

Pengembangan Instrumen Penilaian Bahasa Indonesia Berbantuan *GenAI* untuk Mengukur Kemampuan Berpikir Kritis Peserta Didik SD Kelas IV

Ainur Rofida* & Deri Anggraini

Program Studi Pendidikan Guru Sekolah Dasar, Universitas PGRI Yogyakarta, Sonosewu, Ngestiharjo, Kec. Kasihan, Kabupaten Bantul, Daerah Istimewa Yogyakarta 55184, Indonesia.

*Corresponding Author: aidarfdaa@gmail.com, derianggraini@upy.ac.id

Article History

Received : April 07th, 2026

Revised : April 20th, 2026

Accepted : May 11th, 2026

Abstract: Kemampuan berpikir kritis merupakan salah satu kompetensi penting yang perlu dikembangkan pada peserta didik sekolah dasar, terutama dalam pembelajaran Bahasa Indonesia. Namun, ketersediaan instrumen penilaian yang mampu mengukur kemampuan tersebut secara tepat dan menarik masih terbatas. Penelitian ini bertujuan mengembangkan instrumen penilaian kemampuan berpikir kritis berbantuan *Generative Artificial Intelligence (GenAI)* pada mata pelajaran Bahasa Indonesia untuk peserta didik SD kelas IV. Penelitian menggunakan model pengembangan 4D (*Define, Design, Develop, Disseminate*). Instrumen yang dikembangkan mencakup soal uraian yang disajikan melalui platform *Canva AI*, dilengkapi ilustrasi tematik dari *ChatGPT*, dan audio *text-to-speech* dari *Google AI Studio*. Validasi dilakukan oleh ahli evaluasi pembelajaran dan ahli materi, keduanya menghasilkan persentase 100% dengan kriteria Sangat Layak. Uji validitas empiris melibatkan 30 peserta didik dengan hasil 18 dari 20 butir soal (90%) valid, reliabilitas *Cronbach's Alpha* 0,961, tingkat kesukaran 95% kategori sedang, dan daya pembeda 55% kategori baik dan sedang. Produk akhir terdiri atas 11 butir soal terpilih. Uji coba terbatas pada 25 peserta didik dan 1 guru menunjukkan kepraktisan 84,8% (guru) dan 94,5% (peserta didik) dalam kategori Sangat Praktis. Kesimpulan penelitian ini adalah instrumen penilaian berbantuan *GenAI* yang dikembangkan (*SI PARA*) terbukti layak dan praktis digunakan untuk mengukur kemampuan berpikir kritis peserta didik SD kelas IV.

Keywords: *Artificial Intelligence*, Bahasa Indonesia, Berpikir Kritis, Instrumen Penilaian, Sekolah Dasar.

PENDAHULUAN

Pendidikan adalah pilar mendasar dalam mencetak individu-individu yang unggul dan berdaya saing tinggi. Keterampilan abad ke-21 yang dikenal dengan kerangka 4C (*creative thinking, critical thinking, communication, dan collaboration*) menjadi kompetensi esensial yang wajib dikuasai peserta didik (Partono et al., 2021). Di antara keempat keterampilan tersebut, berpikir kritis menempati posisi yang sangat strategis. Berpikir kritis membantu peserta didik mengaitkan konsep baru dengan pemahaman yang sudah mereka miliki sebelumnya, sehingga proses belajar menjadi lebih mendalam dan tidak sekadar hafalan (Rahmawati et al., 2023). Urgensi pengembangan berpikir kritis semakin diperkuat melalui Kurikulum Merdeka yang menempatkan bernalar kritis sebagai salah satu dari delapan dimensi Profil Pelajar Pancasila.

Secara empiris, kemampuan berpikir kritis peserta didik Indonesia masih memprihatinkan. Hasil PISA 2022 menempatkan Indonesia di

peringkat ke-70 dari 81 negara dengan skor rata-rata membaca 359, matematika 366, dan sains 383, seluruhnya di bawah rata-rata OECD (Nurhasanah & Hidayat, 2024). Temuan (Rofi'ah & Rokhmaniyah, 2024) memperkuat kondisi tersebut, di mana 85% peserta didik kelas V sekolah dasar tergolong dalam kategori tingkat kemampuan berpikir kritis yang sangat rendah. Observasi dan wawancara yang dilakukan peneliti pada 27 Oktober 2025 di kelas IV SD N Mejing 2 juga menunjukkan bahwa sebagian besar peserta didik hanya mampu menjawab "apa" secara permukaan tanpa disertai argumen yang logis. Sementara instrumen penilaian yang digunakan guru masih didominasi ranah kognitif C1-C3 dan belum pernah dirancang khusus untuk mengukur berpikir kritis.

