Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Deep Investigation Using Factorial Analysis and Rasch Model: An Optimized Short-Version of the Computational Thinking Scale (COTS)

Syahrul Alim

syahrul.alim@universitasbosowa.ac.id

Universitas Bosowa

Sirirat Petsangsri

King Mongkut's Institute of Technology Ladkrabang

Dedi Nasruddin

Universitas Bosowa

A. Nur Aulia Saudi

Universitas Bosowa

Susi Susanti

Hasanuddin University

Research Article

Keywords: Computational thinking, psychometrics properties, CFA, Rasch model

Posted Date: July 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10064363/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Computational thinking can be identified as an individual's ability to solve problems in a systematic manner. It represents a skill that has been demonstrated to facilitate effective problem-solving. However, instruments specifically designed to assess computational thinking remain relatively limited. We aim to develop and evaluate a high-quality of a Computational Thinking Skills instrument (COTS) for use in educational contexts with potential for broader application. The study involved students from various universities in Makassar-Indonesia, with 364 participants in the first-study and 434 participants in the second-study. The data analysis was conducted in two stages. The factor structure was examined using Confirmatory Factor Analysis (CFA) with JASP, Followed by Rasch model with j.Metrik. The CFA results indicated that all retained items had factor loadings > 0.50. The model demonstrated good fit across most indices, with values > 0.90, while RMSEA and SRMR were < 0.06, and reliability was > 0.80. The Rasch results showed that the local independence and unidimensionality were adequately satisfied. The rating scale analysis demonstrated a monotonically increasing pattern across categories. The item reliability was 0.97 (excellent), and the person was 0.88 (good). The number of strata was 7.56 for items (excellent) and 3.80 for persons. Moreover, none of the model fit values fell < 0.5 or > 1.5. Finally, item difficulty ranged from − 0.79 (dec3) to 0.82 (algo3), indicating that the instrument is most effective in assessing individuals with moderate levels of CT skills.

INTRODUCTION

Computational thinking (CT) has emerged as a foundational competence in the landscape of 21st-century skills, transcending disciplinary boundaries and educational levels. Initially popularized by Jeannette Wing, CT is now widely recognized as a universal problem-solving framework applicable across fields, rather than being confined to computer science alone (1, 2). It enables individuals to approach complex problems systematically, fostering logical reasoning, abstraction, and algorithmic processes. In educational contexts, CT is increasingly viewed as essential for cultivating higher-order thinking skills, creativity, and adaptability, which are critical for both academic success and future workforce readiness (3–7).

The integration of CT into education has demonstrated substantial benefits. Students with well-developed CT skills tend to exhibit more efficient and structured problem-solving abilities, which positively influence their academic performance and learning experiences (8–10). Moreover, CT supports collaborative and innovative learning environments, enabling learners to engage in meaningful inquiry and solution design (11, 12). Its relevance extends beyond higher education, with evidence suggesting that CT-based interventions can enhance learning outcomes from primary to tertiary levels (13, 14). Consequently, CT is not only a cognitive skill but also a critical educational objective that aligns with the demands of an increasingly digital and complex global society (15–18).

Conceptually, CT encompasses a range of interrelated cognitive processes, including abstraction, decomposition, algorithmic thinking, and pattern recognition or generalization (11, 16, 19, 20). These

components collectively enable individuals to break down complex problems, identify underlying patterns, and construct step-by-step solutions. The theoretical roots of CT can also be traced to constructionist learning theory, particularly the work of Seymour Papert, who emphasized learning through active engagement and programming as a means of developing procedural thinking (21, 22) While Papert's work laid the groundwork, CT has evolved into a broader construct that integrates multiple cognitive and computational processes (2).

Despite its growing importance, the measurement of CT remains a significant challenge. Existing approaches often rely on closed-ended questionnaires; however, the availability of psychometrically robust instruments that account for contextual and cultural differences is still limited (12, 23). This limitation is particularly evident in non-ICT fields, such as the social sciences, where CT has not been extensively operationalized despite its potential applicability. The lack of valid and reliable measurement tools restricts the ability to accurately assess CT skills and, consequently, to design effective educational interventions.

The development of a high-quality measurement instrument necessitates rigorous attention to psychometric properties, particularly validity and reliability. Validity refers to the extent to which an instrument accurately measures the intended construct, whereas reliability reflects the consistency and stability of measurement results (24–26). These properties are fundamental to ensuring that the data generated can be used confidently for research, policy-making, and educational decision-making. Without adequate psychometric quality, measurement outcomes may lead to inaccurate conclusions and ineffective interventions.

To address these challenges, contemporary research increasingly advocates for the integration of multiple analytical approaches in instrument validation. In particular, the combination of CFA and the Rasch measurement model offers a comprehensive framework for evaluating both structural and measurement properties (27–29). CFA enables researchers to test the theoretical factor structure and examine the relationships between observed indicators and latent constructs (29–31), while the Rasch model provides item-level analysis, ensuring unidimensionality, local independence, and invariant measurement (32–35). Importantly, the Rasch model transforms ordinal Likert-scale data into interval-level measurements, thereby enhancing the precision and interpretability of results (36).

