

Analisis Validitas dan Reliabilitas Instrumen Soal IPA Berbasis HOTS dan Literasi Sains Materi Pewarisan Sifat

Muhammad Yusuf¹, Nova E. Ntobuo², Nyoman Sriastuti^{3*}, Nur Azizah Dj. Junus⁴, Risna H. Panyo⁵

^{1,2,3,4}Departement of Natural Science, Universitas Negeri Gorontalo, Gorontalo, Indonesia

Email: ¹muhammad.yusuf@ung.ac.id, ²novantobuo@ung.ac.id

^{3*}nyoman_s2ilmupengetahuanalam@mahasiswa.ung.ac.id,

⁴nur_s2ilmupengetahuanalam@mahasiswa.ung.ac.id,

⁵risna_s2ilmupengetahuanalam@mahasiswa.ung.ac.id

Received: 8 Mei 2026

Revised: 19 Mei 2026

Accepted: 11 Juni 2026

Abstract

This study analyzes the quality of a science test instrument on inheritance, based on Higher Order Thinking Skills (HOTS) and scientific literacy, in terms of validity, reliability, difficulty level, discrimination index, and distractor effectiveness. A descriptive quantitative approach with an ex post facto item analysis design was used. The participants were 28 students, and the instrument consisted of 20 multiple-choice items developed according to HOTS and scientific literacy indicators. Validity was tested using point-biserial correlation, reliability using Kuder-Richardson (KR-20), and item difficulty and discrimination indices were analyzed through correct response proportions and comparisons between upper and lower groups. Results showed that 85% of the items were valid and 15% invalid. The reliability coefficient was 0.8827, indicating high consistency. In terms of discrimination, 60% of items were very good, 25% good, 10% fair, and 5% poor. Most items were moderately difficult (90%), and 10% were difficult, with no easy items. Overall, the instrument demonstrates good quality and is suitable for learning evaluation, although some items require revision to further optimize measurement accuracy.

Keywords: validity, reliability, discrimination index, difficulty level, HOTS, scientific literacy

Abstrak

Penelitian ini dilakukan dengan tujuan untuk menganalisis kualitas instrumen soal IPA berbasis Higher Order Thinking Skills (HOTS) dan literasi sains dalam materi pewarisan sifat ditinjau dari aspek validitas, reliabilitas, tingkat kesukaran dan daya pembeda. Penelitian menggunakan pendekatan kuantitatif deskriptif dengan desain analisis butir soal ex post facto. Subjek penelitian terdiri atas 28 siswa, dengan instrumen berupa 20 butir soal pilihan ganda berbasis HOTS dan literasi sains. Uji validitas dilakukan menggunakan korelasi point biserial, reliabilitas menggunakan rumus Kuder-Richardson (KR-20), serta analisis tingkat kesukaran serta kemampuan pembeda berdasarkan jumlah jawaban benar dan perbandingan kelompok atas dan bawah. Hasil penelitian menunjukkan bahwa 85% butir soal dinyatakan valid, sedangkan 15% tidak valid. Koefisien reliabilitas sebesar 0,8827 tergolong tinggi, sehingga instrumen dianggap konsisten serta andal. Dari aspek daya pembeda, 60% butir berada pada kategori sangat baik, 25% baik, 10% cukup, dan 5% kurang baik. Sementara itu, tingkat kesukaran didominasi kategori sedang (90%) dan sukar (10%), tanpa adanya soal kategori mudah. Secara keseluruhan, instrumen soal IPA berbasis HOTS dan literasi sains pada materi pewarisan sifat menunjukan kualitas yang baik serta layak digunakan sebagai alat evaluasi pembelajaran, walaupun terdapat beberapa butir perlu direvisi untuk meningkatkan mutu pengukuran secara optimal.

Kata-kata kunci: validitas, reliabilitas, daya pembeda, tingkat kesukaran, HOTS, literasi sains.

PENDAHULUAN

Di era globalisasi dan pesatnya perkembangan ilmu pengetahuan serta teknologi, pendidikan menjadi faktor utama dalam membentuk sumber daya manusia yang kritis, adaptif, serta kompetitif. Mutu sumber daya manusia tidak hanya bergantung pada penguasaan konsep, melainkan juga kemampuan menerapkan pengetahuan dalam kehidupan nyata. Dalam konteks ini, literasi sains menjadi kompetensi utama karena memungkinkan individu memahami fenomena ilmiah, menilai informasi berbasis bukti, serta mengambil keputusan secara rasional (Ulfian et al., 2025). Literasi sains mencakup kemampuan berpikir ilmiah, penalaran kritis, dan mengaitkan sains dengan konteks sosial (Mijaya et al., 2021).