Perkembangan *Generative AI (GenAI)* membuka peluang strategis untuk menjawab kesenjangan tersebut. *ChatGPT* dapat menghasilkan ilustrasi tematik kontekstual, *Canva AI* mampu menyajikan instrumen dalam format kuis daring yang interaktif dan estetis,

serta *Google AI Studio* dapat mengonversi teks bacaan menjadi audio *text-to-speech* berkualitas natural. Pemanfaatan AI dalam evaluasi pendidikan terbukti mampu menghasilkan penilaian yang lebih objektif, adaptif, dan efisien serta mendukung pembelajaran yang lebih personal dan inovatif (Sholeh et al., 2024). Namun demikian, pemanfaatan *GenAI* secara spesifik untuk pengembangan instrumen penilaian berpikir kritis di sekolah dasar masih sangat terbatas dikaji, sehingga terdapat kesenjangan penelitian yang mendesak untuk diisi.

Berdasarkan permasalahan tersebut, terdapat tiga isu yang saling berkaitan yaitu rendahnya kemampuan berpikir kritis peserta didik, ketiadaan instrumen penilaian berpikir kritis yang terstruktur di tingkat sekolah dasar, dan besarnya potensi *GenAI* yang belum dioptimalkan. Berdasarkan hal tersebut, penelitian ini bermaksud merancang instrumen penilaian kemampuan berpikir kritis berbantuan *GenAI* yang bersifat interaktif, faktual, dan kontekstual untuk peserta didik kelas IV SD. Instrumen yang dihasilkan diharapkan dapat membantu guru merancang penilaian yang mampu mengungkap kemampuan berpikir tingkat tinggi secara komprehensif, sekaligus menjadi kontribusi akademik dalam pengembangan teknologi pendidikan berbasis kecerdasan buatan generatif.

METODE

Penelitian ini merupakan penelitian pengembangan (*Research and Development*) yang memadukan pendekatan kualitatif dan kuantitatif. Model pengembangan yang digunakan adalah model 4D yang digagas oleh Thiagarajan yang terdiri atas empat tahap utama, yakni *Define* (Pendefinisian), *Design* (Perancangan), *Develop* (Pengembangan), dan *Disseminate* (Penyebaran) (Sugiyono, 2024). Model ini dipilih karena memiliki tahapan yang sederhana, terstruktur, dan efektif untuk pengembangan instrumen penilaian. Produk yang dihasilkan berupa instrumen penilaian kemampuan berpikir kritis berbentuk soal uraian berbantuan *Generative Artificial Intelligence* (*GenAI*) yang diberi nama SI PARA (Sistem Penilaian Paragraf Argumentasi untuk Mengukur Berpikir Kritis).

Penelitian ini dilaksanakan di SD Negeri Mejing 2 yang beralamat di Patukan,

Ambarketawang, Gamping, Sleman, Yogyakarta pada bulan Maret-April 2026. Subjek penelitian terdiri atas dua validator ahli, yaitu ahli evaluasi pembelajaran dan ahli materi, yang bertugas memberikan penilaian terhadap kelayakan instrumen yang dikembangkan. Uji coba instrumen melibatkan 30 peserta didik kelas IV A untuk keperluan analisis butir soal, sedangkan uji terbatas melibatkan 25 peserta didik kelas IV B dan 1 guru kelas.

Prosedur pengembangan dilaksanakan mengacu pada keempat tahap model 4D menurut (Fahma, 2025). Tahap pertama, *Define* (Pendefinisian), mencakup analisis awal melalui wawancara dan observasi pembelajaran, analisis tugas, analisis konsep, serta analisis tujuan pembelajaran. Tahap kedua, *Design* (Perancangan), meliputi penyusunan kisi-kisi instrumen penilaian berdasarkan indikator berpikir kritis, pemilihan media berbasis web interaktif, serta penyusunan rancangan awal instrumen. Tahap ketiga, *Develop* (Pengembangan), mencakup validasi ahli terhadap draft awal instrumen, revisi berdasarkan masukan validator, uji coba instrumen, analisis butir soal, uji terbatas, hingga revisi produk akhir. Tahap keempat, *Disseminate* (Penyebaran), dilaksanakan secara terbatas dengan memperkenalkan produk kepada guru dan peserta didik kelas IV SD Negeri Mejing 2, sekaligus menyebarkan melalui artikel jurnal ilmiah.