Grounded in these methodological considerations, the present study aims to develop and validate a CT measurement instrument with strong psychometric properties. The instrument is designed to achieve high reliability (≥ 0.80) and adequate item discrimination, supported by a clear factorial structure that explains a substantial proportion of variance (25). Furthermore, the integration of Rasch analysis ensures that the instrument meets the requirements of modern measurement theory, including proper scale functioning and accurate estimation of both item difficulty and person ability.

Accordingly, the central research question guiding this study is: *How are the psychometric properties of the computational thinking measurement tool after being analyzed using CFA and the Rasch model?*

Addressing this question is essential for advancing the field of educational measurement and for providing a robust tool that can be used to assess CT skills among university students.

The significance of this study lies in its potential practical implications. A valid and reliable CT measurement instrument can serve as an early diagnostic tool for identifying students' strengths and weaknesses in computational thinking. This information can inform targeted interventions, curriculum design, and instructional strategies aimed at enhancing students' cognitive and problem-solving abilities. Moreover, understanding students' CT levels can help educators design learning environments that promote engagement, motivation, and academic success.

In conclusion, while CT has been widely acknowledged as a critical competency for the modern era, its measurement remains underdeveloped, particularly in non-technical contexts. This study addresses this gap by developing a psychometrically sound instrument grounded in both CFA and Rasch modeling.

Computational thinking concept

CT has acknowledged numerous attention as a necessary skill in our global society for adequately explaining and predicting common events that require computational solutions (37). CT is perceived as an important competency to deal with computing in everyday life (38, 39). It encourages students from all science areas to think in innovative ways (40). Wing (2008) employed the term of CT to define a problem-solving strategy that relies on core computer concepts. It involved three significant foundations: abstraction, algorithms, and automation (19, 41). It is a baseline ability and skill set individuals should have and develop to tackle confusing, complicated, and wide-ranging issues for the future environment in diverse circumstances (42). It comprises the ideas of abstraction, problem-solving, investigation, decomposition, incorporation, algorithmic-thinking, coding, generalization, and debugging (43).

This thinking skill allowed individuals to design computational answers to problems by combining their cognitive abilities and the reasoning processes of computer scientists (44). It is a style of analytical thinking for solving difficulties that entails defining issues using a proper expressive structure, abstracting concepts, and eventually designing computers solutions to the issues (45). It is considered one of the essential 21st century skill for any disciplines (15, 16). CT could be used in a variety of courses across grade levels, offering both obstacles and possibilities (14). CT should be included not just into undergraduate courses at universities, but also at the primary through high school levels of education (2).

In general, CT adheres to the ideas of constructionism, Saymor Papert that a child actively constructs knowledge through contact with the environment, so inclined to provide opportunities to engage in hands-on experiences that fuel the creative process (21, 22). Papert's work with the LOGO programming language, as well as his book Mindstorms: Children, Computers, and Powerful Ideas (1980), demonstrate the principle of CT (21). He defined CT as work on the LOGO programming language and the concept of youngsters handling computers to acquire procedural thinking through programming (21). These two

perspectives share many similarities, however CT is a larger idea that incorporates concepts from Papert (2). CT is a set of skills to help solve problems (41). It is a problem-solving process (46) or thought process (19), algorithm-design and problem-solving strategies that possibly will be employed to address typical computing problems. It included decomposition, pattern recognition (as generalization), algorithmic thinking and abstraction (19, 20, 47, 48).

Hanid et al., (2022) concluded that there were substantial variations in students' computational thinking, visualization abilities, and accomplishment after employing the augmented reality application with CT versus traditional learning. CT is an important aspect in improving thinking scores, visualization skills, and achievement (49). CT aspect successfully improves students mastery and comprehension towards the course (8). Various research have demonstrated that incorporating CT into classrooms might create possibilities for both students and instructors to solve problems cooperatively (11, 12, 50). Moreover, effective treatments based on CT have the potential to contribute on more effective problem solving, and the impact will extend beyond higher education since the same interventions can be deployed in primary and secondary school (13).

CT is important in self-exploration and digital self-efficacy, both of which have an impact on learning satisfaction (51). Student achievement correlated with students CT skill (9). The findings emphasize the significance of students establishing applied CT procedural skills to improve their assignment quality (9). Meanwhile, according to Taub et al., (2015), the benefits of CT from the instructors perspective is to provide significant assistances for the teacher in producing ideas to deal with a variety of challenges.

Shute et al., (2017) and Ezeamuzie & Leung, (2021) identified various popular keywords of CT categories. They classified CT into five categories: (1) Abstraction (2) Decomposition (3) Algorithmic Thinking and (4) Pattern recognition/generalization.

- Abstraction (ignoring unnecessary detail)

It is a mental process used to develop structured thinking (54). Patterning a problem into smaller components and then identifying characteristics are examples of abstraction (11). Ignoring unnecessary elements in favor of concentrating on crucial facts. It allows individuals to focus on the necessary, relevant, and significant aspects (55).

- Decomposition (breaking down into parts)

It entails identifying the structure of the problem and how each of the parts will interact together in the final solution. When decomposition is done correctly, it is easier to alter the solution later by modifying each aspect, and it also allows for the reuse of aspects in other issues' solutions. It is the breaking down of an issue into smaller pieces. Long before the introduction of computational thinking, challenge decomposition was recognized as a generic problem-solving method (56). The complexity of a work can be reduced by recognizing smaller sections that can be addressed independently (40).

