

Pengembangan Instrumen Tes Kemampuan Berpikir Kritis Pada Mata Pelajaran IPAS Sekolah Dasar

Desi Dwi Prastiti

Universitas Negeri Semarang¹

Tri Astuti

Universitas Negeri Semarang²

Wiwi Isnaeni

Universitas Negeri Semarang³

Endang Susilaningsih

Universitas Negeri Semarang⁴

Alamat Korespondensi:

desidwip@students.unnes.ac.id¹

triastuti@mail.unnes.ac.id²

wiwisna@mail.unnes.ac.id³

endangsusilaningsih@mail.unnes.ac.id⁴

ABSTRACT : This study aimed to develop a FRISCO-oriented (Focus, Reason, Inference, Situation, Clarity, Overview) critical thinking test instrument for fourth-grade elementary students in IPAS learning on Indonesian cultural diversity material. The study employed a Research and Development (R&D) method using the ADDIE model. The population consisted of elementary school students, while the sample involved 20 fourth-grade students of SDN Kejawan Putih 1/243 Surabaya selected through purposive sampling. Research instruments included validation sheets and a 20-item multiple-choice critical thinking test. Content validity was assessed by three subject experts and two classroom teachers. Findings showed a content validity score of 0.89 (very valid) and a reliability coefficient of 0.842 (high category). Empirical testing indicated that 80% of items had moderate difficulty and good discrimination power, demonstrating that the instrument is feasible, reliable, and effective for measuring students' critical thinking skills.

Keywords: critical thinking, assessment instrument, elementary school, IPAS

I. Pendahuluan

Perkembangan teknologi dalam dunia pendidikan telah membuka peluang baru dalam memberikan perubahan bagi kehidupan manusia. Seiring dengan perkembangan teknologi yang semakin pesat, terdapat pula kecakapan esensial yang perlu dikembangkan untuk meningkatkan sumber daya manusia (SDM) di Indonesia. Keterampilan tersebut dikenal sebagai 4C di antaranya *creativity* (kreativitas), *critical thinking* (berpikir kritis), *communication* (komunikasi), dan *collaboration* (kolaborasi) (Novianti et al., 2023). Pada perkembangan era revolusi 4.0, kemampuan berpikir kritis menjadi salah satu kemampuan yang perlu dikembangkan sejak dini. Facione (1990) mendefinisikan berpikir kritis sebagai proses kognitif yang melibatkan interpretasi, analisis, evaluasi, inferensi, penjelasan, dan regulasi diri. Kemampuan berpikir kritis perlu dilatih melalui berbagai stimulus dan dukungan lingkungan (Sinaga et al., 2022). Lingkungan sekolah menjadi salah satu wadah yang dapat dimanfaatkan dalam menumbuhkan berbagai karakter dan kompetensi bagi peserta didik seperti berpikir kritis, kreatif, komunikatif, dan kolaborasi (Sari et al., 2025). Pendidikan menjadi salah satu pilar dalam mengembangkan kemampuan berpikir kritis peserta didik melalui proses pembelajaran.

Mata pelajaran Ilmu Pengetahuan Alam dan Sosial (IPAS) merupakan salah satu mata pelajaran yang wajib ditempuh di tingkat sekolah dasar. Mata pelajaran Ilmu Pengetahuan Alam dan Sosial (IPAS) pada Kurikulum Merdeka menjadi salah satu mata pelajaran yang potensial dalam menumbuhkan kemampuan berpikir tingkat tinggi HOTS (*Higher-Order Thinking Skills*), meliputi kemampuan menganalisis, mengevaluasi, dan menciptakan bagi peserta didik. Namun, berbagai studi menunjukkan bahwa kemampuan berpikir kritis peserta didik SD di Indonesia masih relatif rendah (Rahayu et al., 2023; Wahyuni & Kurniawati, 2024). Salah satu faktor yang mempengaruhi pengembangan kemampuan berpikir kritis peserta didik yaitu tipe soal yang disajikan seringkali masih berbasis hafalan atau pemahaman sederhana. Jenis soal pemahaman sederhana kurang mampu merangsang peserta didik untuk berpikir secara mendalam, menganalisis, dan mengevaluasi. Sejalan dengan penelitian Ruci et al. (2023) jenis soal LOTS (*Low Order Thinking Skills*) memiliki efektivitas rendah dalam mendorong kemampuan berpikir kritis peserta didik. Peserta didik perlu dihadapkan dengan soal yang memiliki tingkat pemahaman yang mendalam, sehingga mampu merangsang peserta didik untuk memaksimalkan kemampuan berpikir kritis. Perlu adanya suatu instrumen tes yang dapat menunjang

kemampuan berpikir kritis peserta didik sehingga peserta didik mampu menganalisis dan mengevaluasi secara lebih mendalam.

Pengembangan soal HOTS pada mata pelajaran IPAS menjadi sebuah hal krusial dalam mengatasi permasalahan tersebut. Dalam konteks Taksonomi Bloom yang direvisi, kemampuan berpikir kritis berada pada ranah C4 hingga C6, yaitu menganalisis (*analyze*), mengevaluasi (*evaluate*), dan mencipta (*create*). Sementara itu, Taksonomi SOLO (*Structure of Observed Learning Outcome*) memberikan kerangka struktural untuk mengidentifikasi level pemahaman peserta didik secara hierarkis, mulai dari *prestructural* hingga *extended abstract* (Biggs & Collis, 1982). Pengembangan soal HOTS pada mata pelajaran IPAS masih menghadapi tantangan pada kurangnya pemahaman guru mengenai konsep soal HOTS. Kesulitan dalam menyusun soal HOTS sesuai dengan indikator dan tujuan pembelajaran seringkali masih menjadi hambatan bagi guru (Hulaipah et al., 2023). Berdasarkan hasil wawancara guru kelas IV di SDN Kejawan Putih I/243 Surabaya menunjukkan bahwa minimnya pemahaman guru dalam mengembangkan soal HOTS menyebabkan guru belum yakin untuk menentukan soal yang dirancang sudah berorientasi HOTS atau belum.

