



Kualitas Pengukuran Tes Berpikir Komputasional dan Berpikir Kritis pada Pecahan

Nabilah Fauziyah¹, Tri Astari^{2*}

^{1,2} Universitas Jember, Jember, Indonesia

ARTICLE HISTORY

Received: 28 Juli 2026
Revised: 30 Agustus 2026
Accepted: 6 September 2026

Keywords:

Computational Thinking,
Critical Thinking, Item
Analysis, Fractions, Test
Validity

Kata Kunci:

Berpikir Komputasional,
Berpikir Kritis, Analisis
Butir, Pecahan, Validitas
Tes

DOI:

<https://doi.org/10.31949/ijic.v4i2.19825>

ABSTRACT

Measuring computational and critical thinking skills in elementary mathematics requires tests supported by both theoretical and empirical evidence. This study aimed to evaluate the theoretical feasibility, empirical validity, reliability, discrimination, and difficulty of two constructed-response tests on Grade V fractions. Each test consisted of six items developed from the respective construct indicators. Theoretical validation involved five expert and practitioner validators, while empirical testing involved 28 Grade V students at SDN Kepatihan 2 Jember. The analyses included validation percentages, item-total correlations, split-half reliability using the Spearman-Brown formula, discrimination indices, and item difficulty. The computational thinking test obtained validation scores of 80% for construct, 90% for content, 94% for language, and 77% for practitioner aspects. All items met the empirical validity criterion, with correlations ranging from 0.51 to 0.73, but its reliability coefficient was 0.78, below the 0.80 criterion. The critical thinking test obtained 80%, 90%, 98%, and 74%, respectively; five items met the empirical validity criterion, while Item 6 had a correlation of 0.24 and was classified as very difficult. Its reliability coefficient reached 0.89. The findings provide initial evidence of measurement quality, although further improvement is needed in specific aspects.

ABSTRAK

Pengukuran keterampilan berpikir komputasional dan berpikir kritis dalam matematika sekolah dasar memerlukan tes yang didukung bukti teoretis dan empiris. Penelitian ini bertujuan mengevaluasi kelayakan teoretis, validitas empiris, reliabilitas, daya pembeda, dan tingkat kesulitan dua tes uraian materi pecahan kelas V. Masing-masing tes terdiri atas enam butir berdasarkan indikator konstruk. Validasi teoretis melibatkan lima validator ahli dan praktisi, sedangkan uji empiris dilakukan terhadap 28 siswa kelas V SDN Kepatihan 2 Jember. Analisis mencakup persentase validasi, korelasi butir-total, reliabilitas *split-half* Spearman-Brown, daya pembeda, dan tingkat kesulitan. Tes berpikir komputasional memperoleh validasi 80% pada konstruk, 90% isi, 94% bahasa, dan 77% praktisi. Seluruh butir memenuhi kriteria validitas empiris dengan korelasi 0,51-0,73, tetapi reliabilitasnya 0,78, masih di bawah batas 0,80. Tes berpikir kritis memperoleh 80%, 90%, 98%, dan 74%; lima butir memenuhi kriteria validitas empiris, sedangkan Butir 6 memiliki korelasi 0,24 dan berkategori sangat sulit. Reliabilitas tes ini mencapai 0,89. Temuan menunjukkan bahwa kedua tes memberikan bukti awal kualitas pengukuran, tetapi masih memerlukan penguatan pada aspek tertentu.

1. PENDAHULUAN

Pembelajaran matematika sekolah dasar tidak cukup diarahkan pada ketepatan jawaban, tetapi juga pada kemampuan siswa menafsirkan informasi, memilih strategi, menjelaskan alasan, dan menilai kembali proses penyelesaian. Tantangan ini sangat terlihat pada materi pecahan karena siswa harus memahami relasi bagian-keseluruhan, kesetaraan, urutan, serta operasi pecahan dalam beragam representasi. Studi pada siswa sekolah dasar menunjukkan

bahwa kesulitan pecahan berkaitan dengan lemahnya pemahaman masalah, pemilihan strategi, abstraksi, dan penalaran terhadap hubungan antarpecahan (Agustin et al., 2026; Aulia & Sutarni, 2024; Fitriani & Kowiyah, 2022; Yeo & Webel, 2024). Dengan demikian, asesmen pecahan perlu menangkap proses berpikir yang mendasari jawaban siswa, bukan hanya hasil akhir.

Dua konstruk yang relevan untuk membaca proses tersebut adalah keterampilan berpikir komputasional dan berpikir kritis. Dalam pendidikan matematika dasar, berpikir komputasional diwujudkan melalui dekomposisi masalah, pengenalan pola, abstraksi, dan penyusunan algoritma, sedangkan berpikir kritis berkaitan dengan interpretasi, analisis, evaluasi, inferensi, penjelasan, serta pemeriksaan kembali keputusan. Kajian menunjukkan bahwa integrasi berpikir komputasional dalam matematika dapat memperkuat cara siswa mengorganisasi pemecahan masalah, sementara kemampuan berpikir kritis diperlukan untuk memeriksa informasi dan memberikan alasan matematis yang dapat dipertanggungjawabkan (Alfaris et al., 2025; Candraningtyas & Khusna, 2023; Nordby et al., 2022, 2024; Wahyuningtyas & Widiyono, 2024). Kedua konstruk tersebut beririsan dalam kebutuhan akan proses berpikir terstruktur, tetapi tetap perlu diukur melalui indikator yang jelas agar interpretasi skor tidak tercampur.

Materi pecahan merupakan konteks yang strategis untuk mengukur kedua keterampilan karena penyelesaiannya menuntut siswa memilah informasi, mengenali kesetaraan atau pola, membandingkan representasi, menyusun prosedur, dan mengevaluasi ketepatan hasil. Penelitian pada pembelajaran pecahan menunjukkan bahwa tugas yang mengorganisasi langkah pemecahan masalah dapat memfasilitasi berpikir komputasional, sedangkan tugas yang meminta alasan, perbandingan, dan evaluasi dapat memunculkan kualitas penalaran kritis siswa (Agustin et al., 2026; Tangkui, 2023; Wahyuningtyas & Widiyono, 2024; Yeo & Webel, 2024). Konsekuensinya, kualitas kesimpulan tentang kemampuan siswa sangat bergantung pada kualitas instrumen yang digunakan.

