

Pengembangan instrumen tes uraian yang valid dan reliabel untuk mengukur kemampuan berpikir kritis matematis siswa kelas V SD

Teten Ginanjar Rahayu^{1*}, Alfiana Nurussama², Hafiziani Eka Putri³, Resa Salma Salsabila⁴, Maryam Khaerani⁵, Bernardo Manalu⁶

^{1*,2,3,4,5,6} Pendidikan Guru Sekolah Dasar, Universitas Pendidikan Indonesia, Indonesia

* E-mail corresponding author: tetenginanjarr@upi.edu

Received : 28 Mei 2025, Accepted : 31 Juli 2025, Published : 31 Juli 2025

Abstrak: Penelitian ini bertujuan untuk mengembangkan instrumen tes uraian yang valid, praktis, dan reliabel guna mengukur kemampuan berpikir kritis matematis serta hasil belajar matematika siswa kelas V sekolah dasar. Proses pengembangan instrumen mengikuti tahapan *Research and Development* (R&D) dengan menggunakan model 4-D yang terdiri dari empat tahap utama, yaitu: *Define* (pendefinisian), *Design* (perancangan), *Develop* (pengembangan), dan *Disseminate* (penyebarluasan). Validasi instrumen dilakukan oleh tiga ahli di bidang pendidikan matematika, sedangkan uji coba diberikan kepada 33 siswa kelas V di salah satu SD. Kepraktisan instrumen dievaluasi melalui uji coba terbatas dengan melibatkan guru dan siswa, yang menilai aspek kemudahan penggunaan, dan kejelasan petunjuk. Hasil uji kepraktisan menunjukkan bahwa instrumen mudah digunakan oleh guru dalam proses penilaian dan dipahami oleh siswa tanpa menimbulkan kebingungan atau mengganggu jalannya pembelajaran. Analisis data dilakukan melalui uji validitas butir menggunakan korelasi Pearson dan uji reliabilitas dengan *Alpha Cronbach*. Hasil analisis menunjukkan bahwa semua butir instrumen memenuhi kriteria validitas dengan koefisien antara 0,45-0,82, serta reliabilitas instrumen mencapai 0,87 yang tergolong sangat baik. Berdasarkan hasil tersebut, instrumen tes uraian yang dikembangkan dinyatakan layak digunakan karena telah memenuhi kriteria valid, praktis, dan reliabel, sehingga dapat diandalkan sebagai alat ukur kemampuan berpikir kritis matematis dan hasil belajar matematika pada siswa SD.

Kata kunci: *Pengembangan Instrumen, Instrumen Penilaian; Matematika; Berpikir Kritis Matematis; Siswa Sekolah Dasar.*

Abstract: This study aims to develop an essay test instrument that is valid, practical, and reliable for measuring mathematical critical thinking skills and mathematics learning outcomes of fifth-grade elementary school students. The instrument development process follows the stages of *Research and Development* (R&D) using the 4-D which includes four main phases: *Define*, *Design*, *Develop*, and *Disseminate*. Instrument validation was conducted by three experts in mathematics education, while field trials were carried out with 33 fifth-grade students at an SD. Practicality of the instrument was evaluated through limited trials involving teachers and students, focusing on aspects such as ease of use, and clarity of instructions. The results of the practicality test showed that the instrument was easy for teachers to use in the assessment process and was well understood by students without causing confusion or disrupting classroom activities. Data analysis was performed using item validity tests with Pearson correlation and reliability tests with Cronbach's Alpha. The results indicated that all instrument items met the validity criteria, with correlation coefficients ranging from 0.45 to 0.82, and the instrument's reliability reached 0.87, which is categorized as very good. Based on these findings, the developed essay test instrument is considered feasible for use because it meets the criteria of being valid, practical, and reliable, making it a dependable tool for assessing mathematical critical thinking skills and mathematics learning outcomes in elementary SD.

Kata kunci: *Instrument Development; Assessment Instrument; Mathematics; Mathematical Critical Thinking; Elementary School Students.*

How to Cite: Rahayu, T. G., Nurussama, A., Putri, H. E., Salsabila, R. S., Khaerani, M., & Manalu, B. (2025). Pengembangan instrumen tes uraian yang valid dan reliabel untuk mengukur kemampuan berpikir kritis matematis siswa kelas V SD. *Delta-Pi: Jurnal Matematika dan Pendidikan Matematika*, 14(1), 137-154.

<https://doi.org/10.33387/dpi.v14i1/9973>

A. Pendahuluan

Evaluasi merupakan salah satu tanggung jawab utama guru yang sangat menentukan arah dan kualitas pembelajaran di masa depan. Evaluasi dalam pembelajaran adalah proses yang sistematis dengan mskdu mengukur sejauh mana tujuan dalam pembelajaran telah tercapai serta memberikan dasar bagi guru untuk merancang strategi pengajaran berikutnya (Wangsa et al., 2021). Dalam konteks pendidikan abad ke-21, evaluasi tidak hanya bertujuan menilai hasil belajar, tetapi juga berfungsi sebagai alat refleksi untuk meningkatkan mutu proses pembelajaran (OECD, 2023).

Pada pembelajaran matematika sekolah dasar, kemampuan siswa dalam berpikir kritis menjadi salah satu aspek penting yang harus dikembangkan. Berdasarkan hasil PISA 2022, kemampuan berpikir kritis dan pemecahan masalah matematis siswa Indonesia masih berada di bawah rata-rata negara-negara OECD, di mana hanya 18% siswa Indonesia yang mencapai tingkat kemahiran minimum dalam matematika, jauh lebih rendah dibandingkan rata-rata OECD sebesar 69% (OECD, 2023). Taksonomi Bloom revisi menegaskan pentingnya pengukuran kemampuan berpikir siswa pada kategori tingkat tinggi, yang meliputi menganalisis pada level C4, mengevaluasi pada leverl C5, dan mencipta atau kreasi pada level C6 (Anderson & Krathwohl, 2001; Gunawan & Palupi, 2016). Kompetensi ini sangat relevan dengan tuntutan abad ke-21, di mana diharapkan siswa, siswa dapat mengembangkan kemampuan untuk memecahkan masalah secara kreatif, bekerja sama dengan orang lain, serta berpikir secara kritis (Binkley et al., 2014).

Instrumen evaluasi yang digunakan di banyak sekolah dasar di Indonesia masih didominasi oleh soal-soal yang bersifat prosedural dan belum sepenuhnya mampu mengukur kemampuan berpikir kritis matematis siswa secara komprehensif (Evi & Indarini, 2021; Jaya et al., 2020). Padahal, kemampuan berpikir kritis dalam matematika tidak hanya mencakup pemahaman konsep dasar, tetapi juga kemampuan menganalisis, menyusun argumen, menyelesaikan masalah, dan mengevaluasi informasi secara sistematis (Gulo et al., 2024; Masitoh & Prabawanto, 2016). Dengan demikian, perancangan instrumen evaluasi yang memiliki validitas tinggi menjadi suatu keharusan agar proses penilaian benar-benar dapat mengukur capaian pembelajaran secara akurat, reliabel, dan praktis sangat diperlukan untuk menilai kemampuan siswa secara objektif dan menyeluruh (Fitriana, 2022). Selain itu, sebagian besar instrumen penilaian yang dikembangkan sebelumnya masih terbatas pada validitas isi, tanpa pengujian reliabilitas dan kepraktisan secara menyeluruh, sehingga penggunaannya dalam konteks kelas seringkali kurang optimal (Aini et al., 2023; Fitriana, 2022; Maulana, 2022).