Berdasarkan uraian di atas, maka diperlukan upaya dalam mengintegrasikan soal-soal HOTS untuk menunjang kemampuan berpikir kritis peserta didik. Meskipun penelitian mengenai instrumen HOTS dan kemampuan berpikir kritis telah banyak dilakukan, sebagian besar penelitian hanya berfokus pada level kognitif Taksonomi Bloom tanpa mengintegrasikan indikator berpikir kritis secara spesifik. Kebaruan penelitian ini terletak pada pengembangan instrumen tes yang mengintegrasikan aspek FRISCO (*Focus, Reason, Inference, Situation, Clarity, dan Overview*), level kognitif Taksonomi Bloom revisi (C4–C6), serta konteks materi Kekayaan Budaya Indonesia pada mata pelajaran IPAS kelas IV Sekolah Dasar. Integrasi ketiga komponen tersebut diharapkan menghasilkan instrumen yang lebih komprehensif dalam mengukur kemampuan berpikir kritis peserta didik dibandingkan instrumen HOTS yang hanya berorientasi pada level kognitif. Hal ini diharapkan dapat dijadikan sebagai referensi guru dalam mengembangkan soal HOTS khususnya pada mata pelajaran IPAS. Selanjutnya diharapkan dapat mempersiapkan generasi muda dalam menghadapi tantangan di masa yang akan datang.

2. Metode

Penelitian ini menggunakan metode *Research and Development* (R&D) dengan model pengembangan ADDIE yang terdiri dari lima tahapan: *Analysis, Design, Development, Implementation, dan Evaluation*. Model ADDIE dipilih karena bersifat sistematis dan berorientasi pada proses yang terstruktur dalam pengembangan (Branch, 2009).

Pada tahap *Analysis*, dilakukan analisis kebutuhan dan analisis kurikulum melalui observasi pembelajaran IPAS kelas IV di sekolah dasar, yang bertujuan menentukan urgensi pengembangan dan syarat kelayakan produk. Tahap *Design* pada penelitian ini yaitu membuat rancangan produk dengan membuat kisi-kisi tes. Kisi-kisi instrumen disusun berdasarkan Capaian Pembelajaran (CP) Kurikulum Merdeka, penetapan indikator HOTS, dan penentuan level Taksonomi Bloom (C4, C5, C6). Tahap *Development* meliputi penulisan butir soal, penyusunan rubrik penilaian, dan validasi ahli. Penulisan butir soal berdasarkan kisi-kisi bertujuan menghasilkan produk berupa tes untuk mengukur kemampuan berpikir kritis peserta didik. Instrumen terdiri dari 20 butir soal pilihan ganda dengan empat opsi jawaban (A, B, C, D). Kemudian produk akan divalidasi oleh 5 validator yaitu 3 ahli materi dan 2 guru kelas IV. Masukan dan saran dari kelima validator akan dijadikan acuan dalam memperbaiki produk. Tahap selanjutnya yaitu *Implementation* yaitu produk pengembangan yang telah valid, akan diuji coba. Subjek uji coba terbatas adalah 20 peserta didik kelas IV SDN Kejawan Putih I/243 Surabaya yang dipilih menggunakan teknik *purposive sampling*. Tahap terakhir yaitu *Evaluation* yang mana pada tahap ini akan dilakukan analisis data uji validitas empiris dan reliabilitas yang

telah diperoleh pada tahap implementasi serta dilakukan revisi terhadap produk. Analisis data hasil uji coba dilakukan untuk menentukan kelayakan instrumen secara keseluruhan.

Teknik pengumpulan data dalam penelitian ini berupa angket validasi dan tes. Instrumen yang digunakan berupa lembar validasi dan lembar tes. Teknik analisis data yaitu analisis data kuantitatif dan kualitatif. Data kuantitatif berupa angka dari hasil validasi ahli serta analisis data uji validitas empiris dan reliabilitas. Sedangkan analisis kualitatif berupa saran dan masukan validator. Adapun rumus yang digunakan dalam menganalisis data penelitian dijabarkan sebagai berikut:

2.1 Analisis Data Validasi Ahli

Analisis data validitas butir soal menggunakan rumus *Aiken's Validity* yaitu $V = \frac{\sum s}{n(c-1)}$ dengan V adalah rata-rata validitas, $\sum s$ adalah jumlah skor yang diberikan validator secara keseluruhan, n adalah jumlah validator, dan c adalah jumlah kategori skala penilaian. Kriteria validitas ahli berdasarkan tabel 1 berikut ini.

Tabel 1. Kriteria Validitas Ahli

Nilai V	Kriteria
<0,2	Tidak valid
<0,5	Kurang valid
<0,75	Valid
>0,75	Sangat valid

Sumber : Maulana (2022)

Produk yang dikembangkan dapat dikatakan valid jika memiliki skor rata-rata >0,75 sehingga layak untuk diujicobakan. Sedangkan jika skor yang didapat rata-rata berada pada nilai <0,2 maka diharuskan melakukan revisi dan validasi kembali oleh ahli hingga produk layak untuk diujicobakan.

2.2 Analisis Data Uji Validitas Empiris

Analisis data validitas empiris digunakan untuk mendapatkan hasil tes terkait validitas instrumen tes yang dikembangkan. Uji validitas empiris untuk soal pilihan ganda menggunakan Anates Versi 4. Berikut disajikan kriteria interpretasi koefisien validitas Anates pada tabel 2 berikut ini.

