TEMATIK: Jurnal Konten Pendidikan Matematika
Volume 4, Number 1, Tahun 2026, pp. 83-90
ISSN: 2985-8844
Open Access: http://doi.org/10.55210/jkpm

Kualitas Instrumen Tes Matematika Materi Pola Bilangan Berbasis *Computational Thinking*: Kajian Kesukaran, Daya Pembeda, dan Distraktor

Aulia Zahro Kholili[1], Chika Rahayu[2*], Agung Putra Wijaya[3]

[1,2,3] Universitas Lampung, Indonesia
Email : zahrokholili22@gmail.com[1], chikarahayu@fkip.unila.ac.id[2], agung.wijaya@fkip.unila.ac.id[3]

INFORMASI ARTIKEL
Tersedia Online pada:
29 Juli, 2026

Kata Kunci:
Evaluasi, *Computational Thinking*, Indeks Kesukaran, Daya Pembeda, Distraktor

Keywords:
Evaluation, Computational Thinking, Difficulty Index, Discrimination Index, Distractor

This is an open access article under the CC BY 4.0 license.
Copyright © 2026 by Author. Published by ***Universitas Islam Zainul Hasan Genggong***

Abstrak

Kualitas instrumen tes yang belum dianalisis secara menyeluruh dapat menyebabkan hasil pengukuran kemampuan berpikir siswa menjadi kurang akurat. Oleh karena itu, penelitian ini bertujuan untuk menganalisis kualitas instrumen tes pilihan ganda berbasis *Computational Thinking* (CT) pada materi pola bilangan ditinjau dari indeks kesukaran, daya pembeda, fungsi distraktor, dan reliabilitas instrumen. Penelitian ini menggunakan pendekatan kuantitatif deskriptif. Subjek penelitian terdiri dari 29 siswa kelas VIII SMPN 2 Sungkai Utara yang mengerjakan 10 butir soal pilihan ganda. Instrumen disusun berdasarkan empat indikator berpikir komputasional, yaitu dekomposisi, pengenalan pola, abstraksi, dan algoritma. Data dianalisis menggunakan teori tes klasik, dengan perhitungan indeks kesukaran (P), daya pembeda (D), dan persentase pemilih distraktor. Hasil analisis menunjukkan bahwa 3 butir soal termasuk kategori mudah, 5 butir soal kategori sedang, dan 2 butir soal kategori sukar. Berdasarkan daya pembeda, 7 butir soal memiliki daya pembeda sangat baik, 1 butir soal memiliki daya pembeda baik, dan 2 butir soal memiliki daya pembeda rendah. Analisis fungsi distraktor menunjukkan bahwa 16 dari 30 distraktor (53,3%) berfungsi secara efektif, sedangkan 14 distraktor (46,7%) belum berfungsi secara efektif. Selain itu, instrumen memperoleh koefisien reliabilitas sebesar 0,709, yang menunjukkan konsistensi internal yang baik. Berdasarkan temuan tersebut, instrumen tes pilihan ganda berbasis *Computational Thinking* pada materi pola bilangan memiliki kualitas yang cukup baik ditinjau dari indeks kesukaran, daya pembeda, fungsi distraktor, dan reliabilitas sehingga dapat digunakan untuk mengukur kemampuan berpikir siswa secara sistematis dan terukur.

Abstract

Evaluation is an essential component of mathematics learning as it serves to assess students' competency achievement while ensuring the quality of the test instruments used. A good instrument must not only be content-valid but also possess measurable empirical strength. This study aims to analyze the quality of multiple-choice test items based on Computational Thinking (CT) in the topic of number patterns, reviewed from the aspects of difficulty index, discrimination index, and distractor effectiveness. The research employed a descriptive quantitative approach. The subjects were 29 eighth-grade students from SMPN 2 Sungkai Utara who answered 10 multiple-choice questions. The instrument was developed based on four CT indicators: decomposition, pattern recognition, abstraction, and algorithm. Data were analyzed using classical test theory, by calculating the difficulty index (P), discrimination index (D), and the percentage of distractor selections. The analysis results showed that most items fell into the medium difficulty category and had sufficient to excellent discrimination power. Most distractors also functioned effectively, although several options required improvement. The instrument demonstrated high reliability, indicating its consistency as a measurement tool. These findings suggest that integrating Computational Thinking indicators into multiple-choice mathematics items can be an effective evaluation strategy for assessing students' thinking skills in a structured and measurable way.

