Jurnal Ilmiah Peuradeun
The Indonesian Journal of the Social Sciences

Journal Homepage:
https://journal.scadindependent.org/index.php/jipeuradeun

Vol. 14, No. 2, May 2026
Pages: 773-796
DOI: https://doi.org/10.26811/peuradeun.v14i2.1602

Publisher: SCAD Independent
P-ISSN: 2338-8617 / E-ISSN: 2443-2067

Assessing Students' Computational Thinking Ability Based on Mathematical Reasoning

Muhammad Istiqlal[1]; Heri Retnawati[2]; Widihastuti Widihastuti[3]; Desy Kumala Sari[4]
[1,2,3]Yogyakarta State University, Indonesia
[4]Musamus University, Indonesia
[1]Correspondence Email: muhammadistiqlal.2023@student.uny.ac.id

Article Info

Received: June 11, 2024
Accepted: May 3, 2026
Published: May 30, 2026
Online First: May 30, 2026

Keywords

Assessment;
Computational Thinking;
Mathematical Reasoning;
Multidimensional Item Response Theory.

Abstract

Despite the growing emphasis on twenty-first-century competencies, assessment practices in mathematics education have not sufficiently captured the integration of computational thinking and mathematical reasoning. This study aims to develop and validate the SMART-CT instrument as a multidimensional assessment framework for evaluating computational thinking within mathematical reasoning contexts among senior high school students. The instrument was developed using the ADDIE model and validated through expert judgment, second-order Confirmatory Factor Analysis (CFA), reliability analysis, and Multidimensional Item Response Theory (MIRT) based on data collected from 1,218 students. The results demonstrate that the instrument exhibits strong psychometric qualities, including satisfactory model fit, high internal consistency ($a = 0.861$), and adequate measurement precision within the ability range of $\theta = -0.5$ to $+1.0$, although reduced sensitivity was observed at higher ability levels. By conceptualizing computational thinking as an integral component of mathematical reasoning and operationalizing it within a multidimensional measurement framework, this study contributes to the advancement of more authentic and process-oriented assessment practices in mathematics education.

A. Introduction

The rapid evolution of digital technologies has significantly reshaped the epistemological foundations of knowledge construction and problem-solving in contemporary education. Within the domain of mathematics education, this paradigm shift has redirected instructional emphasis from rote procedural fluency toward more complex, multidimensional cognitive competencies (namely, reasoning, abstraction, and structured, algorithmic problem-solving) (Gunawan et al., 2025; Zaeni et al., 2025; Mujakir et al., 2024; Fauzi et al., 2024). These competencies are no longer peripheral skills but are widely recognized as central pillars of twenty-first-century literacy, closely aligning with global educational priorities and workforce-readiness frameworks (Aba-Oli et al., 2025; Dimov, 2024; Aydin & Birgili, 2023; Jailani et al., 2023). Consequently, the ability to integrate computational logic with mathematical abstraction has emerged as a critical, non-negotiable cognitive asset.

Despite the global consensus on the importance of these higher-order skills, empirical evidence reveals a persistent and troubling gap between aspirational curricula and actual student performance, particularly in the Indonesian context. Indonesian students consistently demonstrate low proficiency in mathematical reasoning, evidenced by the decline in Programme for International Student Assessment (PISA) scores from 379 in 2018 (Schleicher, 2019) to 366 in 2022 (OECD, 2023). This data indicates that the majority of students are confined to performing basic arithmetic operations, with less than 1% capable of navigating complex problems requiring sustained higher-order reasoning. A parallel deficit is observed in computational thinking (CT) performance, as documented by the national Bebras Challenge results, which reveal stagnation in algorithmic logic and pattern recognition abilities among Indonesian learners (Bebras Indonesia, 2023, 2024). This dual deficit suggests that students are struggling not merely with the production of correct answers but, fundamentally, with engaging in structured reasoning and algorithmic thinking.

This macro-level crisis is further exacerbated by micro-level pedagogical and assessment practices that fail to adequately capture or cultivate students' cognitive processes (Novitasari et al., 2024; Villarreal-Lozano et al., 2022). Preliminary findings from Focus Group Discussions (FGDs) conducted with mathematics teachers as part of this study reveal a striking assessment void: 85% of teachers reported not conducting in-depth analyses of students' reasoning pathways, relying almost exclusively on dichotomous scoring (right/wrong) that obscures underlying cognitive processes. Furthermore, 87% of teachers acknowledged rarely providing feedback beyond final scores, and existing rubrics are ill-equipped to account for variations in cognitive complexity or partial understanding. These findings confirm that current assessment approaches are predominantly outcome-oriented and methodologically insufficient to capture the multidimensional nature of students' thinking, thereby masking the specific cognitive bottlenecks that require intervention.

Previous scholarly efforts have attempted to address mathematical reasoning and computational thinking, yet they have largely operated along parallel, disconnected tracks. A substantial body of research on mathematical reasoning has focused on developing context-based assessment instruments and promoting problem-solving heuristics (Herbert et al., 2022; Özaydin & Arslan, 2022). Concurrently, a separate and robust field of inquiry has emerged around computational thinking, emphasizing its definitional frameworks and its role as a transversal twenty-first-century skill (Anđić et al., 2023; Kadijevich, 2023; Pourdavood et al., 2020). While some studies have begun to explore the conceptual relationship or instructional integration between CT and mathematics (Labusch et al., 2019; Liu, 2023; Yadav et al., 2016), the literature remains characterized by a fragmented approach. Existing research primarily addresses correlation or pedagogical alignment rather than developing an integrated, psychometrically rigorous assessment framework situated explicitly within the mathematics education context.

This fragmentation reveals a critical and underexplored limitation in current assessment practices and research methodologies. There exists a conspicuous absence of a standardized assessment instrument that explicitly operationalizes computational thinking as an integral, functional component of mathematical reasoning tasks. Prior approaches have treated CT as a separate variable or an external skill applied to mathematics, rather than a cognitive dimension that emerges within mathematical reasoning activities (Gane, 2020; Román-González & Pérez-González, 2024; Weintrop et al., 2021). This conceptual separation produces an epistemological bias in measurement, whereby the unique synergy between algorithmic decomposition and mathematical abstraction remains unmeasured and, consequently, pedagogically invisible. Moreover, prior studies have relied predominantly on correlational or descriptive designs that treat computational thinking and mathematical reasoning as separate observed variables, thereby failing to model their latent interaction within a unified psychometric framework (Ferreira et al., 2023; Suparman et al., 2025; Ye et al., 2023) This methodological limitation constrains both theoretical advancement and practical assessment utility. Critically, the application of advanced multidimensional measurement models—specifically Multidimensional Item Response Theory (MIRT)—to capture the complex latent interplay between these competencies remains largely unexplored within the Indonesian educational assessment landscape (Jewsbury & van Rijn, 2020; Lai & Ellefson, 2023; Otaya et al., 2025). Addressing this gap is therefore essential to move beyond fragmented assessment paradigms and toward an integrated, psychometrically robust representation of students' higher-order thinking.

