An-Najah University Journal for Research – B

Humanities

مجلــة جامعـــة النجـــاح للأبحــــاث – ب

العلوم الإنسانية

Development and Psychometric Validation of a Computational Thinking Self-Assessment Instrument for Pre-Service Mathematics Teachers

Edi Irawan[1,*], Moh. Khoridatul Huda[2], Retno Widyaningrum[1], Siti Hajaroh[3], & Ratni Purwasih[4]

Type: Full Article. Received: 27July 2025, Accepted: 20 February 2026

Accepted Manuscript, First Online

Abstract: Background: Computational Thinking (CT) is increasingly recognized as an essential skill in mathematics education, particularly for prospective teachers expected to solve complex problems in systematic ways. However, instruments to assess CT in the context of mathematics teacher preparation remain limited and underdeveloped. **Aim:** This study aimed to develop and psychometrically validate the Computational Thinking Self-Assessment (CTSA), a diagnostic tool designed to measure CT skills among prospective mathematics teachers based on four core dimensions: decomposition, abstraction, pattern recognition, and algorithmic thinking. **Method:** A total of 342 prospective mathematics teachers participated in this study. The CTSA underwent expert validation, followed by psychometric testing through a combination of Classical Test Theory and Item Response Theory approaches. Analyses included content validity indexing, Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA), and the Graded Response Model (GRM). **Results:** Content validity results showed excellent agreement (I-CVI and S-CVI = 1.00). EFA supported a four-factor structure explaining 43% of total variance. CFA confirmed acceptable model fit (CFI = 0.86, RMSEA = 0.073, SRMR = 0.062). Most items showed high discrimination and logical thresholds in IRT analysis. Reliability values were acceptable (α = 0.70–0.77; CR = 0.68–0.77), although AVE values remained below the ideal threshold (AVE < 0.50). **Conclusion:** The CTSA instrument provides a valid, reliable, and theoretically grounded measure of computational thinking in the context of mathematics teacher education. Future revisions are needed to improve convergent validity, yet the instrument holds strong potential as a reflective assessment tool in teacher preparation programs.

Keywords: Computational Thinking, Instrument Development, Pre-Service Mathematics Teachers, Self-Assessment, Validity and Reliability

تطوير أداة تقييم ذاتي للتفكير الحسابي والتحقق من صحتها النفسية لمعلمي الرياضيات قبل الخدمة

إيدي إيراوان[1*]، موه .خريدات الهدى[2]، ريتنو ويديانينجروم[1]، ستي هاجاروه[3]، راتني بورواسيه[4]

تاريخ التسليم: 2025/7/27، تاريخ القبول: 2026/2/20

الملخص: خلفية البحث: ،يتزايد الاعتراف بالتفكير الحسابي كمهارة أساسية في تعليم الرياضيات، وخاصةً للمعلمين المُحتملين الذين يُتوقع منهم حل المشكلات المعقدة بطرق منهجية .ومع ذلك لا تزال أدوات تقييم التفكير الحسابي في سياق إعداد معلمي الرياضيات محدودة وغير مُطوّرة. **الهدف** ،نفسياً (CTSA) هدفت هذه الدراسة إلى تطوير وإثبات صحة التقييم الذاتي للتفكير الحسابي وهي أداة تشخيصية مصممة لقياس مهارات التفكير الحسابي بين معلمي الرياضيات المحتملين بناءً على أربعة أبعاد أساسية :التحلل، والتجريد، والتعرف على الأنماط، والتفكير الخوارزمي. **المنهجية**: للتحقق من صحة الخبراء، تلتها اختبارات نفسية قياسية باستخدام مزيج من نظريتي الاختبار الكلاسيكي CTSA شارك في هذه الدراسة 342 معلمًا رياضيات محتملًا .خضعت أداة (GRM) ونموذج الاستجابة المتدرج ،(CFA) وتحليل العوامل التأكيدي ،(EFA) ونظرية الاستجابة للمفردة .وشملت التحليلات فهرسة صلاحية المحتوى، وتحليل العوامل الاستكشافي. **النتائج**: التوافق المقبول للنموذج CFA هيكلًا رباعي العوامل يُفسر 43 %من إجمالي التباين .أكد تحليل EFA دعم تحليل .(I-CVI وS-CVI = 1.00) أظهرت نتائج صلاحية المحتوى توافقًا ممتازًا (CFI = 0.86، RMSEA = 0.073، SRMR = 0.062). أظهرت معظم البنود تمييزًا وعتبات منطقية عالية في تحليل IRT. كانت قيم الموثوقية مقبولة (α = 0.70–0.77؛ CR = 0.68–0.77)، على الرغم من أن قيم AVE ظلت أقل من الحد المثالي (AVE < 0.50). **الاستنتاجات**: مقياسًا صحيحًا وموثوقًا ومبنيًا على أسس نظرية للتفكير الحسابي في سياق CTSA تُقدم أداة إعداد معلمي الرياضيات .هناك حاجة إلى تعديلات مستقبلية لتحسين الصلاح المتقارب، إلا أن الأداة تتمتع بإمكانيات قوية كأداة تقييم تأملية في برامج إعداد المعلمين.

الكلمات المفتاحية: التفكير الحسابي، وتطوير الأدوات، ومعلمو الرياضيات المستقبليون، والتقييم الذاتي، والصلاحية والموثوقية

1Study Program of Mathematics Education, State Islamic University of Kiai Ageng Muhammad Besari Ponorogo, Ponorogo, Indonesia
* Corresponding author: nawariide@uinponorogo.ac.id, orcid: 0000-0003-4600-7075
2 Data Science Department, Faculty of Mathematics and Natural Sciences, State University of Surabaya, Surabaya, Indonesia
3 Study Program of Mathematics Education, State Islamic University of Mataram, Mataram, Indonesia
4 Study Program of Mathematics Education, IKIP Siliwangi, Bandung, Indonesia

1 ،برنامج دراسة تعليم الرياضيات، جامعة كياي أجينج محمد بيساري الإسلامية الحكومية، بونوروغو إندونيسيا
* الباحث المراسل: nawariide@uinponorogo.ac.id
2 برنامج دراسة تعليم الرياضيات، جامعة رادين رحمت الإسلامية الحكومية، مالانج، إندونيسيا
3برنامج دراسة تعليم الرياضيات، جامعة ماتارام الإسلامية الحكومية، ماتارام، إندونيسيا
4باندونغ، إندونيسيا ،IKIP Siliwangi ،برنامج دراسة تعليم الرياضيات

Background

Computational Thinking (CT) has been recognised as an essential skill in 21st-century education and an integral part of the competencies needed in the digital age (Wing, 2006; Yadav, Stephenson, & Hong, 2017). Although originally developed in the context of computer science, CT has strong links with mathematics learning, particularly in developing logical, systematic, and strategic thinking skills. Several studies have shown that the integration of CT in mathematics learning can improve students' critical thinking and problem-solving skills (Angeli & Giannakos, 2020; Weintrop et al., 2015). Therefore, CT skills are very important competencies to be mastered by prospective mathematics teacher students to be able to design learning that is contextual and relevant to the demands of the times.