PENDAHULUAN

Rendahnya kualitas evaluasi pembelajaran matematika, yang tercermin dari rendahnya capaian siswa Indonesia pada asesmen internasional seperti PISA, menunjukkan perlunya instrumen evaluasi yang mampu mengukur kompetensi siswa secara akurat.Evaluasi merupakan komponen penting dalam proses pembelajaran matematika karena berfungsi sebagai alat untuk menilai pencapaian kompetensi siswa, efektivitas pembelajaran, serta dasar perbaikan instruksional. Dalam konteks pendidikan matematika, evaluasi tidak hanya berfokus pada hasil akhir, tetapi juga pada proses berpikir siswa dalam menyelesaikan persoalan matematis (Arikunto, 2012; Retnawati, 2016). Salah satu bentuk evaluasi yang banyak

*Corresponding author.
E-mail addresses: chikarahayu@fkip.unila.ac.id

digunakan di sekolah adalah tes pilihan ganda. Instrumen ini dinilai efisien dalam pelaksanaan dan pengolahan. Namun, untuk dapat dianggap berkualitas, setiap butir soal perlu melalui proses analisis yang sistematis dan objektif, antara lain melalui pengukuran indeks kesukaran, daya pembeda, dan fungsi distraktor (Sudjana, 2016; Haladyna & Downing, 2004).

Seiring dengan tuntutan pembelajaran abad ke-21, pendekatan *Computational Thinking* (CT) mulai diintegrasikan dalam kurikulum matematika. CT meliputi empat komponen utama: dekomposisi, pengenalan pola, abstraksi, dan algoritma (Wing, 2006; Angeli et al., 2016). CT tidak hanya penting dalam konteks informatika, tetapi juga relevan dalam pembelajaran matematika, terutama dalam menyelesaikan masalah berbasis pola bilangan (Cahdriyana & Richardo, 2020; Budiarti et al., 2022). Materi pola bilangan dipilih karena menuntut siswa untuk mengenali keteraturan suatu pola, melakukan generalisasi, dan menyusun langkah penyelesaian secara sistematis, sehingga berkaitan erat dengan komponen pengenalan pola, abstraksi, dan algoritma dalam *Computational Thinking*. Beberapa penelitian menunjukkan bahwa soal matematika yang dikembangkan berdasarkan indikator CT mampu mengungkap cara berpikir logis dan sistematis siswa (Setiawan & Rahayu, 2022; Azizah, 2021). Namun, pengukuran kemampuan CT selama ini lebih banyak dilakukan melalui soal uraian atau observasi, dan belum banyak yang menggunakan soal pilihan ganda secara terstruktur.

Padahal, instrumen pilihan ganda yang disusun berdasarkan prinsip CT berpotensi besar untuk digunakan secara luas dalam evaluasi di kelas. Instrumen semacam ini dapat dianalisis menggunakan teori tes klasik untuk menilai kualitas butir soal melalui indikator validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas distraktor (Haladyna & Downing, 2004). Namun, penelitian terhadap instrumen pilihan ganda berbasis CT masih sangat terbatas, khususnya dalam konteks materi pola bilangan di jenjang SMP. Hal ini menunjukkan adanya celah penelitian, yakni kurangnya data empiris mengenai kualitas soal pilihan ganda matematika berbasis CT yang dianalisis menggunakan parameter statistik klasik.

