

Pengembangan Instrumen Penilaian Hots Berbasis Literasi Geografi Untuk Mengukur Kemampuan Berpikir Kritis Pada Siswa SMA

Mohammad Irfandi Ryaniz¹⁾, Sunar Wahid²⁾, & Julinda Siregar³⁾

^{1,2,3}Universitas Indraprasta PGRI, Jakarta 13760, Indonesia

ABSTRACT

This study aims to develop a Higher Order Thinking Skills (HOTS) assessment instrument based on geography literacy to measure the critical thinking skills of public high school students in Jagakarsa District, South Jakarta. The study utilizes the Research and Development (R&D) method, with a development model encompassing the stages of needs analysis, design, expert validation, small-group trials, and large-group trials. The results of the material and evaluation expert validation showed Aiken's V values ranging from 0,83 to 0,92, placing it in the "highly valid" category. Reliability testing in the small-group trial yielded a Cronbach's Alpha coefficient of 0.87 (high category), while the large-group trial reached 0.91 (very high). Item analysis revealed that 85% of the questions fell into the moderate difficulty category, 10% were easy, and 5% were difficult. The discrimination power was dominated by "good" and "very good" categories at 80%, and distractor effectiveness functioned optimally for more than 75% of the items. The measurement of students' critical thinking skills showed an average score of 78.6, categorized as "good." Based on these findings, the developed geography literacy-based HOTS assessment instrument is declared feasible, reliable, and effective for measuring the critical thinking skills of high school students.

Key Words: Assessment Instrument; HOTS; Critical Thinking Skills; Geography Literacy.

ABSTRAK

Penelitian ini bertujuan untuk mengembangkan instrumen penilaian Higher Order Thinking Skills (HOTS) berbasis literasi geografi guna mengukur kemampuan berpikir kritis peserta didik SMA Negeri di Kecamatan Jagakarsa, Jakarta Selatan. Penelitian menggunakan metode *Research and Development* (R&D) dengan model pengembangan meliputi tahap analisis kebutuhan, perancangan, validasi ahli, uji coba kelompok kecil, dan uji coba kelompok besar. Hasil validasi ahli materi dan evaluasi menunjukkan nilai Aiken's V berkisar antara 0,83–0,92 yang termasuk kategori sangat valid. Uji reliabilitas pada uji coba kelompok kecil memperoleh koefisien Cronbach's Alpha sebesar 0,87 (kategori tinggi), sedangkan pada uji coba kelompok besar sebesar 0,91 (sangat tinggi). Analisis butir soal menunjukkan 85% soal berada pada kategori tingkat kesukaran sedang, 10% mudah, dan 5% sukar. Daya pembeda soal didominasi kategori baik dan sangat baik sebesar 80%, serta efektivitas pengecoh berfungsi optimal pada lebih dari 75% butir soal. Hasil pengukuran kemampuan berpikir kritis peserta didik menunjukkan rata-rata skor 78,6 dengan kategori baik. Berdasarkan temuan tersebut, instrumen penilaian HOTS berbasis literasi geografi yang dikembangkan dinyatakan layak, reliabel, dan efektif untuk mengukur kemampuan berpikir kritis peserta didik SMA.

Kata Kunci: Instrumen Penilaian; HOTS; Kemampuan Berpikir Kritis; Literasi Geografi.

Penulis Korespondensi: (1) Mohammad Irfandi Ryaniz, (2) Universitas Indraprasta PGRI, (3) Jl. Nangka No. 58 C, Tanjung Barat. Kec. Jagakarsa, Jakarta Selatan, Indonesia, (4) Email: 17ryaniz@gmail.com

Copyright © 2026. The Author(s). Published by Universitas Indraprasta PGRI, Jakarta, Indonesia. This is an open access article licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

PENDAHULUAN

Pendidikan pada abad ke-21 menuntut transformasi signifikan dalam proses pembelajaran, di mana fokus utama bergeser dari penguasaan materi secara hafalan menjadi pengembangan karakter dan keterampilan berpikir tingkat lanjut. Sejalan dengan pandangan Hamdayama (2016), pendidik memegang peranan sentral dalam ranah pedagogi untuk memastikan penyampaian pengajaran berkualitas tinggi yang mampu mencetak sumber daya manusia yang kompetitif. Hal ini diperkuat oleh Permendikbud Nomor 21 Tahun 2016 yang mengamanatkan bahwa standar isi pendidikan harus membekali peserta didik dengan kompetensi berpikir kritis, kreatif, dan mandiri. Namun, realitas di lapangan menunjukkan adanya kesenjangan yang mencolok. Berdasarkan analisis terhadap butir soal Penilaian Akhir Semester (PAS) dan Ujian Tengah Semester (PTS) Geografi kelas XI tahun 2024, ditemukan bahwa instrumen evaluasi masih didominasi oleh tingkat Lower Order Thinking Skills (LOTS) dengan persentase mencapai 84% hingga 96%. Kondisi ini mengindikasikan bahwa instrumen yang mampu menstimulasi Higher Order Thinking Skills (HOTS) belum terimplementasi secara efektif di sekolah.