State of the art pengukuran berpikir komputasional pada jenjang dasar menunjukkan pergeseran dari penyusunan soal semata menuju validasi psikometrik yang lebih sistematis. Sejumlah penelitian telah mengembangkan dan menguji instrumen CT menggunakan validasi ahli, teori tes klasik, analisis Rasch, *Item Response Theory*, analisis invariansi, dan pengujian fungsi butir pada kelompok usia yang berbeda (El-Hamamsy et al., 2022, 2025; Inasari et al., 2023; Kong & Lai, 2022; Kong & Wang, 2023; Na et al., 2024; Tsai et al., 2022; Zapata-Cáceres et al., 2024; Zhang & Wong, 2023). Perkembangan tersebut menegaskan bahwa instrumen yang baik perlu menunjukkan kesesuaian konstruk sekaligus fungsi empiris butir pada populasi sasaran.

Kecenderungan serupa terlihat pada pengembangan instrumen berpikir kritis dan asesmen pendidikan. Validasi isi tetap penting untuk memastikan keterwakilan indikator, tetapi bukti tersebut perlu dilengkapi dengan reliabilitas, daya pembeda, tingkat kesulitan, serta analisis fungsi butir agar kelemahan instrumen dapat dideteksi secara lebih tepat (An Nabil et al., 2022; Coppi et al., 2023; Darman et al., 2024; Ernawati et al., 2024; Lukman et al., 2023; Nurhasanah et al., 2024, 2025; Schames Kreitchmann et al., 2024; Sheptian et al., 2024). Studi psikometrik pada siswa sekolah dasar juga menunjukkan bahwa satu indikator statistik tidak selalu cukup untuk menilai kualitas instrumen; informasi tentang validitas, reliabilitas, kesulitan, dan daya pembeda perlu dibaca secara komplementer (Ernawati et al., 2024; Nurhasanah et al., 2025).

Meskipun berbagai instrumen CT dan berpikir kritis telah dikembangkan, kajian yang secara bersamaan mengevaluasi dua tes uraian berbasis pecahan pada siswa sekolah dasar melalui validasi ahli-praktisi, korelasi butir-total, reliabilitas, daya pembeda, dan tingkat kesulitan masih terbatas. Artikel ini menawarkan evaluasi kualitas pengukuran berbasis teori

tes klasik terhadap dua instrumen yang sebelumnya dikembangkan untuk penelitian korelasional, tetapi analisis artikel difokuskan pada kualitas instrumennya sebelum skor digunakan untuk menafsirkan kemampuan siswa. Penelitian bertujuan menganalisis kelayakan teoretis tes berpikir komputasional dan berpikir kritis, mengevaluasi validitas empiris dan reliabilitas masing-masing tes, serta mengidentifikasi karakteristik butir berdasarkan daya pembeda dan tingkat kesulitan. Kontribusi penelitian terletak pada penyajian bukti teoretis dan empiris secara terintegrasi untuk menentukan butir yang telah berfungsi memadai dan butir yang memerlukan revisi, sehingga dapat mendukung pengembangan asesmen keterampilan berpikir tingkat tinggi pada materi pecahan di sekolah dasar.

2. METODE

Penelitian ini menggunakan desain evaluasi instrumen dengan pendekatan kuantitatif berbasis teori tes klasik. Data merupakan data sekunder dari tahap pengembangan dan uji coba dua tes uraian pada penelitian keterampilan berpikir komputasional dan berpikir kritis materi pecahan kelas V. Artikel tidak menganalisis hubungan antara kedua keterampilan pada sampel penelitian utama, tetapi memusatkan analisis pada bukti kelayakan teoretis, validitas empiris, reliabilitas, daya pembeda, dan tingkat kesulitan. Pendekatan multi-bukti ini selaras dengan praktik validasi instrumen pendidikan yang menempatkan penilaian ahli dan bukti fungsi empiris butir sebagai informasi yang saling melengkapi (Ernawati et al., 2024; Nurhasanah et al., 2025; Schames Kreitchmann et al., 2024).

Konteks uji coba adalah SDN Kepatihan 2 Jember pada tahun pelajaran 2025/2026. Populasi terjangkau pada tahap uji coba ialah siswa kelas V di sekolah tersebut. Dokumen perencanaan mencantumkan 31 siswa sebagai sasaran uji coba, sedangkan data hasil yang tersedia dan dianalisis memuat 28 respons lengkap; karena itu, penelitian ini menggunakan 28 siswa sebagai sampel analitik. Teknik sampling uji coba tidak terdokumentasi secara eksplisit pada data sumber, sehingga artikel tidak mengatribusikan teknik sampling tertentu. Sumber data lain berupa hasil penilaian lima validator: satu dosen ahli matematika, dua guru kelas sebagai praktisi, satu dosen bidang asesmen, dan satu dosen bidang bahasa.

Data dikumpulkan melalui lembar validasi dan respons tertulis siswa terhadap dua tes uraian yang diskor menggunakan rubrik bertingkat. Tes berpikir komputasional terdiri atas enam butir yang merepresentasikan *decomposition*, *pattern recognition*, *abstraction*, dan *algorithm design*. Tes berpikir kritis terdiri atas enam butir yang merepresentasikan *interpretation*, *analysis*, *evaluation*, *inference*, *explanation*, dan *self-regulation*. Kedua tes menggunakan materi pecahan kelas V yang mencakup perbandingan, pengurutan, penjumlahan, dan pengurangan pecahan biasa maupun campuran. Kisi-kisi ringkas kedua tes disajikan pada Tabel 1.

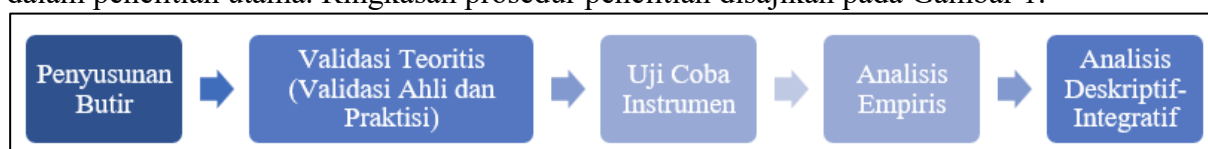
Tabel 1. Kisi-Kisi Tes Keterampilan Berpikir Komputasional Dan Berpikir Kritis

Tes/Butir	Aspek	Indikator tes dan ringkasan tugas	Level Kognitif
CT-1	<i>Decomposition</i>	Menganalisis informasi, memecah masalah menjadi langkah sistematis, dan menentukan sisa waktu belajar dalam konteks pecahan.	C4
CT-2	<i>Pattern recognition</i>	Menganalisis pola perubahan pecahan, memprediksi nilai berikutnya, dan mengurutkan pecahan secara logis.	C4
CT-3	<i>Abstraction</i>	Memilih informasi penting, menggunakan pecahan senilai, dan menentukan nilai paling kecil pada konteks resep.	C3
CT-4	<i>Algorithm design</i>	Menyusun langkah logis untuk mengubah, membandingkan, dan mengurutkan pecahan campuran pada konteks tinggi badan.	C4
CT-5	<i>Algorithm design</i>	Menyusun langkah sistematis penjumlahan pecahan campuran berpenyebut berbeda untuk menentukan total kebutuhan semen.	C4

