



PENGEMBANGAN ASESMEN AUTENTIK PADA LAPORAN PRAKTIKUM UNTUK MENGUKUR HOTS SISWA SMA

Zumrotun Muthohiroh¹, Supriyadi²

UNNES, Semarang, Jawa Tengah^{1,2}

e-mail: zumrotunmuthohiroh@students.unnes.ac.id

Diterima: 03/06/2026; Direvisi: 27/07/2026; Diterbitkan: 03/08/2026

ABSTRAK

Penilaian keterampilan berpikir tingkat tinggi (*Higher Order Thinking Skills/HOTS*) dalam pembelajaran fisika masih didominasi oleh tes tertulis, sehingga belum sepenuhnya mampu mengukur kemampuan analisis, evaluasi, dan kreasi siswa melalui kegiatan praktikum. Di sisi lain, pengembangan asesmen autentik berbasis rubrik analitik pada laporan praktikum, khususnya materi kisi difraksi, masih terbatas sehingga diperlukan instrumen yang valid, reliabel, dan praktis untuk mendukung penilaian HOTS. Penelitian ini bertujuan untuk mengembangkan asesmen autentik berupa rubrik analitik pada laporan praktikum fisika untuk mengukur keterampilan berpikir tingkat tinggi (*Higher Order Thinking Skills/HOTS*) siswa kelas XI pada materi kisi difraksi. Penelitian ini menggunakan metode *Research and Development* (R&D) dengan model ADDIE yang meliputi tahap *analysis*, *design*, *development*, *implementation*, dan *evaluation*. Validitas instrumen diuji melalui *expert judgment* menggunakan koefisien Aiken's V, sedangkan reliabilitas instrumen dianalisis menggunakan *Intraclass Correlation Coefficient* (ICC). Hasil penelitian menunjukkan bahwa asesmen autentik yang dikembangkan memiliki validitas isi yang tinggi dengan nilai $V \geq 0,92$ dan reliabilitas yang baik dengan nilai ICC sebesar 0,74 sehingga layak digunakan sebagai instrumen penilaian. Instrumen yang dikembangkan juga dinilai praktis digunakan dalam pembelajaran fisika. Selain itu, asesmen autentik mampu memotret keterampilan berpikir tingkat tinggi siswa (MA An Nidham, MA Mifathussalam dan MAN 02 Kota Semarang) pada aspek analisis, evaluasi, dan kreasi melalui laporan praktikum. Hasil uji terbatas dan uji lapangan menunjukkan bahwa sebagian besar siswa berada pada kategori kemampuan berpikir tingkat tinggi baik.

Kata Kunci: *Asesmen Autentik, HOTS, Laporan Praktikum*

ABSTRACT

Assessment of Higher Order Thinking Skills (HOTS) in physics learning is still predominantly based on written tests, which have not fully captured students' abilities in analysis, evaluation, and creation through laboratory activities. Moreover, the development of authentic assessment using analytical rubrics for physics laboratory reports, particularly on the topic of diffraction gratings, remains limited. Therefore, there is a need for valid, reliable, and practical assessment instruments to support the evaluation of students' Higher Order Thinking Skills. This study aims to develop an authentic assessment in the form of an analytical rubric for physics lab reports to measure the higher-order thinking skills (HOTS) of 11th-grade students on the topic of diffraction gratings. This study employed the Research and Development (R&D) method using the ADDIE model, which comprises the stages of analysis, design, development, implementation, and evaluation. The validity of the instrument was tested through expert judgment using Aiken's V coefficient, while the reliability of the



instrument was analyzed using the Intraclass Correlation Coefficient (ICC). The results of the study indicate that the authentic assessment developed has high content validity with a V value ≥ 0.92 and good reliability with an ICC value of 0.74, making it suitable for use as an assessment instrument. The developed instrument was also deemed practical for use in physics instruction. Furthermore, the authentic assessment was able to capture students' higher-order thinking skills in the areas of analysis, evaluation, and creation through lab reports. Results from limited testing and field testing indicated that the majority of students fell into the "good" category for higher-order thinking skills.

Keywords: *Authentic Assessment, HOTS, Lab Report*

PENDAHULUAN

Pendidikan di era globalisasi abad ke-21 menuntut transformasi fundamental dalam sistem pembelajaran dan evaluasi, di mana fokus utama kini bergeser dari penguasaan materi secara hafalan menuju pengembangan kemampuan berpikir kritis, kreatif, dan analitis yang terangkum dalam *Higher Order Thinking Skills* (HOTS). Keterampilan ini menjadi kunci bagi peserta didik untuk menghadapi tantangan dunia nyata yang kompleks dan dinamis, namun fakta di lapangan sering kali menunjukkan adanya kesenjangan antara cita-cita kurikulum dengan praktik penilaian yang masih didominasi oleh tes tertulis konvensional berorientasi pada ingatan semata. Urgensi riset ini terletak pada kebutuhan mendesak untuk menghadirkan inovasi instrumen penilaian autentik, khususnya melalui asesmen autentik laporan praktikum, yang mampu memotret proses berpikir ilmiah siswa secara komprehensif dan objektif. Dengan menyediakan alat ukur yang valid dan reliabel, penelitian ini berperan strategis dalam menjembatani kebutuhan guru akan standar penilaian yang tidak hanya mengevaluasi hasil akhir, tetapi juga membina kemandirian berpikir dan literasi sains siswa agar tetap relevan dengan tuntutan zaman.

Melalui proses pendidikan, siswa mengembangkan kemampuan mereka untuk menilai pengalaman mereka, mengevaluasi secara kritis berbagai aspek kehidupan, dan mengembangkan kemandirian berpikir mereka. Dalam beberapa tahun terakhir, keterampilan berpikir tingkat tinggi dengan konotasi inti seperti pemecahan masalah, kreativitas, dan berpikir kritis telah diakui sebagai keterampilan kunci bagi talenta di abad ke-21, dan telah ditekankan serta dimasukkan ke dalam standar kurikulum di banyak negara (Liu et al., 2024). Beberapa organisasi, seperti "*P21 Framework*" dan "*Society for Technology*" dalam bidang Pendidikan, telah mengidentifikasi berbagai jenis keterampilan yang harus dikuasai siswa pada abad ke-21. Kolaborasi, Komunikasi, Berpikir kritis dan pemecahan masalah, dan Kreativitas dianggap sangat penting. Siswa harus memperoleh keterampilan ini karena tiga alasan utama: (a) keterampilan ini sulit diajarkan dan dinilai, sehingga jarang dimasukkan dalam kurikulum; (b) keterampilan ini sangat penting bagi semua siswa di era globalisasi; dan (c) keterampilan ini penting untuk karier apa pun. (Kousloglou et al., 2023)