- Algorithmic thinking (considering the order of steps)

An algorithm is a well-defined technique that describes how to use resources to achieve a particular goal (57). It included creating an algorithm to create step-by-step instructions for addressing the problem. Beginning with what is currently known and going broader, plan how to overcome the problem. It allows individuals to transform their abstract concepts and patterns into a collection of processes.

- Pattern recognition (generalization)

It is an action that seeks possibilities to apply the answer in new circumstances (11). It involves the ability to deduce principles from observations and apply those principles to previously unknown circumstances. Developing components that can be employed in multiple scenarios.

METHOD

The CT instrument used in this study was adapted from the scale developed by Alim (2024), consisting of 22 items designed to assess university students' CT skills. The instrument demonstrated good internal consistency, with a reliability coefficient of 0.88, indicating its suitability for research purposes. Responses were recorded using a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree).

Data were collected anonymously from undergraduate students enrolled in several universities in Makassar, Indonesia. The data collection process was conducted directly in classroom settings to ensure consistency in administration and to minimize potential response bias. Prior to completing the questionnaire, all participants were provided with an informed consent statement explaining the purpose of the study, the voluntary nature of participation, and the confidentiality of their responses.

This study was conducted in two sequential stages of instrument validation. Each stage was completed before proceeding to the next stage, and only items meeting the established psychometric criteria were retained for subsequent analyses.

In the first stage, the factorial validity of the instrument was evaluated using CFA. Data were collected from 364 undergraduate students, and the analysis was performed to examine the extent to which the proposed measurement model fit the observed data. Items with insufficient factor loadings or inadequate psychometric performance were removed to improve the overall construct validity of the instrument.

In the second stage, the refined instrument resulting from the CFA analysis was further evaluated using the Rasch Rating Scale Model (RSM). This stage involved a sample of 434 undergraduate students. Rasch analysis was employed to examine item and category functioning, reliability, separation indices, unidimensionality, local independence, and item difficulty indices. Only items that satisfied both CFA and Rasch measurement criteria were retained in the final version of the CT Instrument.

Confirmatory Factor Analysis (CFA)

CFA is a statistical technique within the framework of Structural Equation Modeling (SEM) used to evaluate the extent to which a theoretically specified measurement model fits the empirical data (59). This analysis is conducted to ensure that the distribution of items aligns with the underlying theoretical structure prior to proceeding with Rasch analysis. In the present study, a total of 22 items were examined using JASP version 0.12.2 (60).

Only high-quality items that adequately represent each component of CT skills were retained for further analysis. The criteria for item selection and model evaluation are as follows:

a. Factor Loadings

Each item must demonstrate a standardized factor loading greater than 0.50 and be statistically significant ($p < 0.05$), indicating that the item adequately represents the intended latent construct (61, 62).

b. Model Fit Indices

The overall model must meet acceptable fit criteria, as indicated by goodness-of-fit indices such as Comparative Fit Index (CFI), Goodness-of-Fit Index (GFI), Tucker–Lewis Index (TLI), Bentler–Bonett Non-Normed Fit Index (NNFI), and other related indices, all exceeding the threshold of 0.90 (61, 63–65).

c. Error of Approximation

The model should demonstrate low levels of error, as reflected by Root Mean Square Error of Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR), both of which should be below 0.08 (Cho et al., 2020; Sathyanarayana & Mohanasundaram, 2024).

d. Reliability

The instrument must exhibit satisfactory internal consistency, with McDonald's ω and Cronbach's α values exceeding 0.80 for the overall scale and above 0.60 for each individual dimension (66–68).

Rasch model analysis

Rasch analysis is one of the item analysis approaches within the framework of Item Response Theory (IRT), which is used to estimate item parameters and individual abilities simultaneously on a common logit scale. This model enables the transformation of ordinal data (e.g., Likert-scale responses) into interval-level measurements, thereby enhancing the precision and interpretability of the results in quantitative terms (32, 69–71). By placing both item difficulty and person ability on the same metric, the Rasch model provides a robust foundation for evaluating the quality of measurement instruments.

In this study, the Rasch analysis was conducted using J.Metrik version 4.1.1 (72). The analysis focuses on several key aspects to ensure the psychometric soundness of the instrument, including: (1) testing fundamental assumptions such as unidimensionality and local independence, (2) evaluating item and person reliability as indicators of measurement consistency, (3) examining model fit through Infit and Outfit Mean Square statistics, (4) analyzing the functioning of the rating scale, and (5) estimating item difficulty to assess the distribution of items across the latent trait continuum.

a. Rasch model assumptions

Local independence refers to the assumption that a respondent's answer to a given item is not influenced by their responses to other items once the underlying latent ability has been controlled for. In other words, any association between items should be fully explained by the latent trait being measured, and there should be no additional dependency among items. This assumption is typically evaluated by examining residual correlations between item pairs. Ideally, these correlations should remain low, indicating that items function independently after accounting for the latent variable. A commonly accepted guideline is that residual correlations should not exceed 0.40, which corresponds to approximately 16% shared variance. When this criterion is satisfied, it suggests that the instrument does not contain redundant or overly similar items that could bias the measurement. Thus, meeting the local independence assumption supports the validity of the Rasch model and ensures that each item contributes uniquely to the assessment of the construct (73–75).