Conceptually, CT consists of four main components: decomposition, abstraction, pattern recognition, and algorithmic thinking. Decomposition refers to the ability to break down complex problems into smaller parts that are easier to analyse and solve (Beecher, 2017; Csizmadia, Standl, & Waite, 2019; Dong et al., 2019; Grover, 2022). Abstraction refers to the process of identifying important information and ignoring irrelevant information in constructing a solution (Beecher, 2017; Cansu & Cansu, 2019; Wing, 2017). The capacity to identify consistent structures, similarities, or recurring patterns within data or problem contexts—commonly referred to as pattern recognition—is a key cognitive process in computational thinking (Beyerer, Richter, & Nagel, 2018; Dong et al., 2019; Irawan, Rosjanuardi, & Prabawanto, 2024). Closely related to this is the development of algorithmic thinking, which emphasizes creating and applying ordered, logical steps to reach effective problem solutions (Csizmadia et al., 2019; Curzon, Bell, Waite, & Dorling, 2019; Dong et al., 2019; Irawan, Rosjanuardi, & Prabawanto, 2025). Both elements, when integrated with abstraction and decomposition, provide a conceptual framework essential for understanding and evaluating computational thinking. These dimensions also offer a structured basis for assessing CT competencies in educational research and practice.

One practical method for evaluating CT is through self-assessment. This approach plays a vital role in helping students develop metacognitive awareness, encouraging reflective thinking, and allowing them to identify their strengths and limitations in applying key concepts and skills (Mok, Lung, Cheng, Cheung, & Ng, 2006; Panadero, Jonsson, & Botella, 2017; Siles-González & Solano-Ruiz, 2016). Within the CT context, self-assessment tools offer prospective mathematics teachers an opportunity to gauge how well they understand and can apply CT principles, both in instructional settings and when tackling mathematical problems.

While several instruments have been designed to measure CT abilities, such as the Computational Thinking Scale (CTS) by Korkmaz et al. (2019; 2017), the digital literacy assessment by Tsai et al. (2021), and tools for early childhood contexts developed by Sung—many of these tools are either generalized or not tailored specifically for use in mathematics teacher education (MTE). Moreover, few studies have incorporated comprehensive psychometric techniques, such as exploratory factor analysis (EFA), confirmatory factor analysis (CFA), and Item Response Theory (IRT), in validating CT instruments within the context of teacher preparation.

To address these limitations, this study developed and validated a new self-assessment instrument, the Computational Thinking Self-Assessment (CTSA), specifically designed for pre-service mathematics teachers. Unlike existing instruments that emphasize programming performance, general digital literacy, or domain-neutral CT skills, the CTSA focuses on the reflective self-assessment of computational thinking processes as they are enacted in mathematical problem-solving and instructional planning. The instrument is grounded in four foundational dimensions of computational thinking—decomposition, abstraction, pattern recognition, and algorithmic thinking (collectively referred to as DAPRAL)—which are conceptually aligned with core mathematical reasoning practices. The validation process emphasized the establishment of content validity, construct validity, and internal reliability, intending to provide a psychometrically sound and pedagogically relevant instrument for use in mathematics teacher education.

This study, therefore, aimed to explore the following research questions: (1) To what extent does the CTSA demonstrate acceptable content and construct validity? (2) Is the factor structure of the CTSA consistent with the theoretical DAPRAL framework? (3) How reliable is the CTSA in measuring the CT skills of future mathematics teachers? The findings are expected to offer a meaningful contribution to the ongoing development of CT assessment tools in education and to support the professional growth of mathematics teacher candidates in integrating computational thinking into their instructional practice.

Method

Study Design: This study used a quantitative approach based on psychometric analysis to develop and validate the Computational Thinking Self-Assessment (CTSA) instrument. This approach is intended to test the validity and reliability of the developed instrument so that it can be used as a valid and reliable measuring instrument for assessing the CT skills of prospective mathematics teachers. The research was conducted in five stages, as shown in Figure 1.

Figure (1). Research Stages

Figure 1 illustrates the five stages of the CTSA instrument development. The preparation of the instrument began with exploring various literature related to the concept of CT, followed by the preparation of the instrument, content validity analysis, instrument testing, and validity and reliability analysis. This stage is a modification of the stages of instrument development used by Daryanes et al. (2025).

Participants: The participants consisted of 342 pre-service mathematics teachers from four universities located in Ponorogo, Malang, Bandung, and Mataram, Indonesia. Participants were recruited using voluntary convenience sampling method. All respondents were enrolled in mathematics teacher education programmes and had completed at least one core mathematics course. Participation was anonymous, and all respondents completed the questionnaire voluntarily. Table 1 presents a summary of the participants' demographic characteristics, including their gender and year of study.

Table (1): Demographic Characteristics of the Participants.

Variable	Category	n	Percentage
Gender	Female	255	74.56%
	Male	87	25.44%
Year of study	1st–2nd year	247	72.22%
	3rd–4th year	95	27.78%

Literature Review: The first stage of this research involved a review of CT definitions, indicators, and dimensions. There are four components of CT used in this study, namely, decomposition, abstraction, pattern recognition, and algorithm, as offered by Dong et al. (2019). The division of CT components into these four categories is more widely used and more relevant to mathematics (Irawan et al., 2025). However, instead of using the PRADA acronym, which has the potential to produce misconceptions about the CT sequence starting from Pattern Recognition, Abstraction, Decomposition, and Algorithm, this study offers a new acronym, namely DAPRAL (Decomposition, Abstraction, Pattern Recognition, Algorithm). This is done to further harmonise that the use of CT, especially those related to mathematics, starts from decomposing the problem, then performing abstraction, finding patterns, and ending with compiling algorithms.

The results of the literature review also produced definitions and indicators for each CT component, as presented in Table 2.

Table (2): Aspects, Definitions, and Indicators of CT Skills

Aspects	Definitions	Indicators	Reference
Decomposition	The ability to break down complex problems into smaller, manageable parts is essential.	• Identify sub-problems of the main problem. • Analyse the problem structure into functional parts.	Barr and Stephenson (2011), Shute et al. (2017)
Abstraction	The ability to filter out important information and ignore irrelevant details to focus on the essence of a problem.	• Identifying important information relevant to solving the problem. • Ignoring irrelevant information:	Wing (2006, 2008) and Yadav et al. (2018)
Pattern Recognition	The ability to recognise regularities, similarities, or trends in data and information to support problem-solving.	• Find recurring patterns in the data or problem structure. • Predict problem-solving procedures based on the identified patterns. • The results were predicted based on the identified patterns.	Grover (2022) and Weintrop et al. (2021)
Algorithm	Ability to design logical and systematic steps to solve problems.	• Organise logical and sequential steps in problem-solving. • Solutions are formulated in the form of algorithms or pseudocode. • Evaluate the efficiency and effectiveness of the solutions.	Shute et al. (2017), Angeli and Giannakos (2020)

Designing Instrument: We based the CTSA instrument on four CT dimensions. Its development followed a five-step procedure, incorporating EFA, CFA, and IRT analysis as listed in Table 2. Each CT component is represented by five items, so there are a total of 20 items in the CTSA. Each question provided five Likert scale answer options: 1 = never, 2 = rarely, 3 = sometimes, 4 = generally, and 5 = always. These answers are in accordance with the purpose of the assessment, which was to assess the ability of CTs to work independently (Korkmaz et al., 2017).

Although several CTSA items are phrased in general terms (e.g., "dividing big problems into smaller parts"), this wording was intentionally retained to ensure applicability across diverse mathematical topics and instructional contexts. Prior research suggests that general cognitive formulations allow respondents to anchor their self-assessments in domain-specific experiences, particularly in mathematics, where problem structures vary widely across content areas (Weintrop et al., 2015). During administration, the participants were instructed to respond based on their experiences in learning and teaching mathematics, thereby contextualizing the items within mathematics teacher education.