Pentingnya instrumen penilaian kognitif yang berkualitas ditegaskan oleh Retnawati (2016) sebagai perangkat yang memenuhi kriteria ilmiah untuk mengukur pencapaian kompetensi siswa. Dalam disiplin Geografi, penilaian tidak boleh sekadar menguji aspek memori, tetapi harus mengintegrasikan literasi geografi atau geo-literacy. Menurut Edelson (2012), literasi geografi mencakup pemahaman sistematis mengenai interaksi, interkoneksi, dan implikasi spasial yang memungkinkan individu mengambil keputusan yang tepat terhadap isu-isu lingkungan. Sejalan dengan itu, Sugiarto (2023) menekankan bahwa penggunaan taksonomi kognitif yang tepat sangat diperlukan untuk mengategorikan tujuan pembelajaran ke dalam ranah intelektual yang lebih tinggi, sehingga siswa mampu mensintesis berbagai ide dan metode untuk mengatasi tantangan keruangan yang kompleks.

Kesenjangan antara regulasi dan praktik evaluasi ini menuntut adanya inovasi dalam pengembangan alat ukur hasil belajar. Sinaga (2020) menyatakan bahwa model pembelajaran dan penilaian HOTS harus dirancang untuk melatih siswa dalam melakukan analisis, evaluasi, dan kreasi. Oleh karena itu, penelitian ini berfokus pada pengembangan instrumen penilaian HOTS berbasis literasi geografi pada materi Keanekaragaman Hayati Indonesia dan Dunia bagi siswa SMA Negeri di Kecamatan Jagakarsa. Melalui metode Research and Development (R&D) yang mengacu pada model pengembangan sistematis seperti ADDIE (Permadi, 2022), penelitian ini bertujuan untuk menghasilkan produk berupa butir soal pilihan ganda dan uraian yang valid dan reliabel. Pengembangan ini diharapkan dapat memberikan sumbangsih teoretis dalam meningkatkan kualitas pembelajaran Geografi serta memberikan rujukan praktis bagi guru dalam menyusun instrumen evaluasi yang secara efektif mampu mengukur dan meningkatkan kemampuan berpikir kritis peserta didik.

Fokus utama dalam pengembangan instrumen ini adalah integrasi indikator berpikir kritis yang diselaraskan dengan materi Keanekaragaman Hayati Indonesia dan Dunia. Sebagaimana dikemukakan dalam landasan teori, kemampuan kognitif tingkat tinggi mengharuskan siswa untuk tidak sekadar mengingat, tetapi mampu mengintegrasikan dan mensintesis berbagai ide serta metode guna mengatasi tantangan yang kompleks (Sugiarto, 2023). Melalui soal-soal pilihan ganda dan uraian yang dikembangkan, peserta didik dilatih untuk melakukan analisis mendalam terhadap fenomena biosfer, sehingga instrumen ini tidak hanya berfungsi sebagai alat ukur hasil belajar, tetapi juga sebagai stimulus untuk meningkatkan kapasitas intelektual dan kemandirian berpikir siswa sesuai dengan target capaian kurikulum yang berlaku.

Secara metodologis, penelitian pengembangan ini menerapkan model ADDIE yang meliputi tahap analisis, perancangan, pengembangan, implementasi, dan evaluasi (Permadi, 2022). Penggunaan model ini menjamin bahwa setiap butir soal yang dihasilkan telah melalui prosedur ilmiah yang ketat, mulai dari identifikasi kebutuhan di lapangan hingga uji validitas dan reliabilitas. Dengan demikian, produk yang dihasilkan diharapkan memiliki tingkat keefektifan, kemenarikan, dan efisiensi yang tinggi dalam proses pembelajaran. Kehadiran instrumen penilaian HOTS yang teruji ini menjadi sangat krusial sebagai rujukan praktis bagi guru geografi dalam menyusun evaluasi yang inovatif, sekaligus memberikan sumbangsih nyata terhadap upaya peningkatan kualitas pendidikan geografi di tingkat SMA.

Penelitian ini bertujuan (1) untuk menganalisis kebutuhan pengembangan instrumen penilaian HOTS berbasis literasi geografi; (2) mengembangkan dan menguji kelayakan instrumen penilaian HOTS berbasis literasi geografi yang sesuai dengan karakteristik materi; dan (3) menguji efektivitas instrumen penilaian HOTS berbasis literasi geografi dalam mengukur kemampuan berpikir kritis siswa.

METODE

Penelitian ini merupakan jenis penelitian pengembangan atau Research and Development (R&D) yang mengadopsi model pengembangan ADDIE, yang terdiri dari tahap Analysis, Design, Development, Implementation, dan Evaluation (Permadi, 2022). Lokasi penelitian ditetapkan di dua sekolah, yaitu SMA Negeri 109 Jakarta dan SMA Negeri 49 Jakarta, dengan subjek penelitian siswa kelas XI tahun ajaran 2025/2026. Prosedur penelitian dimulai dengan analisis kebutuhan untuk mengidentifikasi kesenjangan antara tuntutan kurikulum dengan ketersediaan instrumen HOTS di lapangan, yang kemudian dilanjutkan dengan perancangan draf instrumen penilaian berbasis literasi geografi pada materi Keanekaragaman Hayati Indonesia dan Dunia.

Pada tahap pengembangan, instrumen yang telah dirancang melalui proses validasi oleh ahli materi dan ahli evaluasi untuk memastikan kesahihan isi menggunakan indeks validitas Aiken's V. Selanjutnya, instrumen diuji coba melalui dua tahap, yaitu uji coba kelompok kecil untuk melihat reliabilitas awal dan uji coba kelompok besar untuk menganalisis karakteristik butir soal yang meliputi tingkat kesukaran, daya pembeda, dan efektivitas pengecoh (Sugiarto, 2023). Tahap implementasi dilakukan di kelas nyata untuk mengukur kemampuan berpikir kritis siswa, sementara tahap evaluasi (formatif dan sumatif) dilakukan secara kontinu guna menyempurnakan produk akhir sehingga dihasilkan instrumen penilaian yang layak, reliabel, dan efektif sesuai dengan standar kualitas pengembangan perangkat pembelajaran (Retnawati, 2016).