Tes/Butir	Aspek	Indikator tes dan ringkasan tugas	Level Kognitif
CT-6	<i>Decomposition</i>	Mengidentifikasi informasi penting dan menguraikan langkah pengurangan pecahan campuran untuk menentukan sisa persediaan tepung.	C4
BK-1	<i>Interpretation</i>	Menentukan operasi pecahan sesuai konteks dan menjelaskan makna operasi untuk menentukan sisa cat.	C3
BK-2	<i>Analysis</i>	Mengubah, membandingkan, dan mengurutkan nilai pecahan secara logis pada konteks berat tas.	C4
BK-3	<i>Evaluation</i>	Menilai kebenaran pernyataan melalui perbandingan pecahan dan memberikan alasan pada konteks uang saku.	C5
BK-4	<i>Inference</i>	Menarik kesimpulan logis dari operasi penjumlahan dan pengurangan pecahan pada konteks sisa air.	C5
BK-5	<i>Explanation</i>	Menjelaskan kesalahan penalaran dan memberikan perbandingan pecahan yang benar pada konteks durasi film.	C5
BK-6	<i>Self-regulation</i>	Memeriksa, mengoreksi, dan memperbaiki kesalahan urutan operasi pecahan secara mandiri serta menjelaskan kesalahannya.	C5

Catatan: CT = berpikir komputasional; BK = berpikir kritis.

Berdasarkan kisi-kisi tersebut, evaluasi instrumen dilakukan melalui tahapan yang sistematis, mulai dari penyusunan butir hingga analisis hasil uji coba. Tahapan tersebut bertujuan memperoleh bukti kelayakan teoretis dan empiris sebelum instrumen digunakan dalam penelitian utama. Ringkasan prosedur penelitian disajikan pada Gambar 1.



Gambar 1. Prosedur Evaluasi Kualitas Instrumen

Gambar 1 menunjukkan bahwa prosedur penelitian terdiri atas lima tahap yang dilaksanakan secara berurutan. Tahap pertama adalah penyusunan butir berdasarkan indikator konstruk, tujuan pembelajaran, dan kisi-kisi tes. Tahap kedua adalah validasi teoretis menggunakan lembar penilaian skala 1–5. Validitas isi melibatkan satu dosen ahli matematika dan dua guru kelas, sedangkan validasi ahli melibatkan dosen bidang asesmen dan bahasa. Penilaian mencakup aspek konstruk, isi, bahasa, dan praktisi; skor dikonversi menjadi persentase terhadap skor maksimum. Kategori yang digunakan adalah sangat layak untuk nilai di atas 80% dan layak untuk nilai lebih dari 60% sampai 80%. Validasi ahli digunakan sebagai dasar revisi awal sebelum uji coba, sebagaimana ditekankan dalam pengembangan asesmen terkini bahwa keterwakilan isi perlu ditelaah sebelum fungsi empiris butir diperiksa (An Nabil et al., 2022; Sheptian et al., 2024).

Ketiga, versi hasil revisi diuji kepada siswa dan seluruh respons diskor menggunakan rubrik yang sama. Keempat, validitas empiris setiap butir dianalisis melalui korelasi Product Moment antara skor butir dan skor total; dengan $n=28$ pada taraf 5%, r tabel yang digunakan adalah 0,37. Reliabilitas masing-masing tes diestimasi menggunakan metode split-half dan koreksi *Spearman-Brown*. Daya pembeda dianalisis menggunakan korelasi antara skor butir dan skor total setelah skor butir terkait dikurangkan (*corrected item-total*), sedangkan tingkat kesulitan ditentukan dari proporsi rerata skor butir terhadap skor maksimum. Penggunaan beberapa indikator ini dimaksudkan agar keputusan terhadap butir tidak bergantung pada satu statistik tunggal (Ernawati et al., 2024; Nurhasanah et al., 2025; Riyadi et al., 2024).

Kelima, hasil dianalisis secara deskriptif-integratif. Persentase validasi diringkas per aspek; korelasi butir-total digunakan untuk menentukan pemenuhan kriteria validitas empiris;

koefisien Spearman-Brown digunakan untuk membaca konsistensi tes; sedangkan daya pembeda dan tingkat kesulitan dibaca bersama untuk mengidentifikasi butir yang berfungsi memadai atau memerlukan revisi. Pendekatan ini tetap berada dalam kerangka teori tes klasik; penelitian tidak mengklaim bukti struktur internal berbasis Rasch atau IRT, tetapi menggunakan studi-studi mutakhir tersebut sebagai arah pengujian lanjutan (Darman et al., 2024; Kong & Lai, 2022; Schames Kreitchmann et al., 2024). Seluruh data siswa dianonimkan, digunakan untuk kepentingan akademik, dan dianalisis berdasarkan izin penelitian dari sekolah dengan menjaga kerahasiaan identitas partisipan.