Content Validity: After compiling the CTSA instrument into 20 questions, an expert assessment was conducted. Five experts in mathematics education and instrument development were involved in the content validity. The validation was conducted using a 4 Likert scale rating as proposed by Lynn (1986), namely 1 = not relevant, 2 = less relevant, 3 = quite relevant, and 4 = very relevant. The validation results were then analyzed using the Item-level Content Validity Index (I-CVI) and Scale-level Content Validity Index (S-CVI) according to the guidelines of Polit and Beck (2006; 2007). The I-CVI was calculated based on the proportion of experts who scored 3 or 4 on each item. S-CVI was calculated using the Scale-Level Content Validity Index Averaging Calculation Method (S-CVI/Ave) and the Scale-Level Content Validity Index Universal Agreement Calculation Method (S-CVI/AU). In the validation process involving five experts, an item was declared valid if the I-CVI value was 1.00 (Lynn, 1986).

Instrument Testing: The CTSA instrument, which had been validated and improved according to expert suggestions, was presented online using Google Forms. Before completing the questionnaire, the participants were explained the purpose of the study and consent for participation. Data will be collected from March to April 2025 to ensure sufficient student participation.

Based on the guidelines from Hair et al. (2019) and Comrey and Lee (2013), , the ideal sample size for EFA is 5-10 respondents per item, while for CFA, Kline (2016) recommends a minimum of 200 respondents. In addition, this number is also appropriate for IRT analysis, especially the Graded Response Model (GRM), which requires 150-250 respondents for stable parameter estimation (Embretson & Reise, 2013). Thus, the sample size used in this study met the statistical criteria required to ensure the accuracy and reliability of the analysis results.

Data Analysis: The data obtained from the instrument trials were quantitatively analysed using various data analysis techniques.

***Exploratory Factor Analysis* (EFA)**: An EFA was conducted to identify the factor structure underlying the CTSA instrument. The method used was Principal Axis Factoring (PAF) with Promax rotation to accommodate the possibility of correlation between factors. Decisions regarding the number of factors were based on an eigenvalue > 1 and scree plot analysis. Data eligibility was tested using the Kaiser-Meyer-Olkin (KMO) and Bartlett's Test of Sphericity. KMO values ≥ 0.80 and Bartlett's significance ($p < 0.05$) indicate that the data deserve further analysis through EFA (Hair et al., 2019).

***Confirmatory Factor Analysis* (CFA)**: CFA was used to test the suitability of the factor model obtained from the EFA. Estimation is performed using the Maximum Likelihood Estimation (MLE). The model is considered to have a good fit if it meets the fit index criteria as follows: Comparative Fit Index (CFI) ≥ 0.90, Tucker-Lewis Index (TLI) ≥ 0.90, Root Mean Square Error of Approximation (RMSEA) ≤ 0.08, and Standardised Root Mean Square Residual (SRMR) ≤ 0.08 (Hu & and Bentler, 1999).

Reliability Analysis: Reliability analysis included Cronbach's alpha (CA), Composite Reliability (CR), and Average Variance Extracted (AVE). Cronbach's alpha was used to assess the reliability of internal consistency, both in the overall instrument and in each of the four CT components. Cronbach's alpha values ≥ 0.70 indicated good internal consistency for each dimension. Items with Cronbach's alpha values below 0.30 were considered for deletion or revision.

Convergent Validity: Convergent validity was assessed to determine the extent to which items of the same construct correlated, thus supporting the unidimensionality of each factor in the instrument. To examine convergent validity, two key statistical indicators were calculated for each factor: Composite Reliability (CR) and Average Variance Extracted (AVE). These indicators were calculated using standardised factor loadings and error variances obtained from Confirmatory Factor Analysis (CFA). A CR score ≥ 0.70 indicates acceptable internal consistency reliability (Puhan, Bryant, Guyatt, Heels-Ansdell, & Schünemann, 2005). An AVE score ≥ 0.50 indicates adequate convergent validity, suggesting that more than 50% of the variance in the observed variable is accounted for by the latent construct. Items with low loadings (usually <0.50) were considered for deletion to improve the quality of the instrument.

Discriminant Validity: In addition to convergent validity, discriminant validity was examined using the Fornell–Larcker criterion and the Heterotrait–Monotrait ratio (HTMT). Discriminant validity is supported when the square root of the AVE for each construct exceeds its correlations with other constructs (Fornell & Larcker, 1981), and when HTMT values are below the conservative threshold of 0.90 (Henseler, Ringle, & Sarstedt, 2015). These criteria were applied to ensure that each CT dimension represents a distinct construct despite their theoretical interrelatedness.

Correlation Analysis: In addition to the reliability and construct validity analyses, this study also conducted a correlation analysis among the CT principal components. This analysis aimed to explore the strength and direction of the relationships between dimensions and assess their consistency with the theoretical framework underlying the instrument’s development. Correlations between constructs were assessed using a Pearson correlation matrix, which was supplemented with a significance value (p-value) to determine whether the relationship was statistically significant. This analysis is important for testing the conceptual validity that the four CT components are part of an interrelated framework but still have distinct construct identities.

Item Response Theory: This analysis was used to evaluate item characteristics based on the Graded Response Model (GRM), which is appropriate for Likert scale data (Zein & Akhtar, 2025). The parameters evaluated included the discrimination and difficulty thresholds for each item. An item is considered to meet the criteria if it has a discrimination value ≥ 0.65 (moderate to high) and the thresholds are logically arranged and not extreme (Baker, 2002).

Results

Content Validity: The first stage of analyzing the CTSA instrument is content validity analysis. The results of instrument validation by five experts are presented in Table 3.

Table (3): Results of Content Validation by Experts using I-CVI

Item	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Number of Agreement	I-CVI	Conclusion
DC1	4	4	4	4	4	5	1.00	Excellent
DC2	4	4	4	4	3	5	1.00	Excellent
DC3	4	4	4	4	4	5	1.00	Excellent

Item	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Number of Agreement	I-CVI	Conclusion
DC4	4	4	4	4	3	5	1.00	Excellent
DC5	4	3	4	3	3	5	1.00	Excellent
AB1	4	4	4	3	4	5	1.00	Excellent
AB2	4	4	4	4	4	5	1.00	Excellent
AB3	4	4	4	4	3	5	1.00	Excellent
AB4	4	4	4	3	4	5	1.00	Excellent
AB5	4	4	4	4	3	5	1.00	Excellent
PR1	4	4	4	4	4	5	1.00	Excellent
PR2	4	3	4	3	3	5	1.00	Excellent
PR3	4	4	4	4	3	5	1.00	Excellent
PR4	4	4	4	3	3	5	1.00	Excellent
PR5	4	4	4	4	3	5	1.00	Excellent
AL1	4	4	3	4	4	5	1.00	Excellent
AL2	4	4	4	4	4	5	1.00	Excellent
AL3	4	4	4	4	3	5	1.00	Excellent
AL4	4	4	4	4	3	5	1.00	Excellent
AL5	4	4	4	4	3	5	1.00	Excellent
Number of Agreement	20	20	20	20	20	S-CVI/Ave	1.00	Excellent
Proportion Relevant	1.00	1.00	1.00	1.00	1.00	S-CVI/UA	1.00	Excellent

The results of validation by experts, as presented in Table 3, show that 20 items on the CTSA each have an I-CVI index of 1.00, so they have excellent content validity. The S-CVI values, both S-CVI/Ave and S-CVI/UA, each also had an index of 1.00; therefore, overall, the CTSA instrument had excellent content validity

Exploratory Factor Analysis (EFA): An exploratory factor analysis was conducted to identify the latent structure of the 20-item CTSA. The data feasibility test showed that the data met the assumptions of EFA with a Kaiser-Meyer-Olkin (KMO) value of 0.91 (> 0.7), which was categorised as very good, and the results of Bartlett's Test of Sphericity were significant (χ^2 = 2303.59, df = 190, p = 0.00), which indicated that the data were suitable for factor analysis. Furthermore, the selection of the number of factors was based on the results of the scree plot and parallel analysis, as presented in Figure 2.