Tabel 2. Kriteria Interpretasi Koefisien Validitas Anates

Koefisien	Interpretasi
0,80 – 1,00	Sangat tinggi/sangat baik
0,60 – 0,79	Tinggi
0,40 – 0,59	Cukup
0,20 – 0,39	Rendah
0,00 – 0,19	Sangat rendah
Negatif	Tidak valid/ perlu dibuang/direvisi

Sumber : Arikunto (2018)

Dalam uji validitas empiris menggunakan Anates versi 4, nilai validitas diperoleh dari hasil korelasi antara skor setiap butir soal dengan skor total tes. Pada penelitian ini validitas empiris menggunakan instrumen tes pilihan ganda sebanyak 20 butir soal dengan pilihan jawaban A,B,C, dan D. Setiap butir soal diberikan skor 1 untuk jawaban yang benar dan skor 0 untuk jawaban yang salah.

2.3 Analisis Data Uji Reliabilitas

Analisis reliabilitas digunakan untuk mengetahui tingkat reliabilitas atau konsistensi instrumen dimana butir soal dalam instrumen penelitian ini dapat memberikan hasil pengukuran yang konsisten dan stabil dalam mengukur kemampuan berpikir kritis berdasarkan aspek FRISCO. Analisis reliabilitas menggunakan *Cronbach's Alpha*. Instrumen dikatakan reliabel

jika nilai *Cronbach's Alpha* memenuhi kriteria yang ditetapkan yaitu $>0,90$ kategori sangat tinggi, $0,80-0,90$ kategori tinggi, $0,70-0,79$ kategori cukup/reliabel, $0,60-0,69$ kategori sedang, $0,50-0,59$ kategori rendah, dan $<0,50$ kategori sangat rendah (Fraenkel et al., 2019).

3. Hasil dan Pembahasan

Hasil penelitian mencakup proses pengembangan instrumen tes untuk mengukur kemampuan berpikir kritis peserta didik pada mata pelajaran IPAS yang terdiri dari 5 tahapan yaitu : (1) Tahap *Analysis*, (2) Tahap *Design*, (3) Tahap *Development*, (4) Tahap *Implementation*, dan (5) Tahap *Evaluation*.

3.1 Tahap *Analysis*

Tahap ini terdiri dari analisis kebutuhan awal yaitu dengan mengumpulkan teori ahli yang relevan dengan topik pada penelitian, teori ini kemudian digunakan sebagai rujukan dalam memenuhi kebutuhan penelitian. Menurut Ennis (2011) aspek kemampuan berpikir kritis terdiri dari FRISCO yaitu *Focus, Reason, Inference, Situation, Clarity, dan Overview*. Selanjutnya dilakukan analisis kurikulum yang diterapkan oleh SDN Kejawan Putih I/243 Surabaya, hal ini bertujuan untuk menyesuaikan kurikulum yang digunakan lembaga pendidikan dengan tujuan pembelajaran. Diperoleh hasil bahwa SDN Kejawan Putih I/243 Surabaya menggunakan kurikulum merdeka yang relevan dengan instrumen tes yang akan dikembangkan. Berikut tabel yang menjabarkan aspek kemampuan berpikir kritis peserta didik.

Tabel 3. Aspek Kemampuan Berpikir Kritis

Aspek Kemampuan Berpikir Kritis	Keterangan
F (Focus)	Peserta didik memahami permasalahan pada soal dan mampu mengidentifikasi informasi didalamnya
R (Reason)	Peserta didik memberikan alasan yang logis berdasarkan fakta dalam membuat keputusan/kesimpulan
I (Inference)	Peserta didik mampu menyusun kesimpulan dan menentukan alasan yang logis untuk mendukung kesimpulan yang dibuat
S (Situation)	Peserta didik mampu mengaitkan permasalahan pada soal dengan informasi/pengetahuan yang relevan dengan kehidupan sehari-hari
C (Clarity)	Peserta didik dapat memberikan penjelasan lebih lanjut terhadap kesimpulan yang telah dibuat
O (Overview)	Peserta didik mengevaluasi kembali hasil penyelesaian masalah secara menyeluruh yang dihasilkan pada aspek FRISCO

Tabel 3 menunjukkan aspek kemampuan berpikir kritis yang digunakan sebagai landasan dalam pengembangan instrumen tes, yaitu FRISCO (*Focus, Reason, Inference, Situation, Clarity, dan Overview*). Setiap aspek merepresentasikan tahapan berpikir kritis yang berbeda, mulai dari kemampuan mengidentifikasi informasi penting, memberikan alasan yang logis, menarik kesimpulan, mengaitkan informasi dengan konteks yang relevan, memberikan penjelasan yang jelas, hingga mengevaluasi hasil pemecahan masalah secara menyeluruh. Keenam aspek tersebut dijadikan acuan dalam penyusunan indikator dan butir soal agar instrumen yang dikembangkan mampu mengukur kemampuan berpikir kritis peserta didik secara komprehensif.

3.2 Tahap *Design*

Tahap design dilakukan dengan merancang kisi-kisi yang di dalamnya memuat mata pelajaran, fase, capaian pembelajaran, tujuan pembelajaran, bentuk soal, level kognitif, dan nomor soal. Selanjutnya dilakukan penyajian instrumen tes dengan memperhatikan format huruf, gambar yang disajikan, dan tata letak (*layout*) pada bagian yang termuat dalam lembar instrumen tes yang meliputi tujuan, identitas sekolah, identitas mata pelajaran, alokasi waktu, petunjuk pengisian, dan butir-butir soal. Berikut distribusi soal berdasarkan CP/TP IPAS, indikator, level Taksonomi Bloom, dan butir soal.