Untuk itu perlu dilakukan penelitian dengan fokus pengembangan instrumen untuk mengatasi keterbatasan instrumen evaluasi matematika di tingkat sekolah dasar yang tidak hanya memenuhi kriteria validitas isi, tetapi juga belum secara optimal diuji kepraktisan dan reliabilitasnya sebagai alat ukur kemampuan berpikir kritis matematis secara objektif dan menyeluruh. Berbagai penelitian sebelumnya umumnya berfokus pada Pengembangan instrumen penilaian yang berorientasi pada keterampilan berpikir tingkat tinggi (Higher Order Thinking Skills/HOTS) atau sekadar menguji validitas konten, namun belum banyak yang melakukan uji kepraktisan maupun reliabilitas secara empiris dalam konteks pembelajaran di kelas (Fitriana, 2022; Hapsari & Gularso, 2024; Jaya et al., 2020; Magdalena et al., 2021; Pratiwi & Sukmawati, 2023; Yunita & Oktaviana, 2021). Selain itu, instrumen yang telah dikembangkan seringkali belum diadaptasi secara spesifik untuk karakteristik dan kebutuhan siswa sekolah dasar, sehingga implementasinya di lapangan masih menemui berbagai kendala (Evi & Indarini, 2021; Masitoh & Prabawanto, 2016). Oleh karena itu, pengembangan instrumen penilaian yang tidak hanya memenuhi kriteria validitas isi, namun juga memiliki tingkat kepraktisan yang tinggi dan reliabel serta telah teruji secara empiris di kelas sangat diperlukan untuk mendukung proses evaluasi dan pembelajaran matematika yang difokuskan pada pengembangan kemampuan siswa dalam melakukan berpikir kritis di sekolah (Gunawan & Palupi, 2016; OECD, 2023).

Instrumen evaluasi yang baik tidak sekadar merefleksikan tingkat pencapaian siswa, tetapi juga menjadi dasar bagi guru dalam merancang pembelajaran yang didasarkan pada kebutuhan siswa. Lalu, instrumen yang dikembangkan harus memenuhi prinsip keadilan dan kesetaraan, sehingga dapat memberikan peluang yang sama bagi semua siswa untuk menunjukkan kemampuannya (Wangsa et al., 2021). Proses pengembangan instrumen evaluasi juga harus melalui tahapan validasi oleh ahli dan uji coba lapangan agar instrumen yang dihasilkan benar-benar layak digunakan dalam konteks pendidikan dasar (Fitriana, 2022). Evaluasi yang efektif tidak hanya bermanfaat bagi siswa, tetapi juga memberikan panduan bagi guru dalam menilai dan meningkatkan strategi pembelajaran. Hasil evaluasi yang mengidentifikasi kelemahan tertentu pada siswa dapat menjadi dasar bagi guru untuk mengadopsi pendekatan pembelajaran yang lebih sesuai dengan kebutuhan peserta didik.

Dengan demikian, evaluasi berperan sebagai alat refleksi yang penting dalam pengembangan pendidikan. Pengembangan instrumen evaluasi yang selaras dengan kebutuhan pembelajaran abad ke-21 menjadi sangat relevan, khususnya yang menekankan pada kompetensi berpikir kritis, kolaborasi, dan kreativitas. Penelitian-penelitian terbaru

menunjukkan bahwa instrumen penilaian berbasis kemampuan siswa dalam berpikir kritis dan *Higher Order Thinking Skills/HOTS* telah terbukti validitasnya, praktisitasnya, dan efektifitasnya dalam meningkatkan hasil belajar serta kemampuan siswa dalam berpikir kritis di sekolah dasar (Aini et al., 2023; Hapsari & Gularso, 2024; Pratiwi & Sukmawati, 2023; Yunita & Oktaviana, 2021). Namun, penelitian ini menawarkan kontribusi yang berbeda dari penelitian sebelumnya, karena instrumen penilaian yang dikembangkan tidak hanya valid secara isi, tetapi juga praktis dan reliabel serta telah teruji secara empiris di kelas. Pengembangan instrumen semacam ini sangat diperlukan untuk mendukung proses evaluasi dan pembelajaran matematika secara komprehensif. Atas dasar pertimbangan tersebut, penelitian ini difokuskan untuk mengembangkan instrumen penilaian yang sudah teruji validitasnya, praktisitasnya, serta reliabilitasnya guna mengukur kemampuan berpikir kritis matematis serta hasil belajar matematika siswa kelas V sekolah dasar. Temuan dan hasil dari penelitian ini dimaksudkan agar manfaatnya tidak terbatas pada pembelajaran matematika, tetapi juga menjadi dasar bagi peningkatan kualitas pendidikan di Indonesia secara keseluruhan.

B. Metode Penelitian

Dalam penelitian ini, digunakan model 4-D (*Four-D Model*) yang diperkenalkan oleh Thiagarajan dan rekan-rekannya pada tahun 1974 sebagai dasar pengembangan perangkat pembelajaran. Thiagarajan et al. (1974) menyatakan bahwa "*The model consist of four major stages: define, design, develop, and disseminate*" yang memiliki arti sebuah model pengembangan perangkat pembelajaran yang terdiri atas empat tahap utama, yaitu: *define* (pendefinisian), *design* (perancangan), *develop* (pengembangan), dan *disseminate* (penyebaran). Model ini telah banyak diaplikasikan dalam tahapan pengembangan instrumen pendidikan karena memberikan alur kerja yang sistematis dan terstruktur. Penelitian ini mengadopsi pendekatan desain penelitian pengembangan, atau *Research and Development* (R&D), yang bertujuan untuk menghasilkan produk berupa instrumen yang digunakan untuk mengukur hasil belajar matematika siswa. Selain itu, penelitian ini juga bertujuan untuk menghasilkan instrumen yang dapat digunakan dalam rangka menilai keterampilan berpikir kritis matematis pada siswa kelas V Sekolah Dasar (SD).

Instrumen-instrumen tersebut dirancang untuk dapat memberikan informasi dengan lebih komprehensif terhadap pemahaman peserta didik terhadap materi matematika serta kemampuan mereka dalam menerapkan konsep-konsep matematika secara kritis dalam berbagai situasi yang dihadapi. Dengan demikian, penelitian ini diharapkan dapat berkontribusi

dalam mengembangkan cara-cara penilaian yang lebih efektif dan relevan untuk mengukur perkembangan kemampuan matematis siswa di tingkat sekolah dasar.