Sejalan dengan tuntutan kompetensi abad ke-21 tersebut, sistem pendidikan di Indonesia juga mengarahkan pengembangan kurikulum dan proses pembelajaran untuk membekali peserta didik dengan keterampilan berpikir tingkat tinggi dan kompetensi yang relevan dengan kebutuhan global. Di Indonesia, kurikulum dirancang untuk menghasilkan lulusan yang kompeten dengan kemampuan intelektual yang kuat, terutama keterampilan 4C (berpikir kritis, kreativitas, komunikasi, dan kolaborasi) untuk meraih kesuksesan di abad ke-21. Oleh karena itu, penekanan pada kompetensi abad ke-21 sangat penting untuk mendorong siswa. 'kebutuhan dalam persiapan menghadapi persaingan global, kemajuan teknologi, dan



kejenuhan media (Herlinawati et al., 2024). Pendidikan di abad ke-21 menuntut penguasaan *Higher Order Thinking Skills* (HOTS) sebagai elemen inti dalam membentuk kemampuan berpikir kritis siswa untuk menyelesaikan masalah. Namun, sistem evaluasi tradisional yang berfokus pada hafalan masih mendominasi ruang kelas, sehingga menempatkan siswa pada tahap pembelajaran pasif dan dangkal (Liu et al., 2024). Kondisi ini diperburuk oleh temuan di lapangan yang menunjukkan bahwa instrumen penilaian seperti soal ujian sumatif dan Lembar Kerja Peserta Didik (LKPD) masih memiliki kandungan indikator HOTS yang rendah (Wati et al., 2024; Nurhayati, 2024).

Untuk mengatasi kesenjangan tersebut, penelitian terdahulu menekankan pentingnya pengembangan instrumen penilaian yang valid dan autentik. Penggunaan *autentic assessment* (asesmen autentik) menjadi solusi krusial karena mampu mengukur kecakapan autentik melalui tugas-tugas kompleks dalam situasi nyata (Makmuroh & Pratama, 2022; Xu & Shen, 2025). Agar asesmen ini efektif, pengembangan rubrik penilaian yang mengintegrasikan karakteristik HOTS seperti pemodelan, penalaran, dan pemecahan masalah mutlak diperlukan sebagai panduan umpan balik bagi siswa (Gallardo, 2020; Gradini, 2022). Selain itu, keberhasilan implementasi HOTS sangat bergantung pada ketelitian dalam tahap validasi instrumen, mulai dari aspek materi, konstruksi, hingga bahasa (Suhady et al., 2020). Penggunaan model pengembangan yang sistematis, seperti model Mardapi yang mencakup sembilan tahap pengembangan (Arafah et al., 2021), terbukti mampu menghasilkan perangkat penilaian yang reliabel.

Kebaruan dalam penelitian ini terletak pada pengembangan instrumen evaluasi berupa asesmen autentik yang dirancang khusus untuk mengukur HOTS pada laporan praktikum fisika, khususnya pada materi kisi difraksi. Berbeda dengan rubrik penilaian praktikum pada umumnya yang hanya berfokus pada aspek psikomotorik atau prosedur teknis, instrumen ini mengintegrasikan indikator kognitif tingkat tinggi yang disesuaikan dengan karakteristik penulisan ilmiah siswa. Pengembangan ini juga didasarkan pada data empiris hasil wawancara dan analisis kebutuhan guru di lapangan, sehingga menghasilkan instrumen yang tidak hanya valid secara teoretis tetapi juga praktis untuk diimplementasikan.

Berdasarkan permasalahan tersebut, dapat disimpulkan bahwa pengembangan instrumen evaluasi yang akurat menjadi kebutuhan yang mendesak untuk menjembatani antara kemampuan nyata siswa dengan sistem penilaian yang ada. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan sebuah rubrik analitik berbasis HOTS pada materi kisi difraksi yang valid, reliabel, dan praktis. Kehadiran instrumen ini diharapkan dapat memberikan standar penilaian yang lebih komprehensif bagi guru serta mendorong peningkatan kualitas literasi sains dan kemampuan berpikir kritis siswa dalam pembelajaran fisika.

METODE PENELITIAN

Jenis penelitian ini menggunakan pendekatan Research and Development (R&D) dengan model pengembangan ADDIE. Model ADDIE terdiri atas lima tahap, yaitu analisis (*analysis*), perencanaan (*design*), pengembangan (*development*), implementasi (*implementation*), dan evaluasi (*evaluation*). Penelitian ini bertujuan menghasilkan instrumen untuk mengukur keterampilan berfikir tingkat tinggi siswa SMA pada laporan praktikum. Instrumen yang dikembangkan memiliki kriteria valid, reliabel dan praktis sehingga dapat digunakan dalam kegiatan praktikum fisika. Model ADDIE dipilih karena sangat sistematis, terstruktur dan fleksibel. kerangka kerja ini memastikan setiap tahapan

berjalan logis, meminimalisir kesalahan dan menghasilkan produk yang telah teruji validitas, reliabilitas dan kepraktisan instrumen. Alur model pengembangan ADDIE dapat dilihat dari gambar berikut:



Gambar 1. Alur ADDIE

Pada tahap *analysis* dilakukan analisis kurikulum, wawancara dengan guru mata pelajaran sebagai bagian dari analisis kebutuhan untuk memperoleh informasi mengenai pelaksanaan kegiatan praktikum dan instrumen penilaian yang digunakan, serta studi dokumen terhadap laporan praktikum yang telah disusun oleh peserta didik. Dalam tahapan tersebut terdapat kebutuhan tentang instrumen yang digunakan pada laporan praktikum untuk mengukur keterampilan berfikir tingkat tinggi siswa sesuai dengan tuntutan kurikulum pada abad ke -21. Pada tahap *design* perencanaan fokus pada penyusunan instrumen untuk mengukur keterampilan berfikir tingkat tinggi. Pada tahap *development* pengembangan dilakukan dengan validasi oleh empat ahli sesuai bidangnya. Setelah dilakukan uji validasi dan reliabilitas dilakukan evaluasi terhadap instrumen yang dikembangkan. Dilanjutkan dengan uji coba skala kecil dan evaluasi terhadap produk. Langkah selanjutnya *implementation* dilakukan dengan uji skala besar kemudian hasilnya berupa potret keterampilan berfikir tingkat tinggi siswa. Selanjutnya uji kepraktisan diberikan kepada dua guru pengampu. Hasil uji coba skala besar dan uji kepraktisan kemudian di evaluasi sebagai langkah akhir dalam model pengembangan ADDIE.

Subjek penelitian pada tahap uji coba adalah peserta didik kelas XI SMA yang telah melaksanakan kegiatan praktikum kisi difraksi sehingga memenuhi kriteria sebagai pengguna instrumen yang dikembangkan. Uji coba dilakukan secara bertahap, yaitu uji coba skala kecil yang melibatkan 22 peserta didik dan dilanjutkan dengan uji coba skala besar yang melibatkan 60 peserta didik. Uji kepraktisan dilakukan pada dua guru mapel fisika di dua sekolah. Validator ahli terdiri dari dua dosen dan dua guru mapel fisika yang ahli dalam bidangnya. Teknik pengumpulan data disesuaikan dengan tujuan penelitian, yaitu untuk mengetahui karakteristik validitas, reliabilitas, potret HOTS dan kepraktisan instrumen penilaian yang dikembangkan. Data kevalidan dan reliabilitas diperoleh melalui hasil penilaian validator, data potret HOTS diperoleh dari hasil laporan praktikum peserta didik yang dinilai menggunakan instrumen, data kepraktisan diperoleh dari hasil angket respon siswa dan guru. Analisis data kevalidan dilakukan dengan perhitungan menggunakan Aiken'v. Sedangkan data reliabilitas dianalisis menggunakan ICC (*Intraclass Correlation Coefficient*).