Pengumpulan data dilakukan menggunakan empat instrumen. Pertama, panduan wawancara digunakan untuk menggali informasi awal mengenai kebutuhan instrumen penilaian di lapangan. Kedua, lembar angket validasi ahli evaluasi pembelajaran dan ahli materi digunakan untuk memperoleh penilaian terhadap kelayakan isi dan konstruk instrumen. Ketiga, angket respons guru dengan skala Likert 1–4 digunakan untuk mengukur kepraktisan instrumen dari perspektif pendidik. Keempat, angket respons peserta didik digunakan untuk mengukur kepraktisan instrumen dari perspektif pengguna langsung.

Analisis data dalam penelitian ini meliputi analisis kelayakan produk, analisis kepraktisan produk, dan analisis kelayakan butir soal. Secara keseluruhan, teknik analisis data yang digunakan disajikan pada Tabel 1 berikut:

Tabel 1. Teknik Analisis Data

No	Jenis Analisis	Alat Analisis	Rumus/Acuan	Kriteria
1	Kelayakan produk	Angket validasi ahli	00%	$P > 81,25\%$ = Sangat Layak (Sudijono, 2012)
2	Kepraktisan produk	Angket respons guru & peserta didik	00%	$P > 81,25\%$ = Sangat Praktis (Sudijono, 2012)
3	Validitas empiris	SPSS Statistics 26	Korelasi Pearson	Sig. (2-tailed) $< 0,05$ = Valid (Sugiyono, 2024)
4	Reliabilitas	SPSS Statistics 26	<i>Cronbach's Alpha</i>	$\alpha \geq 0,70$ = Reliabel (Sudijono, 2012)
5	Tingkat kesukaran	Microsoft Excel	$\frac{\text{Rata Skor}}{\text{Maksimal}}$	0,00–0,30 = Sukar 0,31–0,70 = Sedang 0,71–1,00 = Mudah (Arikunto, 2019)
6	Daya pembeda	Microsoft Excel	$\frac{A - \text{Mean } B}{\text{Maksimal}}$	$\leq 0,00$ = Sangat Jelek 0,00–0,20 = Jelek 0,21–0,40 = Sedang 0,41–0,70 = Baik 0,71–1,00 = Sangat Baik (Arikunto, 2019)

HASIL DAN PEMBAHASAN

Hasil Pengembangan Produk

Tahap *Define* dilaksanakan melalui pengamatan langsung dan wawancara bersama guru kelas IV SD Negeri Mejing 2 pada 27 Oktober 2025. Analisis awal menemukan bahwa kemampuan berpikir kritis peserta didik masih tergolong rendah. Peserta didik mampu menjawab "apa" tetapi belum mampu menjelaskan "mengapa" dan "bagaimana". Instrumen penilaian yang digunakan oleh guru masih didominasi ranah kognitif C1–C3 dan sebagian besar bersumber dari LKS, sedangkan instrumen yang dirancang khusus untuk mengukur berpikir kritis belum pernah digunakan sama sekali. Analisis tugas mengidentifikasi Capaian Pembelajaran (CP) dan Tujuan Pembelajaran (TP) Bahasa Indonesia Kelas IV Kurikulum Merdeka pada materi paragraf argumentasi, kemudian ditetapkan dua TP utama yaitu mengidentifikasi opini dan alasan pendukung serta menulis paragraf argumentasi. Analisis konsep menghubungkan materi paragraf argumentasi dengan lima indikator berpikir kritis menurut Robert Ennis, yaitu *elementary clarification*, *basic support*, *inference*, *advanced clarification*, dan *strategies and tactics*.

Tahap *Design* menghasilkan tiga komponen utama. Pertama, kisi-kisi instrumen disusun berdasarkan empat TP dan lima aspek berpikir kritis, sehingga total awal instrumen adalah 20 butir soal. Kedua, pemilihan media mengintegrasikan tiga platform GenAI secara sinergis: *ChatGPT* digunakan untuk

menghasilkan gambar ilustrasi tematik yang merepresentasikan tiga konteks teks bacaan (sampah plastik, hemat energi, dan membaca buku); *Canva AI* dimanfaatkan untuk menyajikan instrumen dalam format kuis daring interaktif yang memadukan teks bacaan, gambar ilustrasi, dan butir soal; serta *Google AI Studio* digunakan untuk menghasilkan audio *text-to-speech* guna mengakomodasi gaya belajar auditori. Ketiga, rancangan awal produk diberi nama SI PARA yang dilengkapi halaman utama, profil pengembang, deskripsi produk, petunjuk pengerjaan, identitas peserta didik, tampilan soal interaktif dengan indikator progres dan tombol navigasi, serta halaman penutup.