To address this critical gap, this study develops the SMART-CT (Mathematical Reasoning–Based Computational Thinking) instrument. The novelty of this study is twofold and warrants explicit articulation. First, conceptually, the SMART-CT instrument moves decisively beyond the fragmented approaches that have long characterized prior scholarship. Rather than treating computational thinking as a discrete skill or an external tool applied to mathematical tasks, this study positions CT as a cognitive substrate intrinsically embedded within mathematical reasoning itself. This integration is operationalized through a multiple-choice multiple-answer format with polytomous scoring, a design choice that enables the capture of partial knowledge and process-oriented reasoning, cognitive nuances systematically obscured by conventional dichotomous assessments. Second, methodologically, this study is among the first in the Indonesian educational assessment landscape, and one of relatively few internationally, to apply a multidimensional GRM-MIRT framework to model the interaction between computational thinking and mathematical reasoning as a unified, hierarchical latent construct. Unlike unidimensional IRT or classical test theory approaches that treat these competencies in isolation, the MIRT framework employed herein enables the simultaneous estimation of item parameters and latent traits across multiple dimensions. This yields a more accurate and psychometrically nuanced representation of how algorithmic logic mediates mathematical abstraction, thereby offering a methodological template that advances both the precision and the theoretical depth of assessment practices in this domain.

Grounded in this identified gap and the proposed novelty, this study is explicitly directed toward: (1) developing and empirically validating the SMART-CT instrument as an integrated assessment framework using a robust Confirmatory Factor Analysis (CFA) and MIRT approach; and (2) establishing a more integrative and accurate psychometric representation of students' higher-order cognitive abilities that captures the intersection of computational logic and mathematical reasoning. By achieving these objectives, this study is positioned not only to fill a critical void in assessment literature but also to provide a foundational framework for advancing assessment practices that are genuinely aligned with the cognitive demands of contemporary mathematics education. By pursuing these objectives, this study establishes a foundation for reconceptualizing how computational thinking is assessed not as an ancillary skill, but as an integral cognitive dimension within the core fabric of mathematics education.

B. Method

This study employed a Research and Development (R&D) framework integrated with a quantitative psychometric approach, combining iterative instrument development via the ADDIE model (Branch, 2009) with validation through second-order Confirmatory Factor Analysis (CFA) and Multidimensional Item Response Theory (MIRT). The research was conducted in five purposively selected senior secondary schools in Sleman Regency, Yogyakarta, Indonesia—a region chosen for its diverse institutional profiles ranging from urban schools with established Bebras Challenge participation to peri-urban schools with limited digital infrastructure, thereby ensuring adequate variance in students' prior exposure to computational thinking. Participants were Grade XI students, with a pilot test of 195 students used for initial item screening and a field test of 1,023 students—a figure exceeding the 500–750 respondents required for stable parameter estimation in polytomous multidimensional Graded Response Models.

Instrument development followed the ADDIE framework operationally. The Analysis phase triangulated curriculum mapping of matrix-domain outcomes from Phase F of the national curriculum, cross-referencing with the Bebras taxonomy to embed decomposition, pattern recognition, abstraction, and algorithmic thinking, and semi-structured interviews with five senior mathematics teachers to identify cognitive bottlenecks. The Design phase produced a Test Specification Blueprint and a three-category polytomous rubric: Score 0 (no correct options or only incorrect), Score 1 (partial credit: at least one correct option alongside incorrect or incomplete selections), and Score 2 (all correct options, no incorrect). During Development, an initial 12-item pool was validated by a seven-member panel comprising three educational measurement experts, two mathematics education scholars, and two senior mathematics teachers. Experts independently rated items on relevance, CT representation, and clarity using a four-point scale; all items achieved Aiken's V coefficients exceeding the 0.80

threshold. Implementation involved supervised paper-based administration with a uniform 90-minute time allocation, guided by a proctor manual specifying verbatim instructions. Following a pilot phase CFA-based refinement, the instrument was administered to the field sample under identical protocols.

Data analysis proceeded sequentially. Content validity was confirmed via Aiken's V (threshold ≥0.80). Construct validity was assessed using second-order CFA with a Robust Maximum Likelihood estimator; model fit was evaluated against CFI/TLI ≥0.95, RMSEA ≤0.06, and SRMR ≤0.08. Internal consistency was measured with Cronbach's alpha (threshold ≥0.70). A compensatory MIRT model with Graded Response Model parameterization estimated item discrimination and threshold parameters, and the Test Information Function assessed measurement precision. All analyses were performed in R version 4.3.1 using the *lavaan*, *semTools*, and *mirt* packages.

Ethical approval for this study was obtained from the Research Ethics Committee of Yogyakarta State University under Approval Number KEPUNY260100822. All participating students and their legal guardians provided written informed consent prior to data collection. Participation was voluntary, and students were informed of their right to withdraw at any time without academic penalty. Data were anonymized using unique alphanumeric codes to delink personal identifiers from response matrices, and all data storage and transfer adhered to institutional data privacy protocols.

C. Results and Discussion

This section presents the results and discussion in a structured, integrated manner. The findings are reported first based on empirical analyses, including content validity, construct validity, reliability, and MIRT modeling, followed by their interpretation in relation to relevant theories and prior studies. This organization ensures a clear and coherent understanding of the SMART-CT instrument's psychometric properties.

1. Results

The SMART-CT instrument was initially developed with 12 items in a multiple-choice, multiple-answer (MCMA) format, designed to reflect the integration of indicators for mathematical reasoning and computational thinking. Each item simultaneously engages aspects of reasoning—such as constructing arguments and drawing conclusions—alongside computational thinking processes, including decomposition, pattern recognition, abstraction, and algorithmic reasoning. The items were distributed across matrix-related topics, including matrix operations, determinants, inverses, and contextual problem-solving, ensuring that the instrument captures a range of cognitive processes across different mathematical contexts. The systematic alignment between cognitive indicators, mathematical reasoning components, and computational thinking components is presented in Table 1.