Validitas Isi (*Content Validity*) dilakukan melalui tinjauan ahli (*expert review*) dengan menghitung *Content Validity Index* (CVI). Ambang batas skor 0,80 menunjukkan bahwa instrumen tersebut memiliki validitas yang dapat diterima. Uji reliabilitas menggunakan ICC. Nilai koefisien ICC yang diperoleh dari hasil perhitungan kemudian diinterpretasikan untuk menentukan sejauh mana instrumen memiliki konsistensi yang stabil, baik secara internal maupun antar-penilai (*inter-rater reliability*). Kriteria tingkat reliabilitas berdasarkan nilai ICC dibagi menjadi empat kategori yaitu ICC < 0,50 reliabilitas rendah, 0,50 – 0,75 reliabilitas sedang, 0,75 – 0,90 reliabilitas baik dan ICC > 0,90 reliabilitas sangat baik. Data kepraktisan instrumen yang diperoleh dari respon guru dan siswa, selanjutnya dilakukan interpretasi terhadap persentase skor kepraktisan untuk menentukan kategori tingkat kepraktisan instrumen yang dikembangkan. Pengelompokan kategori ini bertujuan untuk memberikan gambaran mengenai tingkat kemudahan penggunaan, keterlaksanaan, dan kebermanfaatan instrumen dalam proses pembelajaran. Adapun kriteria kepraktisan instrumen sebagai berikut: 81% - 100% sangat praktis, 61% - 80% praktis, 41% - 60% cukup praktis, 21% - 40% kurang praktis dan 0 – 20% tidak praktis (Jannah & Istianah, 2021). Potret HOTS diperoleh dari hasil nilai laporan praktikum yang dibuat oleh peserta didik menggunakan instrumen yang dikembangkan. Dimana kriteria penilaiannya yaitu 86-100 sangat baik, 71-85 baik, 56-70 cukup dan ≤55 Kurang.

HASIL DAN PEMBAHASAN

Hasil

Fase *analysis* melibatkan identifikasi kebutuhan, tujuan instrumen, dan karakteristik peserta yang akan dinilai (Bhakti et al., 2023). Penelitian ini dimulai dengan menganalisis kebutuhan siswa, kurikulum, dan karakteristik materi pelajaran. Analisis kebutuhan dilakukan dengan observasi ke lapangan berupa studi dokumentasi laporan praktikum fisika dan melakukan wawancara terhadap guru mata pelajaran fisika di lima sekolah. Fokus utama dari analisis ini mencakup beberapa aspek krusial seperti analisis bentuk asesmen, identifikasi kebutuhan instrumen, analisis kendala dan tantangan serta kajian materi yang spesifik. Data yang diperoleh dari fase analisis ini nantinya akan menjadi landasan utama dalam tahap *design* (perancangan) untuk menentukan spesifikasi instrumen evaluasi yang ideal dan sesuai dengan kebutuhan guru di sekolah. Hasil analisis menunjukkan adanya kesenjangan antara pentingnya kemampuan HOTS dengan ketersediaan instrumen penilaian yang relevan pada praktikum Fisika. Hal ini memvalidasi urgensi untuk melanjutkan penelitian ke tahap *Design* (Perancangan) guna mengembangkan instrumen evaluasi untuk HOTS melalui analisis laporan hasil praktikum kisi difraksi.

Setelah melakukan identifikasi kebutuhan pada tahap analisis, langkah selanjutnya adalah melakukan perancangan instrumen. Tahap *Design* ini bertujuan untuk menyusun kerangka instrumen evaluasi yang akan dikembangkan agar sistematis dan mampu mengukur indikator yang telah ditetapkan secara akurat. Pada fase Desain, dibuat rencana terperinci mengenai struktur, isi, dan format instrumen (Bhakti et al., 2023). Pembuatan desain instrumen evaluasi yang sesuai dengan analisis kebutuhan di lapangan. Menyusun kisi-kisi instrumen, menentukan rubrik penilaian, dan menetapkan kriteria penilaian. Terdapat empat tahapan dalam penyusunan kisi-kisi instrumen diantaranya:

1. Penentuan Indikator dan Aspek Kognitif

Instrumen ini dirancang untuk mengukur kemampuan berpikir tingkat tinggi (*Higher Order Thinking Skills* atau HOTS) siswa kelas XI pada materi kisi difraksi.



Pengukuran difokuskan pada tiga tingkat kognitif dalam taksonomi Bloom revisi, yaitu menganalisis (C4), mengevaluasi (C5), dan mencipta (C6). Pada aspek menganalisis (C4), instrumen digunakan untuk menilai kemampuan siswa dalam mengolah data hasil praktikum, mengidentifikasi hubungan antarvariabel, serta menganalisis konsep yang berkaitan dengan fenomena kisi difraksi. Pada aspek mengevaluasi (C5), instrumen bertujuan mengukur kemampuan siswa dalam memberikan argumentasi terhadap hasil pengamatan, menilai ketepatan prosedur maupun hasil eksperimen, serta memvalidasi kesimpulan berdasarkan bukti yang diperoleh. Sementara itu, pada aspek mencipta (C6), instrumen digunakan untuk menilai kemampuan siswa dalam merumuskan solusi, mengembangkan ide baru, atau memodifikasi desain percobaan berdasarkan temuan praktikum sehingga dapat menghasilkan alternatif penyelesaian yang lebih efektif dan ilmiah.

2. Penyusunan Kisi-kisi Instrumen

Perancangan instrumen dimulai dengan penyusunan kisi-kisi yang memetakan keterkaitan antara tujuan pembelajaran, indikator pencapaian kompetensi, dan tingkat kognitif. Kisi-kisi ini berfungsi sebagai pedoman dalam penulisan butir soal atau rubrik penilaian agar tetap relevan dengan materi kisi difraksi yang telah dipelajari oleh siswa.

3. Pemilihan Format Instrumen

Berdasarkan hasil analisis kebutuhan dari praktisi (guru), instrumen dirancang dalam bentuk analisis laporan hasil praktikum. Format ini dipilih karena memungkinkan peneliti untuk melihat proses berpikir kritis siswa melalui dokumentasi data dan narasi yang mereka susun setelah melakukan percobaan. Instrumen ini dilengkapi dengan rubrik penilaian analitik untuk menjamin objektivitas dan ketepatan dalam menilai setiap komponen laporan.

4. Strategi Pengumpulan Data

Rancangan instrumen ini nantinya akan divalidasi oleh ahli materi dan ahli evaluasi untuk memastikan kelayakan konten, konstruk, dan bahasa sebelum diujicobakan pada tahap *Development*.