Kemampuan tersebut selaras dengan keterampilan berpikir tingkat tinggi (HOTS) yang mencakup analisis, evaluasi, dan kreasi dalam taksonomi Bloom revisi (Nissa & Dimas, 2022). Soal berbasis HOTS dirancang untuk mengukur pemikiran kompleks melalui stimulus kontekstual dan pemecahan masalah, bukan sekadar hafalan (Ariska et al., 2024; Indana et al., 2022). Dengan demikian, integrasi HOTS dan literasi sains pada asesmen penting untuk mendukung tujuan pembelajaran IPA secara menyeluruh (Cahyani et al., 2025).

Dalam konteks asesmen IPA, literasi sains mencakup kemampuan memahami dan menerapkan pengetahuan ilmiah dalam kehidupan nyata, sedangkan HOTS meliputi kemampuan menganalisis, mengevaluasi, dan menciptakan yang penting bagi pembelajaran sepanjang hayat (Fatonah et al., 2026). Kolaborasi HOTS dan literasi sains dalam asesmen tercermin pada item soal yang mengintegrasikan kemampuan memahami fenomena ilmiah, menerapkan konsep sains, menganalisis informasi, mengevaluasi bukti, serta memecahkan masalah kontekstual. Integrasi tersebut memungkinkan instrumen memberikan gambaran yang lebih komprehensif mengenai kemampuan siswa dalam memahami, menerapkan, dan menggunakan konsep sains untuk menyelesaikan permasalahan yang berkaitan dengan materi pewarisan sifat.

IPA menuntut pemahaman terhadap konsep, keterampilan proses, serta kemampuan berpikir tingkat tinggi (Adawiyah et al., 2024; Prananda et al., 2021). Walaupun demikian, kemampuan HOTS siswa di Indonesia tergolong rendah karena pembelajaran dan penilaian cenderung berfokus pada hafalan (Dewi et al., 2024). Keterbatasan instrumen yang mampu mengukur HOTS dan literasi sains secara tepat juga menjadi tantangan (Miladanta et al., 2024).

Instrumen penilaian perlu memenuhi syarat validitas dan reliabilitas sehingga hasilnya dapat diandalkan (Dianova & Anwar, 2024). Validitas berkaitan dengan ketepatan mengukur konstruk, sedangkan reliabilitas dengan konsistensi hasil (Sudrajat, 2024). Selain itu, analisis item soal yang mencakup tingkat kesukaran dan daya pembeda diperlukan untuk menjamin setiap soal berfungsi dengan baik (Fajariani & Hanum, 2025). Melalui analisis tersebut, pendidik dapat menentukan soal yang dipertahankan, direvisi, atau tidak digunakan (Mustaqim & Sulisti, 2024). Walaupun soal buatan guru lebih banyak digunakan (Nurhalimah et al., 2022), kualitasnya sering belum dianalisis secara memadai.

Materi pewarisan sifat menuntut kemampuan menganalisis pola persilangan, menginterpretasikan data genetika, dan mengaitkan konsep dengan kehidupan sehari-hari, sehingga sesuai untuk dikembangkan dalam bentuk soal berbasis HOTS dan literasi sains. Oleh karena itu, analisis validitas dan reliabilitas instrumen pada materi ini penting untuk memastikan kelayakan dan konsistensi pengukuran. Penelitian ini dilakukan untuk mengkaji validitas, reliabilitas, tingkat kesukaran, serta daya pembeda item soal IPA berbasis HOTS dan literasi sains pada materi pewarisan sifat guna menentukan kelayakan instrumen dalam mengukur kemampuan siswa secara akurat dan konsisten.

METODE PENELITIAN

Penelitian ini dilakukan melalui pendekatan kuantitatif deskriptif dengan desain *ex post facto* untuk mengkaji kualitas soal IPA berbasis HOTS dan literasi sains pada materi pewarisan sifat. Penilaian difokuskan pada aspek validitas, reliabilitas, tingkat kesukaran

serta daya pembeda dengan memanfaatkan data hasil evaluasi yang telah tersedia. Subjek penelitian berjumlah 28 siswa yang seluruhnya dijadikan sampel (sampling jenuh). Instrumen berupa 20 soal pilihan ganda yang disusun berdasarkan indikator HOTS dan literasi sains, dengan sistem penilaian dikotomik (skor 1 untuk benar dan 0 untuk salah). Data diperoleh melalui dokumentasi dan dianalisis secara kuantitatif menggunakan uji validitas point biserial, reliabilitas KR-20, serta analisis tingkat kesukaran, daya pembeda, dan efektivitas pengecoh berdasarkan kriteria yang telah ditetapkan.

