

Validitas Isi Asesmen Scientific Higher-Order Thinking Berbasis Dual Space Inquiry pada Materi Kinematika

¹Patimah Tussukriyah Caniago, ¹Fuja Novitra, ¹Hidayati, ¹Renol Afrizon

¹Departemen Fisika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Padang, Padang, Indonesia

*Corresponding Author e-mail: patimahcng@gmail.com

Received: June 2026; Revised: July 2026; Published: July 2026

Abstrak

Keterbatasan asesmen fisika yang masih berfokus pada prosedur dan jawaban akhir menyulitkan pengukuran Scientific Higher-Order Thinking (S-HOT), khususnya pada kinematika yang menuntut interpretasi fenomena, data, diagram, dan grafik gerak. Penelitian ini bertujuan memperoleh bukti awal validitas isi asesmen S-HOT berbasis Dual Space Inquiry (DSI) pada materi kinematika. Penelitian menggunakan Educational Design Research dengan model Plomp dan artikel ini melaporkan tahap preliminary research serta development/prototyping. Prototipe mengintegrasikan lima dimensi S-HOT – penalaran ilmiah, berpikir kritis, berpikir kreatif, efikasi diri sains, dan metakognisi – ke dalam fase orientasi, konseptualisasi, eksplorasi, kesimpulan dan asesmen, serta refleksi. Tiga dosen Departemen Fisika Universitas Negeri Padang menilai prototipe secara ahli menggunakan lembar validasi skala lima poin yang mencakup 27 indikator pada aspek isi, bahasa, penyajian, dan kegrafisan. Data dianalisis dengan koefisien Aiken's V. Rata-rata V masing-masing aspek adalah 0,88, 0,90, 0,94, dan 0,94, dengan rata-rata keseluruhan 0,92. Pada tingkat indikator, 19 dari 27 indikator mencapai nilai kritis $V = 0,92$ untuk tiga validator dan lima kategori penilaian, sedangkan delapan indikator memerlukan revisi, terutama pada kedalaman materi, pencegahan miskonsepsi, kejelasan kalimat, rubrik penskoran, dan proporsi tampilan. Temuan menunjukkan bukti awal validitas isi yang cukup kuat untuk melanjutkan revisi prototipe, tetapi belum mendukung klaim reliabilitas, validitas konstruk, kepraktisan, atau efektivitas. Uji keterbacaan, uji coba terbatas, analisis butir, dan pengujian empiris diperlukan pada tahap berikutnya.

Kata kunci: Aiken's V; asesmen; Dual Space Inquiry; kinematika; Scientific Higher-Order Thinking

Content Validity of a Scientific Higher-Order Thinking Assessment Based on Dual Space Inquiry for Kinematics

Abstract

Physics assessments that emphasize procedures and final answers provide limited evidence of students' Scientific Higher-Order Thinking (S-HOT), particularly in kinematics, which requires the interpretation of phenomena, data, motion diagrams, and graphs. This study aimed to obtain initial content-validity evidence for an S-HOT assessment based on Dual Space Inquiry (DSI) for kinematics. The study employed Educational Design Research using the Plomp model; this article reports the preliminary research and development/prototyping phases. The prototype integrated five S-HOT dimensions – scientific reasoning, critical thinking, creative thinking, science self-efficacy, and metacognition – within the DSI phases of orientation, conceptualization, exploration, conclusion and assessment, and reflection. Three lecturers from the Department of Physics, Universitas Negeri Padang, served as expert validators. They evaluated the prototype using a five-point validation sheet comprising 27 indicators across content, language, presentation, and graphic-design aspects. The ratings were analyzed using Aiken's V. The mean V values were 0.88 for content, 0.90 for language, 0.94 for presentation, and 0.94 for graphic design, with an overall mean of 0.92. At the indicator level, 19 of 27 indicators reached the critical value of $V = 0.92$ for three validators and five response categories; eight indicators required revision, particularly those concerning content depth, misconception prevention, sentence clarity, scoring rubrics, and visual proportion. These findings provide initial content-validity evidence for further prototype revision but do not establish reliability, construct validity, practicality, or effectiveness. Readability testing, limited piloting, item analysis, and empirical validation are therefore required.

Keywords: Aiken's V; assessment; Dual Space Inquiry; kinematics; Scientific Higher-Order Thinking

How to Cite: Caniago, P. T., Novitra, F. ., Hidayati, H., & Afrizon, R. . (2026). Validitas Isi Asesmen Scientific Higher-Order Thinking Berbasis Dual Space Inquiry pada Materi Kinematika. *Journal of Authentic Research*, 5(3), 5497-5510. <https://journal-center.litpam.com/index.php/jar/article/view/6289>



<https://journal-center.litpam.com/index.php/jar/article/view/6289>

Copyright© 2026, Caniago et al.

This is an open-access article under the CC-BY-SA License.



PENDAHULUAN

Fisika berperan penting dalam melatih peserta didik untuk menalar secara rasional, kritis, dan sistematis ketika menjelaskan fenomena alam. Pembelajaran fisika tidak cukup berhenti pada penguasaan persamaan, tetapi perlu membantu peserta didik menghubungkan konsep, bukti, dan representasi ilmiah dengan situasi nyata (Lestari, 2021). Kurikulum nasional juga menempatkan kemampuan menganalisis, mengevaluasi, dan menghasilkan gagasan sebagai bagian penting dari capaian pembelajaran, sehingga asesmen perlu memberikan ruang bagi proses berpikir tingkat tinggi, bukan hanya menguji ingatan atau penerapan rutin (BSKP Kemendikbud, 2025). Kemampuan tersebut mendukung pemahaman konseptual dan pemecahan masalah fisika yang lebih mendalam (Putranta et al., 2021). Namun, Higher-Order Thinking Skills (HOTS) masih bersifat umum dan belum sepenuhnya merepresentasikan karakter penalaran sains yang melibatkan bukti, hipotesis, pengolahan data, argumentasi, dan refleksi. Scientific Higher-Order Thinking (S-HOT) menawarkan kerangka yang lebih spesifik melalui lima dimensi yang saling berhubungan, yaitu penalaran ilmiah, berpikir kritis, berpikir kreatif, efikasi diri sains, dan metakognisi (Sun et al., 2022). Dalam pembelajaran fisika, kelima dimensi tersebut perlu diukur melalui tugas yang menuntut peserta didik menjelaskan alasan, membandingkan representasi, mengevaluasi bukti, menghasilkan alternatif solusi, serta merefleksikan strategi berpikirnya. Oleh karena itu, kualitas asesmen menjadi penentu apakah pembelajaran benar-benar memberikan informasi tentang kemampuan ilmiah tingkat tinggi peserta didik. Asesmen yang terstruktur juga membantu guru menggunakan hasil belajar sebagai dasar untuk memberikan umpan balik dan memperbaiki pengalaman inkuiri berikutnya.