Table 1. SMART-CT instrument indicators

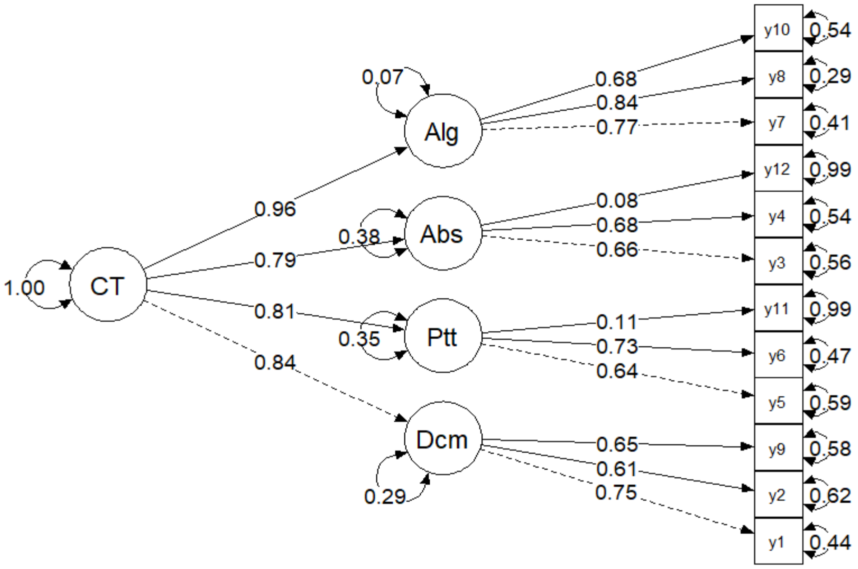
No.	Cognitive Indicators	Mathematical Reasoning Indicators	Computational Thinking Indicators
1.	Explaining matrices and matrix equality using contextual problems.	Constructing mathematical proofs	Pattern Recognition
		Drawing valid conclusions	Decomposition
2.	Performing matrix operations, including addition, subtraction, scalar multiplication, matrix multiplication, and matrix transposition.	Formulating mathematical arguments	Abstraction
		Evaluating mathematical arguments	Algorithm
3.	Analyzing the properties of determinants and inverses of 2 × 2 and 3 × 3 matrices.	Understanding mathematical problems	Decomposition
		Evaluating mathematical arguments	Algorithm
		Drawing valid conclusions.	Pattern Recognition
4.	Solving contextual problems related to matrices and their operations.	Proving mathematical arguments	Abstraction
		Evaluating mathematical arguments	Algorithm
5.	Solving problems related to determinants and inverses of 2 × 2 and 3 × 3 matrices.	Understanding mathematical problems	Pattern Recognition
		Formulating mathematical arguments	Decomposition
		Proving mathematical arguments	Abstraction

Content validity was evaluated using Aiken's V coefficient based on expert judgments from a seven-member panel. As reported in Table 2, the obtained Aiken's V values ranged from 0.8571 to 0.9286. All values exceeded the predetermined threshold of $V \geq 0.82$, indicating strong inter-rater agreement regarding item relevance and construct representation. Consequently, all 12 items were retained for subsequent empirical testing.

Table 2. Aiken's V Indices of SMART-CT Instrument Items

Item	s_1	s_2	s_3	s_4	s_5	s_6	s_7	$\sum s$	V	V_{table}	**Decision**
y1	4	3	4	4	4	4	3	26	0.9286	0.82	Valid
y2	3	4	4	4	4	3	4	26	0.9286	0.82	Valid
y3	3	3	4	3	4	4	4	25	0.8929	0.82	Valid
y4	3	3	3	4	4	4	4	25	0.8929	0.82	Valid
y5	4	3	4	3	3	4	4	25	0.8929	0.82	Valid
y6	3	4	2	4	4	4	3	24	0.8571	0.82	Valid
y7	4	3	4	4	4	3	4	26	0.9286	0.82	Valid
y8	3	4	4	4	3	4	4	26	0.9286	0.82	Valid
y9	3	3	4	4	4	3	4	25	0.8929	0.82	Valid
y10	4	4	3	4	3	4	3	25	0.8929	0.82	Valid
y11	4	3	4	4	4	3	4	26	0.9286	0.82	Valid
y12	3	4	3	3	4	4	3	24	0.8571	0.82	Valid

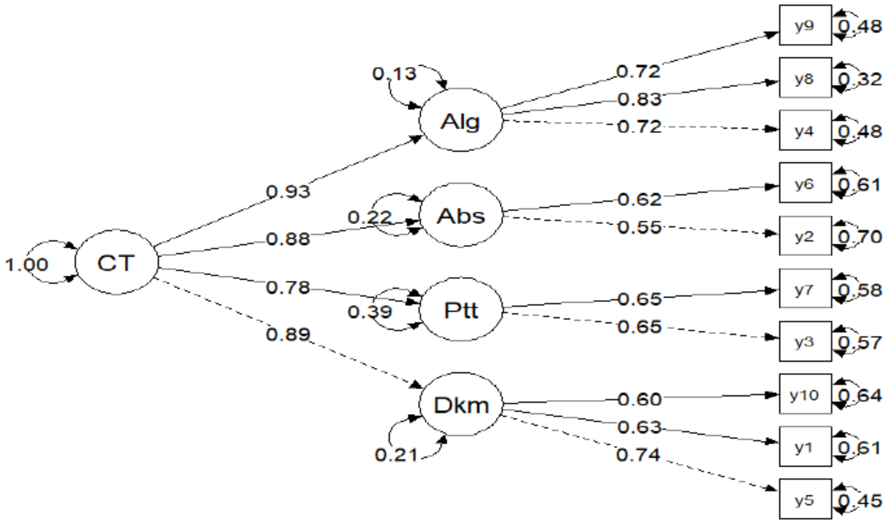
Construct validity was examined using second-order Confirmatory Factor Analysis (CFA) based on pilot test data from 195 students. The initial measurement model, comprising 12 items loading onto four first-order factors (Decomposition, Pattern Recognition, Abstraction, and Algorithm) and one second-order factor (Computational Thinking), was evaluated. As illustrated in Figure 1, two items (y11 and y12) exhibited standardized factor loadings below the acceptable threshold of 0.40 (λ = 0.31 and λ = 0.28, respectively). These items were therefore removed from the model.



*CT: Computational Thinking; Dcm: Decomposition; Ptt: Pattern Recognition; Abs: Abstraction; Alg: Algorithm

Figure 1. Second-order CFA of the initial measurement construct

Following the removal of items y11 and y12, the revised 10-item model was re-estimated. All remaining items demonstrated standardized factor loadings ranging from 0.52 to 0.84, exceeding the acceptable criterion. The fit indices for the revised model are summarized in Table 3. The chi-square test yielded a non-significant p-value of 0.155. Incremental fit indices were CFI = 0.986 and TLI = 0.980, while absolute fit indices were RMSEA = 0.036 (90% CI [0.000, 0.061]) and SRMR = 0.038. Collectively, these values satisfy the criteria for good model fit (CFI/TLI ≥ 0.95; RMSEA ≤ 0.06; SRMR ≤ 0.08).