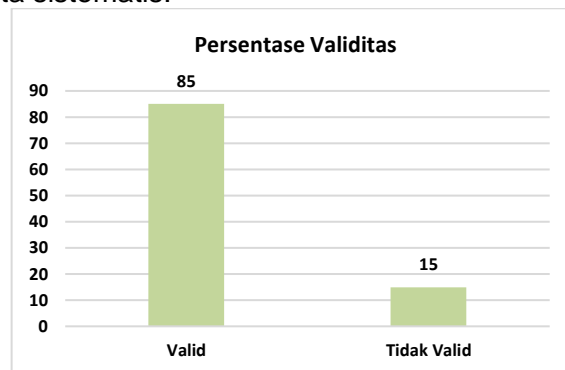
HASIL DAN PEMBAHASAN **Uji Validitas**

Secara teoretis, uji validitas bertujuan menguji konsistensi internal pada butir soal dan hubungannya dengan skor total (Dianova & Anwar, 2024). Pengujian ini dilakukan memastikan bahwa setiap item mengukur konstruk yang sejalan dengan keseluruhan instrumen. Siregar et al. (2024) menyatakan bahwa validitas adalah prosedur untuk menguji ketepatan isi instrumen dalam penelitian. Dengan demikian, validitas menjamin instrumen tidak menyimpang dari tujuan pengukuran. Ketaren et al. (2024) menegaskan bahwa validitas menunjukkan tingkat keakuratan alat ukur terhadap karakteristik objek yang diukur. Sejalan dengan itu, validitas dipahami sebagai tingkat keakuratan instrumen dalam mengukur hal yang seharusnya diukur (Namira et al., 2025). Instrumen yang tidak valid dapat menghasilkan data yang bias dan kurang akurat (Rizki et al., 2024), sehingga tidak mencerminkan kemampuan siswa secara nyata (Rachmawati & Pradana, 2025). Berikut tabel hasil pengujian validitas pada tingkat signifikansi 5% menggunakan nilai r tabel = 0,3739.

Tabel 1. Hasil Uji Validitas Butir Soal

No.	Nomor Soal	Keterangan
1	1, 2, 3, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 19, 20	Valid
2	4, 5, 17	Tidak Valid

Berdasarkan pengujian validitas, dari 20 butir soal, 17 butir dinyatakan valid sedangkan 3 butir lainnya tidak valid pada tingkat signifikansi 5% dengan nilai r tabel 0,3739. Soal yang menunjukkan tingkat korelasi $> 0,3739$ dan valid adalah nomor 1, 2, 3, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 19, dan 20. Sedangkan, soal nomor 4, 5, dan 17 memiliki korelasi $< 0,3739$ sehingga dikategorikan tidak valid. Data hasil uji validitas selanjutnya disajikan dalam bentuk persentase untuk menunjukkan distribusi item soal yang tergolong valid dan tidak valid secara lebih jelas serta sistematis.



Gambar 1. Persentase Uji Validitas

Persentase validitas 85% menunjukkan bahwa instrumen secara umum telah memenuhi kelayakan empiris dan memiliki ketepatan pengukuran yang baik. Dominannya soal valid (85%) menunjukkan bahwa penyusunan kisi-kisi dan integrasi HOTS serta literasi sains telah selaras dengan tujuan pembelajaran. Asesmen literasi sains yang baik mengukur

kemampuan menjelaskan fenomena, menginterpretasi bukti, dan mengaitkan konsep genetika dengan konteks nyata (Rachmawati & Pradana, 2025).

Adapun soal nomor 4, 5, dan 17 perlu direvisi karena kemungkinan terdapat ketidaksesuaian indikator, redaksi kurang jelas, atau tuntutan kognitif belum mencapai HOTS. Jika indikator menuntut analisis (C4) tetapi soal hanya mengukur ingatan (C1–C2), maka korelasinya rendah dan dinyatakan tidak valid (Namira et al., 2025). Stimulus yang kurang kontekstual atau opsi jawaban yang tidak berfungsi juga dapat menurunkan validitas (Rizki et al., 2024).