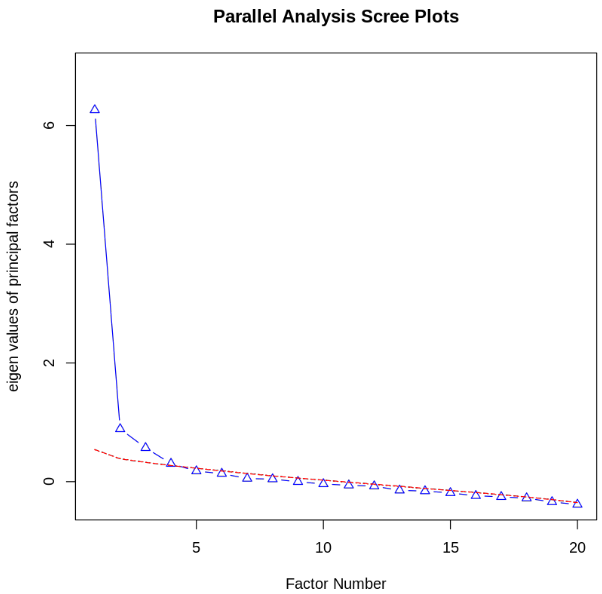

Figure (2). CTSA Scree Plots

Figure 2 shows that the clear elbow point on the fourth factor indicates that the four-factor structure is the most appropriate solution. This decision was also reinforced by the eigenvalue criterion of 1 and the excellent model fit values, including an RMSEA value of 0.039 and TLI of 0.952.

Further analysis using the Principal Axis Factoring method and promax rotation resulted in four main factors that correspond to the theoretical constructs of CT: Decomposition

(DC), abstraction (AB), Pattern Recognition (PR), and Algorithmic Thinking (AL). Each factor consisted of 4-5 items that loaded strongly (loading ≥ 0.40), with little or no significant cross-loading. The cumulative variance explained by these four factors reached 43%, which is acceptable in the context of measurement in education.

Table (4): Factor Loading from Exploratory Factor Analysis

Item	Factor 1 (Algorithm)	Factor 2 (Abstraction)	Factor 3 (Decomposition)	Factor 4 (Pattern Recognition)	h^2	u^2	Complexity
DC1			0.48		0.29	0.71	1.1
DC2			0.28		0.34	0.66	2.4
DC3			0.64		0.44	0.56	1.1
DC4			0.41		0.36	0.64	1.6
DC5			0.44		0.39	0.61	1.9
AB1		0.48			0.29	0.71	1.2
AB2		0.09			0.34	0.66	2.2
AB3		0.51			0.43	0.57	1.4
AB4		0.84			0.65	0.35	1.0
AB5		0.38			0.47	0. 35	2.1
PR1				0.53	0.44	0.56	1.5
PR2				0.79	0.61	0.39	1.0
PR3				0.47	0.46	0.54	1.6
PR4				0.17	0.50	0.50	1.8
PR5				0.27	0.45	0.55	2.3
AL1	0.63				0.53	0.47	1.1
AL2	0.74				0.50	0.50	1.0
AL3	0.63				0.47	0.53	1.1
AL4	0.42				0.41	0.59	1.9
AL5	0.21				0.29	0.71	2.6

As shown in Table 4, the items that make up each factor loaded consistently according to the expected dimensions. For example, items AL1 to AL4 load strongly on Factor 1 (Algorithm), whereas DC1 to DC5 load strongly on Factor 4 (Decomposition). One item (AL5) had a loading below 0.30 and was identified as a candidate for revision in the further instrument development.

Confirmatory Factor Analysis (CFA): After the four-factor structure was obtained through Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA) was conducted to test the suitability of the measurement model with empirical data. The CFA model was designed based on four theoretical constructs, namely Decomposition, Abstraction, Pattern Recognition, and Algorithmic Thinking, each of which was measured through five statement items. The model resulting from the CFA is presented in Figure 3.

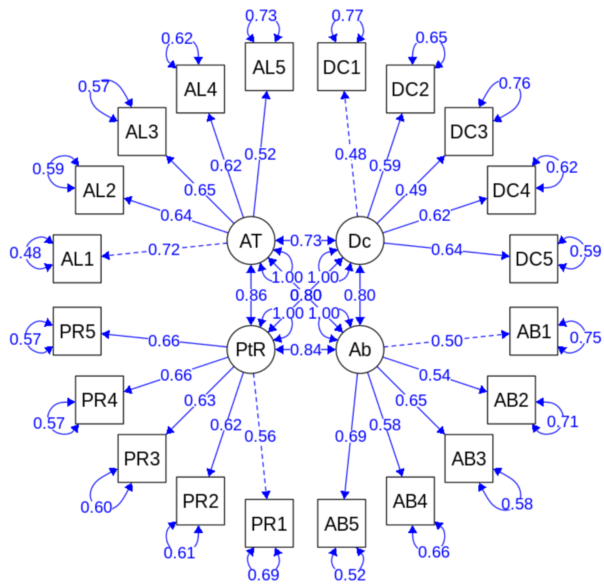

Figure (3). Visualization of CFA paths

Figure 3 shows that the model estimation results indicate that the CFA model has reached convergence, and all factor loadings are statistically significant (p>0.001). The standardised factor loadings ranged from 0.48 to 0.72, with most items having values greater than 0.60. This indicates that each indicator substantially contributes to its respective latent construct. The Algorithmic Thinking dimension had the highest loading value (AL1 = 0.72), while the lowest value was found in item DC1 (0.48), but it was still within acceptable limits (≥ 0.40). In addition, the correlations between latent constructs were in the range of 0.73 to 0.86, which indicated that the four dimensions were strongly correlated and supported the concept that CT is a skill comprising several interrelated aspects. To ensure consistency, factor loadings were interpreted using three qualitative categories: moderate (0.40–0.59), good (0.60–0.69), and very good (≥ 0.70). The loading values of each indicator are presented in Table 5.

Table (5): Factor Loading from Confirmatory Factor Analysis.

Factor	Item	Loading Standard	Interpretation
Decomposition	DC1	0.48	Significant, moderate
	DC2	0.59	Significant, moderate
	DC3	0.49	Significant, moderate
	DC4	0.62	Significant, good
	DC5	0.64	Significant, good
Abstraction	AB1	0.50	Significant, moderate
	AB2	0.54	Significant, moderate
	AB3	0.65	Significant, good
	AB4	0.58	Significant, moderate
	AB5	0.69	Significant, good
Pattern Recognition	PR1	0.56	Significant, moderate
	PR2	0.62	Significant, good
	PR3	0.63	Significant, good
	PR4	0.66	Significant, good
	PR5	0.66	Significant, good
Algorithm	AL1	0.72	Significant, very good
	AL2	0.64	Significant, good
	AL3	0.65	Significant, good
	AL4	0.62	Significant, good
	AL5	0.52	Significant, moderate

Overall, the results of the CFA in this study provide strong evidence for the construct validity of the CTSA developed to measure the CT skills of prospective mathematics teachers. The four-factor measurement model–decomposition, abstraction, pattern recognition, and algorithmic thinking–exhibited a good fit to the data, as indicated by significant standardised loadings and high correlations among latent constructs.

Reliability Analysis: The reliability of the CTSA instrument was tested using Cronbach's alpha (CA) and Average Variance Extracted (AVE). The results of the reliability analysis using these three techniques are presented in Table 6.