Unidimensionality refers to the assumption that all items within an instrument measure a single underlying latent construct. In the context of Rasch analysis, this assumption is typically evaluated using a Principal Component Analysis (PCA) of residuals. Evidence of unidimensionality is supported when the eigenvalue of the first contrast (first residual factor) is less than 3.0, indicating that no substantial secondary dimension is present beyond the primary construct being measured. When this criterion is met, it suggests that the variance explained by potential secondary factors is minimal and does not meaningfully threaten the dominance of the main latent trait. Consequently, the instrument can be considered to have a coherent internal structure, with all items contributing to the measurement of a single construct. This is essential for ensuring the validity and interpretability of the measurement results within the Rasch model framework (73, 75–77).

b. Rating scale threshold

A well-functioning rating scale is characterized by a monotonic increase across response categories, such that each successive category represents a higher level of the underlying construct. This orderly progression indicates that respondents are using the scale in a meaningful and consistent manner, with each category reflecting a distinct level of the latent trait (78–81).

In addition, each response category should be supported by an adequate frequency of use to ensure stable parameter estimation. The average measures for each category are also expected to increase

progressively, demonstrating that higher categories correspond to higher levels of the construct. Furthermore, the threshold parameters (Andrich thresholds) should be ordered logically without any evidence of disordering, indicating that transitions between adjacent categories occur as intended (78, 79, 81). When these criteria are met, the rating scale can be considered to function optimally, providing clear differentiation between response options and contributing to the overall validity and precision of the measurement instrument.

c. Reliability test

Reliability refers to the degree of consistency and stability in measurement, both in distinguishing between individuals (person reliability) and in maintaining a consistent hierarchy of item difficulty (item reliability). Person reliability indicates how well the instrument can differentiate respondents across varying levels of the latent trait, while item reliability reflects the stability of item difficulty estimates across different samples (67).

In the context of Rasch analysis, a reliability coefficient above 0.80 is generally considered desirable, as it indicates that the instrument can reliably and consistently measure differences in respondents' abilities as well as accurately position items along the difficulty continuum. High reliability values suggest that the measurement results are replicable and not substantially influenced by random error (82–84). Consequently, instruments that achieve this threshold are regarded as having strong measurement precision and are suitable for both research and assessment purposes.

d. Model fit estimation

Model fit refers to the degree of correspondence between the observed empirical data and the expectations of the Rasch model, indicating how well individuals' responses to items conform to the probabilistic structure of the model. In other words, it reflects whether the response patterns are consistent with the assumption that item difficulty and person ability interact in a predictable and systematic manner (32, 69, 71, 73).

A commonly used criterion for evaluating model fit in Rasch analysis is based on the Infit and Outfit Mean Square statistics. Acceptable values typically fall within the range of 0.5 to 1.5. Values below 0.5 may indicate redundancy or overpredictability, suggesting that the item provides limited new information. In contrast, values above 1.5 may signal excessive noise or unpredictability in the data, indicating that the item or respondent does not behave consistently with the model expectations (32, 71).

Therefore, items and persons with fit statistics within the acceptable range can be considered to contribute meaningfully to the measurement process, while those outside the range may require further examination, revision, or potential removal to ensure the overall quality and validity of the instrument.

e. Item difficulty estimation

Item difficulty is a parameter that reflects the level of difficulty of an item in measuring a latent construct, typically expressed in logit units. This parameter indicates the point along the latent trait continuum at which a respondent has a 50% probability of endorsing or correctly responding to the item. In the Rasch model framework, item difficulty plays a crucial role in defining the measurement scale and establishing the relative ordering of items from easiest to most difficult (85, 86).

An optimal distribution of item difficulty is expected to be well-spread across the full range of respondents' abilities. Such a distribution ensures that the instrument can effectively capture variability among individuals at low, moderate, and high levels of the latent trait. When items are evenly distributed, the measurement precision is enhanced, as the instrument provides meaningful information across the entire continuum. Conversely, if item difficulty is clustered within a narrow range, the instrument may only be sensitive to a specific segment of respondents, thereby limiting its ability to accurately assess individuals at the extreme ends of the ability spectrum.

RESULT

Confirmatory factor analysis (CFA)

Model fit was evaluated using multiple goodness-of-fit indices, including the Comparative Fit Index (CFI), Goodness-of-Fit Index (GFI), Tucker–Lewis Index (TLI), Bentler–Bonett Non-Normed Fit Index (NNFI), Bollen's Incremental Fit Index (IFI), and the Relative Noncentrality Index (RNI). All of these indices exceeded the recommended threshold of 0.90, indicating that the hypothesized measurement model for CT demonstrates a satisfactory fit to the observed data.

Furthermore, the error-based indices, namely the Root Mean Square Error of Approximation (RMSEA) and the Standardized Root Mean Square Residual (SRMR), were both below 0.06 – see Table 1. These results suggest that the model exhibits a low level of approximation error and residual discrepancy, indicating that the measurement errors are minimal and within acceptable standards. Overall, these findings provide strong evidence that the instrument possesses an adequate model fit and can reliably represent the underlying construct of computational thinking.