Dengan demikian instrumen IPA berbasis HOTS dan literasi sains pada materi pewarisan sifat telah memiliki tingkat ketepatan yang baik, namun revisi pada butir 4, 5, dan 17 tetap diperlukan agar seluruh soal merepresentasikan kemampuan berpikir tingkat tinggi secara akurat dan konsisten.

Reliabilitas

Reliabilitas berkaitan dengan tingkat keandalan terhadap hasil pengukuran. Instrumen dinyatakan reliabel jika menghasilkan data yang relatif konsisten saat digunakan berulang kali dalam kondisi serupa (Siregar et al., 2024). Dianova & Anwar (2024) menjelaskan bahwa reliabilitas menunjuk pada ketetapan, konsistensi, dan stabilitas instrumen dalam pengukuran. Dalam penelitian ini digunakan rumus Kuder-Richardson (K-R 20), yang relevan untuk soal objektif dengan penskoran dikotomi (benar–salah).

Berdasarkan uji validitas yang dilakukan sebelumnya, terdapat 17 item soal valid dan 3 item soal lainnya dinilai tidak valid. Pada uji reliabilitas yang akan digunakan hanya 17 item soal valid, sementara 3 item soal tidak valid tidak akan disertakan. Berikut hasil uji reliabilitas disajikan dalam Tabel 2.

Tabel 2. Hasil Uji Reliabilitas

Nilai	Keterangan
0,882764781	Reliabilitas Tinggi

Hasil pengujian reliabilitas dengan rumus Kuder-Richardson (K-R 20) menunjukkan bahwa koefisien reliabilitas (r_{11}) sebesar 0,8827 dengan kategori reliabilitas tinggi. Nilai yang mendekati 1 menunjukkan bahwa instrumen memiliki tingkat internal yang sangat baik dan item-item soal saling berkaitan dalam mengukur aspek yang serupa secara konsisten.

Ketaren et al. (2024) menyatakan bahwa reliabilitas menilai konsistensi metode ukur; jika hasil pengukuran tetap serupa dalam kondisi yang sama, maka instrumen tersebut andal. Koefisien reliabilitas menunjukkan tingkat konsistensi tersebut. Berdasarkan kriteria Azzahra et al. (2024), koefisien pada kategori tinggi menandakan instrumen sangat layak digunakan. Dengan nilai r_{11} sebesar 0,8827, instrumen ini memiliki kestabilan dan konsistensi pengukuran yang kuat.

Dalam asesmen berbasis HOTS dan literasi sains, reliabilitas sangat penting karena kemampuan seperti menganalisis persilangan gen, mengevaluasi genotipe, dan menginterpretasi rasio fenotipe memerlukan instrumen yang konsisten. Instrumen yang tidak reliabel dapat menghasilkan skor yang tidak stabil dan kurang mencerminkan kemampuan siswa (Rachmawati & Pradana, 2025).

Tingginya koefisien reliabilitas mengindikasikan bahwa sebagian besar item soal yang valid dapat mengukur keterampilan berpikir tingkat tinggi dan literasi sains dengan konsisten, dengan kesalahan pengukuran yang relatif rendah. Secara keseluruhan, reliabilitas yang tinggi hal tersebut memperkuat hasil pengujian validitas sebelumnya, sehingga instrumen memenuhi dua kriteria utama kualitas alat ukur, yaitu ketepatan dan keandalan.

Daya Pembeda

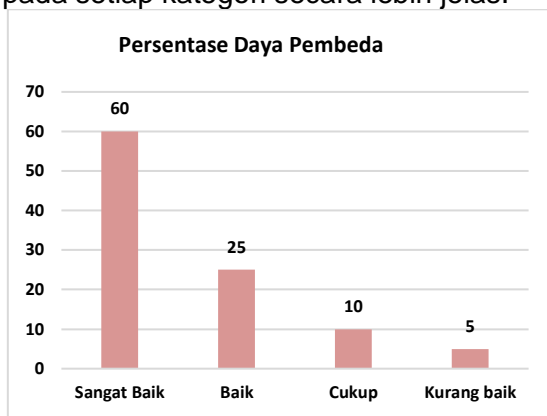
Daya pembeda mengacu pada kemampuan item soal dalam membedakan tingkat kemampuan siswa, baik tinggi maupun rendah (Rachmawati & Pradana, 2025). Nilainya

diperoleh dari selisih jumlah jawaban benar pada kelompok atas dan kelompok bawah. Semakin besar perbedaannya, semakin baik kualitas soal tersebut. Yuliana et al. (2024) menjelaskan bahwa soal yang berkualitas cenderung banyak dijawab benar oleh siswa berkemampuan tinggi dibandingkan siswa berkemampuan rendah. Mustaqim dan Sulisti (2024) juga menyatakan bahwa daya pembeda yang tinggi menunjukkan alat evaluasi mampu mengidentifikasi penguasaan materi secara lebih tepat. Berikut hasil uji daya pembeda tiap butir soal disajikan dalam Tabel 3.