Fase Pengembangan, di mana instrumen sebenarnya dibuat dan diorganisasikan berdasarkan rencana desain (Bhakti et al., 2023). Mengembangkan desain instrumen evaluasi melalui uji validitas ahli (*expert judgment*) untuk mengukur validitas isi dan reliabilitas ahli kemudian diuji coba ke lapangan secara mikro. Instrumen yang dikembangkan merupakan jenis asesmen nontes yang dirancang khusus untuk mengevaluasi kemampuan berpikir tingkat tinggi (*Higher Order Thinking Skills*) siswa melalui analisis laporan hasil praktikum pada materi Kisi Difraksi.

Pada fase ini, instrumen diujicobakan dan digunakan dalam pengukuran aktual (Bhakti et al., 2023). Fase ini dilakukan dengan cara mengujicobakan instrumen pada siswa (skala besar) untuk melihat efektivitasnya dalam mengukur HOTS. Tahap implementasi merupakan langkah nyata untuk menerapkan instrumen evaluasi asesmen kognitif berbasis HOTS yang telah dinyatakan layak oleh para ahli pada tahap sebelumnya. Tujuan utama dari tahap ini adalah untuk melihat efektivitas, praktisitas, dan keterpakaian instrumen dalam situasi pembelajaran yang sebenarnya.

Fase Evaluasi melibatkan analisis data pengukuran dan penilaian efektivitas instrumen, memungkinkan penyesuaian atau perubahan dilakukan jika perlu (Bhakti et al., 2023). Evaluasi hasil produk akhir dengan menganalisis hasil uji coba untuk memperbaiki instrumen. Tahap evaluasi merupakan tahap akhir dari model pengembangan ADDIE yang



bertujuan untuk menilai kualitas keseluruhan dari instrumen evaluasi kognitif berbasis HOTS yang telah dikembangkan. Evaluasi dilakukan secara formatif dan sumatif untuk memastikan bahwa instrumen ini memenuhi kriteria validitas dan reliabilitas dalam mengukur kemampuan berpikir tingkat tinggi siswa pada materi kisi difraksi.

1. Hasil Validitas Isi

Penilaian dilakukan oleh para validator ahli (*expert judgment*) yang memiliki kepakaran di bidang pembelajaran fisika. Masukan kualitatif dari para ahli digunakan sebagai dasar revisi untuk menyempurnakan kualitas instrumen, sementara penilaian kuantitatif dianalisis menggunakan koefisien Aiken's V. Analisis ini dilakukan untuk membuktikan secara empiris bahwa instrumen penelitian yang dikembangkan telah memenuhi kriteria validitas isi yang ketat sebelum diujicobakan kepada subjek penelitian. Pengujian ini melibatkan empat orang pakar (*expert judgment*) yang meninjau relevansi setiap butir pernyataan. Hasil kesepakatan para ahli yang dihitung menggunakan formula koefisien Aiken's V dipaparkan secara rinci pada Tabel 1 berikut:

Tabel 1. Validitas Isi

No Butir	P1	P2	P3	P4	$\sum$ (r-lo)	N (C-1)	V	KET
1	4	4	3	4	11	12	0,92	Valid
2	4	4	3	4	11	12	0,92	Valid
3	4	4	4	4	12	12	1,00	Valid
4	4	4	4	4	12	12	1,00	Valid
5	4	4	4	3	11	12	0,92	Valid
6	3	4	4	4	11	12	0,92	Valid

nilai validitas $\geq 0,92$

Instrumen ini telah memenuhi syarat akurasi empiris yang ketat sebelum diujicobakan kepada subjek penelitian. Dominasi nilai 0,92 membuktikan bahwa secara substansi instrumen telah selaras dengan tujuan penelitian dan dipahami secara objektif oleh para pakar. Secara kritis, hasil ini meminimalisir resiko munculnya bias atau ambiguitas pada saat pengambilan data lapangan, sehingga penelitian ini memiliki landasan teoretis yang kokoh untuk mengklaim bahwa data yang dihasilkan nantinya benar-benar mengukur apa yang seharusnya diukur. Dengan demikian, instrumen ini secara sah dapat dinyatakan siap untuk digunakan dalam tahap penelitian selanjutnya tanpa memerlukan revisi mayor.

2. Hasil Reliabilitas Instrumen

Setelah instrumen asesmen autentik laporan praktikum dinyatakan valid secara isi dan konstruk, langkah selanjutnya adalah menguji reliabilitas instrumen. Uji reliabilitas dilakukan untuk menentukan tingkat konsistensi dan kestabilan instrumen dalam memberikan hasil penilaian. Dalam konteks asesmen autentik seperti laporan praktikum, tantangan utama adalah subjektivitas penilai. Oleh karena itu, instrumen yang reliabel harus mampu meminimalisir bias tersebut sehingga siapapun yang menggunakan instrumen ini akan memberikan skor yang relatif sama untuk kualitas laporan yang sama. Uji reliabilitas juga digunakan untuk menguji konsistensi jawaban responden berdasarkan instrumen penelitian. Semakin reliabel sebuah instrumen penelitian, semakin tinggi pula tingkat konsistensinya (Soesana et al., 2023).



Data skor mentah dan akumulasi penilaian yang menjadi dasar perhitungan koefisien reliabilitas *Intraclass Correlation Coefficient* (ICC). Setelah melakukan perhitungan diperoleh nilai $R = 0,74$ menunjukkan bahwa instrumen tersebut memiliki tingkat reliabilitas yang dapat diterima untuk kategori penelitian pendidikan. Nilai ini mengindikasikan adanya konsistensi yang cukup kuat di antara para rater dalam memberikan penilaian, sehingga subjektivitas individu tidak mendominasi skor akhir. Namun, secara statistik, munculnya angka negatif pada komponen varians dalam proses perhitungan manual mengisyaratkan adanya homogenitas yang sangat tinggi pada data. Hal ini sering terjadi ketika penilaian para validator sangat seragam sehingga variasi antar-skor menjadi sangat minim. Meskipun terdapat anomali pada nilai varians antara, hasil akhir ICC yang melampaui ambang batas moderat ($> 0,70$) memberikan jaminan bahwa instrumen memiliki objektivitas antar-rater yang baik. Dengan demikian, instrumen ini dinyatakan reliabel secara empiris, yang berarti prosedur penilaian telah memiliki keajegan yang cukup untuk meminimalisir kesalahan pengukuran (*measurement error*) saat diaplikasikan di lapangan.