Oleh karena itu, penelitian ini bertujuan untuk menganalisis kualitas instrumen tes matematika materi pola bilangan yang dikembangkan berdasarkan pendekatan *Computational Thinking*. Fokus analisis meliputi indeks kesukaran, daya pembeda, dan fungsi distraktor dari setiap butir soal pilihan ganda. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan instrumen evaluasi yang tidak hanya valid secara isi, tetapi juga memiliki kekuatan statistik dan mampu mengukur kemampuan berpikir komputasional siswa secara efektif.

METODE

Penelitian ini menggunakan pendekatan kuantitatif deskriptif, karena bertujuan untuk menggambarkan karakteristik empiris kualitas soal pilihan ganda berdasarkan data skor siswa (Sugiyono, 2022; El-Hamamsy et al., 2023). Populasi penelitian adalah seluruh siswa kelas VIII SMPN 2 Sungai Utara. Sampel penelitian berjumlah 29 siswa kelas VIII yang dipilih menggunakan teknik purposive sampling karena telah mempelajari materi pola bilangan dan mengikuti tes pilihan ganda berbasis *Computational Thinking*. Data penelitian dikumpulkan melalui pemberian tes pilihan ganda kepada sampel penelitian. Instrumen tes disusun oleh peneliti berdasarkan indikator *Computational Thinking* yang dikemukakan oleh Wing (2006) dan Angeli et al. (2016) serta disesuaikan dengan materi pola bilangan.

Instrumen penelitian berupa 10 butir soal pilihan ganda yang dirancang untuk mengukur empat indikator *Computational Thinking*, yaitu dekomposisi, pengenalan pola, abstraksi, dan algoritma. Sebelum digunakan, instrumen divalidasi melalui validitas isi oleh dua dosen ahli pendidikan matematika. Proses validasi dilakukan dengan menelaah kesesuaian butir soal terhadap indikator *Computational Thinking*, materi pola bilangan, konstruksi soal, serta penggunaan bahasa. Hasil telaah validator digunakan sebagai dasar perbaikan instrumen sebelum diujicobakan kepada siswa (El-Hamamsy et al., 2023).

Teknik analisis data mengikuti teori tes klasik dan didasarkan pada tiga parameter:

1. Indeks kesukaran (P) adalah ukuran seberapa mudah atau sulit suatu soal bagi responden.

Rumus:

$$P = \frac{B}{N}$$

Kategori kesukaran mengikuti skala Anastasi & Urbina (2021) sebagai berikut:

Tabel 1. Kategori Indeks Kesukaran Butir Soal

Indeks Kesukaran (P)	Kategori
0,00 – 0,30	Sukar
0,31 – 0,70	Sedang
0,71 – 1,00	Mudah

2. Daya pembeda (D) Daya pembeda menunjukkan sejauh mana suatu soal mampu membedakan antara siswa yang memiliki kemampuan tinggi dan rendah. Adapun kelompok atas dan kelompok bawah dalam konteks analisis butir soal merujuk pada dua kelompok peserta tes yang memiliki skor tertinggi dan terendah, masing-masing 27% dari total peserta.
   Rumus:

$$D = \frac{B_{atas} - B_{bawah}}{N}$$

   Kategori daya pembeda mengikuti pedoman Sudjana (2022):

Tabel 2. Kategori Rentang Daya Pembeda

Rentang Daya Pembeda (D)	Kategori
D ≥ 0,40	Sangat Baik
0,30 – 0,39	Cukup
0,20 – 0,29	Kurang
D < 0,20	Jelek