Prosedur

Penelitian ini menggunakan metode Research and Development (R&D) dengan model pengembangan ADDIE (Analysis, Design, Development, Implementation, Evaluation) (Permadi, 2022). Variabel utama dalam penelitian ini meliputi instrumen penilaian HOTS (Higher Order Thinking Skills) berbasis literasi geografi sebagai variabel yang dikembangkan, serta kemampuan berpikir kritis peserta didik sebagai variabel yang diukur. Desain penelitian mengikuti alur sistematis yang dimulai dari analisis kebutuhan, perancangan draf soal (pilihan ganda), hingga tahap validasi dan uji coba untuk menghasilkan produk final yang layak dan efektif (Sugiarto, 2023). Proses pengumpulan data dilakukan melalui teknik tes dan non-tes. Teknik non-tes meliputi angket analisis kebutuhan, lembar validasi untuk ahli materi dan ahli evaluasi menggunakan indeks Aiken's V, serta angket respon siswa. Teknik tes digunakan untuk mengukur kemampuan berpikir kritis siswa melalui instrumen yang telah dikembangkan. Penelitian ini dilaksanakan di dua lokasi, yaitu SMA Negeri 109 Jakarta dan SMA Negeri 49 Jakarta, dengan subjek siswa kelas XI tahun ajaran 2025/2026. Waktu penelitian berlangsung selama 11 bulan, terhitung mulai dari April 2025 hingga Februari 2026, yang mencakup seluruh

tahapan mulai dari penyusunan proposal, proses pengembangan di lapangan, hingga sidang tesis. Seluruh prosedur ini dilakukan secara sirkular dan evaluatif untuk menjamin keandalan dan kesahihan instrumen penilaian yang dihasilkan (Retnawati, 2016).

Partisipan

Subjek penelitian ini adalah peserta didik kelas XI tahun ajaran 2025/2026 di dua sekolah negeri Kecamatan Jagakarsa, yaitu SMA Negeri 109 Jakarta dan SMA Negeri 49 Jakarta. Teknik pengambilan sampel menggunakan purposive sampling dengan pertimbangan siswa telah menempuh materi Keanekaragaman Hayati Indonesia dan Dunia. Berdasarkan data sekolah, jumlah sampel yang terlibat terdiri dari sampel dibagi ke dalam dua tahap uji coba, yaitu 20% untuk uji coba kelompok kecil (16 siswa) dan 80% untuk uji coba kelompok besar (59 siswa).

Instrumentasi

Penelitian ini menggunakan beberapa strategi pengumpulan data untuk menganalisis kualitas instrumen penilaian berbasis keterampilan berpikir tingkat tinggi (HOTS) pada materi keanekaragaman hayati di dunia dan Indonesia. Penelitian ini dilakukan pada siswa kelas XI peminatan geografi pada SMA Negeri di kecamatan Jagakarsa. Penilaian kualitas instrumen dibuktikan melalui pengujian teoritis terhadap kualitas soal. Untuk mengumpulkan data butir soal yang memenuhi kriteria tertentu terkait materi, konstruksi, dan bahasa digunakan lembar telaah butir soal. Serangkaian wawancara dilakukan untuk menilai kemahiran siswa dalam memahami materi Keanekaragaman Hayati Indonesia dan Dunia. Data yang diperoleh dari wawancara digunakan untuk menilai pemahaman siswa terhadap keterampilan berpikir tingkat tinggi (HOTS) terkait materi tersebut. Tes mengacu pada kumpulan pertanyaan dan instrumen tambahan yang digunakan untuk menilai kapasitas intelektual, tingkat keahlian, keterampilan, dan bakat yang ditunjukkan oleh individu atau suatu entitas kolektif (Arikunto, 2010 dalam Sugiarto, 2023). Penelitian ini meliputi soal tes pilihan ganda untuk menilai materi Keanekaragaman Hayati Indonesia dan Dunia.

Instrumen untuk mengukur kevalidan digunakan lembar validasi (telaah soal tes) yang dianalisis secara kualitatif. Validasi ditinjau dari tiga aspek, yaitu (a) materi, (b) konstruksi, dan (c) bahasa. Soal tes valid/ layak digunakan berdasarkan penilaian dari validator. Kriteria validasi butir soal menggunakan 5 kriteria penilaian yaitu tidak baik (nilai 1), kurang baik (nilai 2), cukup baik (nilai 3), baik (nilai 4), dan sangat baik (nilai 5).

Secara teoritis, kualitas soal tes bentuk objektif dapat ditelaah dengan memperhatikan hal-hal sebagai berikut: (1) Butir harus sesuai dengan indikator yang ditetapkan; (2) Hanya ada satu jawaban yang benar (3) Pengecoh homogen, dan berfungsi; (4) Pokok soal harus dirumuskan secara jelas; (5) Rumusan pokok soal dan pilihan jawaban harus merupakan pernyataan yang diperlukan saja; (6) Pokok soal jangan memberi petunjuk ke arah jawaban benar; (7) Pokok soal jangan mengandung pernyataan yang bersifat negatif ganda; (8) Pilihan jawaban harus homogen dan logis ditinjau dari segi materi; (9) Panjang rumusan pilihan jawaban relatif sama; (10) Pilihan jawaban yang berbentuk angka atau waktu disusun berdasarkan urutan besar kecilnya angka atau kronologis waktunya; (11) Gambar/grafik/tabel/diagram dan sejenisnya harus jelas dan berfungsi; (12) Butir tes tidak tergantung pada jawaban sebelumnya; (13) Bahasa; (14) Menggunakan bahasa yang sesuai dengan kaidah bahasa Indonesia; (15) Menggunakan bahasa yang komunikatif dan mudah dimengerti; (16) Pilihan jawaban jangan mengulang kata atau frase yang bukan merupakan satu kesatuan pengertian; serta (17) Menggunakan istilah baku.