3. Potret Higher Order Thinking Skill

Data yang diperoleh dari hasil penilaian laporan praktikum menggunakan instrumen yang telah dikembangkan memberikan gambaran yang komprehensif mengenai profil keterampilan berpikir tingkat tinggi (*Higher Order Thinking Skills* atau HOTS) siswa. Analisis profil ini bertujuan untuk memetakan kemampuan siswa kelas XI dalam mengintegrasikan pemahaman teoretis ke dalam aktivitas praktikum pada materi kisi difraksi. Kemampuan tersebut diukur berdasarkan tiga level kognitif dalam taksonomi Bloom revisi, yaitu menganalisis (C4), mengevaluasi (C5), dan mencipta (C6), sehingga hasil penilaian dapat menggambarkan kemampuan siswa dalam mengolah informasi, memberikan penilaian berdasarkan bukti, serta menghasilkan solusi melalui kegiatan praktikum. Profil keterampilan berpikir tingkat tinggi siswa pada tahap uji coba skala kecil disajikan pada Tabel 2.

Tabel 2. Profil Keterampilan Berpikir Tingkat Tinggi (HOTS) Siswa pada Uji Coba Skala Kecil

Kategori	Jumlah Siswa	Persentase (%)
Sangat Baik	2	9
Baik	9	41
Cukup	9	41
Kurang	2	9
Total	22	100

Berdasarkan Tabel 2, hasil uji coba skala kecil menunjukkan bahwa sebagian besar siswa berada pada kategori baik dan cukup dengan persentase gabungan sebesar 82%, sedangkan masing-masing 9% siswa berada pada kategori sangat baik dan kurang. Distribusi tersebut menunjukkan bahwa instrumen asesmen autentik yang dikembangkan mampu memetakan kemampuan berpikir tingkat tinggi siswa secara bervariasi sesuai dengan tingkat penguasaan masing-masing. Sebaran kategori yang diperoleh juga mengindikasikan bahwa instrumen memiliki kemampuan yang baik dalam membedakan tingkat kemampuan siswa pada aspek menganalisis (C4), mengevaluasi (C5), dan mencipta (C6). Hasil ini menunjukkan bahwa instrumen tidak hanya dapat digunakan untuk mengidentifikasi capaian HOTS siswa,



tetapi juga memberikan informasi mengenai variasi kemampuan kognitif peserta didik sebagai dasar penyempurnaan instrumen sebelum diterapkan pada uji lapangan.

Selanjutnya, hasil uji lapangan digunakan untuk menggambarkan profil keterampilan berpikir tingkat tinggi (*Higher Order Thinking Skills* atau HOTS) siswa berdasarkan penilaian laporan praktikum menggunakan rubrik asesmen autentik yang telah dikembangkan. Profil tersebut disajikan dalam bentuk distribusi kategori HOTS yang mencerminkan capaian siswa pada aspek menganalisis (C4), mengevaluasi (C5), dan mencipta (C6). Selain itu, hasil uji lapangan memberikan gambaran mengenai tingkat pencapaian HOTS siswa setelah instrumen diterapkan pada sampel yang lebih luas sehingga dapat menunjukkan kemampuan instrumen dalam mengidentifikasi variasi kemampuan peserta didik. Adapun hasil uji lapangan disajikan pada Tabel 3.

Tabel 3. Profil Keterampilan Berpikir Tingkat Tinggi (HOTS) Siswa pada Uji Lapangan

Kategori HOTS	Jumlah Siswa	Persentase (%)
Sangat Baik (SB)	4	6,67
Baik (B)	37	61,67
Cukup (C)	12	20,00
Kurang (K)	7	11,66
Total	60	100,00

Keterangan: Skor minimum = 50,00; skor maksimum = 96,88; rata-rata skor = 74,17.

Berdasarkan Tabel 3, hasil uji lapangan menunjukkan bahwa kemampuan *Higher Order Thinking Skills* (HOTS) siswa secara umum berada pada kategori baik dengan nilai rata-rata sebesar 74,17. Distribusi kategori menunjukkan bahwa sebagian besar siswa berada pada kategori baik (61,67%), diikuti kategori cukup (20,00%), sedangkan kategori sangat baik dan kurang masing-masing sebesar 6,67% dan 11,66%. Temuan tersebut mengindikasikan bahwa sebagian besar siswa telah mampu memenuhi indikator HOTS dalam penyusunan laporan praktikum, khususnya pada aspek menganalisis (C4), mengevaluasi (C5), dan mencipta (C6). Selain itu, sebaran kategori yang diperoleh menunjukkan bahwa instrumen asesmen autentik yang dikembangkan mampu membedakan tingkat kemampuan berpikir tingkat tinggi siswa secara objektif sekaligus memfasilitasi guru dalam melakukan penilaian terhadap kualitas laporan praktikum. Dengan demikian, instrumen ini tidak hanya berfungsi sebagai alat evaluasi hasil belajar, tetapi juga sebagai sarana untuk mengidentifikasi capaian HOTS siswa melalui kegiatan praktikum. Temuan ini sejalan dengan penelitian Bhakti et al. (2023) yang menyatakan bahwa instrumen penilaian kemampuan berpikir kritis dalam fisika dirancang untuk mengevaluasi kemampuan siswa dalam menganalisis, mengevaluasi, dan mensintesis informasi fisika, mengidentifikasi asumsi, menghubungkan konsep, membuat keputusan berdasarkan bukti, serta memecahkan masalah secara kritis.

4. Kepraktisan Instrumen

Sebuah instrumen asesmen yang ideal tidak hanya dituntut memiliki akurasi dan konsistensi yang tinggi, tetapi juga harus memenuhi aspek kepraktisan dalam implementasinya. Kepraktisan merupakan parameter penting dalam penelitian pengembangan



untuk melihat sejauh mana instrumen yang telah disusun dapat digunakan secara efektif, efisien, dan mudah dioperasikan oleh pendidik dalam situasi pembelajaran yang nyata.

Berdasarkan angket respon guru diperoleh bahwa asesmen autentik untuk mengukur HOTS pada laporan praktikum memperoleh rata-rata persentase kepraktisan sebesar 87,50% dengan kategori "Sangat Praktis". Hasil tersebut menunjukkan bahwa instrumen mudah digunakan oleh guru dalam proses penilaian laporan praktikum. Aspek kejelasan petunjuk memperoleh persentase tertinggi yaitu sebesar 90,63%. Hal ini menunjukkan bahwa rumusan tujuan instrumen, petunjuk penggunaan, serta langkah-langkah penilaian dinilai jelas dan mudah dipahami oleh pengguna. Guru dapat memahami prosedur penggunaan instrumen tanpa mengalami kesulitan yang berarti.

Hasil analisis kepraktisan instrumen asesmen autentik berdasarkan persepsi siswa sebagai pengguna utama di lapangan. Pengukuran kepraktisan ini dilakukan melalui penyebaran angket respon siswa setelah mereka menggunakan instrumen penilaian dalam menyusun laporan praktikum fisika pada materi Kisi Difraksi. Respon siswa terhadap instrumen asesmen autentik pada laporan praktikum fisika materi Kisi Difraksi berada pada kategori "Sangat Baik" atau "Sangat Praktis". Mayoritas siswa menyatakan sangat setuju bahwa instrumen ini membuat proses penilaian menjadi lebih adil, transparan, dan memotivasi mereka untuk menyusun laporan dengan lebih baik.