3. Fungsi distraktor, dievaluasi berdasarkan persentase pemilihan opsi salah oleh siswa. Distraktor dinyatakan efektif jika dipilih minimal 5% siswa (analogi pada Dewi & Putra, 2023; Elgadal & Mariod, 2021), karena distraktor yang baik mampu mendeteksi kesalahan berpikir (PubMed 2023; 2023). Hubungan antara efisiensi distraktor, indeks kesukaran, dan daya pembeda dianalisis secara deskriptif dengan membandingkan hasil ketiga parameter pada setiap butir soal untuk mengetahui karakteristik kualitas butir soal (PubMed 2024).
4. Dalam konteks tes pilihan ganda berbasis teori tes klasik, reliabilitas menunjukkan seberapa stabil dan andal hasil tes jika diberikan pada kelompok yang setara. Dalam penelitian ini, reliabilitas dihitung menggunakan koefisien KR-20 (Kuder-Richardson Formula 20) karena seluruh item bersifat dikotom (jawaban benar atau salah). KR-20 mengukur konsistensi internal, yaitu sejauh mana butir-butir soal dalam satu tes saling berkorelasi.
   Rumus KR-20:

$$r_{11} = \frac{k}{k-1}\left[1 - \frac{\sum pq}{\sigma^2}\right]$$

Pengolahan data dilakukan menggunakan Microsoft Excel untuk menghitung seluruh parameter analisis butir soal. Analisis ini juga memperhatikan pedoman penulisan soal yang bebas dari kesalahan (item writing flaws) karena faktor seperti kualitas bahasa dan distraktor dapat memengaruhi kesukaran serta daya pembeda soal (Schmucker & Moore, 2025).

HASIL DAN PEMBAHASAN

Berdasarkan data hasil tes dari 29 siswa kelas VIII pada instrumen pilihan ganda berbasis Computational Thinking (CT), dilakukan analisis kualitas butir soal berdasarkan teori tes klasik. Hasil analisis menunjukkan bahwa: Analisis meliputi empat aspek, yaitu indeks kesukaran, daya pembeda, fungsi distraktor, dan reliabilitas instrument. Hasil analisis menunjukkan bahwa :

Indeks Kesukaran Butir Soal

Indeks kesukaran dihitung melalui rasio siswa yang menjawab benar terhadap jumlah seluruh peserta. Hasil analisis menunjukkan rentang nilai P = 0,23 – 0,93, yang dikategorikan sebagai berikut:

Tabel 3. Hasil Analisis Indeks Kesukaran Butir Soal

Kategori	Jumlah Soal	Nomor Soal
Mudah (P > 0,70)	3 soal	1, 2, 5
Sedang (0,30 ≤ P ≤ 0,70)	5 soal	3, 4, 6, 7, 8
Sulit (P < 0,30)	2 soal	9, 10

Komposisi soal dalam instrumen tes ini menunjukkan proporsi yang ideal, dengan sebagian besar soal berada pada tingkat kesukaran sedang. Temuan ini selaras dengan panduan Anastasi dan Urbina (2021) yang menyatakan bahwa soal dengan tingkat kesukaran sedang (0,30 ≤ P ≤ 0,70) memiliki potensi maksimal untuk membedakan siswa yang menguasai dan yang belum menguasai konsep. Dalam konteks asesmen berbasis *Computational Thinking* (CT), distribusi seperti ini sangat strategis karena memungkinkan instrumen mengukur kemampuan siswa secara lebih optimal pada berbagai tingkat kemampuan.

Soal-soal dengan tingkat kesukaran rendah berfungsi penting dalam membangun *confidence* atau kepercayaan diri siswa (Birenbaum et al., 2020), karena memberikan peluang kepada semua peserta untuk menjawab dengan benar dan merasa berhasil pada awal pengerjaan. Keberhasilan awal ini penting untuk memotivasi siswa menyelesaikan soal-soal berikutnya yang lebih menantang.

Di sisi lain, soal-soal yang tergolong sukar berperan dalam mengidentifikasi siswa yang memiliki penguasaan konsep secara lebih mendalam. Menurut *Haladyna & Rodriguez (2013)*, butir soal yang sulit namun valid secara konseptual dapat mengungkapkan tingkat berpikir yang lebih tinggi, yang sangat relevan dalam pengukuran kemampuan berpikir komputasional. Hal ini karena *computational thinking* mencakup dimensi seperti *problem decomposition*, *pattern recognition*, dan *algorithmic thinking* yang membutuhkan keterampilan berpikir tingkat tinggi.