*CT: Computational Thinking; Dcm: Decomposition; Ptt: Pattern Recognition; Abs: Abstraction; Alg: Algorithm

Figure 2. Second-Order CFA of the Revised Measurement Construct

Table 3. Fit Indices of the initial second-order CFA measurement construct

Factors	Value	Recommendation	Decision
P-Value (Chi-square)	0.155	> 0.05	Fit
Comparative Fit Index (CFI)	0.986	≥ 0.90	Fit
Tucker-Lewis Index (TLI)	0.980	≥ 0.90	Fit
Root Mean Square Error of Approximation (RMSEA)	0.036	≤ 0.08	Fit
Standardized Root Mean Square Residual (SRMR)	0.038	≤ 0.08	Fit

Internal consistency reliability was assessed using Cronbach's alpha coefficient based on the field test data (N = 1,023). The analysis yielded an alpha coefficient of α = 0.861, with a mean inter-item correlation of 0.38. The scale mean was 1.19 (SD = 0.500). This value exceeds the conventional threshold of α ≥ 0.70, indicating a high degree of internal consistency among the 10 retained items.

Table 4. Reliability estimates of the SMART-CT instrument

	Mean	SD	Cronbach's Alpha
Scale	1.19	0.500	0.861

Descriptive statistics of student scores on the 10-item SMART-CT instrument are presented in Table 5. The observed scores ranged from 0 to 19, with a mean of 10.52 and a standard deviation of 4.58. The distribution exhibited slight negative skewness (-0.31) and kurtosis (-0.47), indicating an approximate normal distribution.

Table 5. Descriptive statistics of test participants

Statistics	Value
Total respondents (N)	1,023
Minimum score	0
Maximum score	19
Mean	10.52
Standard deviation	4.58

Prior to Multidimensional Item Response Theory (MIRT) analysis, the dimensionality of the instrument was examined using parallel analysis and a scree plot. As shown in Figure 3, the eigenvalues exhibited a clear inflection point after the first factor (eigenvalue = 4.21), with subsequent factors showing eigenvalues below 1.0.

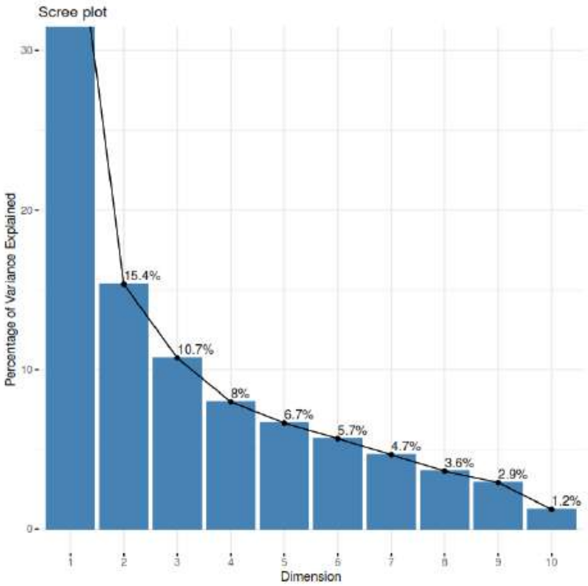


Figure 3. Scree Plot of the dimensionality test

Item parameters were estimated using the Graded Response Model (GRM) within a compensatory MIRT framework. The discrimination parameters (a) and threshold parameters (b_1, b_2) for each of the 10 items are reported in Table 6. Discrimination parameters ranged from a = 0.77 (Item y6) to a = 2.64 (Item y10), indicating variation in the capacity of items to differentiate among students. The threshold parameters b_1 (the point at which a student has a 50% probability of scoring ≥ 1) ranged from -4.75 (Item y4) to 0.24 (Item y9). The threshold parameters b_2 (the point at which a student has a 50% probability of scoring = 2) ranged from -1.18 (Item y4) to 1.52 (Item y9).

Based on these threshold values, items were categorized into effective ability zones. Item y4 operates within a very low ability range, items y1 and y2 function at low ability levels, items y3, y5, y6, y7, and y10 operate within the low-to-moderate range, item y8 functions at moderate-to-high levels, and item y9 targets high ability levels.

Table 6. GRM (Polytomous) Item Parameters with Effective Ability Zones

Item	a	b_1	b_2	Effective θ Zone
y1	1.56	−2.38	−0.29	Low
y2	1.56	−1.81	−0.20	Low
y3	1.98	−0.06	1.19	Moderate
y4	0.86	−4.75	−1.18	Very Low
y5	0.99	−1.85	0.68	Low–Moderate
y6	0.77	−1.94	0.27	Low–Moderate
y7	0.86	−1.80	0.47	Low–Moderate
y8	1.74	−0.48	1.38	Moderate–High
y9	2.06	0.24	1.52	High
y10	2.64	−0.36	0.55	Moderate

Item Characteristic Curves (ICCs) for the 10 items are displayed in Figure 4. Each panel illustrates the probability of selecting a given response category (Score 0, 1, or 2) as a function of the latent trait (θ). The curves exhibit the expected monotonic ordering: as θ increases, the probability of scoring in Category 0 decreases, while the probabilities for Categories 1 and 2 increase sequentially.

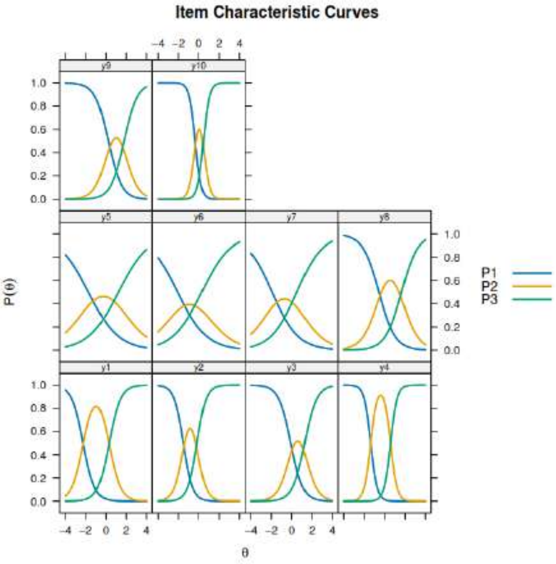


Figure 4. Item characteristics curves

Measurement precision was evaluated using the Test Information Function (TIF) and its associated Standard Error (SE), as depicted in Figure 5. The TIF curve indicates that the SMART-CT instrument provides maximum information (I(θ) > 4.0) within the ability range of θ = -0.5 to +1.0, corresponding to low-to-moderate ability levels. The associated standard error reaches its minimum (SE < 0.50) within this same interval. Beyond θ = +1.5, the information function declines, and the standard error increases, indicating reduced measurement precision for students at higher ability levels.