Tabel 3. Hasil Uji Daya Pembeda

Kategori	Nomor Soal	Jumlah
Sangat Baik	1, 2, 7, 9, 10, 11, 12, 13, 14, 15, 16, 20	12
Baik	3, 6, 8, 18, 19	5
Cukup	4, 5	2
Kurang Baik	17	1

Berdasarkan tabel soal yang termasuk kategori sangat baik adalah nomor 1, 2, 7, 9, 10, 11, 12, 13, 14, 15, 16, dan 20, yang menunjukkan kemampuan sangat efektif dalam membedakan siswa. Kategori baik ada pada nomor 3, 6, 8, 18, dan 19. Sedangkan, nomor 4 serta 5 termasuk kategori cukup, dan nomor 17 berada pada kategori kurang baik. Hasil analisis daya pembeda kemudian disajikan dalam bentuk persentase untuk menunjukkan perbandingan jumlah soal pada setiap kategori secara lebih jelas.



Gambar 2. Persentase Daya Pembeda

Hasil pengujian daya pembeda, dari 20 soal 60% tergolong sangat baik, 25% baik, 10% cukup, dan 5% kurang baik. Hasil ini menandakan bahwa sebagian besar soal sudah mampu mengidentifikasi perbedaan siswa yang memiliki pemahaman baik terhadap materi dengan siswa yang kurang memahami konsep pewarisan sifat.

Soal pada kategori sangat baik memiliki kemampuan membedakan yang sangat jelas sehingga mendukung ketepatan hasil tes. Dominasi kategori ini (60%) menunjukkan bahwa kualitas soal sudah optimal. Rizki et al. (2024) menyarankan agar soal dengan daya pembeda sangat baik dipertahankan karena efektif dalam menilai kemampuan siswa.

Soal kategori baik (25%) juga sudah mampu membedakan kemampuan siswa dengan cukup jelas dan tetap layak digunakan (Rachmawati & Pradana, 2025; Rizki et al., 2024). Sementara itu, kategori cukup (10%) masih dapat digunakan, namun perlu perbaikan pada redaksi atau pilihan jawaban agar lebih tegas dalam membedakan tingkat kemampuan (Mustaqim & Sulisti, 2024; Rizki et al., 2024).

Adapun soal kategori kurang baik (5%) memiliki kemampuan membedakan yang rendah sehingga perlu direvisi. Soal yang memiliki daya pembeda rendah belum mampu memperlihatkan perbedaan yang signifikan antara kelompok atas dan bawah (Yuliana et al., 2024). Rizki et al. (2024) menyarankan perbaikan menyeluruh agar kualitas pengukuran tetap terjaga.

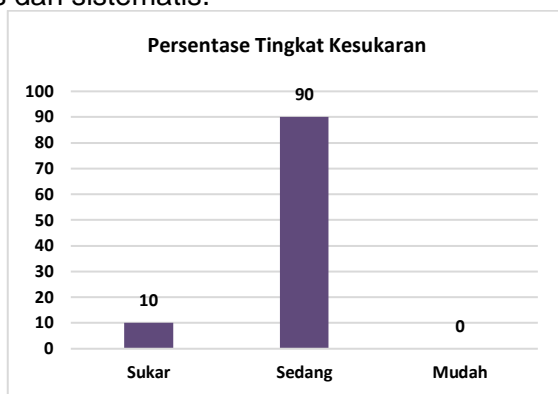
Secara keseluruhan, mayoritas berada pada kategori sangat baik dan baik (85%) memperlihatkan bahwa instrumen IPA berbasis HOTS dan literasi sains pada materi pewarisan sifat telah memiliki kemampuan diskriminatif yang kuat. Temuan ini sesuai

dengan hasil uji validitas dan reliabilitas sebelumnya, sehingga instrumen dinilai mampu mengukur perbedaan tingkat pemahaman siswa secara akurat.