3. HASIL

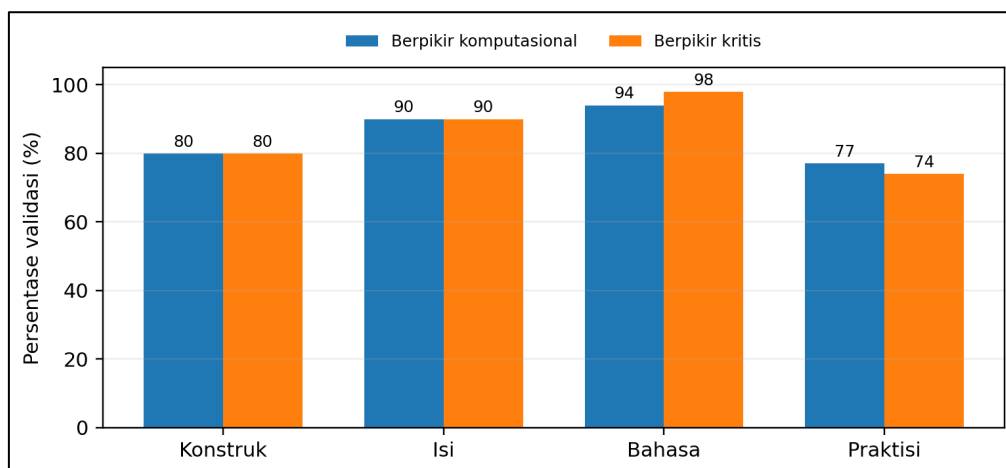
Kelayakan Teoretis Kedua Tes

Tujuan pertama penelitian adalah menganalisis kelayakan teoretis kedua tes. Validasi menunjukkan pola yang serupa. Tes berpikir komputasional memperoleh 80% pada aspek konstruk, 90% pada isi, 94% pada bahasa, dan 77% pada aspek praktisi. Tes berpikir kritis memperoleh 80%, 90%, 98%, dan 74% pada aspek yang sama. Berdasarkan batas kategori yang digunakan, konstruk sebesar 80% dan aspek praktisi berada pada kategori layak, sedangkan aspek isi dan bahasa berada pada kategori sangat layak. Nilai secara rinci disajikan pada Tabel 2 dan perbandingan profilnya divisualisasikan pada Gambar 1.

Tabel 2. Hasil Validasi Teoretis Kedua Instrumen

Aspek	Tes Berpikir Komputasional (%)	Kategori	Tes Berpikir Kritis (%)	Kategori
Konstruk	80	Layak	80	Layak
Isi	90	Sangat layak	90	Sangat layak
Bahasa	94	Sangat layak	98	Sangat layak
Praktisi	77	Layak	74	Layak

Untuk memperjelas perbandingan pola kelayakan kedua instrumen pada setiap aspek, hasil validasi tersebut divisualisasikan pada Gambar 1. Visualisasi ini menunjukkan posisi relatif aspek konstruk, isi, bahasa, dan praktisi antara tes berpikir komputasional dan tes berpikir kritis.



Gambar 1. Profil Validasi Teoretis Kedua Instrumen

Gambar 1 memperlihatkan bahwa kedua instrumen memiliki pola kelayakan yang hampir sama. Aspek isi dan bahasa memperoleh persentase tertinggi, sedangkan aspek praktisi menjadi komponen dengan nilai relatif lebih rendah pada kedua tes. Tes berpikir kritis memiliki persentase bahasa sedikit lebih tinggi daripada tes berpikir komputasional, yaitu 98%

dibandingkan 94%, sementara aspek praktisi masing-masing mencapai 74% dan 77%. Pola ini menunjukkan bahwa secara teoretis kedua instrumen telah memiliki dukungan yang baik, tetapi masih terdapat ruang penyempurnaan terutama dari perspektif keterpahaman dan keterlaksanaan instrumen oleh praktisi. Temuan ini selanjutnya menjadi dasar untuk menelaah apakah kelayakan teoretis tersebut juga didukung oleh fungsi empiris butir dan reliabilitas tes.

Validitas Empiris dan Reliabilitas

Tujuan kedua penelitian adalah mengevaluasi validitas empiris butir dan reliabilitas masing-masing tes. Hasil analisis menunjukkan bahwa seluruh Butir 1–6 pada tes berpikir komputasional memiliki korelasi butir-total yang melampaui $r_{tabel} = 0,37$, dengan koefisien berkisar antara 0,5–0,73. Dengan demikian, seluruh butir tes berpikir komputasional memenuhi kriteria validitas empiris pada tahap uji coba.

Pada tes berpikir kritis, Butir 1-5 memiliki korelasi butir-total sebesar 0,70-0,91 dan memenuhi kriteria validitas empiris. Sebaliknya, Butir 6 memperoleh koefisien sebesar 0,24, lebih rendah daripada $r_{tabel} = 0,37$, sehingga belum memenuhi kriteria validitas empiris. Dengan demikian, lima dari enam butir tes berpikir kritis memenuhi kriteria pada tahap uji coba. Berdasarkan hasil tersebut, Butir 6 selanjutnya menjadi bagian yang memerlukan peninjauan dan revisi sebelum instrumen digunakan pada penelitian utama.

Reliabilitas dianalisis pada tingkat instrumen untuk mengetahui konsistensi skor yang dihasilkan oleh masing-masing tes. Analisis dilakukan menggunakan metode split-half dengan koreksi Spearman-Brown. Berdasarkan kriteria yang digunakan dalam penelitian sumber, koefisien reliabilitas di bawah 0,80 dikategorikan belum memenuhi kriteria reliabel, sedangkan koefisien 0,85-0,89 berada pada kategori reliabilitas tinggi. Hasil analisis menunjukkan koefisien reliabilitas tes berpikir komputasional sebesar 0,78 dan tes berpikir kritis sebesar 0,89. Ringkasan hasil tersebut disajikan pada Tabel 3.

Tabel 3. Reliabilitas *Split-Half* Kedua Instrumen

Instrumen	Koefisien	Interpretasi
Berpikir komputasional	0,78	Belum memenuhi kriteria reliabel
Berpikir kritis	0,89	Reliabilitas tinggi

Hasil pada Tabel 3 menunjukkan bahwa bukti validitas empiris dan reliabilitas kedua tes membentuk pola yang berbeda. Tes berpikir komputasional memiliki seluruh butir yang memenuhi kriteria validitas empiris, tetapi koefisien reliabilitasnya sebesar 0,78 masih berada di bawah batas 0,80 yang ditetapkan dalam penelitian. Sebaliknya, tes berpikir kritis menunjukkan reliabilitas tinggi sebesar 0,89, meskipun satu butir belum memenuhi kriteria validitas empiris. Perbedaan tersebut tidak menunjukkan hasil yang bertentangan karena validitas empiris butir dan reliabilitas memberikan informasi yang berbeda: korelasi butir-total menggambarkan keterkaitan setiap butir dengan skor tes secara keseluruhan, sedangkan reliabilitas menggambarkan konsistensi skor instrumen sebagai satu kesatuan.

Dengan demikian, tujuan kedua menunjukkan bahwa dukungan empiris terhadap kedua instrumen memiliki karakteristik yang berbeda. Tes berpikir komputasional menunjukkan fungsi butir yang memadai, tetapi konsistensi internalnya masih perlu diperkuat, sedangkan tes berpikir kritis memiliki konsistensi yang tinggi tetapi masih memerlukan perbaikan pada satu butir. Oleh karena itu, keputusan terhadap penggunaan instrumen tidak hanya didasarkan pada validitas empiris dan reliabilitas, tetapi perlu dilengkapi dengan analisis karakteristik setiap butir. Analisis selanjutnya memeriksa daya pembeda dan tingkat kesulitan untuk menentukan butir yang dapat dipertahankan serta butir yang memerlukan revisi.