Kebutuhan tersebut belum sepenuhnya terjawab oleh praktik asesmen fisika di sekolah. Kajian instrumen HOTS menunjukkan bahwa banyak soal masih berfokus pada hasil akhir dan belum konsisten mengungkap proses penalaran peserta didik (Desiriah & Setyarsih, 2021). Temuan pada pembelajaran fisika juga memperlihatkan capaian yang rendah pada aspek analisis, evaluasi, dan kreasi; kondisi ini menunjukkan kesulitan peserta didik dalam mengolah informasi, menilai bukti, dan merancang penyelesaian yang tidak rutin (Fitriana et al., 2024). Instrumen yang dominan berupa perhitungan dasar dapat menghasilkan nilai, tetapi tidak cukup memperlihatkan bagaimana peserta didik menafsirkan fenomena, membangun argumen, atau mengevaluasi strategi berpikirnya (Rohmah & Sunarti, 2020; Delvananta Givarin & Ellianawati, 2025). Masalah serupa tampak pada studi pendahuluan penelitian ini melalui angket, observasi, wawancara, dan kajian dokumen pembelajaran: asesmen yang digunakan cenderung mengulang pola contoh, mengubah angka, dan menekankan prosedur. Karena artikel ini berfokus pada validasi ahli, data studi pendahuluan digunakan sebagai dasar kebutuhan pengembangan dan tidak diperlakukan sebagai bukti efektivitas produk. Literatur pengembangan instrumen juga menegaskan bahwa asesmen HOTS harus dibangun melalui indikator yang jelas, stimulus yang relevan, serta prosedur validasi yang transparan (Zhou et al., 2023). Dengan demikian, diperlukan asesmen yang mampu menangkap bukti kognitif dan reflektif secara terpisah: tugas kognitif untuk penalaran, berpikir kritis, dan kreativitas; serta prompt refleksi atau skala respons untuk efikasi diri dan metakognisi.

Dual Space Inquiry (DSI) relevan untuk kebutuhan tersebut karena menggabungkan ruang belajar tatap muka dan digital dalam rangkaian inkuiri yang

terstruktur. Kerangka DSI memuat fase orientasi, konseptualisasi, eksplorasi, kesimpulan dan asesmen, serta refleksi (Novitra et al., 2025). Setiap fase menyediakan konteks berbeda untuk memperoleh bukti S-HOT: orientasi mengaktifkan pengetahuan awal dan identifikasi masalah; konseptualisasi menuntut pertanyaan dan hipotesis; eksplorasi menghasilkan data dan pengujian dugaan; kesimpulan dan asesmen menguji kualitas interpretasi; sedangkan refleksi mengungkapkan metakognisi dan efikasi diri. Studi pengembangan DSI menunjukkan bahwa dukungan digital dapat memperluas fleksibilitas dan partisipasi dalam aktivitas inkuiri (Loreza et al., 2026; Novitra et al., 2026). Namun, penelitian asesmen S-HOT berbasis DSI masih terbatas. Akbar (2026) melaporkan validasi asesmen serupa pada materi listrik arus searah, sehingga pergantian topik semata tidak cukup menjadi kebaruan. Penelitian ini menempatkan kebaruan pada operasionalisasi karakteristik kinematika melalui fenomena gerak, diagram gerak, tabel data, grafik posisi-waktu dan kecepatan-waktu, serta hubungan antarrepresentasi dalam setiap fase DSI. Kinematika menuntut peserta didik berpindah antara representasi verbal, matematis, grafis, dan eksperimental, sehingga struktur bukti yang dinilai berbeda dari topik listrik. Pendekatan inkuiri pada materi gerak juga relevan untuk melatih interpretasi data dan penalaran berbasis bukti (Asyifa, 2026). Integrasi tersebut dirancang agar asesmen tidak hanya menilai jawaban benar, tetapi juga kualitas alasan, proses, dan refleksi.

Berdasarkan gap tersebut, penelitian ini bertujuan menganalisis validitas isi asesmen S-HOT berbasis DSI pada materi kinematika untuk peserta didik SMA. Validitas isi dalam artikel ini dibatasi pada kesesuaian prototipe menurut penilaian tiga ahli terhadap aspek isi, bahasa, penyajian, dan kegrafisan; penelitian belum menguji validitas konstruk, reliabilitas, kepraktisan, maupun efektivitas pada peserta didik. Konstruk S-HOT dioperasionalkan melalui lima dimensi, yaitu penalaran ilmiah, berpikir kritis, berpikir kreatif, efikasi diri sains, dan metakognisi (Sun et al., 2022). Ketiga dimensi kognitif dinilai melalui respons terbuka, interpretasi data, dan argumentasi berbasis bukti, sedangkan efikasi diri dan metakognisi ditempatkan pada bagian refleksi agar tidak disamakan dengan skor benar-salah. Aktivitas asesmen disusun mengikuti lima fase DSI secara konsisten dan mengintegrasikan karakter kinematika yang menuntut penggunaan multirepresentasi. Analisis menggunakan koefisien Aiken's V pada tingkat indikator dan rata-rata aspek. Hasil penelitian diharapkan memberikan bukti awal mengenai relevansi konten prototipe serta menunjukkan bagian yang masih memerlukan revisi sebelum dilakukan uji keterbacaan, uji coba terbatas, analisis butir, dan pengujian empiris. Dengan ruang lingkup tersebut, kontribusi penelitian bukan membuktikan peningkatan S-HOT, melainkan menyediakan rancangan asesmen yang secara teoritis memetakan dimensi S-HOT, fase DSI, karakter materi kinematika, dan bukti respons yang dinilai. Penegasan batas ini penting agar interpretasi hasil tetap sejalan dengan desain penelitian dan bukti yang benar-benar dikumpulkan.