Adapun langkah pengembangan instrument yang digunakan pada penelitian ini adalah menyusun spesifikasi tes, menulis soal tes, menelaah soal tes, melakukan uji coba, menganalisis soal, melakukan revisi terhadap soal, menyusun tes secara menyeluruh, mengadministrasikan tes kepada peserta, serta menganalisis dan menafsirkan hasil tes yang diperoleh. (Cooke et al., 2011; Setiana, 2018). Tes diartikan juga sebagai sejumlah pertanyaan yang membutuhkan jawaban atau tanggapan dengan tumengukur tingkat kemampuan seseorang. Terdapat sembilan tahapan utama yang harus dilalui dalam proses pengembangan tes hasil belajar, diantaranya: 1) menyusun spesifikasi tes; 2) menulis soal tes; 3) menelaah soal tes; 4) melakukan uji coba tes; 5) menganalisis butir soal; 6) memperbaiki tes; 7) merakit tes; 8) melaksanakan tes; dan 9) menafsirkan hasil tes (Mardapi et al., 2011). Lalu sembilan langkah tersebut dikelompokkan ke dalam empat tahapan utama tersebut sehingga dapat dilihat pengelompokkannya sebagai berikut: 1) Tahap *Define* meliputi menyusun spesifikasi tes; 2) Tahap *Design* mencakup menulis soal tes dan menelaah soal tes; 3) Tahap *Develop* meliputi melakukan uji coba tes, menganalisis butir soal, memperbaiki tes, dan merakit tes; 4) Tahap *Disseminate* terdiri atas melaksanakan tes dan menafsirkan hasil tes.

C. Hasil dan Pembahasan

Bagian hasil dan pembahasan dalam artikel ini menyajikan proses dan temuan utama dari pengembangan instrumen tes hasil belajar. Dalam upaya menghasilkan instrumen penilaian yang valid, reliabel, dan efektif, penelitian ini mengikuti sembilan langkah sistematis sebagaimana diuraikan oleh Mardapi et al. (2011), yaitu: 1) menyusun spesifikasi tes; 2) menulis soal tes; 3) menelaah soal tes; 4) melakukan uji coba tes; 5) menganalisis butir soal; 6) memperbaiki tes; 7) merakit tes; 8) melaksanakan tes; dan 9) menafsirkan hasil tes (Cooke et al., 2011; Mardapi et al., 2011; Setiana, 2018). Seluruh langkah tersebut kemudian dikelompokkan ke dalam empat tahapan utama berdasarkan model pengembangan 4-D yang digunakan dalam penelitian, yaitu model pengembangan 4D yang meliputi tahap pendefinisian (*define*), perancangan (*design*), pengembangan (*develop*), serta penyebaran (*disseminate*). Pengelompokan ini memberikan kerangka yang terstruktur dalam proses pengembangan instrumen, sehingga setiap tahap dapat dijelaskan secara rinci untuk memberikan gambaran yang komprehensif mengenai proses pengembangan instrumen serta kualitas hasil yang diperoleh. Diharapkan, paparan ini dapat memberikan kontribusi nyata dalam meningkatkan kualitas penilaian hasil belajar, khususnya pada pembelajaran matematika di sekolah dasar.

1. Tahap *Define*

Tahap awal pengembangan instrumen adalah tahap *Define* meliputi menyusun spesifikasi soal tes menyusun spesifikasi soal tes. Pada tahap *define* ini, peneliti merumuskan tujuan pengukuran serta membuat blueprint atau kisi-kisi tes yang memuat kompetensi dasar, indikator, materi, serta bentuk dan teknik penilaian yang sesuai. Spesifikasi tes menjadi acuan utama dalam penulisan butir-butir soal dilakukan dengan menyesuaikan pada tujuan pengukuran serta karakteristik kompetensi yang hendak diukur (Alam et al., 2019). Berikut adalah indikator beserta deskripsinya yang digunakan dalam pengembangan instrumen tes kemampuan berpikir kritis matematis siswa kelas V SD menurut Facione (dalam Ira et al., 2024; Mastuti et al., 2022; Mufidah & Khoiriyah, 2021).

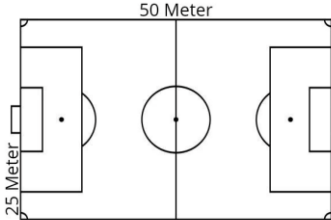
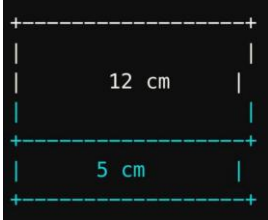
Tabel 1. Indikator Kemampuan Berpikir Kritis Matematis

Indikator	Deskripsi
Interpretasi	Memahami informasi dalam soal, menuliskan apa saja yang diketahui dan ditanyakan secara lebih tepat
Analisis	Menghubungkan informasi, memilih rumus, dan langkah penyelesaian yang tepat
Evaluasi	Menilai ketepatan strategi dan memeriksa hasil perhitungan
Inferensi	Menyimpulkan hasil akhir secara logis dan tepat

2. Tahap *Design*

Berdasarkan spesifikasi yang telah disusun, langkah berikutnya adalah tahap *Design* yang mencakup menulis atau menyusun soal tes dan juga menelaah soal tes menulis butir-butir soal. Penulisan soal dilakukan dengan memperhatikan kaidah penulisan soal yang baik dan benar untuk bentuk soal uraian. Setiap butir soal harus mengukur indikator yang telah ditetapkan seperti pada tabel berikut ini.

Tabel 2. Butir-Butir Soal Uraian Kemampuan Berpikir Kritis Matematis

Indikator Berpikir Kritis Matematis	Soal Tes
Interpretasi	Memahami informasi dalam soal, menuliskan apa yang diketahui dan ditanyakan secara tepat  Sebuah lapangan sepak bola memiliki bentuk persegi panjang dengan ukuran panjang 50 meter dan lebar 25 meter. Tentukan berapa keliling lapangan sepak bola tersebut!
Analisis	Menghubungkan informasi, memilih rumus, dan langkah penyelesaian yang tepat  Ada sebuah persegi panjang yang memiliki panjang 12 cm dan lebar 5 cm. Jika lebar diperpendek menjadi 3 cm, berapa selisih panjang awal dengan yang baru?
Evaluasi	Menilai ketepatan strategi dan memeriksa hasil perhitungan  Ukuran sebuah taman berbentuk persegi panjang memiliki panjang 15 m dan lebar 10 m. Jika panjangnya ditambah 5 m dan lebarnya ditambah 2 m, berapa keliling taman yang baru?
Inferensi	Menyimpulkan hasil akhir secara logis dan tepat Sebuah persegi panjang mempunyai panjang 8 cm dan lebar 5 cm. Jika panjang persegi panjang tersebut diperpanjang 2 cm dan lebarnya diperpendek 1 cm, berapakah perubahan pada kelilingnya?

