata yang berkaitan dengan evaluasi kepraktisan instrumen *authentic assessment* yang telah dievaluasi oleh para praktisi. Berdasarkan statistik Tabel 5 tentang uji kepraktisan alat penilaian kemampuan berpikir kritis, skor rata-rata untuk kemudahan penggunaan adalah 93,75. Selain itu, komponen kegunaan penyajian memperoleh skor rata-rata 98,21, sedangkan komponen daya tarik memperoleh skor rata-rata 75. Perolehan skor rata-rata alat evaluasi adalah 88,98 dengan kriteria sangat tinggi, berdasarkan ketiga skor rata-rata untuk ketiga komponen tersebut. Kriteria penilaian kualitas instrumen penilaian adalah kepraktisan, yang mengukur kemudahan guru dan siswa dalam menggunakan instrumen penilaian yang dikembangkan.

Hasil akhir produk ini berupa instrumen *authentic assessment* kemampuan berpikir kritis peserta didik pada materi hukum newton tentang gerak berbasis PBL yang telah valid, reliabel, dan juga memiliki nilai kepraktisan yang tinggi. Produk disajikan dalam satu perangkat dan dapat digunakan dalam pembelajaran fisika untuk mengukur kemampuan berpikir kritis yang dimiliki oleh peserta didik. Instrumen ini dapat membantu dan mempermudah guru dalam memaksimalkan penilaian terhadap peserta didik dengan menyesuaikan tahapan pembelajaran berbasis PBL. Suatu instrumen dikatakan praktis jika memberikan kemudahan dalam persiapan, penggunaan, interpretasi, perolehan hasil, serta penyimpanan. Instrumen penilaian yang dikembangkan dianggap layak karena memenuhi kriteria validitas, reliabilitas dan kepraktisan.

Hasil validasi menunjukkan instrumen penilaian *authentic assessment* berbasis PBL memenuhi kriteria sangat valid dengan skor rata-rata 82,48%. Validitas tertinggi terdapat pada aspek bahasa (86,10%), diikuti konstruk (83,33%) dan substansi (78,02%). Tingginya validitas bahasa mencerminkan kejelasan instruksi dan rubrik penilaian, sementara skor substansi yang sedikit lebih rendah mengindikasikan perlunya penyesuaian kedalaman materi Hukum Newton. Temuan ini sejalan dengan penelitian Thiagarajan (1974) bahwa validasi multidimensi (konstruk, substansi, bahasa) krusial dalam pengembangan instrumen. Validitas empiris melalui model Rasch juga mengonfirmasi 12 butir soal memenuhi kriteria *fit* statistik (MNSQ, ZSTD, Pt-Measure Corr), memperkuat kelayakan instrumen untuk mengukur berpikir kritis.

Instrumen ini memiliki reliabilitas sangat baik dengan *Cronbach's Alpha* 0,81, melebihi ambang batas 0,7 yang direkomendasikan (Sumintono, 2015). Nilai *person reliability* (0,81) dan *item reliability* (0,89) menunjukkan konsistensi respons siswa dan kemampuan butir soal dalam membedakan tingkat kemampuan peserta. Hasil ini lebih unggul dibandingkan instrumen serupa (Wijaya et al., 2017) ($\alpha=0,79$), terutama karena penggunaan analisis Rasch yang meminimalkan bias. Reliabilitas tinggi ini menjamin instrumen dapat digunakan secara konsisten dalam berbagai konteks pembelajaran PBL.

Skor kepraktisan mencapai 96,16% (sangat praktis), dengan aspek kebermanfaatan tertinggi (96,42%). Guru menilai instrumen mudah digunakan (95,31%) dan menarik (93,75%), berkat rubrik terstruktur dan desain yang jelas. Temuan ini mendukung pernyataan Agustianti et al. (2022) bahwa kepraktisan adalah kunci adopsi alat penilaian di kelas. Tingginya skor

kepraktisan juga mencerminkan kesesuaian instrumen dengan kebutuhan guru dalam Kurikulum Merdeka, khususnya untuk menilai keterampilan berpikir kritis secara otentik.