METODE

Desain Penelitian

Penelitian ini menggunakan Educational Design Research (EDR) dengan model Plomp yang mencakup tiga tahap utama: preliminary research, development or prototyping phase, dan assessment phase (Plomp, 2013). Artikel ini hanya melaporkan preliminary research dan bagian validasi ahli pada development/prototyping phase; assessment phase berupa uji kepraktisan dan efektivitas belum dilakukan. Pada preliminary research, tim mengidentifikasi kebutuhan asesmen melalui kajian literatur, analisis kurikulum, telaah dokumen asesmen, observasi pembelajaran, serta wawancara dengan guru dan peserta didik. Informasi tersebut digunakan untuk merumuskan spesifikasi produk, bukan untuk menguji hipotesis atau menarik generalisasi tentang populasi. Pada tahap pengembangan, konstruk S-HOT, sintaks DSI, capaian pembelajaran kinematika, dan karakter multirepresentasi dipadukan dalam kisi-kisi prototipe. Prototipe kemudian ditelaah oleh ahli, direkap pada tingkat indikator, dan dipetakan ke bagian yang memerlukan perbaikan. Siklus penelitian yang dilaporkan terdiri atas analisis kebutuhan, perancangan prototipe I, validasi ahli, dan penyusunan rencana revisi prototipe II. Karena penilaian ulang setelah revisi belum tersedia, nilai Aiken's V yang dilaporkan merepresentasikan prototipe I. Pembatasan ini ditegaskan agar hasil tidak ditafsirkan sebagai validitas empiris atau keberhasilan penggunaan produk di kelas. Alur penelitian dan posisi validasi ahli dalam siklus EDR ditunjukkan pada Gambar 1. Fokus keluaran penelitian ini adalah bukti awal relevansi isi dan daftar prioritas revisi yang dapat digunakan sebelum prototipe diuji kepada pengguna sasaran.



Gambar 1. Posisi validasi ahli dalam siklus Educational Design Research

Validator dan Konteks Studi Pendahuluan

Subjek pada tahap validasi adalah tiga dosen Departemen Fisika, Universitas Negeri Padang, yang berperan sebagai validator ahli. Pemilihan dilakukan secara purposif karena validasi awal memerlukan penilai yang memahami substansi fisika, pembelajaran, dan pengembangan instrumen. Penilaian diberikan pada prototipe yang sama menggunakan lembar validasi dengan kriteria identik. Setiap validator menilai kesesuaian konsep kinematika, keterkaitan dengan capaian pembelajaran, keterwakilan S-HOT, konsistensi tahapan DSI, keterbacaan bahasa, kelengkapan penyajian, serta kegrafisan. Hasil penilaian tidak digunakan untuk membandingkan performa antarvalidator, tetapi untuk menilai relevansi setiap indikator produk dan mengidentifikasi bagian yang memerlukan revisi. Jumlah tiga validator masih dapat digunakan untuk penelaahan formatif awal, tetapi membatasi ketepatan estimasi dan kekuatan generalisasi. Oleh sebab itu, interpretasi Aiken's V dilakukan secara hati-hati dengan mempertimbangkan nilai kritis yang sesuai dengan jumlah validator dan kategori skor, bukan hanya menggunakan kategori umum. Data studi pendahuluan yang mendasari kebutuhan pengembangan ditempatkan sebagai konteks desain dan tidak dianalisis sebagai hasil utama atau dasar klaim validitas produk.

Produk, Instrumen, dan Prosedur

Produk yang dinilai adalah prototipe asesmen S-HOT berbasis DSI untuk materi kinematika. Organisasi prototipe dibagi ke dalam lima kelompok aktivitas sesuai fase DSI: orientasi, konseptualisasi, eksplorasi, kesimpulan dan asesmen, serta refleksi. Stimulus menggunakan fenomena gerak, narasi kontekstual, diagram gerak, tabel data, dan grafik posisi-waktu atau kecepatan-waktu. Respons kognitif berupa jawaban terbuka, interpretasi representasi, perumusan hipotesis, analisis data, dan argumentasi berbasis bukti; respons efikasi diri dan metakognisi ditempatkan pada prompt refleksi atau skala respons agar tidak diperlakukan sebagai soal benar-salah. Jawaban kognitif direncanakan dinilai menggunakan rubrik analitik yang memisahkan ketepatan konsep, penggunaan bukti, kualitas penalaran, dan kejelasan komunikasi ilmiah. Lembar validasi ahli memuat 27 indikator: 10 indikator kelayakan isi, 5 indikator bahasa, 6 indikator penyajian, dan 6 indikator kegrafisan. Setiap indikator dinilai pada skala 1-5 dan dilengkapi ruang komentar. Prosedur validasi meliputi pemberian prototipe dan lembar penilaian, pengisian oleh validator, rekapitulasi skor, perhitungan Aiken's V, identifikasi indikator di bawah nilai kritis, dan penyusunan prioritas revisi. Kisi-kisi integrasi S-HOT, DSI, materi, dan bukti jawaban disajikan pada Tabel 1. Dalam naskah ini, validasi difokuskan pada struktur prototipe dan kualitas penyajiannya, bukan pada karakteristik psikometrik respons peserta didik.