Tabel 4. Distribusi Butir Soal Berdasarkan Level Taksonomi Bloom

CP/TP IPAS	Indikator	Level Bloom	No. Butir
Mengidentifikasi keragaman budaya Indonesia	Mengidentifikasi informasi penting pada permasalahan budaya	C4	1, 2, 3
Mengelompokkan jenis keragaman budaya	Memberikan alasan logis berdasarkan fakta budaya	C4	4, 5, 6
Mengidentifikasi budaya lokal di lingkungan sekitar	Menarik kesimpulan dari informasi budaya lokal yang disajikan	C4	7, 8, 9
Menjelaskan makna pelestarian budaya	Mengaitkan permasalahan budaya dengan konteks kehidupan nyata	C5	10, 11, 12,
Menjelaskan pentingnya pelestarian budaya	Memberikan penjelasan logis terhadap upaya pelestarian budaya	C5	13, 14, 15
Memberikan contoh tindakan pelestarian budaya	Menjelaskan alasan pemilihan tindakan pelestarian budaya	C5	16, 17, 18
Menyimpulkan dan menentukan alternatif solusi terkait hubungan budaya dan identitas bangsa	Mengevaluasi dan menyimpulkan alternatif solusi terhadap permasalahan budaya	Mengarah pada level C6	19, 20

Mengingat bentuk instrumen yang digunakan adalah pilihan ganda, pengukuran kemampuan mencipta (C6) belum dapat dilakukan secara optimal sebagaimana pada soal uraian atau tugas kinerja. Oleh karena itu, butir nomor 19–20 diposisikan sebagai soal yang mengarah pada kemampuan C6 melalui penyajian alternatif solusi dan pengambilan keputusan.

Tabel 5. Blueprint Instrumen Berbasis FRISCO

Aspek FRISCO	Indikator	No. Soal	Level Bloom	Materi
Focus	Mengidentifikasi informasi penting pada permasalahan budaya	1,2,3	C4	Keragaman budaya Indonesia
Reason	Memberikan alasan logis berdasarkan fakta budaya	4,5,6	C4	Pengelompokan budaya
Inference	Menarik kesimpulan dari informasi budaya lokal yang disajikan	7,8,9	C4	Budaya lokal
Situation	Mengaitkan permasalahan budaya dengan konteks kehidupan nyata	10,11,12	C5	Pelestarian budaya
Clarity	Memberikan penjelasan logis terhadap upaya pelestarian budaya	13,14,15	C5	Makna budaya
	Menjelaskan alasan pemilihan tindakan pelestarian budaya	16,17,18		
Overview	Mengevaluasi dan menyimpulkan alternatif solusi terhadap permasalahan budaya	19,20	Mengarah pada level C6	Budaya dan identitas bangsa

Tabel 5 menunjukkan keterkaitan antara aspek FRISCO dengan indikator, level kognitif Taksonomi Bloom, materi pembelajaran, dan nomor soal yang dikembangkan. Blueprint ini memastikan bahwa setiap butir soal dirancang secara sistematis untuk mengukur aspek berpikir kritis tertentu sesuai dengan tujuan pembelajaran IPAS materi Kekayaan Budaya Indonesia. Dengan demikian, integrasi FRISCO dalam instrumen dapat diverifikasi secara jelas dan terukur.

3.3 Tahap Development

Tahap *development* ini sebagai tahap pengembangan instrumen tes berdasarkan kisi-kisi yang telah dibuat pada tahap perancangan. Instrumen yang telah dikembangkan dilakukan validasi terhadap ahli yang bertujuan untuk memperoleh penilaian serta masukan dari validator. Hal ini bertujuan agar diperoleh kelayakan sebelum diujicobakan kepada peserta didik. Validator ahli terdiri dari 3 ahli materi dan 2 guru kelas IV. Validasi konten menilai setiap butir

soal berdasarkan aspek materi (M), konstruksi soal (K), bahasa (B), dan kunci jawaban (KJ), yang kemudian dipadukan dengan kesesuaian aspek kemampuan berpikir kritis FRISCO.

Tabel 6. Hasil Validasi Ahli Berdasarkan Butir Soal

Aspek	Integrasi dengan FRISCO	Rata-rata (V)
Materi	Mengukur kesesuaian substansi soal dengan kompetensi, indikator pembelajaran, dan indikator kemampuan berpikir kritis FRISCO yang ditetapkan.	0,850
Konstruksi Soal	Mengukur ketepatan penyusunan stimulus, pokok soal, dan alternatif jawaban dalam memfasilitasi proses berpikir kritis sesuai aspek FRISCO.	0,858
Bahasa	Mengukur kejelasan, ketepatan, dan efektivitas penggunaan bahasa agar peserta didik dapat memahami tuntutan berpikir pada aspek FRISCO tanpa bias makna atau multitafsir.	0,867
Kunci Jawaban	Mengukur ketepatan jawaban benar dalam merepresentasikan proses berpikir kritis yang diharapkan pada aspek FRISCO dan memastikan pengecoh berfungsi secara logis.	1,000
Rata-rata skor validitas keseluruhan		0,89
Kriteria Validitas		Sangat Valid

Berdasarkan tabel 6 diperoleh hasil validasi instrumen tes oleh para ahli, secara menyeluruh diperoleh rata-rata validitas sebesar 0,89 yang memiliki kriteria validitas sangat valid, sehingga instrumen tes yang dikembangkan layak untuk dilakukan uji coba kepada peserta didik. Data kualitatif pada tahap ini berupa saran dan masukan dari para ahli terhadap instrumen tes yang dikembangkan. Adapun saran dan masukan para ahli sebagai berikut.