Tahap *Develop* mencakup validasi ahli, revisi produk, uji validitas empiris, dan uji coba terbatas. Validasi dilakukan oleh dua validator: ahli evaluasi pembelajaran (dosen Universitas PGRI Yogyakarta) dan ahli materi (guru kelas IV SD Negeri Tawang Sari), masing-masing melalui tiga putaran validasi–revisi hingga dinyatakan layak. Revisi tahap pertama mencakup pengurangan TP dari empat menjadi dua TP yang lebih terfokus, penyusunan soal secara berurutan mengikuti TP, dan penyederhanaan bahasa soal. Revisi tahap kedua menyempurnakan butir-butir spesifik seperti penggantian frasa abstrak menjadi lebih konkret dan pemecahan soal yang terlalu kompleks menjadi dua sub-pertanyaan. Setelah revisi produk dilanjutkan uji validitas empiris yang melibatkan 30 peserta didik dan uji coba terbatas dilaksanakan pada 25 peserta didik dan 1 guru.

Tahap *Disseminate* dilakukan secara terbatas dengan memperkenalkan instrumen SI PARA kepada guru dan peserta didik kelas IV SD Negeri Mejing 2. Guru diberikan penjelasan mengenai cara penggunaan dan keunggulan instrumen penilaian dalam mendukung proses pembelajaran. Selain itu, hasil pengembangan disebarluaskan melalui publikasi artikel jurnal ilmiah sebagai bentuk kontribusi akademik. Diseminasi pada tahap ini masih bersifat terbatas mengingat keterbatasan waktu dan sumber daya dalam skala skripsi, namun diharapkan dapat berkembang lebih luas pada penelitian selanjutnya.

Hasil Kelayakan Instrumen Penilaian

Validasi ahli menghasilkan persentase 100% pada seluruh aspek penilaian dari kedua validator, sehingga instrumen dinyatakan Sangat Layak. Secara rinci, ahli evaluasi pembelajaran memberikan penilaian Sangat Layak pada aspek tampilan, konstruksi, dan pemrograman dengan skor total 80 (100%). Sedangkan ahli materi memberikan penilaian Sangat Layak pada aspek materi dan bahasa dengan skor total 120 (100%). Hasil validasi ahli selengkapnya disajikan pada Tabel 2.

Tabel 2. Hasil Validasi Ahli

Validator	Aspek	Persentase	Keterangan
Ahli Evaluasi Pembelajaran	Tampilan	100%	Sangat Layak
	Konstruksi	100%	Sangat Layak
	Pemrograman	100%	Sangat Layak
	Jumlah	100%	Sangat Layak
Ahli Materi	Materi	100%	Sangat Layak
	Bahasa	100%	Sangat Layak
	Jumlah	100%	Sangat Layak

Uji validitas empiris pada 30 peserta didik menunjukkan hasil yang memuaskan. Dari 20 butir soal yang diujicobakan, sebanyak 18 butir soal (90%) dinyatakan valid dengan nilai signifikansi $< 0,05$ menggunakan korelasi *Pearson (SPSS Statistics 26)*, sedangkan 2 butir soal (nomor 2 dan 17) dinyatakan tidak valid. Uji reliabilitas menghasilkan nilai *Cronbach's Alpha* sebesar 0,961 lebih dari ambang batas 0,60, sehingga instrumen dikategorikan memiliki reliabilitas sangat tinggi. Uji tingkat kesukaran menunjukkan 19 butir soal (95%) berkategori sedang dan 1 butir (nomor 12) berkategori mudah, serta tidak terdapat soal berkategori sukar. Uji daya pembeda menunjukkan 2 butir soal (10%) berkategori baik, 9 butir (45%) sedang, dan 9 butir (45%) jelek. Rekapitulasi pemilihan butir soal final untuk produk akhir menggunakan 11 butir soal yang telah memenuhi kriteria kevalidan, tingkat kesukaran pada kategori sedang, serta daya pembeda pada kategori baik dan sedang.