Karakteristik Butir dan Keputusan Pemanfaatan

Tujuan ketiga mengidentifikasi karakteristik butir berdasarkan daya pembeda dan tingkat kesulitan sekaligus menelusuri keputusan pemanfaatannya setelah uji coba. Pada tes

berpikir komputasional, daya pembeda berada pada rentang 0,41-0,62 dengan kategori cukup hingga baik; tingkat kesulitan terdiri atas dua butir mudah, dua sedang, dan dua sulit. Karena seluruh Butir 1-6 memenuhi kriteria validitas empiris dan menunjukkan daya pembeda yang memadai, keenamnya dipertahankan untuk penelitian utama. Pada tes berpikir kritis, Butir 1-5 dipertahankan secara langsung. Butir 6 memiliki korelasi butir-total 0,24 dan tingkat kesulitan 1% (sangat sulit), meskipun indeks daya pembedanya 1,00. Butir 6 tidak dihapus karena merupakan satu-satunya representasi indikator *self-regulation* dan dokumentasi uji coba mencatat adanya keterbatasan waktu pada butir terakhir; butir tersebut direvisi dan kemudian tetap digunakan dalam penelitian utama. Keputusan setiap butir dirangkum pada Tabel 4.

Tabel 4. Karakteristik Empiris Dan Keputusan Pemanfaatan Butir

Tes/Butir	r butir-total/status	IDP/kategori	IKES/ kategori	Keputusan
CT-1	0,73 / valid	0,59 / cukup	68% / mudah	Dipertahankan
CT-2	0,73 / valid	0,62 / baik	59% / sedang	Dipertahankan
CT-3	0,65 / valid	0,45 / cukup	49% / sedang	Dipertahankan
CT-4	0,51 / valid	0,41 / cukup	33% / sulit	Dipertahankan
CT-5	0,63 / valid	0,51 / cukup	25% / sulit	Dipertahankan
CT-6	0,72 / valid	0,41 / cukup	62% / mudah	Dipertahankan
BK-1	0,70 / valid	0,61 / baik	56% / sedang	Dipertahankan
BK-2	0,83 / valid	0,69 / baik	60% / sedang	Dipertahankan
BK-3	0,82 / valid	0,73 / baik	58% / sedang	Dipertahankan
BK-4	0,91 / valid	0,84 / sangat baik	49% / sedang	Dipertahankan
BK-5	0,86 / valid	0,80 / baik	35% / sulit	Dipertahankan
BK-6	0,24 / belum valid	1,00 / sangat baik	1% / sangat sulit	Direvisi, lalu dipertahankan

Sumber: Hasil uji coba instrumen terhadap 28 siswa dan dokumentasi keputusan revisi, 2026.

Dengan demikian, instrumen yang digunakan dalam penelitian utama tetap terdiri atas enam butir tes berpikir komputasional dan enam butir tes berpikir kritis. Seluruh butir tes berpikir komputasional dipertahankan karena memenuhi kriteria yang ditetapkan pada tahap uji coba. Pada tes berpikir kritis, Butir 1–5 dipertahankan tanpa revisi, sedangkan Butir 6 tetap digunakan setelah dilakukan perbaikan berdasarkan hasil evaluasi butir. Namun, data sumber tidak mendokumentasikan uji coba ulang secara khusus terhadap Butir 6 setelah revisi. Oleh karena itu, penelitian ini tidak mengklaim bahwa validitas empiris Butir 6 pada versi revisi telah dikonfirmasi kembali melalui pengujian empiris.

4. PEMBAHASAN

Temuan mengenai kelayakan teoretis menunjukkan bahwa kedua instrumen memperoleh dukungan yang lebih kuat pada aspek isi dan bahasa dibandingkan aspek konstruk dan praktisi. Pola tersebut menunjukkan bahwa butir tes telah memiliki kesesuaian yang baik dengan materi yang diukur dan penggunaan bahasa yang sesuai, tetapi penilaian dari perspektif praktisi masih memberikan ruang untuk penyempurnaan. Dalam pengembangan instrumen pendidikan, penilaian ahli dan praktisi diperlukan untuk memastikan bahwa butir tidak hanya merepresentasikan konstruk secara konseptual, tetapi juga dapat dipahami dan digunakan sesuai dengan karakteristik peserta didik. Penelitian pengembangan instrumen mutakhir juga menempatkan validasi ahli dan praktisi sebagai tahap penting sebelum instrumen diuji secara empiris (An Nabil et al., 2022; Astari et al., 2025; Lukman et al., 2023; Nurhasanah et al., 2025; Sheptian et al., 2024). Dengan demikian, persentase penilaian praktisi yang relatif lebih rendah pada kedua tes tidak menunjukkan ketidaksesuaian konstruk, tetapi mengindikasikan perlunya perhatian terhadap kejelasan perintah, kompleksitas tugas, waktu pengerjaan, dan kesesuaian tuntutan soal dengan karakteristik siswa sekolah dasar.

Bukti teoretis tersebut selanjutnya diperkuat melalui analisis empiris pada tingkat butir. Pada tes berpikir komputasional, seluruh Butir 1-6 memenuhi kriteria validitas empiris dengan

korelasi butir-total sebesar 0,51-0,73. Hasil ini menunjukkan bahwa setiap butir memiliki hubungan yang memadai dengan skor tes dan mendukung keputusan untuk mempertahankan seluruh butir. Temuan tersebut sejalan dengan perkembangan asesmen *computational thinking* yang menekankan bahwa keterwakilan indikator perlu disertai bukti mengenai fungsi empiris butir sebelum skor digunakan untuk menginterpretasikan kemampuan siswa (El-Hamamsy et al., 2025; Kong & Lai, 2022; Tsai et al., 2022; Zhang & Wong, 2023).

Meskipun seluruh butir berpikir komputasional memenuhi kriteria validitas empiris, koefisien reliabilitas *split-half* Spearman-Brown sebesar 0,78 masih berada sedikit di bawah batas 0,80 yang digunakan dalam penelitian. Temuan ini memperlihatkan bahwa terpenuhinya validitas empiris pada setiap butir tidak secara otomatis menghasilkan konsistensi skor tes yang memenuhi kriteria reliabilitas yang ditetapkan. Interpretasi nilai 0,78 perlu mempertimbangkan bahwa tes hanya terdiri atas enam butir dan uji coba dilakukan pada 28 siswa. Penelitian mengenai asesmen berpikir komputasional menunjukkan bahwa estimasi reliabilitas dapat dipengaruhi oleh jumlah dan karakteristik butir, tingkat kesulitan, rentang kemampuan peserta, serta karakteristik populasi yang menjadi sasaran pengukuran (El-Hamamsy et al., 2025; Kong & Lai, 2022; Kong & Wang, 2023; Ocampo et al., 2024; Zhang & Wong, 2023). Oleh karena itu, hasil tersebut lebih tepat dipandang sebagai bukti bahwa konsistensi pengukuran tes berpikir komputasional masih perlu diperkuat melalui penyempurnaan atau penambahan butir serta pengujian pada sampel yang lebih besar dan beragam.