Soal-soal uraian tersebut dikembangkan menjadi 10 butir soal yang akan disajikan pada tabel 4 soal tes sebelum dan sesudah revisi dengan tingkat kesukaran dan level kognitif yang sama, namun masing-masing soal disajikan dalam konteks yang berbeda-beda. Dengan demikian, setiap butir soal mengukur kemampuan yang setara, tetapi menggunakan situasi atau permasalahan yang bervariasi untuk menguji pemahaman siswa secara lebih komprehensif. Setelah soal ditulis, dilakukan penelaahan atau telaah soal oleh ahli (*expert judgment*). Penelaahan bertujuan menilai kesesuaian isi, bahasa, dan konstruk soal dengan spesifikasi serta

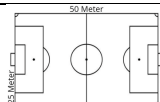
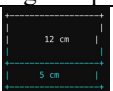
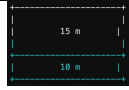
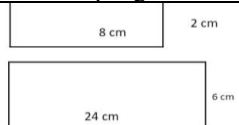
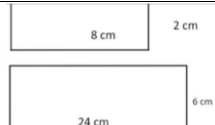
memastikan tidak ada bias atau kekeliruan dalam soal. Hasil telaah digunakan untuk merevisi soal sebelum uji coba. Hasil telaah menunjukkan bahwa soal-soal memiliki catatan sebagai berikut:

Tabel 3. Hasil Telaah Expert

Aspek Penilaian	Saran Perbaikan Terperinci
Kesesuaian indikator	Pastikan setiap soal secara eksplisit mengukur keempat indikator kemampuan berpikir kreatif matematis: kelancaran (banyak cara/jawaban), kelenturan (beragam strategi), keaslian (jawaban unik/orisinil), dan keterincian (penjelasan langkah/detail). Jangan hanya satu atau dua indikator yang terakomodasi.
Redaksi soal	Redaksi soal harus diperjelas dan disesuaikan dengan tingkat perkembangan kognitif siswa SD. Hindari istilah teknis atau kalimat yang ambigu. Gunakan bahasa yang sederhana, lugas, dan mudah dipahami oleh siswa kelas IV.
Jenis soal	Hindari soal yang hanya menguji hafalan atau prosedur rutin. Kembangkan soal yang menuntut penalaran, argumentasi, dan memungkinkan lebih dari satu cara penyelesaian. Soal sebaiknya bersifat terbuka sehingga siswa dapat menunjukkan kreativitasnya.
Tingkat kesulitan	Soal yang terlalu mudah atau terlalu teknis perlu direvisi agar lebih menantang, namun tetap sesuai dengan kemampuan siswa SD. Soal harus dapat mendorong siswa berpikir kritis dan kreatif, bukan sekadar mengulang rumus.
Unsur keterincian	Tambahkan instruksi pada soal agar siswa tidak hanya memberikan jawaban akhir, tetapi juga menjelaskan langkah-langkah, alasan, atau proses berpikir yang digunakan. Hal ini penting untuk mengukur indikator keterincian dengan baik.
Keputusan akhir	Setelah revisi sesuai saran di atas, soal dinyatakan layak digunakan untuk uji coba di lapangan. Jika masih ada kekurangan minor, lakukan revisi kecil sebelum implementasi.

Berikut adalah soal tes sebelum dan sesudah revisi oleh ahli (*expert*).

Tabel 4. Soal Tes Sebelum dan Sesudah Revisi

No Soal	Sebelum Revisi	Sesudah Revisi
1	Jika diketahui panjang persegi di atas adalah 8 cm, tentukan lebar persegi panjang tersebut?	 Sebuah lapangan sepak bola berbentuk persegi panjang dengan panjang 50 meter dan lebar 25 meter. Hitunglah keliling dari lapangan sepak bola tersebut!
2	Diketahui halaman rumah Tasya berbentuk persegi panjang dan memiliki luas dengan ukuran 240 m ² . Jika diketahui salah satu sisi halaman tersebut memiliki panjang 10 m Tentukan berapa lebar halaman rumah tasya?	 Sebuah persegi panjang memiliki panjang 12 cm dan lebar 5 cm. Jika lebar diperpendek menjadi 3 cm, berapa selisih panjang awal dengan yang baru?
3	Memiliki kebun berbentuk persegi panjang dengan ukuran panjang 20 m dan lebar 7 m. Ayah Ingin menjual kebun tersebut dengan harga Rp. 700.000,- per m ² Berapakah uang yang dibutuhkan untuk membeli tanah mil Ayah?	 Sebuah taman berbentuk persegi panjang memiliki panjang 15 m dan lebar 10 m. Jika panjangnya ditambah 5 m dan lebarnya ditambah 2 m, berapa keliling taman yang baru?
4	Sebuah persegi panjang memiliki panjang 8 cm dan lebar 5 cm. Jika panjang persegi panjang tersebut diperpanjang 2 cm dan lebarnya diperpendek 1 cm, berapakah perubahan pada kelilingnya?	Sebuah persegi panjang memiliki panjang 8 cm dan lebar 5 cm. Jika panjang persegi panjang tersebut diperpanjang 2 cm dan lebarnya diperpendek 1 cm, berapakah perubahan pada kelilingnya?
5	Andri mengelilingi lapangan yang berbentuk lingkaran dengan jari-jari 28 m. Ia mengelilingi lapangan tersebut sebanyak 3 kali putaran. Jika Andri berencana melakukan aktivitas serupa di lapangan lain yang memiliki jari-jari 35 m, berapa perbedaan jarak total yang ditempuh jika Andri tetap melakukan 3 kali putaran di kedua lapangan tersebut?	Andri mengelilingi lapangan yang berbentuk lingkaran dengan jari-jari 28 M . Ia mengelilingi Lapangan tersebut sebanyak 3 kali putaran. Jika Andri berencana melakukan aktivitas serupa di lapangan lain yang memiliki 35 m, perbedaan jari-jari berapa perbedaan jarak total yang ditempuh Andri jika Andri tetap melakukan 3 kali putaran kedua di lapangan tersebut?
6	 Gambar di atas menunjukkan 2 persegi panjang memiliki ukuran yang berbeda. Jika kedua persegi panjang tersebut dijumlahkan, perhatikan 2 pernyataan di bawah ini: Menurut Danu luas kedua persegi panjang tersebut jika dijumlahkan adalah 100 cm². Menurut Siti luas kedua persegi panjang tersebut jika dijumlahkan adalah 160 cm² perhitunganmu Jawab apakah yang benar?	 Gambar di atas menunjukkan dua persegi panjang dengan ukuran yang berbeda. Jika luas kedua persegi panjang tersebut dijumlahkan, perhatikan dua pernyataan di bawah ini: Menurut Danu, luas kedua persegi panjang tersebut jika dijumlahkan adalah 100 cm². Menurut Siti, luas kedua persegi panjang tersebut jika dijumlahkan adalah 160 cm². Menurut kalian, pendapat siapakah yang benar?
7	Jika keliling sebuah persegi panjang adalah 30 cm dan lebar persegi panjang tersebut adalah 7 cm, berapakah panjangnya? Jelaskan cara perhitungannya!	Sebuah kolam renang berbentuk persegi panjang memiliki ukuran 25 m x 10 m. Jika kolam renang tersebut dibuat lebih lebar dengan tambahan 5 m pada lebar, berapa keliling kolam renang yang baru?