Tabel 7. Saran dan Masukan Ahli

Ahli	Saran dan Masukan
Validator I	Secara keseluruhan soal baik, namun ada beberapa soal yang masih belum tergolong HOTS seperti soal nomor 3, 11, dan 14.
Validator II	Soal nomor 1 dan 5 memerlukan perbaikan dibagian konstruksi soal.
Validator III	Soal nomor 19 dapat diperjelas dengan menambahkan hubungan budaya dengan identitas bangsa melalui stimulus cerita.
Validator IV	Soal nomor 3 dapat ditingkatkan ke level HOTS dengan memberikan stimulus berupa situasi nyata terkait budaya daerah. Soal nomor 7 bahasa dalam cerita dapat disederhanakan agar mudah dipahami dan sesuai karakter peserta didik kelas IV.
Validator V	Soal nomor 2 ilustrasi soal perlu dibuat lebih menarik dan kontekstual.

Berdasarkan saran dan masukan para ahli, secara keseluruhan instrumen sudah relevan dengan capaian pembelajaran, tujuan pembelajaran, indikator soal HOTS, level kognitif, dan aspek FRISCO. Namun ada beberapa revisi kecil terkait dengan materi yang harus lebih diperkuat dengan stimulus kehidupan nyata untuk meningkatkan kualitas HOTS, serta penyederhanaan bahasa pada beberapa soal tertentu agar lebih komunikatif bagi peserta didik khususnya kelas IV. Saran dan masukan oleh para ahli akan dijadikan sebagai acuan revisi instrumen yang dikembangkan sebelum dilakukan uji coba lapangan.

3.4 Tahap Implementation

Tahap *implementation* ini merupakan tahap uji coba instrumen yang telah dikembangkan dan direvisi berdasarkan saran dan masukan oleh validator. Instrumen diujicobakan secara terbatas kepada 20 peserta didik kelas IV SDN Kejawan Putih I/243 Surabaya. Analisis data uji validitas empiris menggunakan Anates versi 4 yang bertujuan mengetahui tingkat kesukaran (*difficulty index*) dan daya beda (*discrimination index*) setiap butir soal. Analisis reliabilitas menggunakan *Cronbach's Alpha* bertujuan untuk mengetahui tingkat reliabilitas atau konsistensi instrumen dimana butir soal dalam instrumen penelitian ini dapat memberikan hasil pengukuran yang konsisten dan stabil dalam mengukur kemampuan berpikir kritis berdasarkan aspek FRISCO. Instrumen dikatakan reliabel jika nilai *Cronbach's Alpha* memenuhi kriteria yang ditetapkan, di antaranya >0,90 kategori sangat tinggi, 0,80-0,90 kategori tinggi, 0,70-0,79 kategori cukup/reliabel, 0,60-0,69 kategori sedang, 0,50-0,59 kategori rendah, dan <0,50

kategori sangat rendah (Fraenkel et al., 2019). Berikut hasil analisis butir soal instrumen menggunakan Anates versi 4.

Tabel 8. Hasil Analisis Validitas Empiris Per Butir

No. soal	Koefisien validitas	Kategori validitas	Tingkat kesukaran	Daya pembeda	Keputusan revisi
1	0,68	Tinggi	Sedang	0,52 (Baik)	Dipertahankan
2	0,62	Tinggi	Sedang	0,48 (Baik)	Dipertahankan
3	0,51	Cukup	Mudah	0,36 (Cukup)	Dipertahankan
4	0,73	Tinggi	Sedang	0,61 (Baik)	Dipertahankan
5	0,65	Tinggi	Sedang	0,54 (Baik)	Dipertahankan
6	0,58	Cukup	Sedang	0,45 (Baik)	Dipertahankan
7	0,71	Tinggi	Sedang	0,57 (Baik)	Dipertahankan
8	0,55	Cukup	Sedang	0,43 (Baik)	Dipertahankan
9	0,47	Cukup	Mudah	0,38 (Cukup)	Dipertahankan
10	0,76	Tinggi	Sedang	0,64 (Baik)	Dipertahankan
11	0,63	Tinggi	Sedang	0,50 (Baik)	Dipertahankan
12	0,57	Cukup	Sedang	0,46 (Baik)	Dipertahankan
13	0,69	Tinggi	Sedang	0,55 (Baik)	Dipertahankan
14	0,61	Tinggi	Sedang	0,47 (Baik)	Dipertahankan
15	0,54	Cukup	Mudah	0,34 (Cukup)	Dipertahankan
16	0,67	Tinggi	Sedang	0,53 (Baik)	Dipertahankan
17	0,72	Tinggi	Sedang	0,60 (Baik)	Dipertahankan
18	0,59	Cukup	Sedang	0,44 (Baik)	Dipertahankan
19	0,64	Tinggi	Sedang	0,49 (Baik)	Dipertahankan
20	0,49	Cukup	Mudah	0,32 (Cukup)	Dipertahankan

Tabel 9. Rekapitulasi Hasil Validitas Empiris

Kategori	Jumlah Soal	Persentase
Validitas tinggi	11	55%
Validitas cukup	9	45%
Mudah (0,71–1,00)	4	20%
Sedang (0,31–0,70)	16	80%
Sukar (0,00–0,30)	0	0%
Daya Pembeda Baik (>0,40)	16	80%
Daya Pembeda Cukup (0,20–0,40)	4	20%

Hasil analisis empiris butir soal menggunakan Anates versi 4 pada uji coba terbatas terhadap 20 peserta didik kelas IV, menunjukkan 4 butir soal dengan persentase 20% berkategori mudah dan 16 butir soal dengan persentase 80% berkategori sedang. Berdasarkan hasil analisis tidak terdapat kategori soal sukar. Sedangkan analisis pembeda menunjukkan sebanyak 16 butir soal dengan persentase 80% memiliki daya pembeda baik dan 4 butir soal dengan persentase 20% memiliki daya pembeda cukup. Hasil penelitian tersebut menunjukkan sebagian besar instrumen memiliki kualitas yang baik dan mampu membedakan kemampuan peserta didik. Selain tingkat kesukaran dan daya pembeda, analisis validitas empiris menunjukkan bahwa seluruh butir soal memiliki koefisien validitas pada kategori cukup hingga tinggi sehingga layak digunakan pada tahap uji coba terbatas.