Pola yang berbeda ditemukan pada tes berpikir kritis. Koefisien reliabilitas sebesar 0,89 menunjukkan konsistensi yang tinggi pada tingkat tes, sedangkan Butir 1-5 memiliki korelasi butir-total sebesar 0,70-0,91 dan memenuhi kriteria validitas empiris. Namun, Butir 6 hanya memperoleh korelasi sebesar 0,24 sehingga belum memenuhi kriteria yang ditetapkan. Temuan tersebut menunjukkan bahwa reliabilitas yang tinggi pada tingkat instrumen tidak berarti setiap butir di dalamnya memiliki fungsi empiris yang sama kuat. Korelasi butir-total memberikan informasi mengenai hubungan suatu butir dengan skor tes, sedangkan reliabilitas menggambarkan konsistensi skor instrumen sebagai satu kesatuan. Karena itu, keduanya perlu ditafsirkan secara bersamaan dan tidak digunakan secara saling menggantikan (Coppi et al., 2023; Darman et al., 2024; Ernawati et al., 2024; Nurhasanah et al., 2025; Schames Kreitchmann et al., 2024).

Analisis daya pembeda dan tingkat kesulitan memberikan informasi tambahan mengenai fungsi setiap butir. Keenam butir tes berpikir komputasional memiliki indeks daya pembeda sebesar 0,41–0,62 dengan kategori cukup hingga baik. Tingkat kesulitan juga tersebar pada kategori mudah, sedang, dan sulit. Sebaran tersebut menunjukkan bahwa tes tidak hanya terdiri atas butir pada satu tingkat tuntutan, sedangkan daya pembeda yang memadai menunjukkan kemampuan butir untuk membedakan variasi performa siswa pada kelompok uji coba. Kombinasi validitas empiris, daya pembeda, dan variasi tingkat kesulitan tersebut mendukung keputusan untuk mempertahankan seluruh Butir 1–6 tes berpikir komputasional. Analisis semacam ini penting karena kualitas instrumen tidak cukup ditentukan oleh satu karakteristik, tetapi perlu mempertimbangkan bagaimana setiap butir berfungsi dalam rentang kemampuan peserta yang diukur (Ernawati et al., 2024; Nurhasanah et al., 2025; Riyadi et al., 2024).

Perhatian khusus diperlukan pada Butir 6 tes berpikir kritis. Butir tersebut memiliki indeks daya pembeda sebesar 1,00, tetapi tingkat kesulitannya hanya 1% dan korelasi butir-total sebesar 0,24. Kombinasi tersebut menunjukkan bahwa nilai daya pembeda yang sangat tinggi tidak dapat digunakan sendirian untuk menyatakan bahwa suatu butir telah berfungsi dengan baik. Tingkat kesulitan yang ekstrem dan korelasi butir-total yang rendah menunjukkan adanya masalah yang tetap perlu diperhatikan. Hal ini memperkuat pentingnya menggunakan beberapa indikator secara bersamaan ketika mengevaluasi kualitas butir, sebagaimana

ditunjukkan dalam berbagai penelitian pengembangan dan validasi instrumen terkini (Darman et al., 2024; Nurhasanah et al., 2025; Schames Kreitchmann et al., 2024).

Dokumentasi uji coba menunjukkan bahwa Butir 6 ditempatkan pada bagian akhir tes dan sebagian siswa mengalami keterbatasan waktu sehingga mengerjakannya secara terburu-buru. Kondisi tersebut dapat menjadi salah satu faktor situasional yang berkaitan dengan rendahnya performa siswa pada butir tersebut, tetapi bukti yang tersedia belum cukup untuk menetapkan keterbatasan waktu sebagai penyebab tunggal. Selain itu, Butir 6 merupakan satu-satunya butir yang merepresentasikan indikator *self-regulation*. Penghapusan langsung terhadap butir tersebut berpotensi mengurangi keterwakilan konstruk berpikir kritis. Atas pertimbangan tersebut, Butir 6 direvisi dan tetap digunakan dalam penelitian utama. Namun, karena tidak tersedia hasil uji coba ulang secara khusus setelah revisi, kualitas empiris Butir 6 versi perbaikan belum dapat dinyatakan telah terkonfirmasi. Kondisi ini menunjukkan bahwa keputusan mempertahankan suatu butir perlu mempertimbangkan keseimbangan antara bukti statistik dan keterwakilan konstruk, tetapi tetap membutuhkan pengujian kembali setelah revisi.

Rangkaian temuan tersebut menunjukkan bahwa kelayakan teoretis, validitas empiris, reliabilitas, daya pembeda, dan tingkat kesulitan memberikan informasi yang berbeda tetapi saling melengkapi. Validasi ahli dan praktisi memberikan bukti mengenai keterwakilan konstruk, ketepatan isi, bahasa, dan keterterapan instrumen, sedangkan pengujian empiris menunjukkan bagaimana setiap butir dan tes bekerja ketika diberikan kepada siswa. Karena itu, keputusan mengenai kualitas instrumen tidak tepat apabila hanya didasarkan pada satu statistik. Pendekatan berbasis beberapa sumber bukti lebih memungkinkan peneliti mengidentifikasi butir yang dapat dipertahankan, butir yang memerlukan revisi, serta aspek instrumen yang masih membutuhkan penguatan sebelum digunakan secara lebih luas (An Nabil et al., 2022; Ernawati et al., 2024; Nurhasanah et al., 2025; Schames Kreitchmann et al., 2024).