Kombinasi ketiga aspek ini menjadikan instrumen tidak hanya layak secara teoritis (valid dan reliabel), tetapi juga aplikatif di lapangan. Contohnya, validasi konstruk yang tinggi (83,33%) didukung reliabilitas item (0,89), sementara kepraktisan (96,16%) memastikan instrumen tidak rumit digunakan guru. Hal ini mengisi celah penelitian sebelumnya seperti Nurindanasari et al. (2020) yang hanya fokus pada validitas tanpa uji kepraktisan. Integrasi ini juga memperkuat temuan Boone et al. (2014) bahwa instrumen berbasis Rasch perlu diseimbangkan dengan kemudahan penggunaan.

Hasil akhir dari produk ini adalah instrumen *authentic assessment* kemampuan berpikir kritis siswa pada materi hukum Newton tentang gerak berbasis PBL yang sudah valid, dan sangat bermanfaat. Produk ini hadir dalam satu instrumen dan dapat digunakan untuk menilai kemampuan berpikir kritis siswa ketika mereka belajar fisika. Dengan memodifikasi fase-fase pembelajaran berbasis PBL, instrumen ini dapat membantu dan memudahkan guru untuk memaksimalkan penilaian siswa. Sebuah instrumen dianggap praktis jika menawarkan kemudahan dalam persiapan, penggunaan, interpretasi, penerimaan hasil, dan penyimpanan (Kurniawan, 2022). Instrumen ini memberikan dampak praktis bagi guru fisika dalam menerapkan PBL sekaligus penilaian autentik. Namun, uji coba terbatas pada 35 siswa dan 3 validator mengharuskan penelitian lanjutan dengan sampel lebih besar. Saran untuk pengembangan berikutnya mencakup penyederhanaan rubrik untuk meningkatkan skor substansi serta uji coba lintas sekolah untuk menguji generalisasi. Temuan penelitian ini menjadi fondasi pengembangan instrumen sejenis untuk materi sains lainnya.

KESIMPULAN

Temuan penelitian membuktikan bahwa instrumen penilaian autentik berbasis PBL ini telah memenuhi standar kelayakan secara teoritis (validitas dan reliabilitas) maupun praktis, sehingga siap diimplementasikan dalam pembelajaran fisika. Hasil penelitian menunjukkan bahwa instrumen yang dikembangkan tidak hanya secara teoritis layak digunakan tetapi juga aplikatif dalam pembelajaran di kelas. Validasi ahli dan uji empiris membuktikan instrumen ini mampu mengukur konstruk berpikir kritis secara akurat, sementara tanggapan positif dari guru menguatkan nilai praktisnya dalam sistem penilaian Kurikulum Merdeka.

Keberhasilan pengembangan instrumen ini memberikan kontribusi penting dalam bidang asesmen pendidikan, khususnya dalam menghubungkan pendekatan PBL dengan penilaian autentik. Temuan penelitian menawarkan solusi konkret bagi guru dalam melaksanakan penilaian yang berorientasi pada keterampilan berpikir tingkat tinggi. Untuk penelitian selanjutnya, pengembangan instrumen serupa dapat diperluas untuk materi dan jenjang pendidikan lainnya, sekaligus mengeksplorasi dampak penggunaan instrumen ini terhadap peningkatan kualitas pembelajaran secara lebih komprehensif. Dengan demikian, penelitian ini tidak hanya memberikan manfaat langsung bagi praktik pembelajaran di kelas tetapi juga membuka peluang untuk pengembangan penelitian lebih lanjut di bidang asesmen pendidikan.

DAFTAR PUSTAKA

- Agustianti, R., Abyadati, S., Nussifera, L., Irvani, A. I., Handayani, D. Y., Hamdani, D., & Amarulloh, R. R. (2022). *Asesmen Dan Evaluasi Pembelajaran*. Tohar Media.
- Association, N. E. (2002). *Preparing 21st Century Students for a Global Society Great Public Schools for Every Student*.
- Astuti, S., Dania, M., & Anwar, M. (2018). Pengembangan Lkpd Berbasis Pbl (Problem Based Learning) Untuk Meningkatkan Keterampilan Berpikir Kritis Peserta Didik Pada Materi Keseimbangan Kimia. *Chemistry Education Review (CER)*, 1, 90. <https://doi.org/10.26858/cer.v0i1.5614>

- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Dordrecht: Springer. <https://doi.org/10.1007/978-94-007-6857-4>
- Christi, R. Y. D., Jeffry Handhika, & Andista Candra Yusro. (2020). RADIASI : Jurnal Berkala Pendidikan Fisika RADIASI: Jurnal Berkala Pendidikan Fisika. *Jurnal Berkala Pendidikan Fisika*, 12(2), 61–64.
- Depdiknas. (2005). Peraturan Pemerintah RI Nomor 19 Tahun 2005 tentang Standar Nasional Pendidikan. *Jakarta: Depdiknas.*, 2005–2008.
- Indrapangastuti, D. (2023). *Berpikir kritis melalui problem based learning (teori dan implementasi)*. CV Pajang Putra Wijaya.
- Ismiyana, N., Fajriyah, K., & Reffiane, F. (2023). Pengaruh Model Problem Based Learning Terhadap Kemampuan Berfikir Kritis Siswa Pada Mata Pelajaran Ipa Materi Peredaran Darah Kelas V Sd Negeri 1 Juwangi. *Didaktik: Jurnal Ilmiah PGSD STKIP Subang*, 9(2), 5917–5930.
- Joyce, B., & Weil, M. (1972). *Conceptual Complexity, Teaching Style and Models of Teaching*.
- Kurniawan, H. (2022). *Pengantar praktis penyusunan instrumen penelitian*. Deepublish.
- Linacre, J. M. (2006). *A User's Guide to Winstep Ministep RaschModel Computer Programs*. MESA Press.
- Nurindanasari, D. A., Setiawan, D. A., & Yuniasih, N. (2020). Pengembangan Instrument Authentic Assessment Pada Aspek Literasi Membaca di SDN 03 Plaosan Kabupaten Malang. *Seminar Nasional PGSD UNIKAMA*, 4, 123–130.
- Redhana, I. W. (2024). *Literasi Digital: Pedoman Menghadapi Society 5.0*. Samudra Biru.
- Retug, N. (2010). Analisis kebutuhan pengembangan perangkat pembelajaran berbasis masalah pada pembelajaran sains kimia di SMP. *Jurnal Pendidikan Dan Pengajaran*, 43(2), 106–113.
- Riduwan. (2012). *Cara Mudah Menggunakan dan Memaknai Path Analysis (Analisis Jalur)*. Alfabeta.
- Robert H. Ennis. (1996). *Critical Thinking*. New Jersey: Prentice-Hall.
- Rosidin, U. (2017). *Evaluasi dan Asesmen Pembelajaran*. Media Akademi.
- Safitri, K., & Harjono, N. (2021). Pengembangan Instrumen Penilaian Sikap Sosial Aspek Tanggung Jawab Pembelajaran Tematik Terpadu Siswa Kelas 4 SD. *Jurnal Pedagogi Dan Pembelajaran*, 4(1), 111. <https://doi.org/10.23887/jp2.v4i1.33352>
- Sugiyono. (2015). *Metode Penelitian dan Pengembangan Research and Development*. Alfabeta.
- Sumintono, B., dan Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. Tim komunikata.
- Sylvia, I., Anwar, S., & Khairani, K. (2019). Pengembangan Instrumen Penilaian Autentik Berbasis Pendekatan Authentic Inquiry Learning Pada Mata Pelajaran Sosiologi di Sekolah Menengah Atas. *Jurnal Socius: Journal of Sociology Research and Education*, 6(2), 103. <https://doi.org/10.24036/scs.v6i2.162>

- Thiagarajan, S. A. O. (1974). *Instructional Development for Training Teachers of Exceptional Children: A Sourcebook*. Indiana Univ., Bloomington. Center for Innovation in Teaching the Handicapped. National Center for Improvement of Educational System.
- Walsh, L. N., Howard, R. G., & Bowe, B. (2007). Phenomenographic study of students' problem solving approaches in physics. *Physical Review Special Topics - Physics Education Research*, 3(2), 1–12. <https://doi.org/10.1103/PhysRevSTPER.3.020108>
- Wijaya, U. R. B., Sumarni, W., & Haryani, S. (2017). Pengembangan instrumen penilaian berpikir kritis pada pembelajaran kimia berpendekatan sets (Science, environment, technology, and society). *Chemistry in Education*, 6(2), 35–41.
- Wulandari, A., Suyatna, A., Viyanti, V., & Rosidin, U. (2019). Development of CBT- Based Assessment Instruments Using WQC Application to Measure HOTS in Impulse Momentum. *Jurnal Pembelajaran Fisika*, 9(1), 113–128. <https://doi.org/10.23960/jpf.v9.n1.202110>