Pembahasan

Hasil penelitian menunjukkan bahwa pengembangan asesmen autentik pada laporan praktikum untuk mengukur keterampilan berfikir tingkat tinggi siswa SMA berhasil memenuhi tiga kriteria utama, yaitu valid, reliabel dan praktis. Perhatikan kembali Tabel 1 menunjukkan tingkat validitas isi yang sangat kuat, namun perlu dicermati melalui distribusi penilaian para ahli. Nilai koefisien Aiken's V yang mencapai rentang 0,92 hingga 1,00 mengindikasikan bahwa instrumen tersebut memiliki keterwakilan konsep yang hampir sempurna di mata para rater. Hasil ini relevan dengan penelitian Yunitasari (2019) yang menyatakan bahwa validitas isi dengan perolehan koefisien Aiken's antara 0,70–1,00, di mana instrumen dinyatakan valid karena memenuhi validitas isi dan empiris. Temuan tersebut juga didukung oleh An Nabil et al. (2022) yang menjelaskan bahwa validitas isi instrumen yang dianalisis menggunakan indeks Aiken's V dapat digunakan untuk menentukan kelayakan butir instrumen, dengan butir yang memenuhi kriteria nilai Aiken's V dinyatakan valid. Hasil penelitian Bashooir dan Supahar (2018) menunjukkan bahwa lembar pengamatan berupa rubrik penskoran dan penilaian diri terbukti valid dengan koefisien V Aiken 0,75. Temuan tersebut diperkuat oleh Kania et al. (2024) yang menegaskan bahwa indeks Aiken's V merupakan bukti validitas isi yang kuat berdasarkan penilaian para ahli, sehingga instrumen yang memenuhi kriteria Aiken's V dapat dinyatakan valid untuk digunakan dalam proses asesmen. Suhardi (2022) juga menyatakan Rata-rata V hitung Paket 1 0,86 (Sangat Tinggi) perangkat instrumentasi pengukuran validasi pakar menggunakan Formula Aiken's V untuk mengukur validasi konten.

Analisis pada butir 3 dan 4, tercapai konsensus absolut ($V = 1$), yang berarti seluruh ahli sepakat tanpa keraguan mengenai relevansi butir tersebut. Sementara itu, pada butir lainnya (1, 2, 5, dan 6) dengan skor 0,92, terdapat variasi penilaian minor di mana salah satu ahli memberikan skor sedikit di bawah nilai maksimal (skor 3 dari skala 4). Meskipun demikian, deviasi ini tidak signifikan secara statistik untuk menggugurkan validitas butir,



karena nilai tersebut masih jauh melampaui standar umum validitas isi yang biasanya mematok ambang batas minimal di angka 0,70 hingga 0,80.

Reliabilitas yang tinggi ditunjukkan dengan nilai r_{xx} mendekati angka 1. Berdasarkan hasil uji reliabilitas oleh empat validator didapatkan nilai $R=0.74$ hal ini sejalan dengan penemuan oleh Soesana et al. (2023) nilai reliabilitas akan dianggap cukup memuaskan jika memenuhi $R_{xx} \geq 0.700$. Instrumen penelitian akan dinyatakan memiliki tingkat reliabilitas yang memuaskan dan dapat dipercaya apabila koefisien reliabilitas R_{xx} yang dihasilkan menunjukkan nilai mendekati angka 1, dengan kriteria ambang batas minimum sebesar $R_{xx} \geq 0,700$. Penelitian Rahmawati et al. (2022) nilai reliabilitas memiliki nilai average measure 0,778. Hal ini dapat dikatakan bahwa nilai ICC (*Interclass Correlation Coefisient*) yang dinilai oleh 3 rater dikategorikan sedang atau *moderate reliability*.

Analisis hasil uji coba terbatas menunjukkan bahwa sebagian besar siswa telah memiliki kemampuan berpikir tingkat tinggi pada kategori baik dan cukup. Sebanyak 18 dari 22 siswa berada pada kedua kategori tersebut, yang mengindikasikan bahwa mayoritas siswa telah mampu menganalisis fenomena gelombang cahaya, mengolah data hasil pengamatan, serta menghubungkan konsep teoritis dengan hasil praktikum pada materi kisi difraksi. Meskipun demikian, kemampuan tersebut masih perlu ditingkatkan agar siswa mampu mencapai level berpikir yang lebih kompleks, terutama dalam mengevaluasi hasil eksperimen secara kritis dan menyusun argumentasi yang didukung oleh bukti ilmiah.

Berdasarkan hasil pengamatan selama pelaksanaan uji coba terbatas, siswa yang berada pada kategori cukup dan kurang umumnya mengalami kesulitan pada bagian pembahasan dan penarikan kesimpulan laporan praktikum. Kesulitan tersebut menunjukkan bahwa siswa belum sepenuhnya mampu mengembangkan gagasan, memberikan interpretasi yang mendalam terhadap hasil pengamatan, maupun merumuskan solusi berdasarkan temuan praktikum. Kondisi ini berkaitan dengan aspek mencipta (C6) dalam taksonomi Bloom revisi yang merupakan tingkat kognitif tertinggi dan memerlukan pembiasaan melalui aktivitas pembelajaran serta asesmen yang menuntut kemampuan berpikir tingkat tinggi. Temuan ini sejalan dengan pendapat Lubbe et al. (2025) yang menyatakan bahwa peninjauan kembali taksonomi Bloom, khususnya pada tingkat evaluasi dan kreativitas, dapat membantu pendidik mengembangkan penilaian yang lebih efektif dalam mendorong pemikiran kritis serta mendukung siswa menjadi pembelajar mandiri melalui kegiatan praktikum.

Hasil uji lapangan menunjukkan bahwa siswa relatif mampu memahami konsep dan menyajikan hasil pengamatan dengan baik. Namun demikian, aspek pembahasan dan simpulan memperoleh skor yang lebih rendah dibandingkan aspek lainnya. Rendahnya capaian pada aspek pembahasan dan simpulan menunjukkan bahwa siswa masih mengalami kesulitan dalam melakukan interpretasi mendalam dan menyusun argumentasi ilmiah. Kondisi ini mengindikasikan bahwa pembelajaran praktikum di sekolah masih cenderung berfokus pada prosedur dan hasil akhir, sementara kemampuan reflektif dan evaluatif belum berkembang secara optimal. Temuan tersebut sejalan dengan karakteristik HOTS yang menuntut kemampuan berpikir kompleks, seperti menganalisis hubungan antar konsep, mengevaluasi hasil, serta menghasilkan kesimpulan berdasarkan bukti empiris. Oleh karena itu, diperlukan pembelajaran yang lebih mendorong siswa untuk berdiskusi, merefleksikan hasil eksperimen, dan menyusun argumentasi ilmiah secara lebih mendalam. Selain itu, hasil penelitian menunjukkan bahwa asesmen autentik yang dikembangkan tidak hanya mengukur aspek kognitif, tetapi juga aspek keterampilan proses dan sikap ilmiah. Hal ini terlihat dari adanya penilaian pada aspek etika penulisan dan tanggung jawab akademik siswa.