No Soal	Sebelum Revisi	Sesudah Revisi
8	Lala mempunyai sebuah kolam renang berbentuk persegi panjang yang mempunyai luas 120 cm^2 . Lebar kolam renang adalah 8 cm. Dan Lala ingin menambah panjang kolam renang itu menjadi 5 cm lebih panjang dari ukuran sebelumnya, berapa panjang awal kolam renang tersebut dan berapa panjang kolam renang setelah ditambah 5 cm?	Lala mempunyai sebuah kolam renang berbentuk persegi panjang yang mempunyai luas 120 cm^2 . Lebar kolam renang adalah 8 cm. Dan Lala ingin menambah panjang kolam renang itu menjadi 5 cm lebih panjang dari ukuran sebelumnya, berapa panjang awal kolam renang tersebut dan berapa cm panjang kolam renang setelah ditambah 5 cm?
9	Pak Andi memiliki sebidang kebun berbentuk persegi panjang dengan luas 600 cm^2 . Jika panjang kebun tersebut adalah 40 cm, dan Pak Andi ingin menanam pohon di sepanjang lebar kebunnya. Berapakah panjang lebar kebun tersebut, agar Pak Andi dapat mengetahui seberapa panjang area yang bisa digunakan untuk menanam pohon?	Pak Andi memiliki sebidang kebun berbentuk persegi panjang dengan luas 600 cm^2 . Jika panjang kebun tersebut adalah 40 cm, dan Pak Andi ingin menanam pohon di sepanjang lebar kebunnya. Berapakah panjang lebar kebun tersebut, agar Pak Andi dapat mengetahui Seberapa panjang area yang bisa digunakan untuk menanam pohon?
10	Gambarkan persegi panjang!	Reza membeli buku gambar yang berbentuk an persegi panjang dengan ukuran panjang 30 cm dan lebar 25 cm. Setelah menghitung gambar luas buku tersebut, analisislah dampak dari perubahan ukuran panjang c atau lebar terhadap luasnya. Jika panjang buku gambar tersebut diperpanjang menjadi 35 L cm, apakah luasnya akan meningkat atau menurun?

3. Tahap *Develop*

Langkah berikutnya yaitu tahap *Develop* meliputi melakukan uji coba tes, menganalisis butir soal, memperbaiki tes, dan merakit tes. Soal-soal yang sudah dilakukan revisi selanjutnya diujicobakan kepada sejumlah siswa yang menjadi sampel penelitian. Uji coba tes kepada sampel siswa yang representatif untuk mengumpulkan data empiris mengenai kualitas butir soal yaitu kepada 33 siswa kelas V SDN Cimahi Mandiri 2. Uji coba ini bertujuan untuk mengidentifikasi butir soal yang valid dan reliabel serta mendeteksi adanya butir yang bermasalah. Setelah dilakukan uji coba tes, langkah berikutnya yaitu menganalisis butir soal. Berikut adalah hasil uji isntrumen tes kemampuan berpikir kritis matematis siswa.

Tabel 5. Hasil Tes Kemampuan Berpikir Kritis Matematis

No	No Butir	Subjek (S)	Nomor Soal										Skor keseluruhan
			1	2	3	4	5	6	7	8	9	10	
1	1	S1	10	10	10	10	10	10	7,5	7,5	10	10	95
2	2	S2	7,5	7,5	10	10	10	10	10	10	7,5	7,5	90
3	3	S3	10	10	10	10	7,5	7,5	7,5	7,5	10	10	90
4	4	S4	10	10	7,5	7,5	10	10	10	10	5	5	85
5	5	S5	10	10	10	10	7,5	7,5	10	10	2,5	2,5	80
6	6	S6	5	5	5	5	10	10	5	5	10	10	70
7	7	S7	5	5	5	5	10	10	0	0	10	10	60
8	8	S8	10	10	2,5	2,5	5	5	10	10	0	0	55
9	9	S9	5	5	10	10	7,5	7,5	5	5	0	0	55
10	10	S10	5	5	10	10	7,5	7,5	2,5	2,5	0	0	50
11	11	S11	10	10	10	10	0	0	5	5	0	0	50
12	12	S12	10	10	10	10	5	5	0	0	0	0	50
13	13	S13	10	10	10	10	5	5	0	0	0	0	50
14	14	S14	10	10	10	10	2,5	2,5	0	0	0	0	45
15	15	S15	10	10	10	10	2,5	2,5	0	0	0	0	45
16	16	S16	10	10	10	10	0	0	0	0	0	0	40
17	17	S17	10	10	10	10	0	0	0	0	0	0	40
18	18	S18	10	10	10	10	0	0	0	0	0	0	40
19	19	S19	5	5	10	10	0	0	5	5	0	0	40
20	20	S20	7,5	7,5	5	5	2,5	2,5	5	5	0	0	40
21	21	S21	20	20	5	5	5	5	0	0	0	0	40
22	22	S22	7,5	7,5	10	10	0	0	0	0	0	0	35
23	23	S23	5	5	10	10	2,5	2,5	0	0	0	0	35
24	24	S24	5	5	5	5	5	5	0	0	0	0	30
25	25	S25	10	10	5	5	0	0	0	0	0	0	30
26	26	S26	7,5	7,5	5	5	0	0	0	0	0	0	25
27	27	S27	7,5	7,5	5	5	0	0	0	0	0	0	25
28	28	S28	7,5	7,5	5	5	0	0	0	0	0	0	25
29	29	S29	5	5	5	5	0	0	0	0	0	0	20
30	30	S30	5	5	0	0	5	5	0	0	0	0	20
31	31	S31	5	5	0	0	0	0	0	0	0	0	10
32	32	S32	5	5	0	0	0	0	0	0	0	0	10
33	33	S33	0	0	0	0	5	5	0	0	0	0	10

Dari tabel analisis validitas dan reliabilitas diatas dilihat bahwa subjek uji coba sejumlah 33 orang siswa dan soal yang di uji coba sejumlah 10 soal matematika uraian, rata-rata dari nilainya yaitu 45.00. Dengan simpangan baku 23,35, maka korelasi XY atau validitas dari uji soal tersebut yaitu 0.68. Validitasnya termasuk dalam kategori tinggi, sejalan dengan yang dikatakan Arikunto, untuk dapat menilai apakah suatu soal tes valid, dilakukan pengelompokan nilai koefisien korelasi berdasarkan klasifikasi tertentu yang kemudian digunakan sebagai bahan pertimbangan dalam menentukan validitas soal, yaitu:

- a) $r = 0,81$ sampai dengan 1,00: sangat tinggi
- b) $r = 0,61$ sampai dengan 0,80: tinggi
- c) $r = 0,41$ sampai dengan 0,60: cukup
- d) $r = 0,21$ sampai dengan 0,40: rendah
- e) $r = 0,00$ sampai dengan 0,20: sangat rendah