Analisis Data

1. Analisis Kualitatif Soal Tes HOTS

Rekomendasi dan umpan balik yang diberikan oleh instruktur evaluator menjadi landasan untuk menyempurnakan pengembangan produk yang sedang berlangsung. Masukan

dan rekomendasi yang diberikan oleh evaluator yang merupakan dosen dapat mencakup ketidakakuratan materi, rumusan pertanyaan, atau aspek kebahasaan. Selanjutnya, peneliti melanjutkan untuk menyingkat temuan penelitian untuk menginformasikan keputusan untuk perbaikan pada setiap item pertanyaan. Klasifikasi selanjutnya berkaitan dengan hasil penilaian validasi.

Tabel 1. Verifikasi hasil validasi

No	Hasil Review	Kategori
1.	Butir soal tidak ada catatan saran dan masukan	Valid tanpa revisi
2.	Butir soal mendapatkan catatan saran	Valid dengan revisi
3.	Butir soal disarankan diganti	Tidak valid

Berdasarkan hasil rangkuman tabel butir soal dianalisis untuk menentukan keputusan butir soal yang diterima, diperbaiki, atau ditolak. Pedoman yang digunakan untuk menentukan keputusan butir soal diterima, diperbaiki, atau ditolak dapat dilihat pada tabel.

Tabel 2. Rangkuman Penilaian, Penilaian, dan Keputusan

No	Rangkuman Penilaian	Penilaian	Keputusan
1.	Keseluruhan kriteria telaah butir pada aspek materi, konstruksi dan bahasa semua sesuai dengan kaidah dan didukung oleh semua penelaah (81–100%).	Valid tanpa revisi	Butir soal baik (diterima)
2.	Butir tersebut tidak memenuhi sebanyak-banyaknya satu aspek materi, tiga aspek konstruksi serta satu aspek bahasa (70–80%).	Valid dengan revisi	Butir soal diperbaiki sesuai saran validator
3.	Butir tidak memenuhi semua kriteria yang telah ditetapkan (< 69%).	Tidak valid	Butir soal diganti

2. Analisis Kuantitatif Soal Tes HOTS

Teknik analisis data kuantitatif yang digunakan dalam penelitian ini menggunakan aplikasi AnBuso versi 8. Analisis data kuantitatif meliputi:

a. Validitas Isi

Instrumen penilaian tes HOTS mata pelajaran geografi SMA Kelas XI Keanekaragaman Hayati Indonesia dan Dunia. Selanjutnya, proses validasi dilakukan untuk memastikan bahwa instrumen tes secara akurat menangkap konstruk yang ingin dinilai. Pengembangan setiap item soal didasarkan pada grid yang telah merata sesuai dengan uraian mata pelajaran yang termasuk dalam kurikulum. Hal ini memastikan bahwa validasi konten (teoretis) memenuhi persyaratan yang diperlukan. Perhitungan koefisien validitas isi dalam penelitian ini berasal dari penilaian ahli yang diberikan oleh para profesional di bidang pengukuran dan ilmu sosial. Perhitungan hasil expert judgement dilakukan dengan menggunakan rumus Aiken seperti yang dijelaskan oleh Azwar (2016:116).

b. Daya Pembeda Butir Soal

Kapasitas diskriminatif suatu pertanyaan mengacu pada kemampuannya untuk membedakan antara siswa yang telah mencapai kemahiran dalam materi pelajaran yang sedang dipertimbangkan dan mereka yang belum atau telah mencapai tingkat kemahiran yang lebih rendah. Ada beberapa keuntungan yang terkait dengan pembedaan soal tes.

Indeks validitas diskriminan untuk setiap item pertanyaan biasanya direpresentasikan sebagai proporsi. Indeks daya pembeda suatu pertanyaan berkorelasi positif dengan kemampuannya membedakan siswa yang sudah memahami topik dan yang belum. Kisaran indeks kekuatan pembeda adalah dari -1,00 hingga +1,00. Kekuatan atau kualitas suatu soal berbanding lurus dengan tingkat daya pembedanya. Jika nilai diskriminan bernilai negatif (< 0), hal ini menunjukkan bahwa lebih banyak siswa dari kelompok bawah (yakni mereka yang tidak memahami isi) menjawab pertanyaan secara

akurat dibandingkan dengan kelompok atas (yakni siswa dari kelompok atas (yakni siswa dari kelompok atas), mereka yang memiliki pemahaman yang kuat terhadap materi yang disampaikan oleh instruktur).

c. Keefektifan Distraktor/ Pengecoh

Suatu distraktor yang efektif harus dipilih oleh beberapa peserta tes atau minimal dipilih oleh 5% peserta tes (Basuki dan Hariyanto, 2014: 144). Pengecoh dikatakan berfungsi efektif jika banyak dipilih oleh peserta didik tes dari kelompok bawah, sebaliknya jika banyak dipilih oleh kelompok atas, maka distraktor/ pengecoh tersebut tidak berfungsi sebagaimana mestinya.