Dari sisi praktis, kedua instrumen berpotensi digunakan sebagai dasar pengembangan asesmen diagnostik pada materi pecahan karena informasi yang diperoleh tidak terbatas pada benar atau salahnya jawaban akhir. Tes berpikir komputasional menelusuri kemampuan siswa dalam melakukan dekomposisi, mengenali pola, melakukan abstraksi, dan menyusun algoritma, sedangkan tes berpikir kritis mencakup interpretasi, analisis, evaluasi, inferensi, eksplanasi, dan *self-regulation*. Karakteristik tersebut relevan dengan penyelesaian masalah pecahan yang membutuhkan pemahaman hubungan antarrepresentasi, pemilihan strategi, pemeriksaan ketepatan prosedur, dan penjelasan terhadap proses penyelesaian. Temuan penelitian terdahulu juga menunjukkan pentingnya berpikir komputasional dan berpikir kritis dalam pemecahan masalah matematika dan penalaran pecahan pada pendidikan dasar (Agustin et al., 2026; Aulia & Sutarni, 2024; Candraningtyas & Khusna, 2023; Fitriani & Kowiyah, 2022; Tangkui, 2023; Wahyuningtyas & Widiyono, 2024; Yeo & Webel, 2024).

Interpretasi hasil penelitian tetap perlu mempertimbangkan sejumlah keterbatasan. Uji coba hanya melibatkan 28 siswa dari satu sekolah, sedangkan setiap konstruk diukur menggunakan enam butir. Reliabilitas baru diestimasi menggunakan metode *split-half Spearman-Brown*, sementara struktur internal, fungsi informasi tes, dan invariansi pengukuran belum dianalisis. Selain itu, Butir 6 tes berpikir kritis belum diuji kembali secara khusus setelah dilakukan revisi. Oleh karena itu, temuan penelitian ini lebih tepat diposisikan sebagai bukti awal kualitas pengukuran. Penelitian selanjutnya perlu melibatkan sampel yang lebih besar dan beragam, memperluas jumlah butir terutama pada indikator yang hanya direpresentasikan oleh satu tugas, serta melakukan uji ulang terhadap butir yang telah direvisi. Analisis Rasch atau *Item Response Theory* juga dapat dipertimbangkan untuk memperoleh bukti yang lebih kuat mengenai kecocokan butir, distribusi tingkat kesulitan, dan fungsi pengukuran pada rentang kemampuan siswa yang lebih luas (Darman et al., 2024; Ernawati et al., 2024; Inasari et al., 2023; Nurhasanah et al., 2025; Ocampo et al., 2024).

5. KESIMPULAN

Kedua instrumen menunjukkan kualitas yang berbeda. Tes berpikir komputasional memiliki kelayakan teoretis yang baik dan seluruh enam butir memenuhi kriteria validitas empiris, tetapi reliabilitasnya sebesar 0,78 masih berada di bawah batas 0,80 yang digunakan dalam penelitian. Tes berpikir kritis memiliki reliabilitas tinggi sebesar 0,89 dan lima dari enam butir memenuhi kriteria validitas empiris. Seluruh butir tes berpikir komputasional dipertahankan karena memiliki daya pembeda cukup hingga baik dan tingkat kesulitan yang bervariasi. Pada tes berpikir kritis, Butir 1-5 dipertahankan, sedangkan Butir 6 direvisi karena belum memenuhi kriteria validitas empiris dan tergolong sangat sulit, meskipun memiliki daya pembeda tinggi. Butir tersebut tetap digunakan untuk mempertahankan indikator *self-regulation*, tetapi masih memerlukan uji empiris ulang setelah revisi. Temuan ini menegaskan bahwa kualitas instrumen perlu ditentukan melalui keterpaduan kelayakan teoretis, validitas empiris, reliabilitas, daya pembeda, dan tingkat kesulitan. Pengujian lebih lanjut pada sampel yang lebih besar dan beragam tetap diperlukan untuk memperkuat kualitas pengukuran kedua instrumen.

DAFTAR PUSTAKA

- Agustin, H. N., Mariana, N., Siswono, T. Y. E., & Wiryanto. (2026). Elementary Students' Difficulties in Adding Fractions. *Journal of Innovation and Research in Primary Education*, 5(1), 262–274. <https://doi.org/10.56916/jirpe.v5i1.2753>
- Alfaris, R., Sa'dijah, C., Nusantara, T., Pratiwi, N., & Arifin, S. (2025). A Dual-Construct Assessment Model of Creative Mathematical Reasoning and Critical Thinking in Primary Education. *Journal of Innovation and Research in Primary Education*, 4(1), 63–72. <https://doi.org/10.56916/jirpe.v4i1.782>
- An Nabil, N. R., Wulandari, I., Yamtinah, S., Ariani, S. R. D., & Ulfa, M. (2022). Analisis indeks Aiken untuk mengetahui validitas isi instrumen asesmen kompetensi minimum berbasis konteks sains kimia. *PAEDAGOGIA*, 25(2), 184. <https://doi.org/10.20961/paedagogia.v25i2.64566>
- Astari, T., Nurhasanah, Cahyanita, E., & Nuriman. (2025). Pengembangan instrumen literasi sains berbasis kearifan lokal untuk pembelajaran di sekolah dasar. *Prosiding Seminar Nasional Pendidikan Sains (SNPS) 2025. Prosiding Universitas Sebelas Maret*.
- Aulia, S. N. A., & Sutarni, S. (2024). Analysis of Mathematic Critical Thinking Abilities of 4th Class Primary School Students on Fraction Material as well as Gender Differences. *Prima: Jurnal Pendidikan Matematika*, 8(2), 302. <https://doi.org/10.31000/prima.v8i2.10941>
- Candraningtyas, S. R., & Khusna, H. (2023). Computational thinking ability becomes a predictor of mathematical critical thinking ability. *Alifmatika: Jurnal Pendidikan Dan Pembelajaran Matematika*, 5(2), 247–263. <https://doi.org/10.35316/alifmatika.2023.v5i2.247-263>
- Coppi, M., Fialho, I., & Cid, M. (2023). Scientific literacy assessment instruments: A systematic literature review. *Educação Em Revista*, 39. <https://doi.org/10.1590/0102-4698237523-t>
- Darman, D. R., Suhandi, A., Kaniawati, I., Samsudin, A., & Wibowo, F. C. (2024). Development and Validation of Scientific Inquiry Literacy Instrument (SILI) Using Rasch Measurement Model. *Education Sciences*, 14(3), 322. <https://doi.org/10.3390/educsci14030322>
- El-Hamamsy, L., Zapata-Cáceres, M., Marcelino, P., Bruno, B., Dehler Zufferey, J., Martín-Barroso, E., & Román-González, M. (2022). Comparing the psychometric properties of two primary school Computational Thinking (CT) assessments for grades 3 and 4: The