Distribusi tingkat kesukaran yang berimbang, yakni dengan kehadiran soal mudah, sedang, dan sukar, mendukung pendekatan asesmen yang inklusif dan efektif (Brookhart, 2017; Suparman et al., 2025). Hal ini memungkinkan siswa dengan kemampuan beragam tetap dapat menunjukkan kompetensinya. Terlebih dalam konteks pembelajaran matematika berbasis CT, keberagaman tingkat kesukaran mendorong aktivasi proses berpikir dari level dasar (seperti mengenali pola) hingga level kompleks (seperti menyusun algoritma atau strategi penyelesaian masalah).

Daya Pembeda Soal

Daya pembeda dianalisis dengan membandingkan kinerja kelompok atas (27% tertinggi) dan kelompok bawah (27% terbawah). Hasilnya disajikan dalam tabel:

Tabel 4. Hasil Analisis Daya Pembeda Soal

Kategori	Jumlah Soal	Nomor Soal
Sangat Baik (D ≥ 0,40)	7 soal	4, 6, 7, 8, 9, 10
Baik (0,30 ≤ D < 0,40)	1 soal	3
Kurang (D < 0,20)	2 soal	1, 2

Sebanyak 70% butir soal dalam instrumen ini memiliki daya pembeda sangat baik (D ≥ 0,40), menunjukkan bahwa sebagian besar soal mampu secara efektif membedakan antara siswa berkemampuan tinggi dan rendah. Menurut *Arikunto (2019)*, butir soal dengan daya pembeda tinggi merupakan indikator validitas internal yang baik karena soal tersebut tidak hanya mengukur kemampuan, tetapi juga mampu mengelompokkan siswa berdasarkan tingkat penguasaan materi.

Temuan ini menunjukkan bahwa instrumen telah disusun dengan mempertimbangkan karakteristik kognitif yang mendukung pengukuran kemampuan *Computational Thinking* secara komprehensif. Terutama dalam konteks *computational thinking* (CT), yang menekankan kemampuan berpikir tingkat tinggi seperti *decomposition*, *abstraction*, *pattern recognition*, dan *algorithmic thinking* (Wing, 2006). Soal-soal yang dapat membedakan dengan jelas siswa yang memahami konsep CT dari yang belum menguasainya menandakan bahwa instrumen ini bukan hanya sekadar mengukur hafalan, tetapi mengevaluasi proses berpikir kritis dan sistematis.

Sebaliknya, dua soal (nomor 1 dan 2) yang memiliki daya pembeda rendah (< 0,20) perlu ditinjau ulang. Kemungkinan soal tersebut terlalu mudah atau justru tidak sesuai dengan indikator yang hendak diukur. *Nitko & Brookhart (2014)* menyarankan bahwa soal dengan daya pembeda rendah bisa menunjukkan adanya kesalahan dalam penyusunan alternatif jawaban, ketidaktepatan dalam redaksi soal, atau kurangnya keterkaitan dengan indikator kognitif yang ditargetkan.

Dengan daya pembeda yang kuat, sebagian besar soal dalam instrumen ini berpotensi untuk digunakan sebagai alat diagnosis dalam asesmen formatif. *Black & Wiliam (2009)* menekankan bahwa asesmen formatif yang baik harus mampu mengungkap kelemahan dan kekuatan siswa secara akurat, agar guru dapat memberikan intervensi pembelajaran yang tepat. Oleh karena itu, instrumen ini sangat cocok digunakan untuk memetakan pemahaman siswa terhadap materi matematika berbasis CT, serta sebagai dasar dalam perencanaan tindak lanjut pembelajaran.