Tabel 1. Pemetaan S-HOT, Fase DSI, Materi Kinematika, dan Bukti Respons

Fase DSI	Fokus asesmen kinematika	Indikator S-HOT utama	Bukti/jawaban yang dinilai
Orientasi	Fenomena gerak nyata/digital; identifikasi masalah dan informasi relevan	Berpikir kritis; penalaran ilmiah; efikasi diri awal	Rumusan masalah, alasan awal, dan tingkat keyakinan
Konseptualisasi	Perumusan pertanyaan atau hipotesis tentang posisi, kecepatan, dan percepatan	Penalaran ilmiah; berpikir kreatif	Pertanyaan, hipotesis, prediksi, dan representasi alternatif
Eksplorasi	Analisis data eksperimen/simulasi, tabel, diagram gerak, dan grafik	Penalaran ilmiah; berpikir kritis	Data terolah, grafik, pola, dan evaluasi bukti
Kesimpulan dan asesmen	Perbandingan hipotesis dengan bukti dan transfer konsep ke situasi baru	Berpikir kritis; penalaran ilmiah; berpikir kreatif	Klaim-bukti-penalaran, solusi, dan justifikasi
Refleksi	Evaluasi strategi, kesalahan, kemajuan, dan keyakinan diri	Metakognisi; efikasi diri sains	Jurnal refleksi, skala keyakinan, dan rencana perbaikan

Analisis Data

Skor validasi dianalisis menggunakan koefisien Aiken's V untuk menilai tingkat kesepakatan ahli mengenai relevansi setiap indikator (Aiken, 1980, 1985). Koefisien dihitung dengan rumus $V = \sum s / [n(c-1)]$, dengan $s = r - l_0$; r adalah skor validator, l_0 adalah skor terendah (1), c adalah jumlah kategori skor (5), dan n adalah jumlah validator (3). Karena tiga validator dan lima kategori menghasilkan penyebut 12, nilai V bergerak dalam kelipatan 0,083. Keputusan pada tingkat indikator menggunakan nilai kritis sisi kanan $V = 0,92$ pada $\alpha = 0,05$ untuk $n = 3$ dan $c = 5$. Dengan demikian, indikator bernilai 0,92 atau 1,00 diperlakukan sebagai bukti awal yang memenuhi kriteria, sedangkan indikator bernilai 0,83 atau 0,75 diprioritaskan untuk revisi. Rata-rata aspek tetap dilaporkan sebagai ringkasan deskriptif, tetapi tidak menggantikan keputusan tingkat indikator. Untuk memudahkan pembacaan, rata-rata 0,80–1,00 disebut tinggi, 0,60–0,79 sedang, dan di bawah 0,60 rendah berdasarkan klasifikasi deskriptif yang diadaptasi dari Retnawati (2016). Pemisahan antara nilai kritis dan kategori deskriptif dilakukan untuk menghindari inkonsistensi istilah "valid" dan "sangat valid". Interval kepercayaan Aiken's V idealnya dihitung dari matriks skor mentah (Penfield & Giacobbi, 2004); karena matriks tersebut tidak disajikan dalam naskah awal, analisis saat ini dibatasi pada nilai V dan probabilitas kritis. Skala penilaian ditampilkan pada Tabel 2, sedangkan aturan interpretasi ditampilkan pada Tabel 3.

$$V = \sum s / [n(c-1)] \quad \text{dan} \quad s = r - l_0$$

Tabel 2. Skala Penilaian Validasi

Skor	Kriteria
1	Sangat tidak setuju
2	Tidak setuju
3	Netral
4	Setuju
5	Sangat setuju

Tabel 3. Aturan Keputusan dan Ringkasan Deskriptif Aiken's V

Jenis interpretasi	Nilai	Makna
Keputusan indikator	$V \geq 0,92$	Memenuhi nilai kritis ($n = 3$; $c = 5$; $\alpha = 0,05$)
Keputusan indikator	$V < 0,92$	Diprioritaskan untuk revisi
Ringkasan deskriptif	0,80–1,00	Tinggi
Ringkasan deskriptif	0,60–0,79	Sedang
Ringkasan deskriptif	$< 0,60$	Rendah

HASIL DAN PEMBAHASAN

Validitas Kelayakan Isi

Validasi kelayakan isi mencakup 10 indikator yang menilai ketepatan konsep, kesesuaian dengan capaian pembelajaran, konteks kinematika, kedalaman materi, potensi miskonsepsi, dan relevansi soal. Nilai V berada pada rentang 0,75–1,00 dengan rata-rata 0,88, sehingga secara deskriptif aspek isi berada pada kategori tinggi. Namun, keputusan tingkat indikator menunjukkan bahwa enam indikator mencapai nilai kritis 0,92, sedangkan empat indikator belum mencapai kriteria tersebut. Dua indikator bernilai 0,83 berkaitan dengan kesesuaian indikator soal terhadap tujuan pembelajaran dan penyajian konsep/prinsip kinematika secara benar. Dua indikator bernilai 0,75 berkaitan dengan kedalaman serta keluasan materi sesuai perkembangan peserta didik dan pencegahan miskonsepsi. Temuan ini menunjukkan bahwa rata-rata aspek yang tinggi tidak boleh digunakan untuk menutupi kelemahan pada indikator tertentu. Hasil rinci disajikan pada Tabel 4. Indikator dengan $V = 1,00$, yaitu sifat kontekstual dan relevansi fenomena kehidupan sehari-hari, merupakan kekuatan prototipe karena konteks autentik membantu peserta didik menghubungkan bukti dengan konsep. Sebaliknya, skor yang lebih rendah mengisyaratkan perlunya penyesuaian tingkat kesulitan, verifikasi representasi, dan pemeriksaan kemungkinan interpretasi alternatif sebelum uji coba kepada peserta didik.