- Beginners' CT test (BCTt) and the competent CT test (cCTt). *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.1082659>
- El-Hamamsy, L., Zapata-Cáceres, M., Martín-Barroso, E., Mondada, F., Zufferey, J. D., Bruno, B., & Román-González, M. (2025). The Competent Computational Thinking Test (cCTt): A Valid, Reliable and Gender-Fair Test for Longitudinal CT Studies in Grades 3–6. *Technology, Knowledge and Learning*, 30(3), 1607–1661. <https://doi.org/10.1007/s10758-024-09777-8>
- Ernawati, E., Habibah, R. Y., Syarifah, N., Firmansyah, F., & Attamimi, H. R. (2024). Item analysis test of science, Indonesian language, and mathematics using the rasch model in elementary schools. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 28(2), 195–209. <https://doi.org/10.21831/pep.v28i2.75448>
- Fitriani, F., & Kowiyah. (2022). Mathematics Critical Thinking Skills for The Third Grade Elementary School Students on Fractions Material. *MIMBAR PGSD Undiksha*, 10(3), 463–468. <https://doi.org/10.23887/jjpgsd.v10i3.48741>
- Inasari, L., Lidinillah, D. A. M., & Prehanto, A. (2023). Pengembangan instrumen tes computational thinking Siswa Sekolah Dasar melalui analisis RASCH model. *COLLASE (Creative of Learning Students Elementary Education)*, 6(1), 102–110. <https://doi.org/10.22460/collase.v1i1.16188>
- Kong, S.-C., & Lai, M. (2022). Validating a computational thinking concepts test for primary education using item response theory: An analysis of students' responses. *Computers & Education*, 187, 104562. <https://doi.org/10.1016/j.compedu.2022.104562>
- Kong, S.-C., & Wang, Y.-Q. (2023). Monitoring cognitive development through the assessment of computational thinking practices: A longitudinal intervention on primary school students. *Computers in Human Behavior*, 145, 107749. <https://doi.org/10.1016/j.chb.2023.107749>
- Lukman, H. S., Setiani, A., & Agustiani, N. (2023). Validitas Instrumen Tes Kemampuan Berpikir Kritis Matematis Berdasarkan Teori FRISCO. *SJME (Supremum Journal of Mathematics Education)*, 7(1), 55–67. <https://doi.org/10.35706/sjme.v7i1.6960>
- Na, C., Clarke-Midura, J., Shumway, J., van Dijk, W., & Lee, V. R. (2024). Validating a performance assessment of computational thinking for early childhood using item response theory. *International Journal of Child-Computer Interaction*, 40, 100650. <https://doi.org/10.1016/j.ijcci.2024.100650>
- Nordby, S. K., Bjerke, A. H., & Mifsud, L. (2022). Computational Thinking in the Primary Mathematics Classroom: a Systematic Review. *Digital Experiences in Mathematics Education*, 8(1), 27–49. <https://doi.org/10.1007/s40751-022-00102-5>
- Nordby, S. K., Mifsud, L., & Bjerke, A. H. (2024). Computational thinking in primary mathematics classroom activities. *Frontiers in Education*, 9. <https://doi.org/10.3389/educ.2024.1414081>
- Nurhasanah, Astari, T., Cahyanita, E., & Nuriman. (2025). Psychometric properties of a local wisdom-based science literacy instrument for Primary School Students: An Item Response Theory approach. *Jurnal Penelitian Pendidikan IPA*, 11(11), 965–975. <https://doi.org/10.29303/jppipa.v11i11.12973>
- Nurhasanah, Hidayatullah, Z., & Arif, Moh. B. S. (2024). Karakteristik instrumen tes literasi digital ditinjau dari validitas isi dan validitas empiris (kecocokan butir dengan model, reliabilitas, serta tingkat kesukaran butir). *Journal of Classroom Action Research*, 6(4), 916–923. <https://doi.org/10.29303/jcar.v6i4.9650>
- Ocampo, L. M., Corrales-Álvarez, M., Cardona-Torres, S. A., & Zapata-Cáceres, M. (2024). Systematic Review of Instruments to Assess Computational Thinking in Early Years of Schooling. *Education Sciences*, 14(10), 1124. <https://doi.org/10.3390/educsci14101124>

- Riyadi, A., Susongko, P., & Munadi, M. (2024). Model asesmen literasi sains pada peserta didik Sekolah Dasar dengan aplikasi Model Rasch. *Journal of Education Research*, 5(3), 3044–3054. <https://doi.org/10.37985/jer.v5i3.1411>
- Schames Kreitchmann, R., Nájera, P., Sanz, S., & Sorrel, M. Á. (2024). Enhancing content validity assessment with item response theory modeling. *Psicothema*, 36(2), 145–153. <https://doi.org/10.7334/psicothema2023.208>
- Sheptian, R., Ocy, D. R., Triana, D. D., & Wahyuni, L. D. U. (2024). Developing a validated essay-based assessment instrument to measure science literacy in energy Topics. *JISAE: Journal of Indonesian Student Assessment and Evaluation*, 10(2), 88–100. <https://doi.org/10.21009/jisae.v10i2.50762>
- Tangkui, R. Bin. (2023). Integrating Computational Thinking and The Polya's Model In Minecraft: The Effects on Learners' Fractions Achievement. *Indonesian Journal of Science and Mathematics Education*, 6(2), 125–136. <https://doi.org/10.24042/ijsme.v6i2.16033>
- Tsai, M.-J., Chien, F. P., Wen-Yu Lee, S., Hsu, C.-Y., & Liang, J.-C. (2022). Development and Validation of the Computational Thinking Test for Elementary School Students (CTT-ES): Correlate CT Competency With CT Disposition. *Journal of Educational Computing Research*, 60(5), 1110–1129. <https://doi.org/10.1177/07356331211051043>
- Wahyuningtyas, I. N., & Widiyono, A. (2024). The Effect of STEM-Based Discovery Learning Model on Critical Thinking Skills of Elementary School Students on Fraction Topics in the Context of Islamic Education. *Jurnal Tarbiyatuna*, 15(2), 139–157. <https://doi.org/10.31603/tarbiyatuna.v15i2.12632>
- Yeo, S., & Webel, C. (2024). Elementary students' fraction reasoning: a measurement approach to fractions in a dynamic environment. *Mathematical Thinking and Learning*, 26(1), 20–46. <https://doi.org/10.1080/10986065.2022.2025639>
- Zapata-Cáceres, M., Marcelino, P., El-Hamamsy, L., & Martín-Barroso, E. (2024). A Bebras Computational Thinking (ABC-Thinking) program for primary school: Evaluation using the competent computational thinking test. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-023-12441-w>
- Zhang, S., & Wong, G. K. W. (2023). Development and validation of a computational thinking test for lower primary school students. *Educational Technology Research and Development*, 71(4), 1595–1630. <https://doi.org/10.1007/s11423-023-10231-2>