Table (6): Reliability Test Results for Each Factor.

Factor	CA	AVE	CR
Decomposition	0.700	0.310	0.681
Abstraction	0.728	0.347	0.721
Pattern Recognition	0.766	0.394	0.758
Algorithm	0.763	0.398	0.770

Reliability analysis results showed that all dimensions of the CTSA instrument had adequate internal consistency. Cronbach's alpha values ranged from 0.70 to 0.77, with the

Pattern Recognition dimension showing the highest reliability (α = 0.77), followed by Algorithmic Thinking (α = 0.76), abstraction (α = 0.73), and decomposition (α = 0.70). Based on general criteria (Hair et al., 2019), all dimensions had acceptable reliability (α ≥ 0.70). In addition, the Composite Reliability (CR) value also shows a similar pattern, with a range of values between 0.69 and 0.77. The Algorithmic Thinking and Pattern Recognition dimensions had the highest CR (0.77 and 0.76, respectively), while Decomposition was at the minimum limit (CR = 0.69). These results support the notion that each construct forms a homogeneous and reliable concept unit for measuring the CT skills of prospective mathematics teachers.

However, the results of the Average Variance Extracted (AVE) analysis still show limitations in terms of convergent validity. All dimensions had AVE values below the ideal threshold of 0.50, ranging from 0.31 to 0.40. This indicates that although the construct is quite reliable, the proportion of indicator variance successfully explained by the latent construct is still relatively low. In other words, the items in each dimension still leave considerable variance that is not explained by the construct; therefore, it needs to be improved in the next stage of instrument development.

Discriminant Validity: Discriminant validity was examined using the Fornell–Larcker criterion and the HTMT. The Fornell–Larcker criterion indicated that the square roots of the AVE for each construct were comparable to, though in some cases slightly lower than, the inter-construct correlations, reflecting the theoretically interconnected nature of the dimensions of CT. The HTMT values ranged from 0.78 to 0.89, remaining below the recommended threshold of 0.90. These results suggest that, despite relatively high correlations, the four dimensions demonstrate acceptable discriminant validity while functioning as closely related components of a higher-order CT construct.

Correlation between CT Components: Lastly, this study examined the correlation between the four components of CT skills, and the results are presented in Table 7.

Table (7): Correlation between CT Components.

	Decomposition	Abstraction	Pattern Recognition	Algorithm
Decomposition	1.00			
Abstraction	0.80**	1.00		
Pattern Recognition	0.80**	0.84**	1.00	
Algorithm	0.73**	0.79**	0.86**	1.00

The results presented in Table 7 show that Pattern Recognition and Algorithm have the highest correlation (0.86), while Decomposition and Algorithm have the lowest correlation (0.73). Overall, these estimates are based on the correlated factor model proposed in this paper. Thus, there was a significant positive correlation between the four CT skills, especially among prospective mathematics teachers.

Item Response Theory (IRT): The item response theory used in this study is the Graded Response Model (GRM). The GRM produces two main parameters for each CTSA instrument item: discrimination and difficulty. The discrimination parameter (a) indicates the ability of the item to distinguish respondents with different levels of CT ability, while the difficulty parameter (b1-b4) represents the transition points between categories on a 5-point Likert scale. A visualisation of the probability curves of all items from the GRM analysis using R software in Google Colab is shown in Figure 4.

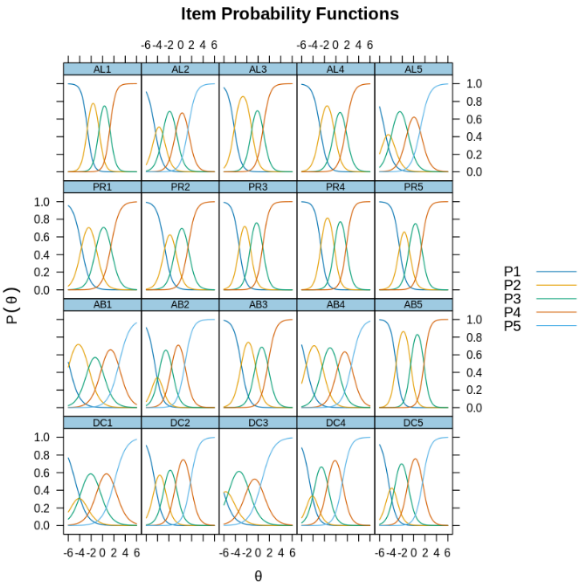

Figure (4). Item Probability Functions for All CTSA Items

In Figure 4, the category curves (P1 to P5) for most items alternate along the latent ability spectrum (θ), indicating that the instrument can distinguish CT ability levels well. In the Algorithmic Thinking dimension (AL1-AL5), items such as AL3 and AL4 show a clear threshold between categories and a balanced distribution of responses. The Pattern Recognition dimension (PR1-PR5) also shows an ideal curve, especially PR2 and PR4 which show smooth transitions between categories. Meanwhile, some items on the Abstraction and Decomposition dimensions, such as AB1, AB5, and DC3, show extreme curves or less spread, which indicates the possibility that the item is too easy, too difficult, or needs to be revised. For example, item DC3 shows a concentration of responses in the agree and strongly agree categories. Nevertheless, the factor loading values for these items remained within an acceptable range (≥ 0.49); therefore, they still met the criteria for construct validity. Conversely, items such as AL1 showed a more even distribution of responses and a higher factor loading (0.72), indicating a stronger ability to discriminate between respondents. These comparisons provide concrete illustrations of items with stronger dispersion and those exhibiting limited variability. More specifically, the discrimination (a) and thresholds (b1-b4) of each CTSA item are presented in Table 8.

Table (8): IRT Parameters for Each Item.

Item	a (Discrimination)	b1	b2	b3	b4
DC1	0.9686	-4.7707	-3.4601	-0.6855	2.0915
DC2	1.4979	-4.4811	-2.7426	-0.7725	1.8084
DC3	0.9521	-6.5561	-4.8558	-1.8454	0.6206
DC4	1.4489	-4.6225	-3.6584	-1.4435	1.1578
DC5	1.5360	-4.4347	-3.2556	-0.9999	1.5835
AB1	1.0560	-5.9427	-2.5286	-0.0750	2.9076
AB2	1.5162	-4.4989	-3.5669	-1.5123	0.8232
AB3	1.5179	-2.9786	-0.4640	1.7531	NA
AB4	1.2163	-5.2611	-2.3891	0.3301	2.7855
AB5	2.0347	-3.1017	-0.5278	1.8086	NA
PR1	1.3476	-3.7214	-1.1174	1.5023	NA
PR2	1.5183	-2.7845	-0.8594	1.4017	NA
PR3	1.8251	-3.3466	-1.3591	0.8218	NA
PR4	1.9286	-2.6569	-0.2855	1.8428	NA
PR5	1.7919	-2.5340	-0.7806	1.3927	NA
AL1	2.0156	-2.6762	-0.6184	1.3090	NA
AL2	1.5044	-4.4869	-2.9917	-0.7557	1.3992
AL3	1.6826	-4.1774	-1.1397	0.8935	NA
AL4	1.5722	-2.7845	-0.3098	1.7771	NA
AL5	1.2641	-5.1754	-3.7463	-1.1005	1.2031

The analysis results, as presented in Table 8, show that most items have high to very high discrimination values, with a values ranging from 0.95 to 2.03. Items such as AB5, PR3, PR4, PR5, and AL1 showed a value above 1.70, indicating that they were highly informative and sensitive to variations in ability level. In contrast, item DC3 has a lower discrimination value (a = 0.95) and extreme thresholds (b1 = -6.56, b4 = 0.62), indicating that the majority of respondents with low ability levels can answer this item easily, signalling that this item is too easy and lacks discrimination. Some items, such as AB3 and AB5, only produced three thresholds (without b4), indicating that not all categories were optimally used by respondents. This could be due to less challenging or less relevant item formulations in the questionnaire. Overall, this instrument can be used to measure the CT skills of prospective mathematics teachers.