Tabel 4. Hasil Validitas Komponen Kelayakan Isi

Komponen kelayakan isi	V	Keputusan
Kesesuaian konsep dan prinsip ilmiah dengan kaidah keilmuan	0,92	Memenuhi
Kesesuaian materi asesmen dengan capaian pembelajaran kinematika	0,92	Memenuhi
Kesesuaian materi dengan fakta kehidupan sehari-hari	0,92	Memenuhi
Kesesuaian indikator soal dengan tujuan pembelajaran	0,83	Revisi
Kedalaman dan keluasan materi sesuai perkembangan peserta didik	0,75	Revisi
Kesesuaian dengan tuntutan Kurikulum Merdeka	0,92	Memenuhi
Asesmen tidak menimbulkan miskonsepsi	0,75	Revisi
Konsep dan prinsip kinematika disajikan secara benar	0,83	Revisi
Relevansi asesmen dengan perkembangan ilmu pengetahuan	0,92	Memenuhi
Soal kontekstual dan relevan dengan fenomena sehari-hari	1,00	Memenuhi
Rata-rata	0,88	Tinggi (deskriptif)

Prioritas revisi pada aspek isi adalah memperjelas hubungan setiap tugas dengan tujuan pembelajaran dan indikator S-HOT, menambah penanda antarrepresentasi, serta meninjau stimulus yang berpotensi memunculkan miskonsepsi tentang jarak-perpindahan, kelajuan-kecepatan, dan kemiringan grafik. Rubrik juga perlu membedakan kesalahan matematis dari kesalahan konseptual agar bukti penalaran tidak direduksi menjadi jawaban akhir. Langkah tersebut sejalan dengan prinsip bahwa instrumen HOTS harus mengukur kemampuan yang dituju melalui tugas yang relevan dan terdefinisi dengan baik (Musayroh, 2021). Studi Sun et al. (2022) memvalidasi struktur S-HOT sebagai lima konstruk yang saling berkaitan, tetapi menggunakan kuesioner dan analisis faktor; penelitian ini berbeda karena mengoperasionalkan konstruk tersebut dalam aktivitas kinematika berbasis inkuiri. Dibandingkan asesmen S-HOT berbasis DSI pada listrik arus searah (Akbar, 2026), prototipe ini menekankan translasi antara representasi gerak, interpretasi kemiringan dan luas grafik, serta konsistensi data eksperimen. Perbedaan ini perlu terlihat dalam kisi-kisi dan rubrik agar kebaruan tidak sekadar berupa pergantian materi. Karena empat indikator isi belum memenuhi nilai kritis dan prototipe belum dinilai ulang, hasil saat ini lebih tepat disebut bukti awal validitas isi, bukan bukti bahwa seluruh butir telah valid.

Validitas Penggunaan Bahasa

Validasi penggunaan bahasa terdiri atas lima indikator. Nilai V berada pada rentang 0,83–1,00 dengan rata-rata 0,90. Tiga indikator mencapai nilai kritis 0,92, yaitu keterbacaan bagi peserta didik, konsistensi istilah fisika, dan kejelasan instruksi. Dua indikator bernilai 0,83, yaitu ketepatan tata bahasa baku serta kejelasan kalimat agar tidak menimbulkan makna ganda. Hasil tersebut menunjukkan bahwa bahasa prototipe secara umum dapat dipahami, tetapi beberapa konstruksi kalimat masih

berpotensi menambah beban baca atau mengaburkan tuntutan soal. Pada asesmen kinematika, ambiguitas kecil dapat mengubah makna, misalnya penggunaan istilah “jarak” untuk “perpindahan”, “kecepatan” untuk “kelajuan”, atau frasa waktu yang tidak menunjukkan interval dengan jelas. Oleh karena itu, penyuntingan perlu dilakukan pada tingkat istilah, struktur kalimat, dan hubungan antara stimulus dengan pertanyaan. Hasil rinci disajikan pada Tabel 5. Perbaikan bahasa tidak hanya bersifat kosmetik; bahasa yang tidak presisi dapat menghasilkan kesalahan respons yang berasal dari salah tafsir, bukan dari keterbatasan S-HOT peserta didik.

Tabel 5. Hasil Validitas Komponen Penggunaan Bahasa

Komponen penggunaan bahasa	V	Keputusan
Bahasa baku dan mudah dimengerti peserta didik	0,83	Revisi
Bahasa terbaca jelas oleh peserta didik	1,00	Memenuhi
Kalimat jelas dan tidak menimbulkan makna ganda	0,83	Revisi
Istilah fisika digunakan secara tepat dan konsisten	0,92	Memenuhi
Instruksi pengerjaan jelas dan sistematis	0,92	Memenuhi
Rata-rata	0,90	Tinggi (deskriptif)

Revisi bahasa diarahkan pada pemendekan kalimat majemuk, penggunaan satu istilah untuk satu konsep, penghapusan kata yang tidak berfungsi, serta penempatan perintah kerja secara eksplisit. Istilah “asesmen”, “berpikir”, “Scientific Higher-Order Thinking”, dan nama fase DSI juga perlu ditulis konsisten di seluruh bagian produk. Temuan ini sejalan dengan Faizah et al. (2022), yang menekankan bahwa keterbacaan dan ketepatan bahasa merupakan prasyarat agar instrumen berbasis HOTS benar-benar menangkap kemampuan berpikir, bukan kemampuan menebak maksud soal. Putri dan Pahlevi (2021) juga menunjukkan bahwa instruksi yang jelas membantu peserta didik memahami tuntutan tugas dan mengurangi kesalahan yang tidak berkaitan dengan konstruk. Dalam prototipe ini, penyuntingan bahasa perlu dilakukan bersamaan dengan peninjauan rubrik agar kata kerja pada pertanyaan sesuai dengan bukti yang diharapkan, misalnya “jelaskan dengan bukti”, “bandingkan dua representasi”, atau “evaluasi konsistensi data”. Karena dua indikator belum mencapai nilai kritis, klaim bahwa seluruh aspek bahasa telah valid harus dihindari sampai revisi dan validasi ulang dilakukan.