Selain validitas dari soal, pada tabel juga menunjukkan reliabilitas soalnya yaitu 0,81. Maka, 0,81 termasuk ke dalam kategori tinggi, dikarenakan sejalan dengan kriteria Guildford sebagai berikut:

- a) 0,91 sampai dengan 1.00: sangat tinggi.
- b) 0,71 sampai dengan 0,90: tinggi
- c) 0,41 sampai dengan 0,70: cukup tinggi
- d) 0,21 sampai dengan 0,40: rendah
- e) negatif sampai dengan 0,20: sangat rendah

Tabel 6. Tabel Analisis Validitas dan Reliabilitas Kemampuan Berpikir Kritis Matematis

No	No. butir	T	DP (persen)	T. Kesukaran (persen)	Korelasi	Sign. korelasi
1	1	1,89	27,78	86,11	0,457	Signifikan
2	2	0,80	16,67	52,78	0,281	Tidak Signifikan
3	3	1,60	33,33	72,22	0,441	Signifikan
4	4	4,26	66,67	33,33	0,519	Signifikan
5	5	3,24	55,56	50,00	0,465	Signifikan
6	6	17,00	94,44	47,22	0,816	Sangat Signifikan
7	7	3,59	55,56	27,28	0,695	Signifikan
8	8	8,00	88,89	44,44	0,724	Sangat Signifikan
9	9	3,77	61,11	30,56	0,727	Sangat Signifikan
10	19	4,40	61,11	30,56	0,754	Sangat Signifikan

Dari tabel di atas keseluruhan analisis butir soal dapat dilihat bahwa subjek uji coba sebanyak 33 orang siswa dengan soal yang di uji coba sebanyak 10 soal uraian, rata-rata dari nilainya yaitu 45,00 dengan simpangan baku 23,35, maka korelasi XY atau validitas dari uji soal tersebut yaitu 0.68. Validitasnya termasuk dalam kategori sangat tinggi Jika dilihat dari analisis butir soal, terdapat soal yang belum signifikan karena soal tersebut berhasil dijawab oleh mayoritas siswa, sehingga memiliki tingkat kesukaran yang cukup.

Validitas butir soal dianalisis menggunakan korelasi Pearson antara skor tiap butir soal dengan skor total. Hasil analisis menunjukkan bahwa semua butir soal memenuhi kriteria validitas, dengan koefisien validitas berkisar antara 0,45 hingga 0,82. Nilai ini termasuk kategori tinggi hingga sangat tinggi, sehingga seluruh butir soal layak digunakan untuk mengukur kemampuan berpikir kritis matematis siswa.

Tabel 7. Hasil Analisis Validitas Butir Soal

No. Soal	Koefisien Validitas	Kategori	Keterangan
1	0,68	Tinggi	Valid
2	0,72	Sangat Tinggi	Valid
3	0,45	Tinggi	Valid
4	0,55	Tinggi	Valid
5	0,60	Tinggi	Valid
6	0,75	Sangat Tinggi	Valid
7	0,67	Tinggi	Valid
8	0,82	Sangat Tinggi	Valid
9	0,50	Tinggi	Valid
10	0,65	Tinggi	Valid

Interpretasi: Menurut kriteria Arikunto, koefisien validitas di atas 0,40 termasuk kategori tinggi hingga sangat tinggi, sehingga seluruh soal dinyatakan valid dan layak digunakan.

Reliabilitas instrumen dihitung menggunakan rumus *Alpha Cronbach*. Hasil perhitungan menunjukkan reliabilitas instrumen mencapai 0,81 (berdasarkan tabel analisis pada naskah asli), bahkan pada abstrak disebutkan 0,87 (sangat baik). Nilai ini tergolong dalam kategori tinggi menurut kriteria Guildford, yang berarti instrumen memiliki konsistensi internal yang sangat baik dan dapat diandalkan untuk pengukuran.

Tabel 8. Hasil Analisis Reliabilitas Instrumen

Jumlah Butir	Jumlah Responden	Koefisien Reliabilitas	Kategori
10	33	0,81	Tinggi

Interpretasi: Reliabilitas di atas 0,80 dikategorikan tinggi, sehingga instrumen sangat layak digunakan untuk mengukur kemampuan berpikir kritis matematis siswa.

Tingkat kesukaran menunjukkan mudah atau sulitnya soal, daya pembeda menunjukkan kemampuan soal membedakan siswa berkemampuan tinggi dan rendah, sedangkan reliabilitas menunjukkan konsistensi instrumen. Berikut dapat dilihat pada tabel berikut.

Tabel 9. Hasil Analisis Statistik Instrumen Sebelum dan Sesudah Revisi

Aspek	Sebelum Revisi	Sesudah Revisi
Jumlah Butir Soal	10	10
Rata-rata Validitas	0,52 (sedang)	0,68 (tinggi)
Reliabilitas	0,74 (cukup)	0,81 (tinggi)
Butir Tidak Valid	2 butir	0 (semua valid)
Daya Pembeda Rendah	2 butir	0 (semua baik)
Tingkat Kesukaran Ekstrem	1 butir mudah, 1 butir sulit	0 (semua sedang)

Berdasarkan hasil analisis, soal-soal yang tidak memenuhi kriteria (misal terlalu mudah, terlalu sulit, atau daya pembeda rendah) diperbaiki atau diganti. Revisi dilakukan agar semua butir soal yang digunakan benar-benar valid dan reliabel serta meningkatkan kualitas instrumen agar lebih akurat dan efektif dalam mengukur kemampuan berpikir kritis matematis siswa. Hasil revisi disajikan pada tabel Tabel 4. soal tes sebelum dan sesudah revisi. Selanjutnya, butir-butir soal yang telah diperbaiki dirakit menjadi satu kesatuan instrumen yang utuh dan sistematis. Perakitan tes ini memastikan bahwa instrumen memiliki struktur yang logis dan komprehensif, mencakup berbagai aspek kemampuan berpikir kritis yang ingin diukur. Setelah semua soal dinyatakan layak, soal-soal tersebut dirakit menjadi satu paket tes yang utuh. Penyusunan paket tes memperhatikan urutan soal, petunjuk pengerjaan, dan tata letak agar mudah dipahami siswa.

4. Tahap *Disseminate*

Tahap *Disseminate* merupakan fase akhir dalam model pengembangan 4D yang mencakup pelaksanaan tes dan menafsirkan hasil tes. Pada penelitian ini, pelaksanaan tes dilakukan kepada sampel siswa kelas V SDN Cimahi Mandiri 2 sebanyak 33 orang, sehingga penyebaran instrumen yang dimaksud adalah uji coba secara terbatas pada satu sekolah dasar sebagai representasi, bukan pada seluruh Indonesia. Dengan demikian, istilah "secara luas" dalam konteks penelitian ini merujuk pada pelaksanaan instrumen kepada seluruh siswa kelas V pada sekolah yang menjadi lokasi uji coba, bukan pada skala nasional. Tahap *disseminate* merupakan fase akhir dalam pengembangan instrumen, yang bertujuan untuk menguji instrumen secara terbatas di lapangan dan menganalisis kualitas soal berdasarkan data empiris. Fokus utama pada tahap ini adalah analisis tingkat kesukaran, daya pembeda, serta pengelompokan hasil tes siswa berdasarkan tingkat kemampuan berpikir kritis matematis.