Discussion

Content Validity: The results of the content validity analysis of the Computational Thinking Self-Assessment (CTSA) instrument showed that all items obtained an Item-Level Content Validity Index (I-CVI) value of 1.00. This value indicates that all items were considered highly relevant by the experts involved in the validation. In addition, the Scale-Level Content Validity Index value in its two approaches, namely S-CVI/Ave and S-CVI/UA, also shows the maximum number, that is, 1.00. According to Polit and Beck (2006), an I-CVI value of 1.00 at five panelists is the highest standard for stating that an item has excellent content validity. Thus, this finding provides strong evidence that all items in the CTSA instrument are in accordance with the theoretical construct being assessed.

Content validity is a crucial first step in the instrument development process because it ensures that each item accurately reflects the intended concept. Lynn (1986) emphasised the importance of expert involvement in assessing the appropriateness of item content with construct indicators to ensure content appropriateness. In this context, the involvement of five experts in mathematics education and computational thinking provided assurance that the items in the CTSA instrument adequately contained the fundamental aspects of computational thinking skills.

The high consistency of assessment between experts also shows that there are no significant differences in perceptions related to the relevance of items, which strengthens the interpretation that this instrument has a high level of interexpert agreement. This is important, given that one of the common challenges in the development of psychometric instruments is rater disagreement on certain items, which can decrease the reliability of the content validity (Yusoff, 2019). The CTSA instrument in this study successfully addressed this challenge with full agreement.

Overall, these findings validate that the CTSA met the criteria for excellent content validity according to internationally recognised scientific criteria. Strong content validity is an important foundation before conducting further construct structure analysis, such as Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA), as it ensures that the items tested represent the dimensions to be measured (DeVellis, 2017). Thus, these results provide a strong empirical basis for the reliability and feasibility of the instrument in the context of assessing the computational thinking of prospective mathematics teachers' CT.

Exploratory Factor Analysis (EFA): The exploratory factor analysis (EFA) results show that the four-factor structure is the most suitable representation for identifying the latent dimensions of the CTSA instrument. This is indicated by the Kaiser-Meyer-Olkin (KMO) value of 0.91, which, according to Kaiser (1974), is categorised as "marvelous" as well as a significant Bartlett's Test of Sphericity result ($p < 0.001$), indicating that the correlation between items is strong enough for factor analysis (Hair et al., 2019). These findings provide initial support that the data meet the assumptions required for EFA and are worth analysing to reveal the instrument's latent structure.

The identification of four main factors emerging from the Promax rotation results-namely, decomposition, abstraction, pattern recognition, and algorithmic thinking–is highly consistent with the theoretical framework of Computational Thinking as proposed by Wing (2006) and Brennan and Resnick (2012). These four dimensions have been widely recognised as core cognitive components in the development of computational thinking skills, particularly in the context of mathematics education. The compatibility between the empirical factor structure and the conceptual model indicates that the CTSA instrument has a good fit for the domain of the construct being measured.

The factor loading table shows that almost all items have loading values above 0.40, which is the minimum limit suggested in psychometric studies to express the strength of an item's contribution to the factor (Costello & Osborne, 2005). Item AL5 which has a loading below the threshold is of particular concern and is recommended to be revised or reconsidered at a later stage of development. However, there was no significant cross-loading between factors, indicating that each item was quite specific to its origin factor. This is an additional indicator that the instrument has a clean and stable structure.

Overall, the four factors explained about 43% of the total variance, which is within the acceptable range in the fields of social sciences and education (Tabachnick, Fidell, & Ullman, 2019). Thus, these EFA results provide strong evidence that the CTSA instrument has an empirically valid and theoretically sound factor structure. The recommended next step is to conduct a confirmatory factor analysis (CFA) to test the suitability of this measurement model more formally and confirm the consistency of the structure in a wider population.

Confirmatory Factor Analysis (CFA): The results of the CFA confirm that the four-factor measurement model–decomposition, abstraction, pattern recognition, and algorithmic thinking–has a good fit with the empirical data. All indicators in the model showed statistically significant factor loadings ($p < 0.001$), with standardised loading values ranging from 0.48–0.72. Most indicators have loadings above 0.60, which, according to Hair et al. (2019) indicates a substantial contribution to the latent construct. This finding provides empirical support that the indicators compiled in the CTSA instrument actually measure important aspects of computational thinking (CT) skills, as theoretically formulated in previous literature (Brennan & Resnick, 2012; Román-González, Pérez-González, & Jiménez-Fernández, 2017).

Furthermore, the correlations between factors were in the range of 0.73 to 0.86, indicating that the four dimensions have a close relationship while still representing unique constructs. This finding is in line with the multidimensional approach to CT, which emphasises that this skill consists of several distinct yet mutually supportive thinking processes (Shute et al., 2017). Although the correlation values between factors were relatively high, this did not pose a multicollinearity problem, but rather reinforced the interpretation that each dimension was an integral part of the overall computational thinking skill. Thus, the CFA results in this study provide a strong basis for concluding that the factor structure of the CTSA instrument has been statistically confirmed and is theoretically valid.

Reliability Analysis: The results of the reliability analysis show that all constructs in the CTSA instrument have adequate internal consistency, as indicated by the Cronbach's alpha (CA) value which is in the range of 0.70 to 0.77. This value is within acceptable limits according to Hair et al. (2019), who state that a CA value ≥ 0.70 is an indicator that a construct has good internal reliability in the context of social and educational research. This indicates that each dimension of the CTSA instrument–decomposition, abstraction, pattern recognition, and algorithmic thinking–consistently measures the same attribute and is reliable in assessing the computational thinking ability of prospective mathematics teachers. This finding is also in line with previous studies that confirm the importance of reliability testing in cognitive construct-based instruments such as CT (KaleliOğlu, Gülbahar, & Kukul, 2016; Román-González et al., 2017).

In addition to Cronbach's Alpha, Composite Reliability (CR) analysis also showed consistent results with CA. The CR for the four factors was in the range of 0.681 to 0.770, which is above the recommended minimum threshold of 0.70 (Fornell & Larcker, 1981), although the decomposition dimension was slightly below but still tolerable. These results reinforce previous findings that the constructs in the CTSA are stable and reliable. However, the Average Variance Extracted (AVE) analysis showed that none of the dimensions reached the ideal value of 0.50. AVE values ranging from 0.31 to 0.40 indicate that, although the construct is reliable, the variance explained by the construct against its indicators is still relatively low. Therefore, the convergent validity of the instrument needs to be improved through the revision or rearrangement of some items (Henseler et al., 2015).

Correlations between dimensions showed strong and significant relationships ($p < 0.001$), with the highest correlation between Pattern Recognition and Algorithmic Thinking ($r = 0.86$) and the lowest correlation between Decomposition and Algorithmic Thinking ($r = 0.73$). These correlations support the conceptual framework that CT skills are not stand-alone entities but consist of interconnected and mutually reinforcing components (Shute et al., 2017; Weintrop et al., 2015). The strong relationship between these components indicates that students who have high abilities in one aspect of CT tend to have good abilities in other aspects as well, especially in the context of developing higher-order thinking skills in learning mathematics and technology.