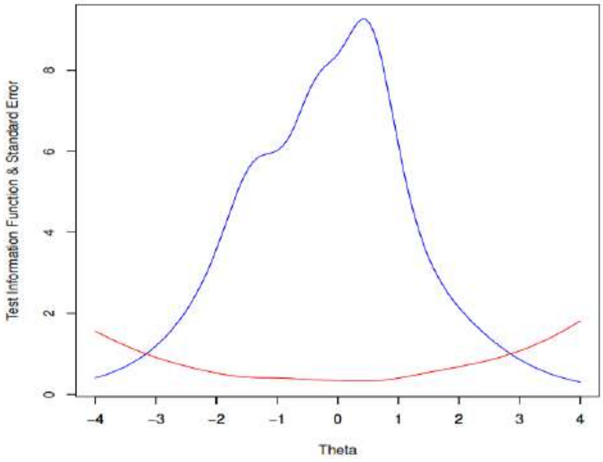


Figure 5. Test Information Function & Standard Error

2. Discussion

This study positions the SMART-CT instrument as a significant advancement in the assessment of higher-order thinking in mathematics education by demonstrating that computational thinking can be operationalized as an integrated dimension within mathematical reasoning. The central meaning of the findings is that computational thinking should not be treated as an external or supplementary competency, but rather as a cognitive process that is structurally embedded in mathematical reasoning activities. This position shifts the conceptualization of higher-order thinking from a fragmented framework to a unified cognitive system, in which algorithmic logic and mathematical abstraction operate in a coordinated manner during problem-solving (Dagiene & Dolgopolovas, 2022; Dagienė & Sentance, 2016; Shute et al., 2017; Jeannotte & Kieran, 2017; Pradana et al., 2025).

Empirical evidence supporting this position is consistently demonstrated across multiple analytical stages. The instrument shows strong psychometric quality, as indicated by content validity indices (Aiken's V) exceeding 0.85, a well-fitting second-order CFA model (CFI = 0.986; RMSEA = 0.036), high internal consistency ($\alpha = 0.861$), and stable parameter estimates derived from MIRT analysis. In addition, the instrument provides optimal measurement precision within the ability range of $\theta = -0.5$ to $+1.0$. These results collectively confirm that SMART-CT is not only statistically robust but also capable of capturing meaningful variation in students' cognitive processes within mathematical reasoning contexts, supporting the use of multidimensional approaches for assessing complex cognitive constructs (Immekus et al., 2019; Tan, 2024; Assegaf et al., 2025).

The significance of these findings lies in explaining why this integration emerges empirically. The confirmation of a second-order factor structure indicates that decomposition, pattern recognition, abstraction, and algorithmic thinking do not function independently in mathematical problem solving. Instead, they operate as interdependent processes that collectively shape how learners interpret and solve mathematical problems. The strong loadings of first-order factors onto the higher-order construct (ranging from 0.68 to 0.91) suggest a high degree of cognitive interdependence, indicating that these processes reinforce one another during reasoning activities. This integration can be interpreted as a form of cognitive synergy, in which decomposition supports abstraction, pattern recognition guides inference, and algorithmic reasoning structures solution pathways, reflecting the interconnected nature of computational thinking dimensions described in previous studies (Dagiene & Dolgopolovas, 2022; Dagienė & Sentance, 2016; Shute et al., 2017). Furthermore, the prominent role of algorithmic thinking within this integrated structure is consistent with arguments emphasizing its importance in linking procedural and conceptual aspects of mathematical cognition (Kadijevich, 2023; Pradana et al., 2025).

The MIRT findings further deepen this interpretation by revealing how the integrated dimensions of computational thinking function within the measurement process. The variation in discrimination parameters (a), ranging from moderate (0.77) to very high (2.64), indicates that certain items are more sensitive in differentiating students' abilities across the measured construct. Notably, items with the highest discrimination parameters (y9, y10) correspond to tasks requiring both abstraction and algorithmic reasoning. This finding suggests that assessment tasks incorporating these dimensions provide stronger diagnostic information for distinguishing among students with different levels of performance. While the present study does not establish causal relationships among computational thinking dimensions, the results indicate that abstraction and algorithmic reasoning contribute substantially to the measurement of higher-order thinking within the SMART-CT framework. Such findings provide empirical support for the argument that algorithmic thinking plays an important role in connecting procedural knowledge with conceptual understanding in mathematical problem solving (Kadijevich, 2023; Pradana et al., 2025; Ratnawulan et al., 2025).

These findings both align with and extend existing theoretical perspectives. Previous research has conceptualized computational thinking as a multidimensional construct composed of interconnected processes (Dagiene & Dolgopolovas, 2022; Dagienė & Sentance, 2016; Shute et al., 2017), while mathematical reasoning has been described as a multifaceted cognitive activity involving argumentation, inference, abstraction, and problem solving (Jeannotte & Kieran, 2017; Pradana et al., 2025). The present study supports these perspectives by demonstrating that these dimensions converge within a unified latent structure when operationalized in authentic mathematical tasks. More importantly, the findings extend prior work by showing that the integration between computational thinking and mathematical reasoning is not only theoretically plausible but can also be empirically represented and measured through a multidimensional psychometric framework. This result reinforces arguments that complex cognitive competencies are more appropriately modeled through interconnected latent dimensions rather than isolated constructs, particularly when the objective is to capture the multidimensional nature of higher-order thinking (Immekus et al., 2019; Tan, 2024).

At the same time, the findings raise important questions regarding dominant assessment approaches that continue to treat computational thinking and mathematical reasoning as separate constructs. Existing instruments frequently assess computational thinking within decontextualized environments, particularly in programming and computer science settings (Román-González et al., 2018; Dimov, 2024), whereas assessments of mathematical reasoning often emphasize argumentation, inference, and problem-solving outcomes without explicitly incorporating computational thinking dimensions (Jeannotte & Kieran, 2017; Safkolam et al., 2022; Pradana et al., 2025). While these approaches have contributed substantially to their respective fields, they may provide only a partial representation of students' higher-order thinking processes. The present findings suggest that when computational thinking dimensions are embedded within authentic mathematical tasks, they can be represented within a unified latent structure. This indicates that the interaction between computational thinking and mathematical reasoning can be meaningfully captured through integrated assessment frameworks. Furthermore, the application of multidimensional modelling addresses limitations associated with traditional assessment approaches that rely on unidimensional constructs and dichotomous scoring, which may not adequately represent partial understanding, intermediate reasoning processes, or variations in cognitive performance (Immekus et al., 2019; Tan, 2024; Hamdi et al., 2026).

Within this context, the contribution of the present study becomes increasingly evident. Conceptually, the findings provide empirical support for an integrated assessment framework in which computational thinking is represented as a meaningful component of mathematical reasoning. Rather than positioning these domains as independent competencies, the results suggest that they can be operationalized and measured within a coherent multidimensional structure. This perspective contributes to the growing body of literature advocating a closer integration between computational thinking and mathematics education (Dagiene & Dolgopolovas, 2022; Dagienė & Sentance, 2016; Shute et al., 2017). Methodologically, the study contributes an assessment framework that combines MCMA item formats, polytomous scoring, second-order CFA, and MIRT modeling. The integration of these methodological components enables a more nuanced representation of students' cognitive performance by capturing variations in reasoning quality rather than relying solely on dichotomous response outcomes. As a result, the study contributes not only a validated instrument but also a multidimensional approach for assessing complex cognitive constructs within mathematics education (Immekus et al., 2019; Tan, 2024).