Hasil Kepraktisan Instrumen Penilaian

Uji coba terbatas dilaksanakan pada 25 peserta didik dan 1 guru kelas IV B SD Negeri Mejing 2. Respons guru menunjukkan persentase kepraktisan keseluruhan sebesar 84,8% dengan kategori Sangat Praktis: aspek kemudahan penggunaan 82,1% dan aspek kebermanfaatan 87,5%. Respons peserta didik menunjukkan persentase kepraktisan keseluruhan yang lebih tinggi, yaitu 94,5% dengan kategori Sangat Praktis: aspek kemudahan penggunaan 93% dan aspek kebermanfaatan 96%. Kelemahan yang teridentifikasi adalah kurang jelasnya petunjuk penggunaan pada setiap butir soal (nilai 0,84 dari peserta didik), yang kemudian menjadi dasar revisi produk dengan menambahkan petunjuk pengerjaan yang lebih eksplisit. Hasil angket kepraktisan selengkapnya disajikan pada Tabel 3.

Tabel 3. Hasil Uji Kepraktisan

Responden	Aspek	Persentase	Keterangan
Guru	Kemudahan Penggunaan	82,1%	Sangat Praktis

Responden	Aspek	Persentase	Keterangan
Peserta Didik	Kebermanfaatan	87,5%	Sangat Praktis
	Rata-rata	84,8%	Sangat Praktis
	Kemudahan Penggunaan	93%	Sangat Praktis
	Kebermanfaatan	96%	Sangat Praktis
	Rata-rata	94,5%	Sangat Praktis

Pembahasan

Proses Pengembangan Instrumen Penilaian

Proses pengembangan instrumen SI PARA dengan model 4D berjalan secara sistematis dan berbasis bukti. Pada tahap *Define*, temuan analisis awal menunjukkan bahwa rendahnya kemampuan berpikir kritis peserta didik dipengaruhi oleh dua faktor, yakni internal berupa rendahnya motivasi dan kebiasaan membaca, serta eksternal berupa lingkungan keluarga yang kurang mendukung budaya bernalar secara kritis. Hal ini selaras dengan (Hayati & Setiawan, 2022) yang membuktikan bahwa rendahnya kemampuan berbahasa dan bernalar berkorelasi langsung dengan rendahnya kemampuan berpikir kritis peserta didik SD. Ketiadaan instrumen berpikir kritis yang terstruktur juga diperkuat oleh (Nugroho & Airlanda, 2020) yang menyimpulkan bahwa instrumen demikian masih sangat terbatas ketersediaannya di tingkat SD.

Pada tahap *Design*, pemilihan soal uraian sebagai bentuk instrumen merupakan keputusan yang tepat secara pedagogis. (Mahanal, 2019) menegaskan bahwa tes subjektif lebih efektif dalam mengungkap kemampuan berpikir tingkat tinggi dibandingkan tes objektif, karena menuntut peserta didik mengonstruksi jawaban secara mandiri dan menunjukkan proses berpikirnya. Integrasi tiga platform *GenAI* juga mencerminkan prinsip pembelajaran multimodal. (Marinda et al., 2025) membuktikan bahwa platform digital interaktif mampu meningkatkan keterlibatan dan motivasi peserta didik secara signifikan. Fitur audio *text-to-speech* khususnya menjadi solusi inklusif sesuai prinsip *Universal Design for Learning* (UDL) yang mengutamakan fleksibilitas representasi materi (Mujiono et al., 2018). Pemilihan indikator berpikir kritis Robert Ennis sebagai landasan kisi-kisi juga tepat karena kerangkanya bersifat komprehensif, operasional, dan telah banyak digunakan dalam penelitian pendidikan (Ritdamaya & Suhandi, 2020).

Pada tahap *Develop*, pola validasi berulang tiga putaran mencerminkan pendekatan

pengembangan yang reflektif dan responsif. Revisi bahasa soal dari frasa abstrak ke frasa konkret berkaitan langsung dengan validitas konstruk. Bahasa yang ambigu berpotensi menimbulkan *construct-irrelevant variance*, yaitu varians skor yang muncul akibat faktor di luar konstruk yang hendak diukur (Sihotang et al., 2025). Dalam konteks peserta didik SD, perbaikan bahasa juga relevan secara perkembangan kognitif karena membantu mengurangi beban linguistik sehingga peserta didik dapat memusatkan proses berpikirnya pada analisis dan argumentasi (Wicaksono et al., 2020). Tahap *Disseminate* yang dilakukan secara terbatas tetap bermakna sebagai langkah awal penyebarluasan dan diharapkan berkembang lebih luas melalui pelatihan guru, platform digital, maupun integrasi dalam sistem pembelajaran sekolah.