Efektifitas HOTS peserta didik. Hasil analisis dari respon jawaban peserta didik memuat sejumlah informasi mengenai kemampuan peserta didik yang berupa skor. Hasil kemampuan peserta didik disajikan dalam bentuk grafik distribusi frekuensi kemampuan. Kategori kemampuan higher order thinking skills peserta didik diperoleh berdasarkan dari nilai peserta didik yang dikelompokkan pada kelas interval. Dalam memperoleh hasil efektifitas peningkatan berpikir kritis pada dalam materi yang diujikan.

HASIL

Hasil Analisis Kebutuhan

Hasil analisis kebutuhan menunjukkan saat proses penilaian pembelajaran di kelas XI Peminatan Geografi masih didominasi oleh instrumen evaluasi yang berorientasi pada kemampuan berpikir tingkat rendah (LOTS). Berdasarkan hasil studi pendahuluan melalui wawancara dengan guru, sebagian besar soal yang digunakan masih menekankan pada kemampuan mengingat, memahami, dan menyebutkan konsep atau soal LOTS, sementara tuntutan Kurikulum Merdeka mengharuskan peserta didik menguasai kemampuan berpikir tingkat tinggi (HOTS). Kondisi ini menyebabkan peserta didik belum terbiasa menganalisis fenomena geografis secara kritis, mengaitkan konsep dengan konteks nyata, serta menarik simpulan berbasis data spasial dan lingkungan. Dengan demikian, diperlukan pengembangan instrumen penilaian yang tidak hanya mengukur penguasaan konsep, tetapi juga kemampuan analitis, evaluatif, dan reflektif peserta didik.

Dengan keberagaman peserta didik, pembelajaran geografi pada kelas peminatan ini dapat lebih dioptimalkan dengan pendekatan geoliterasi. Menggunakan geoliterasi ini membuat guru semakin mudah dalam meningkatkan pemahaman peserta didik terkait gejala-gejala keruangan dengan pengalaman pada setiap individu peserta didik. Menelaah kebutuhan instrumen soal yang berbasis HOTS di kelas XI Peminatan Geografi materinya adalah Keanekaragaman Hayati Indonesia dan Dunia. Maka disusunlah kisi-kisi pengembangan soal berbasis HOTS dengan mengoptimalkan pendekatan geoliterasi. Berdasarkan kisi kisi dan batasan yang sudah disusun dalam kebutuhan instrumen soal HOTS berbasis literasi geografi pada materi yang dikembangkan, tersusun 20 butir soal HOTS berbasis literasi geografi dengan sebaran 45% merujuk pada literasi geografi Edelson komponen interaksi, 45% komponen literasi interkoneksi dan 10% komponen literasi implikasi.

Pengembangan instrumen penilaian HOTS berbasis literasi geografi dilakukan melalui beberapa tahap penyusunan model. Model draf 1 merupakan rancangan awal instrumen yang disusun berdasarkan hasil analisis kebutuhan, indikator pembelajaran, dan kisi-kisi soal. Instrumen dikembangkan dalam bentuk soal pilihan ganda yang memuat stimulus kontekstual terkait fenomena keanekaragaman hayati Indonesia dan dunia. Setiap butir soal dirancang untuk mengukur kemampuan berpikir tingkat tinggi peserta didik, khususnya pada aspek analisis, evaluasi, dan pemecahan masalah yang terintegrasi dengan komponen literasi geografi, yaitu interaksi, interkoneksi, dan implikasi.

Selanjutnya, model draf 2 merupakan hasil revisi dari draf awal berdasarkan masukan validator yang mencakup aspek materi, konstruksi, dan bahasa. Perbaikan dilakukan pada kejelasan stimulus, ketepatan konsep geografi, serta penyempurnaan alternatif jawaban dan pengecoh. Tahap berikutnya menghasilkan model final setelah dilakukan uji coba empiris melalui analisis butir soal yang meliputi validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Hasil analisis tersebut digunakan untuk melakukan revisi lanjutan sehingga diperoleh instrumen penilaian HOTS berbasis literasi geografi yang layak digunakan untuk mengukur kemampuan berpikir kritis peserta didik.

Kelayakan Model (Teoritik dan Empiris)

1) Kelayakan Model Teoritik

Telaah kuantitatif dianalisis menggunakan formula Aiken's V guna menentukan Content Validity Coefficient (CVC). Kemudian sebelum tahap analisis kuantitatif dengan formula Aiken's V, peneliti akan melakukan telaah validasi ahli dengan validator guru pengampu geografi dari dua sekolah yang menjadi objek penelitian. Tahap validasi ahli bertujuan untuk melihat kelayakan instrumen soal HOTS berbasis literasi geografi yang sudah disusun. Hasil dari tahap validasi ahli ini akan menghasilkan saran, masukan, serta penilaian instrumen soal HOTS berbasis literasi geografi expert judgment untuk kesimpulan layak atau tidak soal untuk digunakan. Berikut disajikan hasil analisis validasi ahli yang dianalisis kuantitatif menggunakan formula Aiken's V:

Tabel 3. Analisis Kuantitatif Ahli

No soal	Aiken's V	Kategori	No soal	Aiken's V	Kategori
1	0,67	Soal diganti	11	0,88	Valid tanpa revisi
2	0,85	Valid tanpa revisi	12	0,88	Valid tanpa revisi
3	0,82	Valid tanpa revisi	13	0,82	Valid tanpa revisi
4	0,97	Valid tanpa revisi	14	0,87	Valid tanpa revisi
5	0,72	Valid dengan revisi	15	0,82	Valid tanpa revisi
6	0,83	Valid tanpa revisi	16	0,88	Valid tanpa revisi
7	0,83	Valid tanpa revisi	17	0,88	Valid tanpa revisi
8	0,83	Valid tanpa revisi	18	0,88	Valid tanpa revisi
9	0,88	Valid tanpa revisi	19	0,90	Valid tanpa revisi
10	0,87	Valid tanpa revisi	20	0,88	Valid tanpa revisi

Hasil dari validator yaitu terdapat 1 butir soal (No. 1) yang memperoleh nilai Aiken's V sebesar 0,67 sehingga dikategorikan sebagai "Soal diganti." Selain itu, terdapat 2 butir soal (No. 5 dan No. 15) yang memiliki nilai Aiken's V masing-masing 0,72 dan dikategorikan "Valid dengan revisi." Adapun 17 butir soal lainnya memperoleh nilai Aiken's V $\geq 0,80$ dan dikategorikan "Valid tanpa revisi," dengan nilai tertinggi terdapat pada butir soal No. 4 sebesar 0,97. Hal ini menunjukkan bahwa secara umum kualitas validitas isi instrumen berada pada kategori tinggi hingga sangat tinggi.

2) Kelayakan Model Empiris

Kelayakan model draft 1 yang sudah di revisi oleh validator yang kemudian model draft 2 ini diuji coba pada kelompok kecil. Hasil uji coba lapangan awal pada kelompok kecil atau dilakukan pada 2 sekolah tersebut. Uji coba lapangan awal dikerjakan secara daring menggunakan google form Tahap uji coba awal ini dilakukan dengan mengerjakan soal pilihan ganda sebanyak 20 soal. Sampel uji coba kelompok kecil ini diambil hanya 20% sampel dari keseluruhan sampel sebanyak 16 siswa, 10 siswa dari SMA Negeri 109 dan 6 siswa dari SMA Negeri 49. Siswa yang mengerjakan uji coba ini merupakan siswa yang berbeda kelas dari siswa uji coba lapangan kelompok besar. Analisis yang digunakan untuk mendapatkan

informasi kualitatif setiap butir soal yang sudah dikembangkan, peneliti olah menggunakan analisis butir soal AnBuso versi 8.

b. Efektivitas Model

Kualitas instrumen HOTS berbasis literasi geografi pada materi yang diujikan, ditelaah berdasarkan karakteristik setiap butir soal yang menunjukkan tingkat kesukaran, daya beda, dan revisi pengecoh. Berikut disajikan ringkasan hasil interpretasi analisis instrumen soal HOTS berbasis literasi geografi tahap uji coba lapangan pada kelompok besar. Hasil analisis butir soal tentang tingkat kesukaran soal menggunakan analisis Anbuso versi 8, dapat dilihat pada tabel berikut ini:

Tabel 4. Tingkat kesukaran soal

Kategori	Nomor Soal	Jumlah Soal	Persentase (%)
$P < 0,3$; <i>Sukar</i>	6, 7, 10, 17	4	20
$0,3 \leq p \leq 0,7$; <i>Sedang</i>	1, 4, 5, 9, 11, 12, 13, 14, 16, 18	10	50
$p > 0,7$; <i>Mudah</i>	2, 3, 8, 15, 19, 20	6	30

Tabel 5. Daya pembeda soal

Koefisien	Nomor Soal	Jumlah Soal	Persentase (%)	Ket
$>0,7$	1, 2, 3, 4, 5, 6, 14, 16	8	40	Sangat Baik
$0,4-0,7$	7, 8, 9, 10, 12, 13, 17, 18	8	40	Baik
$0,2 - 0,4$	11, 15, 19, 20	4	20	Cukup Baik
$< 0,2$	-	0	0	Jelek
Negatif	-	0	0	Tidak Baik

Tabel 6. Pengecoh Pilihan ganda

Kategori	Nomor Soal	Jumlah Soal	Persentase (%)
Sangat Baik	2	1	5
Baik	1, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 16, 17, 18	15	75
Perlu Revisi	8, 15, 19, 20	4	20

Berdasarkan hasil uji lapangan kelompok besar, instrumen penilaian HOTS berbasis literasi geografi yang dikembangkan terbukti efektif dalam mengukur kemampuan berpikir kritis peserta didik. Hal ini ditunjukkan oleh dominasi butir soal dengan daya pembeda pada kategori baik dan sangat baik, sebaran tingkat kesukaran yang relatif proporsional, serta efektivitas pengecoh yang umumnya berfungsi secara optimal. Karakteristik empiris tersebut mengindikasikan bahwa instrumen mampu membedakan peserta didik berdasarkan tingkat kemampuan berpikir tingkat tinggi, khususnya dalam menganalisis dan mengevaluasi fenomena keanekaragaman hayati dalam konteks keruangan dan lingkungan. Dengan demikian, instrumen model final tidak hanya layak digunakan, tetapi juga efektif sebagai alat ukur kemampuan berpikir kritis berbasis literasi geografi.