Tingkat Kesukaran

Tingkat kesukaran adalah presentase siswa yang dapat menjawab soal dengan benar, dengan indeks antara 0–1. Nilai mendekati 0 menandakan soal sangat sukar, sedangkan mendekati 1 menandakan soal sangat mudah (Rachmawati & Pradana, 2025). Oleh karena itu, tingkat kesukaran harus seimbang sehingga soal tidak terlalu mudah maupun terlalu sulit. Namira et al. (2025) menegaskan bahwa keseimbangan ini penting untuk menghasilkan pengukuran yang efektif dan akurat. Item soal yang terlalu mudah tidak mampu menunjukkan perbedaan kemampuan siswa, di sisi lain soal yang terlalu sukar dapat membuat hasil tes tidak mencerminkan kemampuan sebenarnya. Hasil uji tingkat kesukaran disajikan pada tabel berikut.

Berdasarkan tabel rekapitulasi indeks kesukaran, kebanyakan item soal berada pada kategori sedang, yaitu sejumlah 18 soal yang meliputi nomor 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 19, dan 20. Sementara itu, kategori sukar terdapat pada nomor 11 dan 18 dengan jumlah 2 soal, dan tidak ada soal pada kategori mudah. Data selanjutnya disajikan dalam bentuk persentase untuk memperlihatkan distribusi tingkat kesukaran pada setiap kategori secara lebih jelas dan sistematis.



Gambar 3. Persentase Tingkat Kesukaran

Hasil pengujian tingkat kesukaran dari 20 butir soal, 90% tergolong sedang, 10% sukar, dan tidak ditemukan ada soal kategori mudah (0%). Hal ini menunjukkan bahwa instrumen memiliki tingkat kesukaran yang proporsional untuk mengukur kemampuan siswa pada materi pewarisan sifat.

Dominannya kategori sedang (90%) menunjukkan bahwa mayoritas soal berada pada tingkat kesulitan yang ideal dan dapat memberikan informasi pengukuran secara optimal (Namira et al., 2025). Sementara itu, 10% soal berkategori sukar. Soal sukar umumnya dijawab oleh siswa berkemampuan tinggi dan dapat digunakan untuk tujuan selektif, namun tetap perlu ditelaah agar tidak terdapat kesalahan redaksi (Fitriani, 2021). Dalam konteks HOTS, keberadaan soal sukar dalam jumlah terbatas relevan karena menuntut kemampuan analisis dan penalaran yang lebih kompleks. Tidak adanya soal kategori mudah menunjukkan bahwa instrumen lebih menekankan kemampuan analitis dan evaluatif dibandingkan sekadar hafalan.

Dengan demikian instrumen telah menunjukkan tingkat kesukaran yang tepat dan sejalan dengan tujuan pengukuran. Hasil ini juga mendukung temuan validitas, reliabilitas, dan daya pembeda sebelumnya, sehingga instrumen dinilai tepat dijadikan alat untuk menilai kemampuan berpikir tingkat tinggi dan literasi sains secara objektif dan efektif.

SIMPULAN

Berdasarkan hasil analisis, instrumen soal IPA berbasis HOTS dan literasi sains pada materi pewarisan sifat secara umum telah memenuhi kriteria kualitas yang baik. Sebanyak 85% butir soal dinyatakan valid, mengindikasikan bahwa mayoritas soal telah mampu mengukur konstruk yang relevan dengan indikator pembelajaran. Koefisien reliabilitas sebesar 0,8827 tergolong tinggi, dengan demikian instrumen memiliki konsistensi dan keandalan yang baik. Dari aspek daya pembeda, 85% item soal tergolong sangat baik dan baik, yang berarti mampu membedakan secara efektif siswa berkemampuan tinggi dan rendah. Sementara itu, tingkat kesukaran didominasi kategori sedang (90%), menunjukkan bahwa soal memiliki tingkat kesulitan yang proporsional dan dapat mengukur kemampuan siswa secara efektif. Secara keseluruhan, instrumen layak dijadikan alat evaluasi pembelajaran, meskipun masih ada beberapa butir yang perlu direvisi untuk meningkatkan kualitasnya secara optimal.