Fungsi Distraktor

Tabel 5. Hasil Analisis Fungsi Distraktor

No. Soal	Kunci	Distraktor Efektif	Distraktor Tidak Efektif
1	C	A (6.67%), B (6.67%)	D (3.33%)
2	C	B (10%), C (80%)	A (3.33%), D (6.67%)
3	C	B (6.67%), D (3.33%)	A (0%)
4	B	A (6.67%)	C (3.33%), D (3.33%)
5	C	A (3.33%), C (90%)	B (3.33%), D (3.33%)
6	C	A (3.33%), C (90%)	B (3.33%), D (3.33%)
7	A	D (6.67%)	B (3.33%), C (0%)
8	C	B (6.67%), D (6.67%)	A (0%)
9	B	A (6.67%), C (6.67%)	D (6.67%)
10	C	A (3.33%), C (90%), D(6.67%)	B (0%)

Sebanyak 16 dari 30 distraktor (53,3%) dalam instrumen ini dikategorikan efektif karena berhasil menarik perhatian peserta didik, yaitu dipilih oleh minimal 5% dari total peserta. Ini berarti lebih dari separuh distraktor telah berfungsi sebagaimana mestinya sebagai pengecoh yang menggugah proses berpikir siswa. Namun, masih terdapat 14 distraktor (46,7%) yang tidak efektif, karena dipilih oleh kurang dari 5% peserta atau bahkan tidak dipilih sama sekali. Sebagai contoh, pilihan A pada soal nomor 3 dan pilihan C pada soal nomor 7 tidak menarik perhatian siswa, sehingga kehilangan fungsi diagnosisnya.

Distraktor yang tidak efektif umumnya terjadi karena opsi tersebut tidak cukup logis, terlalu jelas salahnya, atau tidak sesuai dengan pola kesalahan umum yang dibuat oleh siswa. Menurut Haladyna & Downing (1989), distraktor yang baik harus menarik bagi siswa yang belum memahami konsep, mencerminkan miskonsepsi yang umum, dan cukup masuk akal

sehingga dapat menimbulkan keraguan. Ketika sebuah distraktor gagal menarik perhatian, itu artinya peluang untuk mengetahui letak kesalahan pemikiran siswa hilang.

Temuan ini sejalan dengan penelitian Suparman et al. (2025) yang menunjukkan bahwa butir soal dengan daya pembeda yang baik lebih efektif dalam membedakan kemampuan siswa dan mendukung kualitas instrumen dalam mengukur *Computational Thinking*. Dengan daya pembeda yang kuat, sebagian besar soal dalam instrumen ini berpotensi digunakan sebagai alat diagnosis dalam asesmen formatif. Black dan Wiliam (2009) menekankan bahwa asesmen formatif yang baik harus mampu mengungkap kelemahan dan kekuatan siswa secara akurat sehingga guru dapat memberikan intervensi pembelajaran yang tepat. Dalam konteks *computational thinking* (CT), kualitas distraktor menjadi lebih krusial. Distraktor yang efektif seharusnya tidak hanya menjadi pilihan acak, melainkan merepresentasikan kesalahan logis atau prosedural yang mungkin dilakukan siswa saat menerapkan konsep seperti *decomposition*, *pattern recognition*, *abstraction*, atau *algorithmic thinking*. Grover & Pea (2018) menekankan bahwa kesalahan dalam berpikir komputasional sering kali bersifat sistemik, sehingga distraktor harus dirancang untuk menangkap pola kesalahan tersebut.

Peningkatan kualitas distraktor menjadi hal yang esensial dalam upaya memperkuat validitas butir soal. Seperti dinyatakan oleh *Suah & Ong (2012)*, distraktor yang berkualitas tidak hanya memperbaiki fungsi butir soal, tetapi juga meningkatkan keakuratan dalam membedakan tingkat pemahaman siswa. Oleh karena itu, dalam pengembangan soal selanjutnya, penulis disarankan untuk menganalisis kesalahan siswa dalam latihan atau tes sebelumnya sebagai dasar penyusunan opsi pengecoh yang lebih representatif terhadap pola pikir yang keliru namun umum.