Tabel 10. Hasil Analisis Reliabilitas *Cronbach's Alpha*

No. Soal	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
1	0,548	0,834
2	0,517	0,836
3	0,401	0,844
4	0,621	0,829
5	0,573	0,832
6	0,468	0,838
7	0,592	0,831
8	0,446	0,839
9	0,387	0,845

No. Soal	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
10	0,654	0,827
11	0,534	0,835
12	0,472	0,838
13	0,581	0,832
14	0,501	0,836
15	0,398	0,844
16	0,556	0,833
17	0,617	0,830
18	0,453	0,839
19	0,528	0,835
20	0,382	0,846

Hasil analisis reliabilitas menggunakan *Cronbach's Alpha* dengan jumlah responden 20 peserta didik dan jumlah soal 20 butir menunjukkan hasil 0.842. Nilai tersebut berada pada kategori tinggi karena berada pada rentang 0,80-0,90. Hasil ini menunjukkan bahwa instrumen memiliki konsistensi internal yang baik dalam mengukur kemampuan berpikir kritis peserta didik berdasarkan aspek FRISCO. Analisis lebih lanjut menunjukkan bahwa nilai *Corrected Item-Total Correlation* seluruh butir berada di atas 0,30, yang mengindikasikan bahwa setiap butir soal memberikan kontribusi positif terhadap pengukuran konstruk yang sama. Selain itu, nilai *Cronbach's Alpha if Item Deleted* pada seluruh butir tidak menunjukkan peningkatan yang signifikan dibandingkan nilai alpha keseluruhan (0,842). Dengan demikian, tidak terdapat butir soal yang perlu dihapus berdasarkan hasil analisis reliabilitas. Temuan ini didukung oleh hasil validitas empiris yang menunjukkan seluruh butir soal memiliki kategori validitas cukup hingga tinggi, tingkat kesukaran didominasi kategori sedang (80%), serta daya pembeda baik (80%). Oleh karena itu, instrumen tes berpikir kritis berbasis FRISCO yang dikembangkan dapat dinyatakan memiliki reliabilitas tinggi dan layak digunakan. Namun, penelitian ini hanya melibatkan 20 peserta didik dalam satu sekolah, maka dapat dikatakan penelitian ini masih berada pada uji coba terbatas. Sehingga hasil validitas empiris, reliabilitas, tingkat kesukaran, dan daya pembeda yang diperoleh belum dapat digeneralisasikan secara luas dan perlu diuji kembali pada sampel yang lebih besar serta beragam.

Pembahasan

Analisis validasi ahli berdasarkan kelayakan konten dan kesesuaian materi menunjukkan rata-rata validitas sebesar 0,89 dengan kriteria sangat valid. Pencapaian skor menunjukkan bahwa instrumen telah memenuhi standar kelayakan untuk diimplementasikan dalam pengukuran kemampuan berpikir kritis peserta didik. Kesesuaian materi dengan aspek FRISCO (*Focus, Reason, Inference, Situation, Clarity, Overview*) terdapat pada desain soal yang tidak hanya mengukur pemahaman faktual tentang keragaman budaya Indonesia, namun juga menuntut peserta didik untuk melakukan proses kognitif yang lebih kompleks. Pada level Focus, soal dirancang untuk mengarahkan perhatian peserta didik pada aspek kritis dalam materi. Soal dirancang pada tingkat Taksonomi Bloom C4-C6 untuk memastikan peserta didik terlibat dalam proses berpikir yang lebih mendalam. Aspek konstruksi soal memperoleh skor validasi 0,858, menunjukkan ketepatan penyusunan stimulus, pokok soal, dan alternatif jawaban dalam memfasilitasi proses berpikir kritis sesuai dengan aspek FRISCO. Konstruksi soal yang baik memungkinkan peserta didik melakukan *reasoning* (penalaran) yang sistematis. Stimulus yang dirancang dengan cermat, memungkinkan peserta didik mengidentifikasi hubungan antar konsep (*inference*), menganalisis situasi budaya dalam konteks yang relevan (*situation*), dan merumuskan kesimpulan logis (*overview*). Berdasarkan saran dan masukan para ahli, masih diperlukan revisi kecil terkait beberapa konstruksi soal, dimana beberapa soal masih belum memenuhi kriteria soal HOTS. Perlu adanya pendalaman stimulus berbasis situasi kehidupan

nyata untuk memperkuat elemen *situation* dalam aspek FRISCO, khususnya soal yang mengukur interpretasi dan analisis mendalam.

Kejelasan bahasa dan aksesibilitas kognitif menunjukkan skor validitas 0,867 mengenai kejelasan, ketepatan, dan efektivitas penggunaan bahasa dalam instrumen tes. Kejelasan bahasa sangat penting bertujuan untuk memastikan bahwa butir soal dalam instrumen tidak memiliki penafsiran ganda dan tidak bias, sehingga tidak menghalangi peserta didik dalam berpikir kritis. Dalam hal ini, elemen *clarity* dalam aspek FRISCO terpenuhi melalui perumusan soal yang komunikatif dan memastikan bahwa butir soal mudah dipahami peserta didik. Berdasarkan saran dan masukan para ahli, terdapat beberapa butir soal yang membutuhkan penyederhanaan bahasa dan peningkatan kontekstualisasi visual agar peserta didik dapat fokus pada proses berpikir kritis tanpa mengurangi kompleksitas kognitif yang diukur. Ketepatan kunci jawaban dan logika pengecoh pada instrumen tes kemampuan berpikir kritis memperoleh skor validitas 1,000. Kategori skor ini menunjukkan bahwa jawaban benar dalam instrumen tepat merepresentasikan kemampuan berpikir kritis yang diharapkan dalam aspek FRISCO, dan pengecoh (*distractor*) berfungsi secara logis dalam merangsang analisis peserta didik. Dalam soal HOTS, pengecoh yang logis menjadi penting terutama dalam elemen *inference* dan *reason* dimana peserta didik perlu membedakan antara kesimpulan valid dan tidak valid.