Table 1. Fit indices information	
Index	Value
Comparative Fit Index (CFI)	0.922
Goodness of fit index (GFI)	0.918
Tucker-Lewis Index (TLI)	0.908
Bentler-Bonett Non-normed Fit Index (NNFI)	0.908
Bollen's Incremental Fit Index (IFI)	0.923
Relative Noncentrality Index (RNI)	0.922
Root mean square error of approximation (RMSEA)	0.057
Standardized root mean square residual (SRMR)	0.046

Furthermore, the results of the factor loading analysis showed that the values ranged from 0.52 (abs6) to 0.75 (algo20) – see Table 2. These findings indicate that each item demonstrates an adequate level of association with its respective latent construct, suggesting that all items function as intended in representing the underlying dimensions. In addition, the number of items varies across the dimensions. The first component, abstraction, consists of three items, while the last component, generalization, comprises five items. Meanwhile, both the decomposition and algorithmic thinking dimensions consist of six items each. This distribution reflects the structural composition of the instrument, with each dimension being represented by a sufficient number of indicators to capture the construct adequately.

Table 2
Factor loadings and reliability

Factor	Indicator	Estimate	Reliability		
			McDonald's ω	Cronbach's α	Overall
Abstraction	abs1	0.579			
	abs4	0.686	0.622	0.617	0.89
	abs5	0.517			
Decomposition	dec6	0.659			
	dec7	0.564			
	dec8	0.552	0.788	0.787	
	dec9	0.689			
	dec10	0.615			
	dec11	0.623			
Algorithmic	algo12	0.684			
	algo13	0.749			
	algo14	0.684	0.855	0.855	
	algo15	0.701			
	algo16	0.742			
	algo17	0.720			
Generalization	gen18	0.622			
	gen19	0.713			
	gen20	0.599	0.788	0.786	
	gen21	0.611			
	gen22	0.632			

Furthermore, the overall reliability coefficient reached 0.89, as indicated by both McDonald’s ω and Cronbach’s α, demonstrating a high level of internal consistency across the instrument. This consistency suggests that the items collectively provide a stable and reliable measurement of the underlying construct. At the dimensional level, all subscales exceeded the acceptable cutoff values. Specifically, the abstraction dimension yielded a reliability coefficient of 0.62, which can be considered acceptable for an exploratory or newly developed scale. The decomposition dimension showed a reliability of 0.79, while the algorithmic thinking dimension demonstrated a strong reliability of 0.85. Similarly, the generalization

dimension achieved a reliability coefficient of 0.79. These findings, as presented in Table 2, indicate that each dimension contributes meaningfully to the overall construct and exhibits adequate to strong internal consistency.

However, at this stage, two items were excluded as they did not meet the required factor loading criteria, namely items 3 and 4 within the first dimension, abstraction. The removal of these items was necessary to ensure that only indicators with adequate psychometric properties were retained in the measurement model. As a result, a total of 20 items remained and were considered suitable for further analysis. The items that met the validity criteria were subsequently carried forward to the next stage, namely the Rasch analysis. This step was intended to further evaluate the measurement properties of the instrument at the item and person levels.

The model plot presented on Fig. 1 provides a visual representation of the distribution of items across each dimension, accompanied by their respective parameter estimates. This visualization facilitates a clearer understanding of how each item contributes to the latent constructs and supports the interpretation of the overall measurement structure.

Rasch model analysis

a. Rasch assumption

The Rasch model analysis revealed that all 20 items satisfied the established model fit criteria, indicating that each item functioned consistently with the expectations of the Rasch measurement model. This finding suggests that the items contribute meaningfully to the measurement of the underlying construct and do not exhibit substantial misfit that would compromise the validity of the scale. Furthermore, the assessment of the unidimensionality assumption showed that the eigenvalue of the first contrast was 2.54, which is below the commonly accepted threshold of 3.0 (32, 87). This result provides evidence that the instrument predominantly measures a single latent dimension. In this context, the findings support the conclusion that the scale is unidimensional and specifically captures the construct of CT skills without the presence of significant secondary dimensions. Overall, these results indicate that the instrument demonstrates a satisfactory internal structure, with items that align well with the Rasch model and collectively represent a coherent and unidimensional measurement of CT abilities.

The analysis of the local independence assumption indicated that the highest observed residual correlation was 0.400 (pat4 and pat5), which remains within the acceptable limit for evaluating local independence in the Rasch model framework (69, 70, 88–91). These findings suggest that the responses to each item were not influenced by responses to other items after controlling for the underlying latent trait. In other words, there is no evidence of substantial local dependency among the items. Therefore, it can be concluded that the assumption of local independence is satisfactorily met, indicating that each item contributes independently to the measurement of the construct without redundancy or overlap beyond what is explained by the latent variable.

b. Model fit and Rating scale analysis

Moreover, none of the Infit and Outfit Mean Square statistics for the items fell below the lower bound of 0.5 or exceeded the upper threshold of 1.5, indicating that all items fit well within the expectations of the Rasch model – see Table 3. This result suggests that the response patterns are consistent with the probabilistic assumptions of the model, and that no items exhibit substantial noise or redundancy.