The implications of these findings extend across multiple levels of educational practice. At the classroom level, the SMART-CT instrument provides educators with diagnostic information regarding students' performance across different dimensions of computational thinking embedded within mathematical reasoning. Such information may support more targeted instructional interventions and assist teachers in identifying specific areas requiring further development. At the curricular level, the findings reinforce ongoing efforts to integrate computational thinking into mathematics education, supporting the design of learning experiences that simultaneously foster reasoning, abstraction, pattern recognition, and algorithmic thinking (Dagiene & Dolgopolovas, 2022; Dagienė & Sentance, 2016; Shute et al., 2017). At the assessment level, the results highlight the value of multidimensional and polytomous approaches for evaluating higher-order thinking. By recognizing partial knowledge and different levels of reasoning complexity, such approaches provide a richer representation of student learning than conventional assessment systems that rely exclusively on correct-or-incorrect classifications (Immekus et al., 2019; Tan, 2024; Suarsana et al., 2025).

Beyond the Indonesian context, this study contributes to global discussions on assessment reform and twenty-first-century competencies. Educational frameworks such as the OECD Future of Education and Skills 2030 and the UNESCO AI Competency Framework emphasize the importance of assessing the integration of computational and domain-specific thinking to prepare learners for increasingly complex technological and knowledge-based environments (OECD, 2019; UNESCO, 2021; Balulu et al., 2025). The findings of this study contribute to this agenda by providing empirical evidence that computational thinking and mathematical reasoning can be represented and assessed within a unified multidimensional framework. By offering a psychometrically validated model that operationalizes the interaction between computational and disciplinary thinking in authentic mathematical contexts, the SMART-CT framework extends current efforts to integrate computational thinking into mathematics education (Ye et al., 2023; Ferreira et al., 2023; Weintrop et al., 2021).

The international significance of this contribution lies in its response to a persistent challenge in educational assessment. Although educational systems increasingly emphasize higher-order thinking, problem solving, and transferable competencies, assessment practices often continue to evaluate these abilities through fragmented and independent constructs (Aba-Oli et al., 2025; Aydin & Birgili, 2023). The present findings suggest that such separation may overlook important cognitive interactions. By demonstrating that computational thinking dimensions can be integrated within mathematical reasoning through a second-order multidimensional structure, this study offers a methodological contribution that may support future assessment development aimed at capturing complex cognitive performance more comprehensively (Immekus et al., 2019; Jewsbury & van Rijn, 2020; Lai & Ellefson, 2023).

Moreover, the study demonstrates that methodological innovation in educational assessment can emerge from a Global South context and contribute to international scholarship. Rather than merely adapting existing frameworks, the SMART-CT model provides empirical evidence of how computational thinking and mathematical reasoning can be integrated within a coherent assessment system, thereby enriching global discussions on higher-order thinking and educational measurement (Ye et al., 2023; Yadav et al., 2016; Weintrop et al., 2021). This contribution is also relevant to Sustainable Development Goal 4 (Quality Education), which emphasizes not only access to education but also the improvement of learning quality and assessment practices. By applying a multidimensional GRM-MIRT approach, the study offers a framework for capturing complex cognitive competencies that remain difficult to represent in large-scale assessments such as PISA and TIMSS, thereby contributing to ongoing international efforts to strengthen educational measurement and learning outcomes (OECD, 2019; UNESCO, 2021; Schleicher, 2019; OECD, 2023; Immekus et al., 2019; Tan, 2024).

Despite these contributions, several limitations must be acknowledged. First, the SMART-CT instrument was developed within a specific mathematical domain, namely matrices, which limits the generalizability of the findings to other mathematical topics such as algebra, geometry, or data analysis. Second, the cross-sectional design restricts the ability to examine how the relationship between computational thinking and mathematical reasoning evolves over time. Third, although the sample size is large and heterogeneous within the Yogyakarta region, it may not fully represent the broader Indonesian student population, particularly regarding access to technology and prior exposure to computational thinking activities. Fourth, the reduced measurement precision at higher ability levels limits the instrument's effectiveness in distinguishing high-achieving students. These limitations indicate that while the SMART-CT framework provides a strong foundation for integrated assessment, further refinement is necessary to enhance its scope, sensitivity, and applicability across diverse educational contexts.

D. Conclusion

This study demonstrates that computational thinking can be validly operationalized and assessed as an integral dimension of mathematical reasoning through the SMART-CT framework. The findings indicate that the instrument possesses strong psychometric qualities, evidenced by satisfactory content validity, a well-fitting second-order factor structure, high internal consistency ($\alpha = 0.861$), and stable item parameter estimates derived from the multidimensional Graded Response Model. In addition, the instrument provides optimal measurement precision within the low-to-moderate ability range, confirming its suitability for diagnostic and formative assessment purposes. These findings support the view that computational thinking is not merely an additional skill associated with mathematics learning but a cognitive dimension that functions within mathematical reasoning processes.

This study contributes to the growing body of research seeking to integrate computational thinking and mathematics education. Conceptually, the findings provide empirical evidence that decomposition, pattern recognition, abstraction, and algorithmic thinking can be represented within a unified latent structure of mathematical reasoning. Methodologically, the study offers an assessment framework that combines MCMA item formats, polytomous scoring, second-order CFA, and MIRT modelling, enabling a more comprehensive representation of students' higher-order thinking than conventional dichotomous assessment approaches. The findings also contribute to ongoing international efforts to develop assessment systems capable of capturing complex cognitive competencies required in contemporary education.

Building on the limitations identified in this study, future research is encouraged to validate the SMART-CT framework across broader mathematical domains, examine its applicability in more diverse educational settings, employ longitudinal designs to investigate the development of computational thinking and mathematical reasoning over time, and refine higher-difficulty items to improve measurement accuracy for high-achieving learners.

This study provides evidence that computational thinking and mathematical reasoning can be assessed within an integrated multidimensional framework. By offering a psychometrically validated assessment model, the SMART-CT framework contributes to the advancement of more authentic, process-oriented, and cognitively meaningful assessment practices in mathematics education.

Acknowledgments

The authors would like to express sincere gratitude to the Indonesia Endowment Fund for Education (LPDP), Ministry of Finance of the Republic of Indonesia, for the financial support that made this research possible.

Declaration of Competing Interest

The authors declare that they have no known competing financial or non-financial interests that could have appeared to influence the work reported in this paper.

Declaration of Generative AI

During the preparation of this manuscript, the authors used Gemini to improve the clarity and readability of the text. Output generated by the tool was carefully reviewed and edited by the authors, who take full responsibility for the content of this article. All substantive intellectual contributions, including conceptualization, analysis, interpretation of data, and final decisions regarding content, arguments, and conclusions, were carried out solely by the authors. The authors take full responsibility for the integrity, originality, and academic quality of this article.