Hasil analisis empiris menggunakan Anates versi 4 menunjukkan distribusi tingkat kesukaran soal sebagai berikut 4 butir soal dengan persentase 20% memiliki kategori mudah dan 16 butir soal dengan persentase 80% memiliki kategori sedang. Distribusi soal secara menyeluruh menunjukkan bahwa instrumen dirancang dengan tingkat yang sesuai dengan karakteristik peserta didik kelas IV. Berdasarkan perspektif FRISCO, distribusi soal mencerminkan keseimbangan antara aksesibilitas kognitif dan kedalaman proses berpikir. Soal dengan kategori sedang menuntut peserta didik melakukan operasi kognitif kompleks seperti *fokus*, *reason*, dan *inference*. Sementara soal dengan kategori mudah tetap mempertahankan kompleksitas berpikir kritis, tetapi dengan stimulus lebih langsung atau konteks yang lebih familiar bagi peserta didik kelas IV. Hasil analisis daya pembeda menunjukkan bahwa 16 butir soal dengan persentase 80% memiliki daya pembeda baik dan 4 butir soal dengan persentase 20% memiliki daya pembeda cukup. Daya pembeda yang baik ini sangat signifikan karena menunjukkan bahwa instrumen mampu membedakan tingkat kemampuan berpikir kritis peserta didik. Dalam konteks FRISCO, peserta didik dengan kemampuan tinggi diharapkan dapat melakukan *focus* lebih presisi, menghasilkan *reason* yang lebih solid, melakukan *inference* yang lebih akurat, menganalisis *situation* dengan perspektif yang luas, mempertahankan *clarity* dalam penjelasan, dan mencapai *overview* yang lebih komprehensif. Sebaliknya, peserta didik dengan kemampuan rendah mungkin mengalami kesulitan pada salah satu atau beberapa komponen FRISCO. Dalam hal ini, soal dengan daya pembeda 80% telah mampu menciptakan perbedaan tersebut.

Analisis reliabilitas dan konsistensi internal menggunakan *Cronbach's Alpha* menunjukkan nilai 0,842 yang berada pada kategori tinggi. Nilai ini menandakan bahwa instrumen memiliki konsistensi internal yang baik. Artinya, butir-butir soal dalam instrumen memiliki hubungan yang cukup kuat dan konsisten dalam mengukur kemampuan berpikir kritis berdasarkan aspek FRISCO. Nilai *Cronbach's Alpha* sebesar 0,842 mencerminkan bahwa setiap butir soal berkontribusi secara efektif dalam mengukur dimensi FRISCO secara terpadu. Reliabilitas yang tinggi menunjukkan bahwa instrumen tes ini mampu mengukur kemampuan berpikir kritis peserta didik secara stabil dan dapat digunakan secara berulang kali untuk mengukur progres peserta didik dalam mengembangkan kemampuan berpikir kritis berbasis FRISCO.

Berdasarkan analisis menyeluruh terhadap validasi ahli, validasi empiris, dan reliabilitas, dapat disimpulkan bahwa instrumen tes berpikir kritis yang dikembangkan memiliki kualitas yang sangat baik dan siap untuk diimplementasikan. Kualitas instrumen dan kesiapan

implementasi dilihat dari validasi ahli dengan skor 0,89 memiliki kategori sangat valid, menunjukkan materi, konstruksi soal, bahasa, dan kunci jawaban telah sesuai dengan aspek FRISCO. Validitas empiris sebesar 20% mudah dan 80% sedang, menunjukkan tingkat kesulitan soal yang sesuai untuk peserta didik kelas IV SD. Daya pembeda dengan hasil 80% baik dan 20% cukup, menunjukkan instrumen tes kemampuan berpikir kritis efektif dalam membedakan kemampuan berpikir kritis tinggi dan rendah. Serta reliabilitas dengan skor 0,842 memiliki kategori tinggi, menunjukkan konsistensi internal yang baik dalam mengukur kemampuan berpikir kritis peserta didik sesuai dengan aspek FRISCO. Meski begitu, penelitian ini hanya melibatkan 20 peserta didik dalam satu sekolah, maka dapat dikatakan penelitian ini masih berada pada uji coba terbatas. Sehingga hasil validitas empiris, reliabilitas, tingkat kesukaran, dan daya pembeda yang diperoleh belum dapat digeneralisasikan secara luas dan perlu diuji kembali pada sampel yang lebih besar serta beragam.