In addition, the rating scale analysis demonstrated that participants were able to effectively utilize all five response categories. Each category showed a monotonically increasing pattern in the average measures, indicating a logical progression across the scale. This pattern confirms that respondents could clearly distinguish between the different rating categories, which in turn supports the accuracy and validity of their responses. Consequently, the functioning of the rating scale can be considered optimal, as it allows for meaningful differentiation in participants' levels of the underlying construct.

c. Reliability and Separation Index

Additionally, the analysis demonstrated that the item reliability coefficient reached 0.98, which falls within the excellent category, indicating a very high level of consistency in the hierarchical ordering of item difficulty across samples. This suggests that the set of items is well-targeted and sufficiently diverse to define a stable measurement continuum. Meanwhile, the person reliability was 0.87, which can be considered good, reflecting the instrument's ability to consistently distinguish between individuals with different levels of the underlying trait. This level of reliability indicates that the scale is effective in capturing variability in respondents' abilities.

Furthermore, the separation index, expressed in terms of strata, showed that the number of strata for items was 9.88, categorized as excellent. This implies that the instrument is capable of differentiating item difficulty into multiple distinct levels. For persons, the strata value was 3.80, which falls into the moderate category, indicating that the instrument can classify respondents into approximately four distinct ability groups. Overall, these results provide strong evidence of the measurement precision and discriminative power of the instrument, both in terms of item calibration and person differentiation – see Table 3.

d. Item difficulty

The distribution of item difficulty ranged from − 0.79 (dec3) to 0.82 (algo3), indicating that the items were primarily clustered within the moderate range of the latent trait. This pattern suggests that the instrument is well targeted for medium levels of CT skills. However, the absence of more difficult items implies that individuals with higher levels of ability may not be measured with optimal precision, as the instrument provides limited coverage at the upper end of the continuum, an issue that is also showed in Fig. 2.

Table 3. Summary of the Item Information

Item	Difficulty	WMS	UMS	Reliability		Number of Strata	
				Items	Persons	Items	Persons
abs1	-0.56	0.95	1.06	0.98	0.87	9.89	3.80
abs4	-0.27	1.01	1.10				
abs5	0.46	0.89	0.92				
dec1	-0.68	0.85	0.83				
dec2	-0.34	0.88	0.88				
dec3	-0.79	1.44	1.48				
dec4	-0.03	0.90	0.89				
dec5	-0.56	1.19	1.13				
dec6	-0.76	1.14	1.06				
alg1	0.08	0.85	0.83				
alg2	0.15	0.76	0.75				
alg3	0.82	1.00	1.08				
alg4	0.09	0.86	0.85				
alg5	0.26	0.90	0.89				
alg6	-0.02	0.82	0.82				
pat1	0.47	1.07	1.10				
Pat2	0.41	1.12	1.12				
Pat3	-0.30	1.22	1.17				
Pat4	0.80	1.29	1.36				
Pat5	0.78	1.05	1.09				

Overall, the instrument can be effectively used to assess the level of CT among university students, particularly those who fall within the moderate category of the construct. This is reflected in the distribution of item difficulty, where most items are concentrated within the low to moderate difficulty range, while relatively few items represent the moderate to high difficulty levels.

As a result, the instrument is more sensitive in capturing variations among individuals with average levels of computational thinking, as it provides more information within this range of the latent trait. Conversely, its ability to accurately differentiate individuals with higher levels of CT is somewhat limited due to the scarcity of more challenging items. Therefore, the measurement precision of this instrument

is maximized when the target respondents are predominantly within the moderate category of CT ability. The finalized CT instrument presented in Table 4 is intended for further use in research and practical applications.

Table 4
Validated Items Based on CFA and Rasch Measurement Model

Aspect	English	Indonesian
Abstraction	♣ I can determine the distinction between information that is useful and that is not.	♣ Saya bisa membedakan antara informasi yang bermanfaat dan yang kurang/tidak bermanfaat
	♣ I concentrate on things that are directly related to a problem I am dealing with.	♣ Saya memperhatikan hal-hal yang berkaitan langsung dengan masalah yang saya hadapi
	♣ I can solve a problem by finding the best solution.	♣ Saya dapat menemukan cara yang tepat untuk menyelesaikan suatu masalah.
Decomposition	♣ I consider that large problems can be solved if they are described in detail.	♣ Saya menganggap bahwa masalah yang besar dapat diselesaikan jika diuraikan secara rinci/detail
	♣ I will concentrate on the primary factors that contribute to a problem.	♣ Saya fokus pada penyebab utama terjadinya suatu masalah
	♣ I see big problems as a collection of small problems.	♣ Saya melihat bahwa masalah besar terbentuk dari tumpukan masalah-masalah kecil
	♣ If I describe a problem, I can quickly understand it.	♣ Saya dapat memahami setiap masalah dengan mudah jika saya menguraikan masalah tersebut.
	♣ There are no complicated problems if we can see them from different perspectives.	♣ Tidak ada masalah yang rumit jika kita mampu melihat dari sudut pandang yang berbeda
	♣ If we understand the details of the problem, heavy problems will seem light.	♣ Masalah yang berat akan terasa ringan jika kita faham secara detail masalah tersebut
Algorithmic thinking	♣ Before I solve a problem, I organize the steps to solve the problem.	♣ Sebelum menyelesaikan suatu masalah, saya terlebihdahulu menyusun langkah-langkah penyelesaian masalah tersebut.
	♣ I will stick to the problem-solving strategy I created earlier.	♣ Saya akan mengikuti pola penyelesaian masalah yang telah saya rencanakan sebelumnya
	♣ I am used to solving problems in a systematic manner.	♣ Saya terbiasa menyelesaikan masalah secara sistematis
	♣ I am used to working through problems step by step.	♣ Saya menyelesaikan masalah secara bertahap
	♣ Before attempting to solve a problem, I devise a strategy for doing so.	♣ Sebelum menyelesaikan masalah, saya terlebihdahulu menyusun rencana penyelesaian masalah tersebut