Validitas Kelayakan Penyajian

Aspek penyajian terdiri atas enam indikator dan memperoleh nilai V antara 0,83–1,00 dengan rata-rata 0,94. Lima indikator mencapai nilai kritis 0,92, yaitu keruntutan sesuai sintaks DSI, kejelasan petunjuk, kelengkapan komponen, relevansi stimulus pendukung, dan fasilitasi S-HOT. Satu indikator bernilai 0,83 adalah kejelasan pedoman penilaian. Temuan ini menunjukkan bahwa struktur umum prototipe telah mengikuti alur DSI secara konsisten, tetapi rubrik belum cukup rinci untuk menjamin keseragaman penskoran. Dalam asesmen respons terbuka, rubrik yang terlalu umum dapat menyebabkan skor lebih banyak dipengaruhi interpretasi penilai daripada

kualitas bukti peserta didik. Oleh karena itu, setiap tugas perlu dilengkapi kriteria analitik, deskripsi level performa, contoh respons, dan aturan pemberian skor parsial. Hasil rinci disajikan pada Tabel 6. Struktur penyajian yang runtut merupakan kekuatan karena peserta didik dapat mengikuti proses dari identifikasi fenomena hingga refleksi, sementara guru memperoleh jejak bukti yang terorganisasi pada setiap fase.

Tabel 6. Hasil Validitas Komponen Kelayakan Penyajian

Komponen kelayakan penyajian	V	Keputusan
Penyajian sistematis dan runtut sesuai sintaks DSI	1,00	Memenuhi
Petunjuk pengerjaan jelas dan mudah dipahami	1,00	Memenuhi
Komponen asesmen lengkap dan sistematis	0,92	Memenuhi
Pedoman penilaian jelas	0,83	Revisi
Ilustrasi, tabel, atau stimulus relevan	1,00	Memenuhi
Asesmen memfasilitasi dimensi S-HOT	0,92	Memenuhi
Rata-rata	0,94	Tinggi (deskriptif)

Penyajian asesmen yang sistematis memudahkan peserta didik memahami alur kerja dan membantu guru menghubungkan tugas dengan tujuan pembelajaran (Sirajuddin et al., 2021). Erwan Effendi (2023) juga menegaskan bahwa kualitas instrumen HOTS tidak hanya ditentukan oleh tingkat kognitif soal, tetapi juga oleh kejelasan petunjuk dan konsistensi sistem penskoran. Pada prototipe ini, kontribusi DSI terlihat ketika setiap fase menghasilkan bukti berbeda: orientasi menghasilkan identifikasi masalah, konseptualisasi menghasilkan pertanyaan atau hipotesis, eksplorasi menghasilkan data, kesimpulan dan asesmen menghasilkan argumen, sedangkan refleksi menghasilkan evaluasi strategi dan keyakinan diri. Pemetaan tersebut perlu ditampilkan pada kisi-kisi agar pembaca dapat melihat keterwakilan konstruk, bukan hanya tampilan produk. Revisi utama pada aspek penyajian adalah memperluas rubrik analitik dan menambahkan contoh respons jangka. Karena penilaian ulang belum dilakukan, peningkatan kualitas setelah revisi belum dapat diklaim. Nilai rata-rata 0,94 hanya menunjukkan penilaian ahli yang kuat terhadap organisasi prototipe I.

Validitas Kegrafisan

Aspek kegrafisan memuat enam indikator dengan nilai V antara 0,83–1,00 dan rata-rata 0,94. Lima indikator mencapai nilai kritis 0,92, meliputi konsistensi desain, kesesuaian jenis dan ukuran huruf, penempatan gambar atau video, fungsi ilustrasi untuk memahami konsep, dan keterbacaan perpaduan warna. Satu indikator, yaitu daya tarik dan proporsi tampilan, memperoleh V = 0,83. Skor tersebut menunjukkan bahwa elemen visual telah mendukung informasi, tetapi keseimbangan ruang, kepadatan teks, atau proporsi antara stimulus dan pertanyaan masih perlu diperbaiki. Pada materi kinematika, desain grafis memiliki fungsi substantif karena diagram

gerak, sumbu grafik, satuan, skala, dan legenda merupakan bagian dari informasi ilmiah yang harus ditafsirkan. Kesalahan visual dapat mengubah tuntutan kognitif atau memicu jawaban yang tidak mencerminkan kemampuan peserta didik. Hasil rinci disajikan pada Tabel 7. Revisi perlu memastikan konsistensi ukuran grafik, keterbacaan label, jarak antarelelemen, dan penggunaan visual yang benar-benar diperlukan.

Tabel 7. Hasil Validitas Komponen Kegrafisan Asesmen

Komponen kegrafisan asesmen	V	Keputusan
Tampilan asesmen menarik dan proporsional	0,83	Revisi
Desain keseluruhan rapi dan konsisten	0,92	Memenuhi
Jenis dan ukuran huruf sesuai	1,00	Memenuhi
Penempatan gambar, ilustrasi, atau video tepat	0,92	Memenuhi
Ilustrasi mendukung pemahaman konsep kinematika	1,00	Memenuhi
Perpaduan warna tidak mengganggu keterbacaan	1,00	Memenuhi
Rata-rata	0,94	Tinggi (deskriptif)

Kualitas visual yang proporsional dapat meningkatkan kejelasan informasi dan kemudahan penggunaan instrumen (Zulfa et al., 2025). Namun, daya tarik visual tidak boleh diperlakukan sebagai bukti bahwa konstruk S-HOT telah terukur. Dalam penelitian ini, kegrafisan merupakan kualitas pendukung yang membantu peserta didik membaca stimulus dan mengekspresikan respons, sedangkan keterwakilan konstruk tetap ditentukan oleh isi tugas dan rubrik. Revisi visual diprioritaskan pada pengurangan kepadatan halaman, penyelarasan posisi gambar dengan pertanyaan, penggunaan tipografi yang konsisten, dan penyediaan ruang jawaban yang sebanding dengan kompleksitas tugas. Integrasi ruang digital dalam DSI juga menuntut tampilan yang dapat dibaca pada perangkat berbeda tanpa mengubah struktur informasi. Karena hanya satu indikator kegrafisan berada di bawah nilai kritis, aspek ini relatif kuat dibandingkan aspek isi dan bahasa, tetapi masih memerlukan pemeriksaan keterpakaian langsung melalui uji keterbacaan dan uji coba pengguna. Penilaian tiga ahli belum cukup untuk menyimpulkan bahwa tampilan praktis bagi semua peserta didik.