Reliabilitas Instrumen

Tabel 6. Realibilitas Instrumen

Jumlah Varians	Varians Skor Total	Realibilitas (*α Cronbach*)
1,838	5,082	0,709

Berdasarkan hasil analisis, diperoleh varians total skor sebesar 5,082 dan jumlah varians butir soal sebesar 1,838. Kedua nilai ini digunakan dalam rumus *Cronbach's Alpha* untuk mengukur reliabilitas internal suatu instrumen, yaitu sejauh mana butir-butir soal dalam tes tersebut saling konsisten dalam mengukur konstruk yang sama. Hasil perhitungan menunjukkan bahwa koefisien reliabilitas Cronbach's Alpha (α) sebesar 0,709.

Nilai reliabilitas ini telah melewati ambang batas minimal 0,70 yang banyak digunakan dalam penelitian bidang sosial dan pendidikan (Nunnally & Bernstein, 1994). Artinya, instrumen ini dapat dikategorikan *reliable*—memiliki tingkat konsistensi internal yang cukup baik—dan dapat memberikan hasil pengukuran yang relatif stabil dan dapat dipercaya apabila digunakan dalam kondisi serupa, baik pada waktu berbeda maupun pada kelompok siswa yang sebanding.

Tingkat reliabilitas 0,709 juga menunjukkan bahwa sebagian besar butir soal memiliki kontribusi yang positif terhadap total skor. Namun demikian, nilai tersebut belum mencapai tingkat optimal (misalnya 0,80 ke atas), yang umumnya diharapkan dalam pengukuran keterampilan berpikir tingkat tinggi seperti *computational thinking*. Oleh karena itu, instrumen ini masih memiliki ruang untuk perbaikan.

Berdasarkan hasil analisis, butir soal nomor 4 dan 9 yang memiliki varians sangat rendah perlu ditelaah kembali karena menunjukkan kemampuan yang kurang optimal dalam membedakan kemampuan siswa. Temuan ini sejalan dengan Tavakol dan Dennick (2011) yang menyatakan bahwa butir soal dengan varians rendah cenderung menurunkan reliabilitas instrumen karena memberikan informasi pengukuran yang terbatas. Hasil tersebut juga konsisten dengan penelitian Suparman et al. (2025) yang menunjukkan bahwa kualitas butir soal berkontribusi terhadap reliabilitas instrumen.

Dalam konteks pengembangan tes berbasis *computational thinking*, memastikan reliabilitas yang tinggi sangat penting karena konsep-konsep seperti dekomposisi dan algoritmik memerlukan pengukuran yang akurat terhadap pemahaman mendalam. Dengan

demikian, revisi terhadap butir-butir soal yang kurang bervariasi menjadi langkah strategis untuk meningkatkan kualitas instrumen secara menyeluruh.

Berdasarkan keempat aspek yang telah dibahas, dapat disimpulkan bahwa kualitas instrumen pengukuran kemampuan *computational thinking* tidak hanya ditentukan oleh satu aspek saja, tetapi merupakan hasil keterkaitan antara dekomposisi, pengenalan pola, abstraksi, dan algoritmik. Keempat aspek tersebut saling melengkapi dalam menggambarkan kemampuan siswa dalam memahami, menyederhanakan, dan menyelesaikan masalah secara sistematis. Oleh karena itu, instrumen yang digunakan perlu memiliki butir soal yang mampu mengukur keempat aspek tersebut secara proporsional dan bervariasi. Dengan demikian, revisi terhadap butir soal yang kurang mewakili salah satu aspek menjadi langkah penting untuk meningkatkan kualitas instrumen secara menyeluruh, sehingga hasil pengukuran dapat memberikan gambaran yang lebih akurat mengenai kemampuan *computational thinking* siswa.