Aspect	English	Indonesian
	♣ My mind guides me step by step in problem solving.	♣ Pikiran saya menuntun saya dalam memecahkan masalah secara bertahap
Pattern recognition	♣ I can see similarities in every problem I encounter.	♣ Saya dapat melihat kesamaan pada setiap masalah yang saya hadapi
	♣ I tend to look for general patterns in every problem I face	♣ Saya mencari pola yang sama pada setiap masalah yang saya hadapi
	♣ I solve a problem by applying the solution to a previous problem.	♣ Saya menggunakan pengalaman sebelumnya untuk menyelesaikan masalah saat ini.
	♣ I believe that every problem is connected.	♣ Saya menganggap setiap masalah saling memiliki keterkaitan
	♣ Current and previous problems often have something in common.	♣ Masalah saat ini dan masalah yang telah berlalu cenderung memiliki kesamaan

DISCUSSION

The present study offers a comprehensive evaluation of the psychometric properties of the CT instrument by integrating two complementary analytical frameworks, namely CFA and the Rasch measurement model. This dual-approach strategy strengthens the validity argument by combining covariance-based evidence of structural validity with item-level probabilistic modeling, as recommended in contemporary measurement literature (32, 71, 92, 93).

From the CFA perspective, the measurement model demonstrated a consistently satisfactory fit across a wide range of goodness-of-fit indices. Specifically, the Comparative Fit Index (CFI), Goodness-of-Fit Index (GFI), Tucker–Lewis Index (TLI/NNFI), Incremental Fit Index (IFI), and Relative Noncentrality Index (RNI) all exceeded the recommended threshold of 0.90. These findings indicate that the hypothesized factor structure is well supported by the empirical data and that the model adequately reproduces the observed covariance matrix. Such results align with widely accepted guidelines suggesting that values above 0.90 reflect acceptable fit, while values approaching 0.95 indicate a more robust model fit (30, 94).

Moreover, the error-based fit indices further reinforce this conclusion. The Root Mean Square Error of Approximation (RMSEA) and the Standardized Root Mean Square Residual (SRMR) were both below 0.06, indicating a low level of approximation error and minimal residual discrepancies. These findings suggest that the model does not suffer from substantial misspecification and that the measurement errors are within acceptable limits. Collectively, the CFA results provide strong evidence of structural validity, indicating that the instrument is capable of capturing the latent construct of CT in a theoretically coherent manner.

The factor loading estimates, ranging from 0.52 to 0.75, further support the adequacy of the measurement model. Loadings above 0.50 are generally considered practically significant, indicating that each item contributes meaningfully to its respective latent (95). The removal of two items from the abstraction dimension was undertaken to ensure the psychometric quality of the instrument. This decision reflects adherence to established measurement principles, which emphasize that each item should contribute meaningfully to the representation of the underlying latent construct and demonstrate a substantial relationship with the dimension it is intended to measure (96). Specifically, Item 2 *(In order to solve a problem, I will only focus on the most important points)* and Item 3 *(I only think about things that I consider useful)* demonstrated factor loadings below the recommended threshold of 0.50, indicating a weak association with the abstraction construct. According to Hair et al. (2022), factor loadings below 0.50 suggest that an item explains less than 25% of the variance in the latent variable, thereby reducing its contribution to construct measurement and potentially compromising the validity of the scale.

A plausible explanation for the low loadings lies in both the wording and conceptual focus of the items. Item 2 primarily emphasizes the act of prioritizing important information when solving a problem. Although this behavior is related to information selection, it does not fully capture the essence of abstraction as defined within computational thinking. Abstraction involves reducing complexity by identifying underlying patterns, relationships, and essential characteristics, and then constructing generalized representations that can be applied across different situations (41). Consequently, respondents may interpret this item merely as a strategy for focusing attention on salient information rather than as an indication of their ability to abstract and generalize from complex problems.

Similarly, Item 3 emphasizes attending only to information that respondents personally perceive as useful. The phrase “I consider useful” introduces a subjective evaluative component that may vary considerably across individuals, making responses more reflective of personal preferences, interests, or decision-making styles than of abstraction ability. As a result, the item may capture selective attention or information-filtering behaviors rather than the cognitive process of abstraction itself. This interpretation is problematic because abstraction in CT requires individuals not only to ignore irrelevant information but also to create simplified conceptual models, identify common structures, and transfer these representations across contexts (4, 11).