Sintesis Validitas dan Batas Interpretasi

Secara keseluruhan, rata-rata Aiken's V pada empat aspek adalah 0,92. Ringkasan pada Tabel 8 menunjukkan bahwa 19 dari 27 indikator memenuhi nilai kritis 0,92, sedangkan delapan indikator perlu direvisi. Distribusi ini memberikan gambaran yang lebih informatif daripada hanya menyajikan rata-rata aspek. Kekuatan utama prototipe terletak pada kontekstualitas, keruntutan sesuai fase DSI, dukungan stimulus, dan fungsi visual. Prioritas perbaikan terletak pada kedalaman materi, pencegahan miskonsepsi, keterkaitan indikator dengan tujuan, kejelasan bahasa, rubrik penskoran, dan proporsi tampilan. Validitas menggambarkan sejauh mana

bukti dan teori mendukung interpretasi skor untuk tujuan tertentu, sehingga penilaian ahli hanya merupakan salah satu sumber bukti (Sugiyono, 2023). Karena belum ada respons peserta didik, penelitian ini belum dapat menilai tingkat kesukaran, daya pembeda, reliabilitas, struktur konstruk, atau konsekuensi penggunaan asesmen.

Tabel 8. Ringkasan Validitas Empat Aspek

Aspek	Jumlah indikator	Memenuhi $V \geq 0,92$	Perlu revisi	Rata-rata V
Isi	10	6	4	0,88
Bahasa	5	3	2	0,90
Penyajian	6	5	1	0,94
Kegrafisan	6	5	1	0,94
Keseluruhan	27	19	8	0,92

Kontribusi teoretis penelitian ini terletak pada pemetaan eksplisit antara dimensi S-HOT, fase DSI, karakter multirepresentasi kinematika, dan bukti respons. Kerangka DSI sebelumnya dikembangkan untuk memfasilitasi pembelajaran fleksibel melalui integrasi ruang fisik dan digital (Novitra et al., 2025), sedangkan studi selanjutnya menunjukkan hubungan fase DSI dengan pengembangan S-HOT (Novitra et al., 2026). Penelitian ini memperluas arah tersebut ke desain asesmen, tetapi belum menguji hubungan sebab akibat. Oleh karena itu, frasa “meningkatkan S-HOT”, “efektif”, atau “dapat diterapkan” diganti dengan klaim yang lebih terbatas: prototipe memperoleh bukti awal validitas isi berdasarkan penilaian tiga ahli. Keterbatasan lain adalah jumlah validator yang kecil, belum tersedianya interval kepercayaan dari matriks skor mentah, belum adanya validasi ulang setelah revisi, dan belum dilampirkannya seluruh butir serta rubrik. Keterbatasan tersebut harus menjadi dasar tahapan penelitian berikutnya, bukan ditutupi oleh rata-rata nilai yang tinggi.

KESIMPULAN

Asesmen S-HOT berbasis DSI pada materi kinematika memperoleh bukti awal validitas isi berdasarkan penilaian tiga ahli. Rata-rata Aiken’s V adalah 0,88 untuk isi, 0,90 untuk bahasa, 0,94 untuk penyajian, dan 0,94 untuk kegrafisan, dengan rata-rata keseluruhan 0,92. Pada tingkat indikator, 19 dari 27 indikator mencapai nilai kritis $V = 0,92$, sedangkan delapan indikator masih memerlukan revisi. Temuan ini menunjukkan bahwa struktur prototipe telah memadukan fase DSI, dimensi S-HOT, dan karakter multirepresentasi kinematika secara menjanjikan, tetapi belum memberikan bukti tentang validitas konstruk, reliabilitas, kepraktisan, atau efektivitas. Oleh karena itu, prototipe belum dapat dinyatakan meningkatkan S-HOT atau siap digunakan secara luas sebelum revisi dan pengujian empiris dilakukan.

REKOMENDASI

Penelitian berikutnya perlu: (1) merevisi delapan indikator yang belum mencapai nilai kritis dan melakukan validasi ulang dengan panel ahli yang lebih besar; (2) melaporkan profil keahlian validator, matriks skor mentah, serta interval kepercayaan

Aiken's V; (3) melaksanakan uji keterbacaan dan uji coba terbatas untuk menilai kejelasan stimulus, alokasi waktu, dan konsistensi rubrik; (4) menganalisis tingkat kesukaran, daya pembeda, reliabilitas, dan validitas konstruk; serta (5) menguji kepraktisan dan efektivitas melalui implementasi di kelas. Kisi-kisi lengkap, butir asesmen, stimulus, dan rubrik penskoran sebaiknya disertakan sebagai lampiran atau bahan tambahan agar proses pengembangan dapat direplikasi.

ACKNOWLEDGMENT

Penulis menyampaikan terima kasih kepada para validator dan pihak-pihak yang mendukung proses pengembangan prototipe asesmen ini.