DISKUSI

Berdasarkan hasil analisis kebutuhan yang dilakukan melalui wawancara dengan guru geografi, telaah dokumen soal penilaian, serta hasil survei MGMP Geografi, diperoleh temuan bahwa instrumen penilaian yang digunakan selama ini masih didominasi oleh soal-soal yang menuntut kemampuan berpikir tingkat rendah (Lower Order Thinking Skills/LOTS). Hal ini terlihat dari proporsi soal HOTS yang masih sangat rendah, yaitu berkisar antara 4–16% dari keseluruhan butir soal yang digunakan dalam Penilaian Tengah Semester dan Penilaian Akhir Semester. Selain itu, guru mengalami kesulitan dalam merancang soal yang mampu mengukur kemampuan berpikir kritis siswa secara kontekstual, khususnya yang mengintegrasikan literasi geografi. Keterbatasan contoh soal, panduan penyusunan instrumen HOTS, serta belum

tersedianya instrumen penilaian HOTS yang terstandar untuk mata pelajaran geografi menjadi faktor utama yang melatarbelakangi kebutuhan pengembangan instrumen ini.

Faktor pendukung lainnya adalah adanya relevansi konteks soal dengan materi geografi kelas XI dengan materi Keanekaragaman Hayati di Indonesia dan Dunia. Butir-butir soal yang dikembangkan memberikan nilai autentik pada instrumen ini. Hal ini sejalan dengan prinsip asesmen autentik dalam Kurikulum Merdeka yang menekankan keterkaitan antara pembelajaran di kelas dengan realitas kehidupan peserta didik.

Faktor penghambat pengembangan model terletak pada keterbatasan pengalaman melakukan tahapan penelitian RnD ini. Tahapan penelitian RnD ini menjadi hambatan adalah saat dalam tahapan validasi ahli yang harus dilakukan sebelum uji coba empiris. Instrumen pada tahap draf 1 belum melalui proses telaah pakar, sehingga beberapa kelemahan konseptual dan redaksional baru teridentifikasi setelah dilakukan uji coba awal dan uji coba lapangan. Kondisi ini berdampak pada rendahnya kualitas awal beberapa butir soal, khususnya dalam hal kejelasan stimulus, ketepatan pengecoh, dan kesesuaian tingkat kesukaran.

Model instrumen yang dikembangkan ini memiliki beberapa kekuatan utama. Pertama, model ini dirancang secara kontekstual dengan mengintegrasikan isu-isu lingkungan yang berkaitan dengan keanekaragaman hayati di Indonesia dan dunia dan keruangan yang relevan dengan isu-isu terkini. Hal ini meningkatkan validitas isi dan relevansi instrumen. Kedua, model menekankan integrasi antara HOTS dan literasi geografi, sehingga tidak hanya mengukur penguasaan konsep, tetapi juga kemampuan berpikir kritis, analitis, dan reflektif dalam konteks spasial. Integrasi ini merupakan nilai tambah dibandingkan instrumen konvensional yang cenderung mengukur aspek kognitif rendah. Ketiga, hasil analisis butir soal memberikan dasar empiris yang kuat untuk perbaikan berkelanjutan. Data tingkat kesukaran, daya pembeda, dan efektivitas pengecoh menjadi landasan objektif dalam merevisi dan menyempurnakan model.

Model ini juga memiliki sejumlah kelemahan. Kelemahan utama terletak pada kualitas awal beberapa butir soal yang belum optimal akibat tidak dilakukannya validasi ahli pada tahap awal. Hal ini menyebabkan perlunya revisi signifikan setelah uji coba kelompok kecil. Selain itu, masih terdapat butir soal yang daya pengecohnya kurang efektif, sehingga tidak mampu membedakan secara jelas antara peserta didik berkemampuan tinggi dan rendah. Kelemahan ini menunjukkan bahwa penyusunan pengecoh merupakan aspek krusial yang memerlukan perhatian khusus dalam pengembangan instrumen pilihan ganda berbasis HOTS. Kelemahan lainnya adalah keterbatasan generalisasi. Model instrumen yang dikembangkan dan diuji dalam konteks tertentu, sehingga penerapannya di konteks sekolah atau daerah lain memerlukan adaptasi lebih lanjut seperti pada table berikut ini:

Tabel 7. Kualitas Butir Soal Berdasarkan Daya Kesukaran, Daya Pembeda, dan Efektivitas Pengecoh

No Soal	Daya Kesukaran	Predikat	Daya Pembeda	Predikat	Efektivitas Pengecoh
1	0,63	Sedang	1,00	Sangat baik	Sangat baik
2	0,80	Mudah	0,74	Sangat baik	Sangat baik
3	0,73	Mudah	0,79	Sangat baik	Sangat baik
4	0,49	Sedang	0,79	Sangat baik	Sangat baik
5	0,46	Sedang	0,89	Sangat baik	Sangat baik
6	0,28	Sukar	1,00	Sangat baik	Sangat baik
7	0,26	Sukar	0,42	Baik	Baik
8	0,71	Mudah	0,53	Baik	Baik
11	0,60	Sedang	0,34	Cukup baik	Cukup baik
12	0,68	Sedang	0,42	Baik	Baik
13	0,66	Sedang	0,39	Baik	Baik
14	0,61	Sedang	0,87	Sangat baik	Sangat baik
15	0,78	Mudah	0,26	Cukup baik	Cukup baik
16	0,59	Sedang	0,79	Sangat baik	Sangat baik

17	0,29	Sukar	0,55	Baik	Baik
18	0,67	Sedang	0,61	Baik	Baik
19	0,83	Mudah	0,26	Cukup baik	Cukup baik
20	0,83	Mudah	0,26	Cukup baik	Cukup baik