Bibliography

Aba-Oli, Z., Koyas, K., & Husen, A. (2025). Higher-Order Thinking Skills-Oriented Problem-Based Learning Interventions in Mathematics: A Systematic Literature Review. *School Science and Mathematics*, *125*(3), 214–231. https://doi.org/10.1111/ssm.12676

Anđić, B., Mumcu, F., Tejera, M., Schmidthaler, E., & Lavicza, Z. (2023). Unplugging Math: Integrating Computational Thinking into Mathematics Education Through Poly-Universe. *Advances in Sustainability Science and Technology*, 247–263. https://doi.org/10.1007/978-981-99-5540-4_15

Assegaf, A. R., Khaulasari, H., Thohir, M., Ali, A., & Pangastuti, R. (2025). The Effectiveness of Peaceful Education Learning Strategies for Students at Emergency Madrasas Following Natural Disasters in Sidoarjo. *Jurnal Pendidikan Agama Islam*, *22*(1), 57–80. https://doi.org/10.14421/jpai.v22i1.8369

Aydin, U., & Birgili, B. (2023). Assessing Mathematical Higher-Order Thinking Skills: An Analysis of Turkish University Entrance Examinations. *Educational Assessment*, *28*(3), 190–209. https://doi.org/10.1080/10627197.2023.2202311

Balulu, N., Isnawati, I., Supriyadi, S., Jatmiko, B., Ahmad, Z., Limatahu, I., & Takda, A. (2025). Critical Thinking Development in Physics Education Through The Implementation of the BW-ExPort Model at Madrasah Aliyah. *Jurnal Pendidikan Islam*, *12*(1), 35–47. https://doi.org/10.15575/jpi.v12i1.49792

Bebras Indonesia. (2023). *Pengumuman Hasil Bebras Indonesia Challenge 2023*. https://bebras.or.id/v3/pengumuman-hasil-bebras-indonesia-challenge-2023/

Bebras Indonesia. (2024). *Pengumuman Hasil Bebras Indonesia Challenge 2024*. https://bebras.or.id/v3/pengumuman-hasil-bebras-indonesia-challenge-2024/

Branch, R. M. (2009). *Instructional Design: The ADDIE Approach*. Springer US. https://doi.org/10.1007/978-0-387-09506-6

Dagiene, V., & Dolgopolovas, V. (2022). Short Tasks for Scaffolding Computational Thinking by the Global Bebras Challenge. *Mathematics*, *10*(17), 3194. https://doi.org/10.3390/math10173194

Dagienė, V., & Sentance, S. (2016). It's Computational Thinking! Bebras Tasks in the Curriculum. In A. Brodnik & F. Tort (Eds.), *Informatics in Schools: Improvement of Informatics Knowledge and Perception* (Vol. 9973, pp. 28–39). Springer International Publishing. https://doi.org/10.1007/978-3-319-46747-4_3

Dimov, P. S. (2024). An Online Tool for Misinformation Detection: Enhancing Media Literacy and Critical Thinking Skills. *Online Learning In Educational Research (OLER)*, *4*(2), 61–69. https://doi.org/10.58524/oler.v4i2.413

Fauzi, M. R., Hamami, T., & Kim, H. J. (2024). Islamic Religious Education Curriculum Innovation: Fethullah Gülen's Perspective. *Jurnal Pendidikan Agama Islam*, *21*(1), 186–200. https://doi.org/10.14421/jpai.v21i1.7089

Ferreira, T. D., Santos, J. S., Medeiros, R. de A., Andrade, W. L., Brunet, J., & Melo, M. R. A. (2023). Exploring the Relationship of Mathematical Reasoning and Computational Thinking. *Anais Do XXIX Workshop de Informática Na Escola (WIE 2023)*.

Gane, B. (2020). Developing Computational Thinking Assessments from Learning Trajectories: Design Approach and Preliminary Validity Evidence. *Proceedings of the 2020 AERA Annual Meeting*. https://doi.org/10.3102/1582248

Gunawan, M. A., Amalia, F., Setiawan, A., & Ab Ghani, H. H. (2025). Critical Thinking in Math: 10th-Grade Analysis using Cognitive Diagnostic Modeling. *REID (Research and Evaluation in Education)*, *11*(1), 89–100. https://doi.org/10.21831/reid.v11i1.88074

Hamdi, S., Yuliana, L., Oktarina, A. D., Hidayati, K., Arliani, E., & Mz, N. M. (2026). Authentic Mathematics Assessment Using an Integrated Deep Learning and Adiwiyata Testlet Model for Elementary Schools and Madrasah Ibtidaiyah. *Nazhruna: Jurnal Pendidikan Islam*, *9*(1), 187–206. https://doi.org/10.31538/nzh.v9i1.418

Herbert, S., Vale, C., White, P., & Bragg, L. A. (2022). Engagement with a Formative Assessment Rubric: A Case of Mathematical Reasoning. *International Journal of Educational Research*, *111*, 101899. https://doi.org/10.1016/j.ijer.2021.101899

Immekus, J. C., Snyder, K. E., & Ralston, P. A. (2019). Multidimensional Item Response Theory for Factor Structure Assessment in Educational Psychology Research. *Frontiers in Education*, *4*, 434763. https://doi.org/10.3389/feduc.2019.00045/text

Jailani, J., Retnawati, H., Rafi, I., Mahmudi, A., Arliani, E., Zulnaidi, H., Abd Hamid, H. S., & Prayitno, H. J. (2023). A Phenomenological Study of Challenges that Prospective Mathematics Teachers Face in Developing Mathematical Problems that Require Higher-Order Thinking Skills. *Eurasia Journal of Mathematics, Science and Technology Education*, *19*(10), em2339. https://doi.org/10.29333/ejmste/13631

Jeannotte, D., & Kieran, C. (2017). A Conceptual Model of Mathematical Reasoning for School Mathematics. *Educational Studies in Mathematics*, *96*(1), 1–16. https://doi.org/10.1007/s10649-017-9761-8

Jewsbury, P. A., & van Rijn, P. W. (2020). IRT and MIRT Models for Item Parameter Estimation with Multidimensional Multistage Tests. *Journal of Educational and Behavioral Statistics*, *45*(4), 383–402. https://doi.org/10.3102/1076998619881790

Kadijevich, D. M. (2023). Computational/Algorithmic Thinking in School Mathematics. *European Congress of Mathematics*, 749–769. https://doi.org/10.4171/8ecm/40

Labusch, A., Eickelmann, B., & Vennemann, M. (2019). Computational Thinking Processes and Their Congruence with Problem-Sol ving and Information Processing. In *Computational Thinking Education* (pp. 65–78). Springer Singapore. https://doi.org/10.1007/978-981-13-6528-7_5

Lai, R. P. Y., & Ellefson, M. R. (2023). How Multidimensional is Computational Thinking Competency? A Bi-Factor Model of the Computational Thinking Challenge. *Journal of Educational Computing Research*, *61*(2), 259–282. https://doi.org/10.1177/07356331221121052