DAFTAR PUSTAKA

- Adawiyah, R., Maya Istyadji, & Ratna Yulinda. (2024). Pengembangan Bahan Ajar IPA SMP Berbasis Literasi Sains Topik Lapisan Bumi Dan Bencana. *Lencana: Jurnal Inovasi Ilmu Pendidikan*, 2(2), 64–75. <https://doi.org/10.55606/lencana.v2i2.3555>
- Ariska, M., Ismet, I., Andriani, N., & Saparini, S. (2024). Pelatihan Penyusunan Soal-soal Literasi Sains Berbasis HOTS bagi Guru IPA SMP di Kabupaten Muaraenim. *Jurnal Pendidikan dan Pengabdian Masyarakat*, 7(4), 235–241. <https://doi.org/10.29303/jppm.v7i4.7669>
- Azzahra, W., Diana, S., & Nuraeni, E. (2024). Analisis Validitas Dan Reliabilitas Angket Sikap Matematis Biologis Siswa Sma Pada Materi Hereditas Manusia. *Borneo Journal of Biology Education (BJBE)*, 6(1), 14–22. <https://doi.org/10.35334/bjbe.v6i1.5169>
- Cahyani, D. K., Nugroho, A. S., Nizaruddin, & Hayat, M. S. (2025). Analisis Kemampuan Literasi Sains dan Berpikir Kritis Siswa SMP pada Pembelajaran IPA: Analisis Kemampuan Literasi Sains dan Berpikir Kritis Siswa SMP pada Pembelajaran IPA. *PENDIPA Journal of Science Education*, 8(3), 593–600. <https://doi.org/10.33369/pendipa.8.3.593-600>
- Dewi, C., Hatibe, A., & Werdhiana, I. K. (2024). Penenerapan Model PBL pada Konsep Hidrostatika IPA Terpadu Type Connected terhadap HOTS Siswa SMP. *Ideguru: Jurnal Karya Ilmiah Guru*, 10(1), 619–625. <https://doi.org/10.51169/ideguru.v10i1.1333>
- Dianova, F. R., & Anwar, N. (2024). Analisis Butir Uji Validitas, Reliabilitas, Tingkat Kesukaran, dan Daya Pembeda Soal Sumatif Bahasa Arab SD Islam. *Jurnal Bahasa Daerah Indonesia*, 1(3), 13. <https://doi.org/10.47134/jbdi.v1i3.2863>
- Fajariani, K., & Hanum, E. (2025). Evaluasi Kualitas Instrumen Tes Pilihan Ganda pada Mata Pelajaran IPA Kelas V Sekolah Dasar Berdasarkan Analisis Butir Soal: Penelitian. *Jurnal Pengabdian Masyarakat Dan Riset Pendidikan*, 4(2), 9224–9229. <https://doi.org/10.31004/jerkin.v4i2.2984>
- Fatonah, F., Juriah, J., & Jodion Siburian. (2026). Literatur Review: Efektivitas Model Pembelajaran Problem Based Learning (PBL) dalam Mengembangkan Literasi Sains dan Higher Order Thinking Skills (Hots) Peserta Didik. *JURNAL PENDIDIKAN MIPA*, 16(1), 45–57. <https://doi.org/10.37630/jpm.v16i1.3625>
- Fitriani, N. (2021). Analisis Tingkat Kesukaran, Daya Pembeda, dan Efektivitas Pengecoh Soal Pelatihan Kewaspadaan Kegawatdaruratan Maternal dan Neonatal. *Paedagoria : Jurnal Kajian, Penelitian Dan Pengembangan Kependidikan*, 12(2), 199. <https://doi.org/10.31764/paedagoria.v12i2.4956>
- Indana, S., Wasis, W., & Jauhari, M. N. R. (2022). Pelatihan Pembelajaran Dan Penilaian Higher Order Thinking Skills (HOTS) dan Literasi Sains Bagi Guru IPA SMP. *Jurnal Pemberdayaan Masyarakat*, 1(2), 78–83. <https://doi.org/10.46843/jmp.v1i2.275>
- Ketaren, M. A., Girsang, K., Manurung, M., & Ginting, E. R. B. (2024). Uji Validitas Dan Uji Daya Beda Soal Buatan Pilihan Ganda Dengan Tes Sumatif Siswa Kelas IV UPT SD