Kelayakan Instrumen Penilaian

Hasil validasi ahli dengan persentase 100% pada seluruh aspek mengonfirmasi kelayakan instrumen secara teoritik. Nilai tersebut dicapai setelah dua kali revisi yang substantif, menunjukkan bahwa proses pengembangan berbasis masukan validator menghasilkan peningkatan kualitas yang nyata. Secara empiris, nilai *Cronbach's Alpha* 0,961 melampaui temuan (Maryono, 2022) yang memperoleh 0,903 dan (Siti et al., 2021) yang memperoleh 0,87 dalam pengembangan instrumen berpikir kritis serupa, mengindikasikan konsistensi internal yang sangat tinggi. Dua butir soal yang gugur (nomor 2 dan 17) tepat untuk dieliminasi karena butir soal yang tidak memenuhi standar validitas tidak akan merepresentasikan konstruk yang diukur secara tepat (Trianingsih, 2023). Keduanya juga bertepatan dengan butir berkategori daya pembeda jelek, mengonfirmasi secara ganda bahwa eliminasi merupakan tindakan yang benar.

Distribusi tingkat kesukaran yang didominasi kategori sedang (95%) sejalan dengan pendapat (Arikunto, 2019) bahwa instrumen berkualitas didominasi butir sedang karena

mampu mendiskriminasi kemampuan peserta didik secara lebih optimal. Meski demikian, ketiadaan soal berkategori sukar perlu mendapat perhatian. Dalam konteks pengukuran berpikir kritis yang menuntut proses kognitif C4-C6, ketiadaan soal sukar berpotensi menimbulkan *ceiling effect* apabila instrumen digunakan pada populasi dengan kemampuan lebih tinggi (Azhar et al., 2025). Hal ini menjadi rekomendasi pengembangan selanjutnya untuk menambahkan butir soal dengan indeks kesukaran lebih tinggi. Daya pembeda 55% kategori baik dan sedang menunjukkan bahwa instrumen mampu mengidentifikasi perbedaan antara peserta didik yang memiliki kemampuan tinggi dan rendah, sebuah karakteristik penting bagi instrumen berpikir kritis yang mendorong diferensiasi pembelajaran (Azizah, 2023)

Kepraktisan Instrumen Penilaian

Kepraktisan SI PARA dikonfirmasi dari dua sudut pandang yang berbeda. Respons guru (84,8%) yang lebih rendah dibandingkan respons peserta didik (94,5%) dapat dijelaskan dari perspektif psikologi motivasi belajar. Peserta didik menilai dari pengalaman langsung menggunakan instrumen yang interaktif dan menyenangkan, jauh berbeda dari soal kertas konvensional. Sementara guru menilai dengan mempertimbangkan aspek teknis yang lebih luas seperti kemudahan integrasi dengan metode pembelajaran yang sudah ada. Meski demikian, keduanya sama-sama masuk kategori Sangat Praktis, yang merupakan validasi ganda penting sebelum implementasi lebih luas. Tingginya respons peserta didik sejalan dengan (Husein et al., 2020) yang membuktikan bahwa instrumen berbantuan teknologi interaktif mampu meningkatkan keterampilan berpikir kritis sekaligus motivasi peserta didik secara bersamaan.

Satu-satunya kelemahan yang teridentifikasi yaitu kejelasan petunjuk penggunaan (nilai 0,84) langsung ditindaklanjuti dengan revisi berbasis prinsip *user-centered design*. Fakta bahwa instrumen dinilai efisien dari segi biaya oleh guru juga sangat relevan dalam konteks pendidikan Indonesia, karena platform Canva AI yang digunakan bersifat gratis dan dapat diakses melalui berbagai perangkat sehingga meminimalkan hambatan implementasi (Muslimin & Fatimah, 2024). Fitur-fitur SIPARA seperti indikator progres, tombol navigasi fleksibel, gambar ilustrasi tematik, dan audio pembacaan teks berkontribusi pada peningkatan

keterlibatan kognitif peserta didik, sejalan dengan simpulan (Qian & Clark, 2016) bahwa antarmuka digital interaktif meningkatkan engagement secara signifikan. Dari perspektif teori konstruktivisme Piaget, stimulus visual dan audio kontekstual yang disajikan SI PARA sangat membantu peserta didik kelas IV (usia 9-10 tahun) yang berada dalam tahap operasional konkret untuk membangun pemahaman dan mengembangkan berpikir kritis.