Dengan demikian, asesmen autentik pada laporan praktikum dapat menjadi alternatif instrumen penilaian yang efektif untuk mengukur keterampilan berpikir tingkat tinggi siswa SMA. Instrumen yang dikembangkan mampu memberikan gambaran yang lebih komprehensif mengenai kemampuan siswa dibandingkan penilaian konvensional yang hanya berfokus pada hasil tes tertulis. Hal ini sejalan dengan penelitian Sutadji et al. (2021) bahwa penilaian autentik memiliki keunggulan karena memberikan gambaran yang sebenarnya tentang kondisi belajar siswa serta memberikan informasi yang lebih komprehensif mengenai kekuatan dan kelemahan siswa. Melalui penerapan asesmen autentik, guru juga dapat mengidentifikasi keterampilan proses sains (KPS) terintegrasi yang dimiliki siswa secara lebih menyeluruh berdasarkan kinerja nyata selama proses pembelajaran.

Hasil uji kepraktisan berdasarkan angket respon guru menunjukkan bahwa instrumen asesmen autentik yang dikembangkan memperoleh persentase sebesar 87,50% pada aspek pelaksanaan dengan kategori sangat praktis. Hasil tersebut menunjukkan bahwa instrumen mudah dipahami oleh guru sebagai pengguna, mudah diterapkan dalam proses penilaian, efisien dari segi waktu pelaksanaan, serta mampu membantu guru dalam menilai kemampuan *Higher Order Thinking Skills* (HOTS) siswa. Ditinjau dari butir-butir pernyataan pada angket, kepraktisan instrumen tercermin melalui dua aspek utama, yaitu kejelasan petunjuk dan pelaksanaan. Instrumen memiliki petunjuk penggunaan yang jelas, sistematis, dan mudah dipahami sehingga memudahkan guru dalam memahami tujuan serta langkah-langkah penggunaannya. Selain itu, instrumen membantu guru mengidentifikasi kemampuan HOTS pada aspek menganalisis (C4), mengevaluasi (C5), dan mencipta (C6), memberikan penilaian secara lebih objektif, menyusun umpan balik yang spesifik, serta mempercepat proses penilaian dan konversi nilai akhir. Tingginya tingkat kepraktisan tersebut menunjukkan bahwa instrumen telah memenuhi salah satu karakteristik instrumen yang berkualitas, yaitu mudah digunakan oleh praktisi tanpa memerlukan pelatihan yang rumit, sehingga layak diterapkan dalam pembelajaran fisika di kelas.

Berdasarkan hasil pengolahan data angket respon siswa, diperoleh temuan bahwa instrumen asesmen autentik mendapat tanggapan yang sangat positif dari peserta didik. Pada aspek peningkatan keterampilan berpikir tingkat tinggi (*Higher Order Thinking Skills*), siswa memberikan rata-rata skor sebesar 3,42 dari skala 4,00 yang menunjukkan bahwa instrumen mampu mendorong mereka untuk berpikir lebih kritis melalui proses analisis, evaluasi, dan penyusunan laporan praktikum, bukan sekadar mengikuti prosedur praktikum. Selain itu, aspek transparansi dan kejelasan penilaian memperoleh skor tertinggi, yaitu 3,67 pada indikator pemahaman terhadap aspek yang dinilai dan harapan penggunaan instrumen pada pembelajaran selanjutnya. Temuan ini menunjukkan bahwa instrumen memberikan panduan yang jelas bagi siswa dalam menyusun laporan praktikum sesuai dengan kriteria penilaian. Pada aspek kemampuan analisis dan evaluasi, siswa juga memberikan respon positif dengan rata-rata skor 3,50, yang mengindikasikan bahwa instrumen membantu mereka meningkatkan kemampuan menganalisis hasil praktikum sekaligus mengevaluasi kesalahan atau kekurangan laporan secara mandiri. Hasil tersebut menunjukkan bahwa asesmen autentik yang dikembangkan tidak hanya berfungsi sebagai alat penilaian, tetapi juga mendukung perkembangan keterampilan metakognitif dan kemampuan berpikir tingkat tinggi siswa selama proses pembelajaran.

Berdasarkan keseluruhan hasil analisis, dapat disimpulkan bahwa asesmen autentik pada laporan praktikum untuk mengukur keterampilan berpikir tingkat tinggi siswa tergolong sangat layak digunakan dalam proses pembelajaran praktikum. Instrumen ini terbukti valid,



reliabel, sangat praktis, dan efektif dalam mengukur keterampilan berfikir tingkat tinggi siswa. Dengan demikian, instrumen ini tidak hanya berfungsi sebagai alat ukur akan tetapi berfungsi sebagai gambaran berfikir tingkat tinggi siswa sehingga dapat mendorong siswa untuk meningkatkan kompetensi kognitif dalam membuat laporan praktikum. Hal ini sejalan dengan pendapat Bhakti et al. (2023) bahwa Pengukuran kemampuan berpikir kritis dalam pembelajaran fisika harus dilakukan dengan menggunakan alat atau instrumen pengukuran yang tepat, mengingat pentingnya pengukuran kemampuan berpikir kritis dalam fisika

KESIMPULAN

Penelitian ini berhasil mengembangkan instrumen asesmen autentik berupa rubrik analitik untuk menilai laporan praktikum fisika yang terbukti valid, reliabel, dan praktis digunakan dalam pembelajaran. Instrumen yang dikembangkan mampu mengukur keterampilan berpikir tingkat tinggi siswa pada aspek analisis, evaluasi, dan kreasi melalui kegiatan praktikum serta memberikan informasi yang lebih komprehensif mengenai capaian kemampuan siswa dibandingkan penilaian yang hanya berfokus pada hasil akhir. Implementasi instrumen menunjukkan bahwa sebagian besar siswa telah memiliki keterampilan berpikir tingkat tinggi dalam kategori baik, meskipun kemampuan pada aspek interpretasi hasil dan penyusunan simpulan masih memerlukan penguatan. Temuan ini menunjukkan bahwa instrumen asesmen autentik yang dikembangkan tidak hanya layak digunakan sebagai alat penilaian, tetapi juga dapat memberikan umpan balik yang bermanfaat bagi guru dalam mengidentifikasi aspek keterampilan berpikir tingkat tinggi yang masih perlu ditingkatkan.

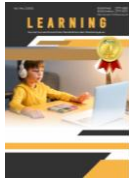
Penelitian selanjutnya disarankan untuk mengimplementasikan instrumen ini pada materi fisika lain, jenjang pendidikan yang berbeda, serta melibatkan jumlah sampel yang lebih luas agar diperoleh bukti empiris yang lebih kuat mengenai efektivitas dan generalisasi penggunaannya. Selain itu, pengembangan asesmen autentik dapat dipadukan dengan platform digital untuk meningkatkan efisiensi proses penilaian dan mendukung pembelajaran yang berorientasi pada keterampilan abad ke-21. Penelitian berikutnya juga dapat menguji pengaruh penggunaan instrumen ini terhadap peningkatan keterampilan berpikir tingkat tinggi siswa melalui desain penelitian eksperimen atau kuasi eksperimen. Dengan demikian, instrumen yang dikembangkan diharapkan dapat menjadi salah satu alternatif penilaian autentik yang mendukung implementasi pembelajaran fisika yang berorientasi pada pengembangan kompetensi abad ke-21.