- Negeri 065013 Medan. *Jurnal Pengabdian Masyarakat Bangsa*, 1(11), 3278–3283. <https://doi.org/10.59837/jpmba.v1i11.706>
- Mijaya, N. P. A. P., Sudiarmika, A. A. I. A. R., & Suardana, I. N. (2021). Pengembangan E-Modul Pembelajaran Ipa Smp Kelas Vii Berbasis Model Pembelajaran Levels of Inquiry Untuk Meningkatkan Literasi Sains Siswa. *Quantum: Jurnal Inovasi Pendidikan Sains*, 12(2), 220. <https://doi.org/10.20527/quantum.v12i2.11258>
- Miladanta, A. N., Nuryantini, A. Y., Farida, I., & Cahyanto, T. (2024). Pengembangan Instrumen Higher Order Thinking Skills (HOTS) Materi Alat Optik melalui Validitas, Reliabilitas, Daya Pembeda dan Tingkat Kesukaran. *Jurnal Penelitian Sains Dan Pendidikan (JPSP)*, 4(2), 148–158. <https://doi.org/10.23971/jpsp.v4i2.7828>
- Mustaqim, M., & Sulisti, H. (2024). Analisis Butir Soal Pas Matematika Peminatan: Daya Pembeda, Tingkat Kesukaran, Dan Kualitas Pengecoh. *Al-'Adad: Jurnal Tadris Matematika*, 3(1), 44–56. <https://doi.org/10.24260/add.v3i1.3011>
- Namira, C. M., Banjarnahor, J., Lahera, L. B., Nainggolan, R., & Gultom, S. M. (2025). Analisis Soal UTS Fisika Siswa Kelas XI SMA Negeri 1 Batang Kuis Berbasis Comparative Thinking: Validitas, Reliabilitas, Kesukaran, dan Daya Pembeda. *SAKOLA: Journal of Sains Cooperative Learning and Law*, 2(1), 422–433. <https://doi.org/10.57235/sakola.v2i1.5694>
- Nissa, N. A., & Dimas, A. (2022). Analisis Buku Ajar IPA SMP pada Materi Hukum Newton Ditinjau dari Aspek Higher Order Thinking Skills (HOTS). *Jurnal Tadris IPA Indonesia*, 2(1), 74–81. <https://doi.org/10.21154/jtii.v2i1.527>
- Nurhalimah, S., Hidayati, Y., Rosidi, I., & Hadi, W. P. (2022). Hubungan Antara Validitas Item Dengan Daya Pembeda Dan Tingkat Kesukaran Soal Pilihan Ganda Pas. *Natural Science Education Research*, 4(3), 249–257. <https://doi.org/10.21107/nser.v4i3.8682>
- Prananda, Y., Rahayu, S., & Nida, S. (2021). Analisis kebutuhan bahan ajar ipa smp dengan mengeksplisitkan hakikat sains (NOS) dan berpikir kritis topik energi dalam sistem kehidupan. *Jurnal MIPA dan Pembelajarannya*, 1(3). <https://doi.org/10.17977/um067v1i3p189-194>
- Rachmawati, D., & Pradana, A. B. (2025). Analisis butir soal mata pelajaran ekonomi: Validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. *Jurnal Pendidikan Ekonomi (JUPE)*, 13(3), 273–284. <https://doi.org/10.26740/jupe.v13n3.p273-284>
- Rizki, A., Halini, H., & Indriani, T. (2024). Analisis Daya Pembeda Butir Tes Pilihan Ganda Materi Persamaan Trigonometri Berdasarkan Teori Tes Klasik. *JURNAL MathEdu (Mathematic Education Journal)*, 7(2), 44–49. <https://doi.org/10.37081/mathedu.v7i2.5961>
- Siregar, T. M., Agustina, N., Siahaan, F. B., Sinaga, L. L., Prabawa, I., & Sagala, R. Z. (2024). Analisis Validitas dan Reliabilitas Butir Soal Pilihan Ganda Dalam Mata Pelajaran Matematika Kelas 12 Semester Genap di SMAN 1 Adiankoting. *ARRUMMAN: Journal of Education and Learning Evaluation*, 1(2), 725–730. <https://doi.org/10.57235/arrumman.v1i2.4398>
- Sudrajat, S. (2024). Analisis Kualitas Instrumen Tes untuk Mengukur Kemampuan Komunikasi Matematika Siswa SMP. *ELIPS: Jurnal Pendidikan Matematika*, 5(2), 210–221. <https://doi.org/10.47650/elips.v5i2.1398>
- Ulfian, U., Syamsu, S., & Mansyur, J. (2025). Pengaruh Model Discovery Learning Terhadap Kemampuan Literasi Sains Siswa Kelas VIII SMP Negeri 1 Banawa Selatan. *JPFT (Jurnal Pendidikan Fisika Tadulako Online)*, 12(1), 34–38. <https://doi.org/10.22487/jpft.v12i1.3460>
- Yuliana, Y., Zuhdi, M., Gunada, I. W., & Taufik, M. (2024). Uji Instrumen Penguasaan Konsep Fisika Pada Pokok Bahasan Elastisitas Dan Hukum Hooke. *Kappa Journal*, 8(1), 57–61. <https://doi.org/10.29408/kpj.v8i1.24874>