Liu, Z. (2023). Measuring Computational Thinking Through Modeling Problem-Solving Processes in Game-Based Learning. *AERA 2023*. https://doi.org/10.3102/ip.23.2013871

Mujakir, M., Nurmalahayati, N., Safrijal, S., Salsabil, P., Fatma, E., & Zainuddin, Z. (2024). Efforts to Improve Scientific Literacy Capabilities in Indonesia: Systematic Literature Review. *Online Learning In Educational Research (OLER)*, *4*(1), 49–59. https://doi.org/10.58524/oler.v4i1.395

Novitasari, W., Herwin, H., Supartinah, S., Wulandari, P., & Budiharti, B. (2024). Number Sense Profile of Prospective Elementary School Teachers in Blended Mathematics Learning. *REID (Research and Evaluation in Education)*, *10*(1), 1–17. https://doi.org/10.21831/reid.v10i1.51394

OECD. (2019). *OECD Future of Education and Skills 2030*. https://www.oecd.org/content/dam/oecd/en/about/projects/edu/education-2040/1-1-learning-compass/OECD_Learning_Compass_2030_Concept_Note_Series.pdf

OECD. (2023). PISA 2022 Results (Volume I): The State of Learning and Equity in Education. In *PISA*. OECD. https://doi.org/10.1787/53f23881-en

Otaya, L. G., Yahiji, K., Rasimin, R., Rahmawati, R., Imtihan, N., & Muhamad, N. Bin. (2025). Assessment of Teachers' Ability to Develop HOTS based Test Items: a GRM Analysis. *Jurnal Ilmiah Peuradeun*, *13*(3), 1995–2018. https://doi.org/10.26811/peuradeun.v13i3.1555

Özaydin, Z., & Arslan, Ç. (2022). Assessment of Mathematical Reasoning Competence in Accordance with PISA 2021 Mathematics Framework. *Kuramsal Eğitimbilim*, *15*(3), 453–474. https://doi.org/10.30831/akukeg.1027601

Pourdavood, B. R., McCarthy, K., & McCafferty, T. (2020). The Impact of Mental Computation on Children's Mathematical Communication, Problem Solving, Reasoning, and Algebraic Thinking. *Athens Journal of Education*, *7*(3), 241–254. https://doi.org/10.30958/aje.7-3-1

Pradana, K. C., Noer, S. H., & Sutiarso, S. (2025). Enhancing Critical Thinking in Mathematics through Android-Based Multimedia and PjBL-STEM. *Online Learning In Educational Research (OLER)*, *5*(1), 81–93. https://doi.org/10.58524/oler.v5i1.534

Ratnawulan, T., Ahmad, A. C., Effendi, Z. R., Syam, R. Z. A., & Achmad, W. (2025). Teachers' Strategies in Developing Scientific Literacy for Children with Special Needs in Inclusive Early Childhood Education. *Nazhruna: Jurnal Pendidikan Islam*, *8*(3), 682–698. https://doi.org/10.31538/nzh.v8i3.433

Román-González, M., & Pérez-González, J.-C. (2024). Computational Thinking Assessment: A Developmental Approach. In H. Abelson & S.-C. Kong (Eds.), *Computational Thinking Curricula in K–12* (pp. 121–142). The MIT Press. https://doi.org/10.7551/mitpress/14041.003.0009

Román-González, M., Pérez-González, J.-C., Moreno-León, J., & Robles, G. (2018). Can Computational Talent be Detected? Predictive Validity of the Computational Thinking Test. *International Journal of Child-Computer Interaction*, *18*, 47–58. https://doi.org/10.1016/j.ijcci.2018.06.004

Safkolam, R., Nuangchalerm, P., Ahmad Zaky El Islami, R., & Saleah, P. (2022). Students' Understanding of Nature of Science in Islamic Private School. *Jurnal Pendidikan Islam*, *9*(1), 1–14. https://doi.org/10.15575/jpi.v0i0.21308

Schleicher, A. (2019). PISA 2018: Insights and interpretations. *Oecd Publishing*.

Shute, V. J., Sun, C., & Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational Research Review*, *22*, 142–158. https://doi.org/10.1016/j.edurev.2017.09.003

Suarsana, I. M., Herman, T., Nurlaelah, E., Suryadi, D., Irianto, I., Jupri, A., Nandiyanto, A. B. D., & Ahzan, Z. N. (2025). Computational Thinking and Mathematical Problem Solving in Madrasah Tsanawiyah Students. *Jurnal Pendidikan Islam*, *12*(1), 75–88. https://doi.org/10.15575/jpi.v12i1.42091

Suparman, S., Juandi, D., Turmudi, T., Martadiputra, B. A. P., & Diana, N. (2025). *Assessing the Quality of Mathematics Test— Enumeration Rules to Measure Students' Computational Thinking Skills*. *28*(3), 412–428. https://doi.org/10.20961/Paedagogia.v28i3.109581

Tan, T. K. (2024). Evaluating Assessment via Item Response Theory Utilizing Information Function with R. *The Quantitative Methods for Psychology*, *20*(1), 33–49. https://doi.org/10.20982/tqmp.20.1.p033

UNESCO. (2021). *AI and Education: Guidance for Policy-Makers*. UNESCO. https://doi.org/10.54675/pcsp7350

Villarreal-Lozano, R. J., Morales-Martinez, G. E., Garcia-Collantes, A., & Barrientos-Amador, M. E. (2022). Cognitive Assessment of Motivation to Perform Classroom or Online Math Tasks among Engineering Students. *International Journal of Evaluation and Research in Education (IJERE)*, *11*(4), 1903. https://doi.org/10.11591/ijere.v11i4.22008

Weintrop, D., Rutstein, D., Bienkowski, M., & McGee, S. (2021). Assessment of Computational Thinking. *Computational Thinking in Education*, 90–111. https://doi.org/10.4324/9781003102991-6

Yadav, A., Hong, H., & Stephenson, C. (2016). Computational Thinking for All: Pedagogical Approaches to Embedding 21st Century Problem Solving in K-12 Classrooms. *TechTrends*, *60*(6), 565–568. https://doi.org/10.1007/s11528-016-0087-7

Ye, H., Liang, B., Ng, O. L., & Chai, C. S. (2023). Integration of Computational Thinking in K-12 Mathematics Education: A Systematic Review on CT-Based Mathematics Instruction and Student Learning. In *International Journal of STEM Education, 10*(1). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1186/s40594-023-00396-w

Zaeni, A., Kartono, Mulyono, & Sukestiarno, Y. L. (2025). Analysis of Higher Order Thinking Skills in a TPACK Based Flipped Classroom Supported by Dynamic Assessment. *Hipotenusa: Journal of Mathematical Society*, *7*(2), 271–281. https://doi.org/10.18326/hipotenusa.V7I2.5554