KESIMPULAN

Penelitian ini berhasil mengembangkan instrumen penilaian kemampuan berpikir kritis berbantuan *Generative AI* yang diberi nama SI PARA (Sistem Penilaian Paragraf Argumentasi untuk Mengukur Berpikir Kritis) menggunakan model 4D. Instrumen dikembangkan dalam format kuis daring interaktif melalui *Canva AI* yang dilengkapi ilustrasi tematik (*ChatGPT*) dan audio *text-to-speech* (*Google AI Studio*). Hasil penelitian menunjukkan bahwa instrumen terbukti layak dengan persentase validasi ahli 100% (Sangat Layak), reliabilitas 0,961 (sangat tinggi), 18 dari 20 butir valid (90%), serta 11 butir final memenuhi kriteria valid, tingkat kesukaran sedang, dan daya pembeda baik dan sedang. Instrumen juga terbukti praktis dengan persentase kepraktisan guru 84,8% dan peserta didik 94,5%, keduanya dalam kategori Sangat Praktis. Penelitian ini merekomendasikan agar pengembangan selanjutnya mempertimbangkan penambahan butir soal berkategori sukar untuk mencegah *ceiling effect*, serta memperluas diseminasi melalui pelatihan guru dan integrasi dalam sistem pembelajaran sekolah secara lebih luas.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Universitas PGRI Yogyakarta yang telah memberikan dukungan pendanaan melalui Anggaran LPPM Tahun 2025–2026. Dukungan tersebut sangat berperan dalam kelancaran pelaksanaan dan penyelesaian penelitian ini. Penulis juga menyampaikan apresiasi kepada seluruh pihak yang telah membantu, baik secara langsung maupun tidak langsung, sehingga penelitian ini dapat terselesaikan dengan baik.