Analisis butir soal yang mencakup daya kesukaran, daya pembeda, dan efektivitas pengecoh memberikan gambaran empiris mengenai kualitas instrumen asesmen yang dikembangkan. Hasil analisis menunjukkan bahwa distribusi tingkat kesukaran soal relatif proporsional, dengan sebagian besar butir berada pada kategori sedang (50%), diikuti kategori mudah (30%) dan sukar (20%). Komposisi ini menunjukkan bahwa instrumen mampu mencakup variasi kemampuan peserta didik secara lebih komprehensif. Ditinjau dari daya pembeda, mayoritas butir soal berada pada kategori baik hingga sangat baik, yang menunjukkan kemampuan instrumen dalam membedakan peserta didik berkemampuan tinggi dan rendah. Butir soal dengan daya pembeda tinggi umumnya memiliki tingkat kesukaran sedang hingga sukar dan menuntut peserta didik untuk menganalisis stimulus, mengaitkan informasi, serta menggunakan penalaran dalam menjawab permasalahan kontekstual,

Meskipun demikian, beberapa butir soal masih memiliki daya pembeda pada kategori cukup baik serta tingkat kesukaran yang relatif mudah, sehingga kontribusinya dalam mengukur kemampuan berpikir tingkat tinggi masih terbatas. Hal ini menunjukkan bahwa sebagian butir soal masih perlu disempurnakan, terutama pada redaksi pertanyaan dan kompleksitas tuntutan kognitifnya. Dari aspek efektivitas pengecoh, sebagian besar alternatif jawaban berada pada kategori baik hingga sangat baik, yang menunjukkan bahwa pengecoh mampu berfungsi secara optimal dalam menarik respons peserta didik yang belum menguasai konsep secara mendalam. Secara keseluruhan, hasil analisis menunjukkan bahwa instrumen asesmen yang dikembangkan telah memiliki kualitas yang baik dalam mendukung pengukuran HOTS dan literasi geografi, meskipun masih memerlukan beberapa perbaikan untuk meningkatkan kualitas pengukurannya.

Secara keseluruhan, hasil pengembangan, kelayakan, dan efektivitas instrumen penilaian HOTS berbasis literasi geografi yang telah dipaparkan pada Bab IV menunjukkan bahwa tujuan penelitian telah tercapai. Instrumen yang dikembangkan tidak hanya menjawab kebutuhan nyata dalam praktik asesmen pembelajaran geografi, tetapi juga telah melalui proses pengembangan yang sistematis serta pengujian teoretik dan empiris yang memadai.

SIMPULAN

Berdasarkan hasil penelitian pengembangan instrumen penilaian HOTS berbasis literasi geografi pada materi Keanekaragaman Hayati, dapat disimpulkan bahwa:

1. Kebutuhan pengembangan instrumen penilaian HOTS berbasis literasi geografi pada siswa kelas XI Peminatan Geografi SMA Negeri di Kecamatan Jagakarsa, Jakarta Selatan tergolong tinggi. Hal ini ditunjukkan oleh masih dominannya penggunaan soal berorientasi berpikir tingkat rendah serta keterbatasan guru dalam menyusun soal HOTS yang terintegrasi dengan literasi geografi.
2. Instrumen penilaian HOTS berbasis literasi geografi berhasil dikembangkan menggunakan model ADDIE dan dinyatakan layak digunakan berdasarkan hasil validasi ahli (materi, konstruksi, dan bahasa) serta analisis validitas, reliabilitas, tingkat kesukaran, dan daya pembeda butir soal.
3. Instrumen yang dikembangkan efektif dalam mengukur kemampuan berpikir kritis siswa melalui soal yang menuntut kemampuan analisis, evaluasi, dan pemecahan masalah dalam konteks fenomena keanekaragaman hayati Indonesia dan dunia.

REFERENSI

- Azwar. (2016). *Penyusunan tes kemampuan kognitif*. Yogyakarta: Pustaka Pelajar
- Basuki, I., & Hariyanto. (2014). *Asemen pembelajaran*. Bandung: Remaja Rosdakarya.
- Hamdayama, Jumanta. (2016). *Metodologi pengajaran*. Bumi Aksara.
- Permadi, M. S. (2022). *Pengembangan Media Pembelajaran Aplikasi Android PBL-GIS Berbasis Geoliteracy Untuk Meningkatkan Spatial Thinking Kelas X Materi DAS Dan Banjir*. Tesis. Surakarta: Universitas Sebelas Maret.
- Retnawati, H. (2016). *Validitas, Reliabilitas Dan Karakteristik Butir*. Yogyakarta: Nuamedika.
- Sinaga, Fadhillatu Jahra. (2020). *Pengembangan Instrumen Penilaian Berbasis Hots (Keterampilan Berpikir Tingkat Tinggi)*. Prosiding Seminar Nasional PBSI-III Tahun 2020.
- Sugiarto, Freddyansyah. (2023). *Pengembangan Instrumen Penilaian HOTS Berbasis Literasi Geografi untuk Meningkatkan Kemampuan Berpikir Kritis Siswa SMP Insan Cendekia Al Mujtaba Tahun Pelajaran 2022/2023 (Materi: Maritim dan Agrikultur Indonesia)*. Tesis. Surakarta: UNS.
- Yusuf, M. E., dkk. (2022). Analisis Butir Soal Penilaian Tengah Semester Gasal Mata Pelajaran Pendidikan Jasmani Olahraga dan Kesehatan Tingkat Keterampilan Berpikir Kritis pada Kelas VI SD Se-Kota Banjarbaru. *Stabilitas: Jurnal Pendidikan Jasmani dan Kesehatan*, 3(1).