4. Kesimpulan

Pengembangan instrumen tes kemampuan berpikir kritis yang berorientasi pada aspek FRISCO (*Focus, Reason, Inference, Situation, Clarity, Overview*) untuk peserta didik kelas IV SD pada materi Kekayaan Budaya Indonesia telah menghasilkan instrumen yang memiliki validitas konten dengan skor 0,89 (sangat valid), mencakup pengukuran pada dimensi materi (0,850), konstruksi soal (0,858), bahasa (0,867), dan kunci jawaban (1,000). Hal ini menunjukkan bahwa instrumen layak digunakan pada tahap uji coba terbatas dan perlu diuji pada sampel yang lebih luas untuk memastikan validitas eksternal dan generalisasi hasil dalam mengukur kemampuan berpikir kritis peserta didik kelas IV SD. Integrasi aspek FRISCO dalam setiap butir soal memastikan bahwa pengukuran tidak hanya berfokus pada kemampuan *recall* atau pemahaman sederhana, melainkan mendorong peserta didik untuk melakukan proses kognitif yang kompleks mulai dari mengidentifikasi informasi penting (*focus*), memberikan penalaran logis (*reason*), menarik kesimpulan (*inference*), menganalisis konteks situasional (*situation*), mempertahankan kejelasan konseptual (*clarity*), hingga menyusun pandangan umum yang holistik (*overview*).

Uji coba terbatas terhadap 20 peserta didik memperoleh hasil validitas empiris sebesar 80% butir soal yang memiliki tingkat kesukaran sedang dan daya pembeda baik. Hal ini menunjukkan bahwa instrumen tes ini efektif dalam membedakan peserta didik yang memiliki kemampuan berpikir kritis tinggi dan rendah. Reliabilitas instrumen tes berpikir kritis sebesar 0,842 dengan kategori tinggi menunjukkan adanya konsistensi internal yang kuat, mengindikasikan bahwa semua butir soal bekerja secara sinergis dalam mengukur kemampuan berpikir kritis peserta didik. Instrumen tes ini diharapkan dapat berfungsi sebagai alat diagnostik yang efektif bagi guru dalam mengidentifikasi level kemampuan berpikir kritis peserta didik, sekaligus menjadi referensi dalam mengembangkan soal-soal HOTS (*Higher Order Thinking Skills*) pada mata pelajaran IPAS khususnya dan mata pelajaran lainnya secara umum. Dengan demikian, instrumen tes ini berkontribusi pada peningkatan kualitas pembelajaran dan pengembangan kompetensi berpikir kritis peserta didik sesuai dengan tuntutan pendidikan abad ke-21 dan visi kurikulum merdeka untuk mempersiapkan generasi muda yang mampu menghadapi tantangan masa depan dengan pemikiran kritis, analitis, dan reflektif.

5. Referensi

- Arikunto, S. (2018). *Dasar-Dasar Evaluasi Pendidikan* (3rd ed.). Bumi Aksara.
- Biggs, J., & Collis, K. (1982). *Evaluating the quality of learning: The SOLO Taxonomy (Structure of the Observed Learning Outcome)*. New York: Academic Press.
- Branch, R. M. (2009). *Instructional Design: The ADDIE Approach* (1st ed.). New York : Springer. <https://doi.org/https://doi.org/10.1007/978-0-387-09506-6>

- Ennis, R. H. (2011). *Critical Thinking* (2nd ed.). Boston: Pearson Education.
- Facione, P. A. (1980). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. Millbrae, CA: California Academic Press.
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2019). *How to design and evaluate research in education* (10th ed.). McGraw-Hill Education.
- Hulaipah, A., Syukri, M., & Indraswati, D. (2023). Analisis Kesulitan Guru Kelas IV Dan V Dalam Menyusun Soal HOTS Pada Mata Pelajaran IPAS Di SDN 2 Perampuan. *Jurnal Ilmiah Profesi Pendidikan*, 8(4), 2450–2460. <https://doi.org/10.29303/jipp.v8i4.1768>
- Maulana, A. (2022). Analisis Validitas, Reliabilitas, Dan Kelayakan Instrumen Penilaian Rasa Percaya Diri Siswa. *Jurnal Kualita Pendidikan*, 3(3), 133–139. <https://doi.org/10.51651/jkp.v3i3.331>
- Novianti, B. A., Nitiasih, P. K., & Riastini, P. N. (2023). Study Of STEM-Based Learning Against 4C Skills (Critical, Creative, Communication, and Collaboration) In Science. *Jurnal Ilmiah Profesi Pendidikan*, 8(3), 1917–1921. <https://doi.org/10.29303/jipp.v8i3.1455>
- Rahayu, S., Kurniawati, I., & Purnomo, H. (2023). Analisis kemampuan berpikir kritis peserta didik SD dalam pembelajaran berbasis digital: Studi kasus di Jawa Timur. *Jurnal Pendidikan Dan Kebudayaan*, 8(2), 134–149. <https://doi.org/10.24832/jpnk.v8i2.2841>
- Ruci, M., Huda, C., & Suneki, S. (2023). Implementasi LKPD Berbasis Saintifik Untuk Membangun Kemampuan Berpikir Kritis Siswa SD. *Didaktik: Jurnal Ilmiah PGSD STKIP Subang*, 9(2), 2808–2822. <https://doi.org/10.36989/didaktik.v9i2.948>
- Sari, D. N., Adira, H. F., Sukiman, & Silfiani, H. (2025). Integrasi PjBl (Project Based Learning) dalam Kurikulum Merdeka untuk Pengembangan Keterampilan 4C Siswa Sekolah Dasar. *Action Research Journal Indonesia (ARJI)*, 7(4), 3452–3467. <https://doi.org/10.61227>
- Sinaga, P., Setiawan, W., & Liana, M. (2022). The Impact of Electronic Interactive Teaching Materials (EITMs) in E-Learning on Junior High School Students' Critical Thinking Skills. *Thinking Skills and Creativity*, 46, 101066. <https://doi.org/10.1016/j.tsc.2022.101066>
- Wahyuni, R., & Kurniawati, D. (2024). Implementasi media AI storytelling dalam peningkatan kemampuan analitis siswa kelas IV SD. *Elementary School Journal PGSD FIP UNIMED*, 14(1), 88–102. <https://doi.org/10.24114/esjgds.v14i1.51289>