REFERENSI

- Akbar, T. A. (2026). Analisis validitas asesmen Scientific Higher Order Thinking berbasis Dual Space Inquiry untuk materi listrik arus searah. *Jurnal Pendidikan MIPA*, 16(3), 346–356.
- Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*, 40(4), 955–959. <https://doi.org/10.1177/001316448004000419>
- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45(1), 131–142. <https://doi.org/10.1177/0013164485451012>
- Asyifa, S. (2026). Pengembangan modul ajar Kurikulum Merdeka berbasis proyek dengan pendekatan STEAM pada materi gerak lurus berubah beraturan. *Prosiding Seminar Nasional Inovasi dan Riset Multidisiplin*, 1(1), 241–248.
- Badan Standar, Kurikulum, dan Asesmen Pendidikan. (2025). Capaian pembelajaran pada pendidikan anak usia dini, jenjang pendidikan dasar, dan jenjang pendidikan menengah (Permendikdasmen No. 11 Tahun 2025). <https://peraturan.bpk.go.id/Details/322487/permendikdasmen-no-11-tahun-2025>
- Delvananta Givarin, & Ellianawati, E. (2025). Analisis pembuatan soal HOTS untuk mengembangkan kemampuan berpikir tingkat tinggi siswa dalam pemecahan masalah fisika. *Journal of Social Humanities and Education*, 1(1), 1–7. <https://doi.org/10.65310/av5pjr10>
- Desiriah, E., & Setyarsih, W. (2021). Tinjauan literatur pengembangan instrumen penilaian HOTS fisika di SMA. *Jurnal Hasil Kajian, Inovasi, dan Aplikasi Pendidikan Fisika*, 7(1), 79–89.
- Effendi, E. (2023). Pengembangan instrumen tes Higher Order Thinking Skill (HOTS) pada materi statistika SMP. *Jurnal Pendidikan dan Konseling*, 4(6), 5076–5080.
- Faizah, F., Yuliana, I. F., & Rohmah, R. S. (2022). Development of objective instruments with HOTS-based reasons reviewed from students' critical thinking skills. *Chemistry Education Practice*, 5(2). <https://doi.org/10.29303/cep.v5i2.3518>
- Fitriana, Sitompul, S. S., & H., M. M. S. (2024). Analisis kemampuan kognitif peserta didik dalam menyelesaikan soal HOTS fisika materi getaran harmonis. *Jurnal Dunia Pendidikan*, 5(2), 553–565.
- Hamidah, M. H., & Wulandari, S. S. (2021). Pengembangan instrumen penilaian berbasis HOTS menggunakan aplikasi Quizizz. *Efisiensi: Kajian Ilmu Administrasi*, 18(1), 105–124. <https://doi.org/10.21831/efisiensi.v18i1.36997>

- Lestari, S. (2021). Pengembangan orientasi keterampilan abad 21 pada pembelajaran fisika melalui pembelajaran PjBL-STEAM berbantuan Spectra-Plus. *Ideguru: Jurnal Karya Ilmiah Guru*, 6(3), 272-279. <https://doi.org/10.51169/ideguru.v6i3.243>
- Loreza, Y., Novitra, F., Mufit, F., & Suherman, D. S. (2026). Pengaruh e-modul berbasis Dual Space Inquiry terhadap keterampilan berpikir kritis siswa pada materi gelombang mekanik. *Science and Education Journal*, 5(2), 2541-2552.
- Musayaroh, T. (2021). Development of a HOTS-based chemical literacy test instrument reviewed from content validity by experts. *Journal of Chemical Education*, 10(3).
- Novitra, F., Abdullah, M. N. S., Özdemir, E., Riyasni, S., Emiliannur, Festiyed, & Metra, P. (2025). Design of Dual Space Inquiry framework for facilitating flexible learning in digital technology era. *International Journal of Educational Research Open*, 8, 100424. <https://doi.org/10.1016/j.ijedro.2024.100424>
- Novitra, F., Sundari, P. D., Muslim, Suherman, D. S., Anshari, R., Hidayati, & Dwiridal, L. (2026). Mobile learning-enhanced Dual Space Inquiry framework for scientific higher-order thinking: A mixed methods investigation. *Social Sciences & Humanities Open*, 13, 103032. <https://doi.org/10.1016/j.ssaho.2026.103032>
- Penfield, R. D., & Giacobbi, P. R., Jr. (2004). Applying a score confidence interval to Aiken's item content-relevance index. *Measurement in Physical Education and Exercise Science*, 8(4), 213-225. https://doi.org/10.1207/S15327841MPEE0804_3
- Plomp, T. (2013). Educational design research: An introduction. In T. Plomp & N. Nieveen (Eds.), *Educational design research* (pp. 10-51). SLO.
- Putranta, H., Supahar, Setiyatna, H., Choiriyah, S., Dwandaru, W. S. B., Widodo, & Warsono. (2021). The effect of smartphone usage intensity on high school students' higher order thinking skills in physics learning. *Journal of Turkish Science Education*, 18(3), 421-438. <https://doi.org/10.36681/tused.2021.82>
- Putri, R. A. H., & Pahlevi, T. (2021). Pengembangan instrumen penilaian berbasis HOTS berbantuan Google Form pada mata pelajaran kearsipan kelas X jurusan OTKP SMKN 2 Kediri. *Journal of Office Administration*, 1(2), 138-152.
- Retnawati, H. (2016). Analisis kuantitatif instrumen penelitian. Parama Publishing.
- Rohmah, U. A., & Sunarti, T. (2020). Profil Higher Order Thinking Skills (HOTS) peserta didik SMA pada materi kalor. *IPF: Inovasi Pendidikan Fisika*, 9(3), 466-472. <https://doi.org/10.26740/ipf.v9n3.p466-472>
- Sugiyono. (2023). Metode penelitian kuantitatif, kualitatif, dan R&D (2nd ed.). Alfabeta.
- Sun, H., Xie, Y., & Lavonen, J. (2022). Exploring the structure of students' scientific higher order thinking in science education. *Thinking Skills and Creativity*, 43, 100999. <https://doi.org/10.1016/j.tsc.2022.100999>
- Zhou, Y., Gan, L., Chen, J., Wijaya, T. T., & Li, Y. (2023). Development and validation of a higher-order thinking skills assessment scale for pre-service teachers. *Thinking Skills and Creativity*, 48, 101272. <https://doi.org/10.1016/j.tsc.2023.101272>
- Zulfa, N. F., Hufri, Darvina, Y., Emiliannur, Hidayati, & Dwiridal, L. (2025). Analysis of the validity of problem-based learning physics modules to facilitate scientific literacy in high school students. *Jurnal Penelitian Pembelajaran Fisika*, 11(2), 170-186. <https://doi.org/10.24036/jppf.v11i2.24>