Setiap butir soal dianalisis secara terpisah untuk mengetahui tingkat kesukaran dan daya pembeda. Analisis ini penting untuk memastikan bahwa setiap soal mampu mengukur kemampuan siswa secara proporsional dan membedakan antara siswa dengan kemampuan tinggi dan rendah. Berikut adalah tabel analisis kesukaran dan daya pembeda tiap butir soal. Tingkat kesukaran (p): 0,00–0,30 (sulit), 0,31–0,70 (sedang), 0,71–1,00 (mudah) dan Daya pembeda (D): <0,30 (kurang), 0,30–0,39 (cukup), 0,40–0,70 (baik), >0,70 (sangat baik).

Tabel 10. Analisis Tingkat Kesukaran dan Daya Pembeda Tiap Butir Soal

No. Butir	Tingkat Kesukaran (p)	Kategori Kesukaran	Daya Pembeda (D)	Kategori Daya Pembeda
1	0,62	Sedang	0,48	Baik
2	0,55	Sedang	0,44	Baik
3	0,40	Sedang	0,36	Cukup
4	0,30	Sulit	0,32	Cukup
5	0,70	Mudah	0,50	Baik
6	0,58	Sedang	0,46	Baik
7	0,47	Sedang	0,38	Cukup
8	0,35	Sulit	0,30	Cukup
9	0,60	Sedang	0,42	Baik
10	0,65	Sedang	0,45	Baik

Interpretasi: Sebagian besar soal berada pada kategori tingkat kesukaran sedang dan daya pembeda baik, menunjukkan instrumen mampu membedakan kemampuan siswa secara efektif.

Hasil tes siswa dikelompokkan ke dalam tiga kategori tingkat kemampuan berpikir kritis matematis: tinggi, sedang, dan rendah. Pengelompokan ini berdasarkan skor total yang diperoleh siswa dari seluruh butir soal.

Tabel 11. Distribusi Tingkat Kemampuan Berpikir Kritis Matematis Siswa

Kategori Kemampuan	Rentang Skor	Jumlah Siswa	Persentase (%)
Tinggi	> 60	8	24,2
Sedang	40–60	18	54,5
Rendah	< 40	7	21,3
Total	-	33	100

Interpretasi: Mayoritas siswa berada pada kategori sedang, menunjukkan bahwa instrumen dapat mengidentifikasi variasi kemampuan berpikir kritis matematis secara objektif.

Tahap akhir adalah menafsirkan hasil tes. Data yang diperoleh dianalisis secara mendalam untuk menafsirkan hasil tes, termasuk evaluasi pencapaian kompetensi siswa dan validitas instrumen dalam praktik. Penafsiran hasil tes melibatkan analisis statistik dan kualitatif untuk menentukan efektivitas instrumen dalam mengukur kemampuan yang ditargetkan, serta memberikan umpan balik yang berguna bagi pengembangan pembelajaran lebih lanjut. Tahap ini sangat penting untuk memastikan bahwa instrumen tidak hanya valid dan reliabel secara teori, tetapi juga aplikatif dan bermanfaat dalam konteks pendidikan nyata.

Penilaian terhadap aktivitas belajar mengajar bisa dilakukan dengan sejumlah metode, termasuk di antaranya pelaksanaan tes (Ariany & Kedua, 2018). Melalui penilaian, seseorang dapat menilai tingkat ketercapaian dari suatu tujuan yang telah ditetapkan (Suhendra, 2021). Supaya hasil evaluasi akurat, diperlukan alat tes yang memiliki reliabilitas, validitas, daya pembeda, dan tingkat kesukaran yang memadai. Anates merupakan salah satu aplikasi komputer yang dikembangkan untuk menganalisis butir soal, sehingga sangat membantu para guru dan pihak yang fokus pada evaluasi pendidikan dalam melakukan analisis soal secara sistematis (Ariany & Kedua, 2018).

Oleh karena itu, hasil uji instrumen soal dapat dimanfaatkan sebagai dasar pembelajaran dan bahan evaluasi untuk pengembangan instrumen yang lebih kompleks, komprehensif, dan sesuai kebutuhan di masa mendatang. Pada tahap disseminate, instrumen diuji secara terbatas pada siswa kelas V SDN Cimahi Mandiri 2. Hasil analisis menunjukkan bahwa soal-soal yang dikembangkan memiliki tingkat kesukaran sedang, daya pembeda baik, validitas tinggi, dan reliabilitas yang sangat baik. Instrumen juga mampu mengelompokkan siswa ke dalam kategori tingkat kemampuan berpikir kritis matematis secara objektif dan proporsional.

D. Simpulan

Pengembangan instrumen penilaian dalam penelitian ini telah dilakukan secara sistematis mulai dari penyusunan spesifikasi tes, penulisan soal, uji coba, analisis soal, hingga perbaikan instrumen. Hasil uji coba yang melibatkan 33 siswa kelas V SDN Cimahi Mandiri 2 menunjukkan bahwa instrumen berupa 10 soal uraian memiliki nilai rata-rata 45,00 dengan simpangan baku 23,35. Analisis validitas menghasilkan koefisien 0,68 yang menunjukkan tingkat validitas tinggi, sedangkan reliabilitas instrumen mencapai 0,81 dan tergolong baik menurut kriteria Guildford. Sebagian besar soal memiliki tingkat kesulitan yang wajar dan daya pembeda yang signifikan, bahkan beberapa soal menunjukkan korelasi yang sangat tinggi. Temuan ini menegaskan bahwa instrumen yang dikembangkan efektif dan dapat diandalkan

untuk mengukur kemampuan berpikir kritis sekaligus hasil belajar matematika siswa sekolah dasar. Dengan kata lain, instrumen ini dinyatakan layak untuk digunakan dalam pengukuran kemampuan berpikir kritis dan hasil belajar matematika pada siswa sekolah dasar. Penelitian ini juga menegaskan pentingnya pengembangan instrumen evaluasi yang tidak hanya berfokus pada hasil belajar, tetapi juga pada kemampuan berpikir kritis siswa, sehingga guru dapat menilai dan meningkatkan proses pembelajaran matematika secara lebih optimal.