Item Response Theory (IRT): Analysis using the Graded Response Model (GRM) in this study showed that most of the items in the CTSA instrument performed very well in measuring the Computational Thinking (CT) skills of prospective mathematics teachers. The discrimination parameter (a) on most items ranged from 0.95 to 2.03, which, according to Baker (2001), can be categorised as high (1.35-1.69) to very high (≥1.70). This indicates that the items can effectively differentiate students based on their CT ability levels. For example, items such as PR4, PR5, AL1, and AB5 had a-values above 1.7, indicating high sensitivity to changes in respondents' latent ability level (θ). This finding supports the construct quality of the instrument in providing accurate and precise information regarding individual differences in CT ability.

Nevertheless, some items showed potential for improvement. Item DC3, for example, showed a lower discrimination value (a = 0.95) and an extreme difficulty threshold (b1 = -6.56), meaning that even respondents with very low ability could still give high-category answers. This suggests that the item is too easy and has limited discriminating power, as Hambleton, Swaminathan, and Rogers found that items with extreme thresholds can cause distortions in ability estimates. In addition, items such as AB3 and AB5 only produced three thresholds (without b4), which may indicate that respondents did not effectively use all Likert scale options. This phenomenon could be due to a lack of variety in responses or less challenging wording.

The visualisation of item probability functions (Item Probability Curves) in Figure 4 also reinforces these quantitative results. The curves that cross proportionally between categories (P1 to P5) show that most items can capture the transition between ability levels well. In particular, the Algorithmic Thinking and Pattern Recognition dimensions showed balanced and well-distributed curve patterns, supporting the use of these dimensions as effective measures in the CTSA instrument. In contrast, items in the Abstraction and Decomposition dimensions showed sharp or extreme overlapping curves, indicating the possibility of non-uniform interpretation by respondents or the need to reformulate some items.

Overall, the results of the IRT analysis with GRM support the validity and accuracy of the CTSA instrument in measuring the computational thinking ability of prospective mathematics teachers. As suggested by Embretson and Reise (2013), the use of IRT provides more in-depth information than classical analysis because it not only shows how well items function, but also at which ability level they are most informative for the test-taker. Therefore, these results provide a strong empirical foundation for the further development of the CTSA instrument, recommending the revision of the marginally performing items, as well as retaining those items that proved to be highly discriminative and well spread across the CT ability spectrum.

Conclusion

This study developed and validated the Computational Thinking Self-Assessment (CTSA) instrument, tailored specifically for preservice mathematics teachers. The final version comprises 20 items across four core dimensions—Decomposition, Abstraction, Pattern Recognition, and Algorithmic Thinking—aligned with established theoretical models of CT.

The multistage validation process, which included expert review, Exploratory and Confirmatory Factor Analysis, reliability testing, and Item Response Theory analysis, yielded robust empirical support for the instrument's quality. Content validity indices (I-CVI and S-CVI) demonstrated complete expert consensus. Factor analyses confirmed the proposed four-factor structure, while internal consistency metrics indicated acceptable to high reliability across all subscales. Although Average Variance Extracted (AVE) values fell below the conventional threshold, suggesting moderate convergent validity, the instrument still meets key psychometric standards. IRT analysis further established the discriminatory power and threshold alignment of the items, supporting their sensitivity to differing CT proficiency levels.

Taken together, these findings validate the CTSA as a theoretically grounded, psychometrically sound instrument for assessing computational thinking skills in the context of mathematics teacher education. Its application holds promise not only for research but also for instructional improvement and reflective learning within teacher preparation programs.

Disclosure Statement

- **Ethical approval and consent to participate:** This study was conducted according to the strictest ethical guidelines; however, no formal consent was sought as the type of study we conducted did not require such consent. All participants completed the form voluntarily and agreed to the use of their data for scientific purposes.
- **Availability of data and materials:** The data supporting the findings of this study are available from the corresponding author upon reasonable request.
- **Author contribution:** E.I. designed the study, communicated with the key people, and wrote the manuscript. M.K.H, S.H., & R.P. collected the data and did the statistical analysis. R.W. has reviewed the data and the final manuscript for approval. The author(s) read and approved the final manuscript.
- **Conflict of interest:** The authors declare that they have no competing interests.

- **Funding:** This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sector.
- **Acknowledgments:** Thanks to the five experts who validated the CTSA instrument and all students who were willing to participate in this study.

List of Abbreviations

AB: Abstraction

AL: Algorithm

AVE: Average Variance Extracted

CFA: Confirmatory Factor Analysis

CR: Composite Reliability

CTS: Computational Thinking Scale

CTSA: Computational Thinking

CTSA: Computational Thinking Self-Assessment

DC: Decomposition

EFA: Exploratory Factor Analysis

GRM: Graded Response Model

IRT: Item Response Theory

KMO: Kaiser-Meyer-Olkin

MTE: Mathematics Teacher Education

PR: Pattern Recognition

Open Access

References

Angeli, C., & Giannakos, M. (2020). Computational Thinking Education: Issues and Challenges. *Computers in Human Behavior, 105*(106185), 1–8. https://doi.org/10.1016/j.chb.2019.106185

Baker, F. B. (2002). *The Basics of Item Response Theory* (2nd ed). College

Park: ERIC Clearinghouse on Assessment and Evaluation.

Barr, V., & Stephenson, C. (2011). Bringing Computational Thinking to K-12: What Is Involved and What Is the Role of the Computer Science Education Community? *ACM Inroads*, *2*(1), 48–54. https://doi.org/10.1145/1929887.1929905

Beecher, K. (2017). *Computational Thinking: A Beginner's Guide to Problem-Solving and Programming*. Swindon: BCS Learning & Development Limited.

Beyerer, J., Richter, M., & Nagel, M. (2018). *Pattern Recognition: Introduction, Features, Classifiers and Principles*. Walter de Gruyter GmbH: Berlin/Boston.

Brennan, K., & Resnick, M. (2012). New Frameworks for Studying and Assessing the Development of Computational Thinking. *Proceedings of the 2012 Annual Meeting of the American*, 25.

Cansu, F. K., & Cansu, S. K. (2019). An Overview of Computational Thinking. *International Journal of Computer Science Education in Schools*, *3*(1), 17–30. https://doi.org/10.21585/ijcses.v3i1.53

Comrey, A. L., & Lee, H. B. (2013). *A First Course in Factor Analysis* (0 ed.). Psychology Press. https://doi.org/10.4324/9781315827506

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, *10*(7). https://doi.org/10.7275/JYJ1-4868

Csizmadia, A., Standl, B., & Waite, J. (2019). Integrating the Constructionist Learning Theory with Computational Thinking Classroom Activities. *Informatics in Education*, *18*(1), 41–67. https://doi.org/10.15388/infedu.2019.03

Curzon, P., Bell, T., Waite, J., & Dorling, M. (2019). Computational Thinking. In S. A. Fincher & A. V. Robins (Eds.), *The Cambridge Handbook of Computing Education Research* (1st ed., pp. 513–546). Cambridge University Press. https://doi.org/10.1017/9781108654555.018

Daryanes, F., Zubaidah, S., Mahanal, S., & Sulisetijono, S. (2025). Development and Validation of Research Skills Instruments for Pre-Service Biology Teachers. *Participatory Educational Research*, *12*(2), 75–96. https://doi.org/10.17275/per.25.20.12.2

DeVellis, R. F. (2017). *Scale development: Theory and applications* (Fourth edition). Los Angeles: SAGE.