REFERENSI

Arikunto, S. (2019). *Prosedur Penelitian: Suatu*

- Pendekatan Praktik*. Rineka Cipta.
- Azhar, A., Herdiansyah, Nurfauzi, D., Khoirunnisa, N. A., Putri, H. E., & Rahayu, T. G. (2025). 2. Pengembangan instrumen untuk mengukur kemampuan berpikir kritis siswa melalui konsep operasi hitung perkalian dan pembagian pecahan pada siswa kelas vi sd 3 babakan sari. *Jurnal Review Pendidikan Dan Pengajaran*, 8(1), 191–200.
- Azizah, I. (2023). Analisis Kualitas Butir Soal Penilaian Harian Bersama Ifisika Kelas X SMA Negeri 1 Patikraja. *Jurnal Pendidikan Fisika*, 10(November), 90–104.
- Fahma, A. (2025). *Pengembangan instrumen tes keterampilan berpikir kritis terintegrasi keislaman mengenai konservasi di Sekolah Menengah Atas*.
- Hayati, N., & Setiawan, D. (2022). Dampak Rendahnya Kemampuan Berbahasa dan Bernalar terhadap Kemampuan Berpikir Kritis Siswa Sekolah Dasar. *Jurnal Basicedu*, 6(5), 8517–8528. <https://doi.org/10.31004/basicedu.v6i5.3650>
- Husein, S., Herayanti, L., & Gunawan. (2020). Pengaruh penggunaan multimedia interaktif terhadap penguasaan konsep dan keterampilan berpikir kritis siswa pada materi suhu dan kalor. *Jurnal Pendidikan Fisika Dan Teknologi*, 1(3), 221–225.
- Mahanal, S. (2019). Asesmen Keterampilan Berpikir Tingkat Tinggi. *Jurnal Penelitian Dan Pengkajian Ilmu Pendidikan: E-Saintika*, 3(2), 51–73.
- Marinda, A. B., Dewanti, A. H., & Irama, D. (2025). Inovasi Pembelajaran Berbasis Teknologi Digital untuk Meningkatkan Kompetensi Belajar Peserta Didik di SMP Muhammadiyah 1 Curup. *Jurnal Literasiologi Volume*, 14(3), 1–10.
- Maryono. (2022). Uji Validasi dan Reliabilitas Instrumen Penilaian Kemampuan Berpikir Kritis Siswa Kelas VI Sekolah Dasar. *Jurnal Kualita Pendidikan*, 3(3), 146–151.
- Mujiono, Degeng, I. N. S., & Praherdhiono, H. (2018). Pengembangan Pembelajaran Sistem Blended Berbasis Universal Design for Learning untuk Kelas Inklusif. *Jurnal Pendidikan*, 3(6), 758–763.
- Muslimin, T. P., & Fatimah, A. A. B. (2024). Kompetensi dan Kesiapan Guru Sekolah Dasar Terhadap Tantangan Pendidikan di Era Society 5 . 0. *CJPE : Cokroaminoto Journal of Primary Education*, 7(4), 55–72.
- Nugroho, A. N., & Airlanda, G. S. (2020). Pengembangan Instrumen Penilaian Keterampilan Berpikir Kritis Pembelajaran IPA Kelas 4 SD. *Jurnal Ilmiah Pendidikan Profesi Guru*, 3(3), 400. <https://doi.org/10.23887/jippg.v3i3.29712>
- Nurhasanah, A., & Hidayat, A. R. (2024). *Mata pelajaran sejarah demokrasi liberal kelas xii ipa 3 di sma negeri 7 kota serang*. 4(1), 60–73. <https://doi.org/10.22437/jejak.v4i1.29345>
- Partono, Wardhani, H. N., Setyowati, N. I., Tsalitsa, A., & Nurrahayu, S. (2021). Strategi Meningkatkan Kompetensi 4C Creativity, Communication, & Collaborative) (Critical Thinking, Strategies to Improve 4C Competencies (Critical Thinking, Creativity, Communication & Collaborative). *Jurnal Penelitian Ilmu Pendidikan*, 14(1), 41–52.
- Qian, M., & Clark, K. R. (2016). Game-based Learning and 21st century skills: A review of recent research. *Computers in Human Behavior*, 63, 50–58. <https://doi.org/10.1016/j.chb.2016.05.023>
- Rahmawati, H., Pujiastuti, P., & Cahyaningtyas, A. P. (2023). Kategorisasi Kemampuan Berpikir Kritis Siswa Kelas Empat Sekolah Dasar di SD se-Gugus II Kapanewon Playen, Gunung Kidul. *Jurnal Pendidikan Dan Kebudayaan*, 8(1), 88–104. <https://doi.org/10.24832/jpnk.v8i1.3338>
- Ritdamaya, D., & Suhandi, A. (2020). Konstruksi Instrumen Tes Keterampilan Berpikir Kritis Terkait Materi Suhu dan Kalor. *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, 02(2), 87–96. <https://doi.org/10.21009/1.02212>
- Rofi'ah, S., & Rokhmaniyah. (2024). Analisis Kemampuan Berpikir Kritis Siswa dalam Memecahkan Masalah pada Mata Pelajaran IPAS kelas V Sekolah Dasar. *Social, Humanities, and Education Studies (SHEs): Conference Series*, 7(3), 1763–1770.
- Sholeh, M. I., Sokip, Syafi'I, A., & Sulistyorini. (2024). Artificial Intelligence in Education : Transforming Assessment and Personalized Learning. *JELS: Journal of Education and Learning Sciences*, 04(01), 50–62.
- Sihotang, A. E., Hutagalung, S. H., Saragih, S., & Akbar, R. F. (2025). Analisis Validitas Isi Dan Konstruksi Pada Soal Ujian Akhir Semester Bahasa Indonesia Kelas VIII

- SMP Harapan 1 Medan. *JKIP : Jurnal Kajian Ilmu Pendidikan*, 6(2), 536–544.
- Siti, N., Gumilang, R., & Tsurayya, A. (2021). Pengembangan Instrumen Kemampuan Berpikir Kritis dan Kreatif Matematika Peserta Didik Kelas VII SMP Development of Mathematics ' Critical and Creative Thinking Instruments for Grade VII on Secondary School. *Jurnal Pendidikan Matematika Dan Sains*, 9(2), 89–98.
- Sudijono, A. (2012). *Pengantar Evaluasi Pendidikan* (Raja Grafindo Persada (ed.)).
- Sugiyono. (2024). Metode penelitian kuantitatif kualitatif dan R&D. In *Penerbit Alfabeta*.
- Trianingsih, L. (2023). Analisis Kualitas Tes dan Butir Soal Integral pada Evaluasi Formatif Matematika Teknik. *IJCEE*, 9(2), 44–57.
- Wicaksono, A. A., Sunarti, T., & Zainuddin, A. (2020). Validitas dan Reliabilitas Instrumen Penilaian Kinerja Praktikum Pada Materi Alat Optik Kelas XI SMA. *IPF : Inovasi Pendidikan Fisika*, 09(03), 325–331.