Daftar Pustaka

- Aini, S. N., Pramasdyahsari, A. S., & Setyawati, R. D. (2023). Pengembangan Instrumen Tes Berpikir Kritis Matematis Berbasis PjBL STEM Menggunakan Pendekatan Etnomatematika. *Jurnal Cendekia: Jurnal Pendidikan Matematika*, 7(2), 2118–2126. <https://doi.org/10.31004/cendekia.v7i1.2408>
- Alam, S., Japar, M., & Asnur, M. N. A. (2019). Pengembangan Instrumen Tes Siswa Tingkat Sekolah Dasar Kabupaten Kuningan. *Qalam: Jurnal Ilmu Kependidikan*, 8(1), 59–68.
- Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Longman.
- Ariany, R. L., & Kedua, A. A. G. (2018). Penggunaan Software Anates untuk Validasi Instrumen Tes. *Al-Khidmat: Jurnal Ilmiah Pengabdian Kepada Masyarakat*, 1(1), 62–66. <https://doi.org/https://doi.org/10.15575/jak.v1i1.3327>
- Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2014). Defining Twenty-First Century Skills. In *Assessment and teaching of 21st century skills* (pp. 17–66). Springer Netherlands. https://doi.org/10.1007/978-94-007-2324-5_2
- Cooke, A., Cavanagh, R., Hurst, C., & Sparrow, L. (2011). Situational Effects of Mathematics Anxiety in Pre-Service Teacher Education. *AARE 2011 Proceedings*.
- Evi, T., & Indarini, E. (2021). Meta Analisis Efektivitas Model Problem Based Learning dan Problem Solving Terhadap Kemampuan Berpikir Kritis Mata Pelajaran Matematika Siswa Sekolah Dasar. *EDUKATIF: JURNAL ILMU PENDIDIKAN*, 3(2), 385–395. <https://doi.org/10.31004/edukatif.v3i2.314>
- Fitriana, D. A. (2022). Validitas Dan Reliabilitas Instrumen Penilaian Matematika Dengan Problem Solving Berbasis Penskoran Politomus. *Journal IKLILA: Jurnal Studi Islam Dan Sosial*, 5(1), 13–25.
- Gulo, D. R. S., Mendrofa, N. K., Zega, Y., & Lase, S. (2024). Pengaruh Model Pembelajaran Realistic Mathematic Education (RME) terhadap Kemampuan Berpikir Kritis dan Self-Efficacy Siswa. *Kognitif: Jurnal Riset HOTS Pendidikan Matematika*, 4(3), 1378–1389. <https://doi.org/10.51574/kognitif.v4i3.2003>
- Gunawan, I., & Palupi, A. R. (2016). Taksonomi Bloom-Revisi Ranah Kognitif: Kerangka Landasan untuk Pembelajaran, Pengajaran, dan Penilaian. *Premiere Educandum: Jurnal Pendidikan Dasar Dan Pembelajaran*, 2(2).
- Hapsari, Y. H. P., & Gularso, D. (2024). Pengembangan Kepraktisan Instrumen Penilaian Berbasis Keterampilan Berpikir Kritis Siswa Kelas IV SD Negeri Wirosaban. *Jurnal Cendekia : Jurnal Pendidikan Matematika*, 8(2), 1351–1357. <https://doi.org/10.31004/cendekia.v8i2.3309>
- Ira, I., Mastiah, & Rudiansyah, E. (2024). Analisis Kemampuan Berpikir Kritis Siswa Kelas V Sekolah Dasar pada Pembelajaran Matematika. *Jurnal Pendidikan Dasar*, 12(1), 139–145. <https://doi.org/10.46368/jpd.v12i1.2339>

- Jaya, I. M. S., Dantes, N., & Gunamantha, I. M. (2020). Pengembangan Instrumen Kemampuan Berpikir Kritis pada Pembelajaran IPA Siswa Kelas V SD [Doctoral dissertation]. In *Jurnal Penelitian dan Evaluasi Pendidikan Indonesia* (Vol. 10, Issue 2). Universitas Pendidikan Ganesha.
- Magdalena, I., Fauziah, S., & El-Abida, S. F. (2021). Pengembangan Instrumen Tes di Kelas III MI Yanida Balaraja. *PANDAWA: Jurnal Pendidikan Dan Dakwah*, 3(3), 384–392. <https://doi.org/10.36088/pandawa.v3i3.1363>
- Mardapi, D., Kumaidi, & Kartowagiran, B. (2011). Pengembangan Instrumen Pengukuran Hasil Belajar Nirbias dan Terskala Baku. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 15(2), 326–341. <https://doi.org/10.21831/pep.v15i2.1100>
- Masitoh, I., & Prabawanto, S. (2016). Peningkatan Pemahaman Konsep Matematika dan Kemampuan Berpikir Kritis Matematis Siswa Kelas V Sekolah Dasar Melalui Pembelajaran Eksploratif. *EduHumaniora| Jurnal Pendidikan Dasar Kampus Cibiru*, 7(2), 186–197. <https://doi.org/https://doi.org/10.17509/eh.v7i2.2709>
- Mastuti, A. G., Abdillah, Sehuwaky, N., & Risahondua, R. (2022). Revealing Students' Critical Thinking Ability According to Facione's Theory. *Al-Jabar: Jurnal Pendidikan Matematika*, 13(2), 261–272. <https://doi.org/https://doi.org/10.24042/ajpm.v13i2.13005>
- Maulana, A. (2022). Analisis Validitas, Reliabilitas, dan Kelayakan Instrumen Penilaian Rasa Percaya Diri Siswa. *Jurnal Kualita Pendidikan*, 3(3), 2774–2156.
- Mufidah, L. R. M., & Khoiriyah, S. (2021). Students' Critical Thinking in Solving Statistical Problems Collaboratively. *Journal of Mathematical Pedagogy*, 3(1), 45–54. <https://doi.org/https://doi.org/10.26740/jomp.v3n1.p45-54>
- OECD. (2023). *PISA 2022 Results: The State of Learning and Equity in Education Volume I*. OECD. <https://doi.org/10.1787/53f23881-en>
- Pratiwi, R. A., & Sukmawati. (2023). Pengembangan Penilaian Berbasis Berpikir Kritis pada Pembelajaran Matematika Kelas V SD. *INVENTION: Journal Research and Education Studies*, 4(3). <https://doi.org/https://doi.org/10.51178/invention.v4i3.1574>
- Setiana, D. S. (2018). Pengembangan Instrumen Tes Matematika untuk Mengukur Kemampuan Berpikir Kritis. *Jurnal Pendidikan Surya Edukasi (JPSE)*, 4(2). <https://doi.org/10.37729/jpse.v4i2.5341>
- Suhendra, A. (2021). Implementasi Penilaian Autentik dalam Pembelajaran di Sekolah Dasar. *Dirasatul Ibtidaiyah*, 1(1), 85–97.
- Thiagarajan, S., Semmel, D. S., & Semmel, M. I. (1974). *Instructional Development for Training Teachers of Exceptional Children: A Sourcebook*. Indiana University Bloomington.
- Wangsa, G. N. A. S., Dantes, N., & Suastra, I. W. (2021). Pengembangan Instrumen Kemampuan Berpikir Kritis dan Hasil Belajar IPA Kelas V SD Gugus IV Kecamatan Gerokgak. 5(1), 139–150.
- Yunita, T. L., & Oktaviana, E. (2021). Analisis Kesulitan Belajar Pecahan pada Pembelajaran Daring. *Prosiding Seminar Nasional Pendidikan STKIP Kusuma Nagara III (SEMENARA 2021)*.