KESIMPULAN DAN SARAN

Secara keseluruhan, instrumen tes berbasis *Computational Thinking* menunjukkan kualitas yang cukup baik ditinjau dari indeks kesukaran, daya pembeda, fungsi distraktor, dan reliabilitas. Sebagian besar soal memiliki tingkat kesukaran sedang dan daya pembeda yang baik, sedangkan 14 dari 30 distraktor (46,7%) belum berfungsi efektif karena dipilih kurang dari 5% siswa atau tidak dipilih sama sekali. Hal ini menunjukkan bahwa beberapa distraktor belum cukup logis atau belum merepresentasikan kesalahan umum siswa. Instrumen memiliki reliabilitas sebesar 0,709, sehingga secara umum dapat digunakan untuk mengukur kemampuan berpikir komputasional siswa, dengan perbaikan pada distraktor dan butir soal yang memiliki daya pembeda rendah.

DAFTAR PUSTAKA

Anastasi, A., & Urbina, S. (2021). Psychological Testing (7th ed.). Pearson Education.

Angeli, C., Voogt, J., Fluck, A., Webb, M., Cox, M., Malyn-Smith, J., & Zagami, J. (2016). A K-6 Computational Thinking Curriculum Framework: Implications for Teacher Knowledge. Journal of Educational Technology & Society, 19(3), 47–57.

Arikunto, S. (2012). Dasar-Dasar Evaluasi Pendidikan (Ed. Revisi). Bumi Aksara.

Budiarti, A., Wibowo, S. C., & Nugraheni, A. S. (2022). Analisis Kemampuan Computational Thinking Siswa SMP dalam Menyelesaikan Soal Pola Bilangan. Jurnal Pendidikan Matematika, 10(2), 123–130.

Cahdriyana, D., & Richardo, H. (2020). Pengembangan Soal Kemampuan Berpikir Komputasional Siswa. Jurnal Pendidikan dan Pembelajaran Matematika, 6(1), 12–20.

Dewi, A. R., & Putra, W. D. (2023). Analisis Fungsi Distraktor dalam Soal Pilihan Ganda Mata Pelajaran Matematika. Jurnal Evaluasi Pendidikan, 14(1), 55–63.

Elgadal, A. A., & Mariod, A. A. (2021). Distractor Effectiveness in Multiple Choice Questions: A Case Study in Mathematics. Journal of Educational Measurement and Evaluation, 4(3), 190–199.

El-Hamamsy, L., Alabdulqader, E., & Alghamdi, A. (2023). A Computational Thinking-Based Assessment Model in Mathematics Learning. International Journal of Instruction, 16(2), 275–290.

Haladyna, T. M., & Rodriguez, M. C. (2013). Developing and Validating Test Items (3rd ed.). Routledge.

Mubarokah, L., Ramdhani, M. A., & Puspita, N. (2023). Kemampuan Berpikir Komputasional Siswa SMP dalam Menyelesaikan Masalah Matematika. Jurnal Pendidikan Matematika dan Sains, 8(1), 40–49.

Schmucker, D. G., & Moore, A. B. (2025). The Impact of Item-Writing Flaws on Student Assessment: An Empirical Study. Journal of Educational Measurement, 61(1), 18–34.

Sudjana, N. (2022). Penilaian Hasil Proses Belajar Mengajar. Remaja Rosdakarya.

Sugiyono. (2022). Metode Penelitian Kuantitatif, Kualitatif, dan R&D. Alfabeta.

Suparman, S., Juandi, D., Turmudi, Martadiputra, B. A. P., & Diana, N. (2025). Assessing the Quality of Mathematics Test–Enumeration Rules to Measure Students' Computational Thinking Skills. PAEDAGOGIA, 28(3), 412–428.

Wing, J. M. (2006). Computational Thinking. Communications of the ACM, 49(3), 33–35.