Dong, Y., Catete, V., Jocius, R., Lytle, N., Barnes, T., Albert, J., … Andrews, A. (2019). PRADA: A Practical Model for Integrating Computational Thinking in K-12 Education. *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, 906–912. Minneapolis: ACM. https://doi.org/10.1145/3287324.3287431

Embretson, S. E., & Reise, S. P. (2013). *Item Response Theory* (0 ed.). Psychology Press. https://doi.org/10.4324/9781410605269

Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, *18*(1), 39–50. https://doi.org/10.1177/002224378101800104

Grover, S. (2022). Computational Thinking Today. In A. Yadav & U. D. Berthelsen (Eds.), *Computational Thinking in Education a Pedagogical Perspective* (pp. 19–40). New York: Routledge Taylor & Francis Group.

https://doi.org/10.4324/9781003102991-2

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis* (Eighth edition). Andover, Hampshire: Cengage.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Hu, L., & and Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Irawan, E., Rosjanuardi, R., & Prabawanto, S. (2024). Research Trends of Computational Thinking in Mathematics Learning: A Bibliometric Analysis from 2009 to 2023. *Eurasia Journal of Mathematics, Science and Technology Education*, *20*(3), 1–16. https://doi.org/10.29333/ejmste/14343

Irawan, E., Rosjanuardi, R., & Prabawanto, S. (2025). How Computational Thinking Can Be Integrated in Statistical Learning: A Cuboid Framework. *Journal on Mathematics Education*, *16*(2). Retrieved from http://jme.ejournal.unsri.ac.id/index.php/jme/article/view/3477

KaleliOğlu, F., Gülbahar, Y., & Kukul, V. (2016). A Framework for Computational Thinking Based on a Systematic Research Review. *Baltic J. Modern Computing*, *4*(3), 14.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (Fourth edition). New York London: The Guilford Press.

Korkmaz, Ö., & Bai, X. (2019). Adapting Computational Thinking Scale (CTS) for Chinese High School Students and Their Thinking Scale Skills Level. *Participatory Educational Research*, *6*(1), 10–26. https://doi.org/10.17275/per.19.2.6.1

Korkmaz, Ö., Çakir, R., & Özden, M. Y. (2017). A Validity and Reliability Study of the Computational Thinking Scales (CTS). *Computers in Human Behavior*, *72*, 558–569. https://doi.org/10.1016/j.chb.2017.01.005

Lynn, M. R. (1986). Determination and Quantification of Content Validity. *Nursing Research*, *35*(6), 382.

Mok, M. M. C., Lung, C. L., Cheng, D. P. W., Cheung, R. H. P., & Ng, M. L. (2006). Self-assessment in higher education: Experience in using a metacognitive approach in five case studies. *Assessment & Evaluation in Higher Education*, *31*(4), 415–433. https://doi.org/10.1080/02602930600679100

Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, *22*, 74–98. https://doi.org/10.1016/j.edurev.2017.08.004

Polit, D. F., & Beck, C. T. (2006). The Content Validity Index: Are You Sure You Know What's Being Reported? Critique and Recommendations. *Research in Nursing & Health*, *29*(5), 489–497. https://doi.org/10.1002/nur.20147

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an Acceptable Indicator of Content Validity? Appraisal and Recommendations. *Research in Nursing and Health*, *30*(4), 459–467. https://doi.org/10.1002/nur.20199

Puhan, M. A., Bryant, D., Guyatt, G. H., Heels-Ansdell, D., & Schünemann, H. J. (2005). Internal consistency reliability is a poor predictor of responsiveness. *Health and Quality of Life Outcomes*, *3*(1), 33. https://doi.org/10.1186/1477-7525-3-33

Román-González, M., Pérez-González, J.-C., & Jiménez-Fernández, C. (2017). Which Cognitive Abilities Underlie Computational Thinking? Criterion Validity of the Computational Thinking Test. *Computers in Human Behavior*, *72*, 678–691. https://doi.org/10.1016/j.chb.2016.08.047

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

Siles-González, J., & Solano-Ruiz, C. (2016). Self-assessment, reflection on practice and critical thinking in nursing students. *Nurse Education Today*, *45*, 132–137.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2019). *Using Multivariate Statistics* (Seventh edition). NY, NY: Pearson.

Tsai, M.-J., Liang, J.-C., & Hsu, C.-Y. (2021). The Computational Thinking Scale for Computer Literacy Education. *Journal of Educational Computing Research*, *59*(4), 579–602. https://doi.org/10.1177/0735633120972356

Weintrop, D., Beheshti, E., Horn, M., Orton, K., Jona, K., Trouille, L., & Wilensky, U. (2015). Defining Computational Thinking for Mathematics and Science Classrooms. *Journal of Science Education and Technology*, *25*(1), 127–147. https://doi.org/10.1007/s10956-015-9581-5

Weintrop, D., Rutstein, D., Bienkowski, M., & McGee, S. (2021). Assessment of Computational Thinking. In A. Yadav & U. D. Berthelsen, *Computational Thinking in Education* (1st ed., pp. 90–111). New York: Routledge. https://doi.org/10.4324/9781003102991-6

Wing, J. M. (2006). Computational Thinking. *Communications of the ACM*, *49*(3), 33–35. https://doi.org/10.1145/1118178.1118215

Wing, J. M. (2008). Computational Thinking and Thinking About Computing. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *366*(1881), 3717–3725. https://doi.org/10.1098/rsta.2008.0118

Wing, J. M. (2017). Computational Thinking's Influence on Research and Education for All. *Italian Journal of Educational Technology*, *25*(2), 7–14. https://doi.org/10.17471/2499-4324/922

Yadav, A., Krist, C., Good, J., & Caeli, E. N. (2018). Computational Thinking in Elementary Classrooms: Measuring Teacher Understanding of Computational Ideas for Teaching Science. *Computer Science Education*, *28*(4), 371–400. https://doi.org/10.1080/08993408.2018.1560550

Yadav, A., Stephenson, C., & Hong, H. (2017). Computational Thinking for Teacher Education. *Communications of the ACM*, *60*(4), 55–62. https://doi.org/10.1145/2994591

Yusoff, M. S. B. (2019). ABC of content validation and content validity index calculation. *Education in Medicine Journal*, *11*(2), 49–54. https://doi.org/10.21315/eimj2019.11.2.6

Zein, R. A., & Akhtar, H. (2025). Getting Started with the Graded Response Model: An Introduction and Tutorial in R. *International Journal of Psychology*, *60*(1). https://doi.org/10.1002/ijop.13265

Appendix

The final list of questions used in the CTSA instrument.

Code	Question
DC1	I am used to dividing big problems into small parts to make them easier to solve.
DC2	I can identify the main elements of a problem before finding a solution.
DC3	I think that solving a problem gradually is more effective than trying to solve it all at once.
DC4	I often categorise related information to make it easier to understand.
DC5	I analyse the relationship between small parts of a problem before solving it.
AB1	I can determine the most important information in a problem and ignore less relevant details.
AB2	I am used to finding the core of a problem before attempting to solve it.
AB3	I can simplify a problem without losing its core.
AB4	I can recognise general patterns in various situations without focusing too much on specific details.
AB5	I can categorise the information based on similar main characteristics.
PR1	I often look for patterns or similarities between the problems I face and previous problems.
PR2	I can use the patterns I find to solve similar problems in the future.
PR3	I find it easier to solve problems if I can identify the underlying patterns.
PR4	I can recognise the relationships between concepts in a problem.
PR5	I often use patterns or strategies that have proven successful in solving new problems.
AL1	I usually take systematic steps before starting to solve a problem.
AL2	I am used to following a clear sequence of steps to solve a problem.
AL3	I often double-check my steps to ensure that the solution is effective.
AL4	I can identify unnecessary steps in a procedure to improve efficiency of the procedure.
AL5	I think that following a logical sequence in solving a problem can help avoid unnecessary mistakes.