DAFTAR PUSTAKA

- An Nabil, N. R., Wulandari, I., Yamtinah, S., Ariani, S. R. D., & Ulfa, M. (2022). Analisis indeks Aiken untuk mengetahui validitas isi instrumen asesmen kompetensi minimum berbasis konteks sains kimia. *PAEDAGOGIA: Jurnal Penelitian Pendidikan*, 25(2), 184–191. <https://doi.org/10.20961/paedagogia.v25i2.64566>
- Arafah, K., Amin, B. D., Sari, S. S., & Hakim, A. (2021). The development of higher-order thinking skills (HOTS) instrument assessment in physics study. *Journal of Physics: Conference Series*, 1899(1), Article 012140. <https://doi.org/10.1088/1742-6596/1899/1/012140>
- Bashooir, K., & Supahar, S. (2018). Validitas dan reliabilitas instrumen asesmen kinerja literasi sains pelajaran Fisika berbasis STEM. *Jurnal penelitian dan evaluasi pendidikan*, 22(2), 219-230. <https://doi.org/10.21831/pep.v22i2.19590>



- Bhakti, Y. B., Arthur, R., & Supriyati, Y. (2023). Development of an assessment instrument for critical thinking skills in physics: A systematic literature review. *Journal of Physics: Conference Series*, 2596(1), Article 012067. <https://doi.org/10.1088/1742-6596/2596/1/012067>
- Gallardo, K. (2020). Competency-based assessment and the use of performance-based evaluation rubrics in higher education: Challenges towards the next decade. *Problems of Education in the 21st Century*, 78(1), 61–79. <https://doi.org/10.33225/pec/20.78.61>
- Gradini, E. (2022). Development of rubric of higher order thinking skills assessment on mathematics learning. *JTAM (Jurnal Teori dan Aplikasi Matematika)*, 6(1), 61–74. <https://doi.org/10.31764/jtam.v6i1.5342>
- Herlinawati, H., Marwa, M., Ismail, N., Junaidi, J., Liza, L. O., & Situmorang, D. D. B. (2024). The integration of 21st century skills in the curriculum of education. *Heliyon*, 10(15), Article e35148. <https://doi.org/10.1016/j.heliyon.2024.e35148>
- Jannah, D. R. M., & Istianah, F. (2021). Pengembangan media pembelajaran pop-up book pada materi siklus air kelas V sekolah dasar. *Jurnal Penelitian Pendidikan Guru Sekolah Dasar*, 9(3), 1824–1836. <https://ejournal.unesa.ac.id/index.php/jurnal-penelitian-pgsd/article/view/39718>
- Kania, N., Kusumah, Y. S., Dahlan, J. A., Nurlaelah, E., Gürbüz, F., & Bonyah, E. (2024). Constructing and providing content validity evidence through the Aiken's V index based on the experts' judgments of the instrument to measure mathematical problem-solving skills. *REID (Research and Evaluation in Education)*, 10(1), 64–79. <https://doi.org/10.21831/reid.v10i1.71032>
- Kousloglou, M., Petridou, E., Molohidis, A., & Hatzikraniotis, E. (2023). Assessing students' awareness of 4Cs skills after mobile-technology-supported inquiry-based learning. *Sustainability*, 15(8), Article 6725. <https://doi.org/10.3390/su15086725>
- Liu, J., Liu, Z., Wang, C., Xu, Y., Chen, J., & Cheng, Y. (2024). K-12 students' higher-order thinking skills: Conceptualization, components, and evaluation indicators. *Thinking Skills and Creativity*, 52, Article 101551. <https://doi.org/10.1016/j.tsc.2024.101551>
- Lubbe, A., Marais, E., & Kruger, D. (2025). Cultivating independent thinkers: The triad of artificial intelligence, Bloom's taxonomy and critical thinking in assessment pedagogy. *Education and Information Technologies*, 30, 17589–17622. <https://doi.org/10.1007/s10639-025-13476-x>
- Makmuroh, U., & Pratama, H. (2022). The implementation of performance-based assessment technique to assess students' analysing skill in English learning. *English Education Journal*, 12(4), 578–586. <https://doi.org/10.15294/eej.v12i4.65365>
- Nurhayati, N. (2024). Analysis of higher order thinking skills indicators on physics student worksheet class XI semester II. *Physics Learning and Education*, 2(2), Article 70. <https://doi.org/10.24036/ple.v2i2.70>
- Rahmawati, R., Wulan, A. R., & Kusnadi, K. (2022). Pengembangan asesmen kinerja keterampilan *inquiry laboratory* pada permasalahan biologi abad ke-21. *Jurnal Paedagogy: Jurnal Penelitian dan Pengembangan Pendidikan*, 9(4), 763–771. <https://doi.org/10.33394/jp.v9i4.5667>
- Soesana, A., Subakti, H., Karwanto, K., Fitri, A., Kuswandi, S., Sastri, L., Falani, I., Aswan, N., Hasibuan, F. A., & Lestari, H. (2023). *Metodologi penelitian kuantitatif*. Yayasan Kita Menulis.



- Suhady, W., Roza, Y., & Maimunah, M. (2020). Pengembangan soal untuk mengukur *higher order thinking skill* (HOTS) siswa. *Jurnal Gantang*, 5(2), 143–150. <https://doi.org/10.31629/jg.v5i2.2518>
- Suhardi, I. (2022). Perangkat instrumen pengembangan paket soal jenis pilihan ganda menggunakan pengukuran validitas konten Formula Aiken's V. *Jurnal Pendidikan Tambusai*, 6(1), 4158–4170. <https://jptam.org/index.php/jptam/article/view/3519>
- Sutadji, E., Susilo, H., Wibawa, A. P., Jabari, N. A. M., & Rohmad, S. N. (2021). Authentic assessment implementation in natural and social science. *Education Sciences*, 11(9), Article 534. <https://doi.org/10.3390/educsci11090534>
- Wati, A., Qadar, R., & Sulaeman, N. F. (2024). High school physics assessments in Bontang City: Fostering higher-order thinking. *Vidya Karya*, 39(1), 63–69. <https://doi.org/10.20527/jvk.v39i1.17675>
- Xu, D., & Shen, Q. (2025). The practical application of performance-based assessment in primary school Chinese teaching. *Education Reform and Development*, 7(1), 118–123. <https://doi.org/10.26689/erd.v7i1.9608>
- Yunitasari, I. (2019). *Pengembangan instrumen penilaian authentic untuk menilai keterampilan observing, inferring, dan predicting pada pembelajaran IPA* (Tesis magister, Universitas Negeri Yogyakarta). Repositori Universitas Negeri Yogyakarta. <https://eprints.uny.ac.id/69114>