From a psychometric perspective, the low factor loadings suggest that the two items share relatively little common variance with the abstraction dimension compared to other indicators. Their weak contribution may stem from the fact that they assess peripheral problem-solving behaviors rather than the core attributes of abstraction. Consequently, retaining these items could introduce construct-irrelevant variance and reduce the conceptual coherence of the dimension. Their removal therefore strengthens the dimensional integrity of the abstraction construct, improves the overall measurement quality of the instrument, and enhances the validity and interpretability of the resulting scores. This refinement process is consistent with best practices in scale development, which recommend retaining only those items that demonstrate both statistical adequacy and theoretical alignment with the intended construct (96, 97).

All the items chosen in each factor were the best and correspond to the theoretical concept of CT skills. Previous researchers, such as Gertsen et al. (2022), who developed a moral authorship questionnaire measurement tool, and Dang et al. (2023), who constructed the school level environment questionnaire, used the same criteria to increase the model's fit, all items with poor psychometric properties were deleted. This approach will lead to a reduction in the number of items, retaining only the best-performing ones. This is consistent with Alim's study, which conducted a short-version construction of the big-five measurement tool and achieved high property quality despite the restricted number of items (100). This refinement process is consistent with best practices in scale development, where poorly performing indicators are excluded to enhance construct representation and internal consistency (101).

In terms of reliability, the instrument demonstrated a high level of internal consistency, with both McDonald's ω and Cronbach's α reaching 0.89. This level exceeds the commonly accepted threshold of 0.70, indicating that the items consistently measure the same underlying construct (102). At the subscale level, reliability coefficients ranged from 0.62 to 0.85. While the abstraction dimension yielded a relatively lower reliability (0.62), it remains acceptable given the limited number of items and the exploratory nature of this dimension (68). The other dimensions, decomposition, algorithmic thinking, and generalization demonstrated moderate to strong reliability, further confirming the multidimensional integrity of the construct.

Building on the CFA findings, the Rasch model analysis provides deeper insights into the measurement functioning at both item and person levels. All 20 retained items satisfied the Rasch model fit criteria, with Infit and Outfit Mean Square statistics falling within the acceptable range of 0.5 to 1.5. This indicates that the observed response patterns align well with model expectations and that no items exhibit excessive noise or redundancy (103). Such results support the internal construct validity of the instrument from a probabilistic measurement perspective.

The assessment of unidimensionality revealed that the eigenvalue of the first contrast (2.54) was below the threshold of 3.0, suggesting that the scale predominantly measures a single latent trait. This finding is particularly noteworthy given that the CFA results support a multidimensional structure. The apparent convergence between these findings can be interpreted within the framework of essential unidimensionality, where multiple related dimensions collectively reflect a dominant overarching construct (32, 104). In this case, abstraction, decomposition, algorithmic thinking, and generalization can be viewed as interrelated facets of computational thinking, justifying the use of a unidimensional Rasch model.

The local independence assumption was also satisfied. Residual correlation analysis indicated that most item pairs exhibited low residual correlations. Only one item pair (Pat4–Pat5) showed a residual correlation slightly above the commonly used threshold, suggesting minor local dependence. Previous studies have suggested that residual correlations exceeding 0.30–0.40 may indicate potential local item dependence and therefore warrant further examination (32, 71). However, because both items belonged to the same theoretical dimension and no systematic pattern of elevated residual correlations was

observed across the instrument, the assumption of local independence was considered adequately satisfied. This interpretation is consistent with recommendations that evidence of local dependence should be evaluated in conjunction with substantive theory and the overall pattern of residual correlations rather than relying solely on a single item pair (105).

In terms of rating scale functioning, the analysis demonstrated that all five response categories operated as intended. The monotonically increasing average measures and well-defined category probability curves indicate that respondents were able to distinguish meaningfully between response options. This finding is critical, as poorly functioning rating scales can compromise measurement precision (103). The current results confirm that the Likert scale structure is appropriate and contributes to the overall validity of the instrument.

The reliability evidence from the Rasch model further complements the classical reliability findings. The item reliability coefficient of 0.98 indicates excellent stability in the ordering of item difficulty, while the person reliability of 0.87 reflects a strong ability to differentiate individuals based on their CT ability. Additionally, the separation indices (9.89 for items and 3.80 for persons) suggest that the instrument can distinguish multiple strata of item difficulty and respondent ability. These results demonstrate a high level of measurement precision and align with recommendations for well-functioning Rasch instruments (36).

However, the analysis of item difficulty distribution reveals an important limitation. The items are primarily concentrated within the low to moderate difficulty range (− 0.79 to 0.82), indicating that the instrument is optimally targeted for individuals with moderate levels of computational thinking. While this enhances sensitivity within this range, it limits the instrument's ability to accurately measure individuals at higher ability levels. This issue of targeting is well-documented in Rasch measurement, where the alignment between item difficulty and person ability is crucial for maximizing measurement information (71). These findings were in accordance with investigation undertaken by Denovan et al (2022), revealed that Ego Resiliency Scale-Revised demonstrated relatively robust psychometric properties within a general population sample. Future research should therefore consider developing more challenging items to improve coverage at the upper end of the latent continuum.

The integration of CFA and Rasch measurement modeling provides a robust and complementary framework for validating the CT instrument. While CFA and Rasch analysis are grounded in different measurement paradigms, their combined use enables a more comprehensive evaluation of construct validity than either approach alone. CFA contributes evidence regarding the theoretical structure of the instrument by examining the relationships between observed indicators and their underlying latent dimensions. The satisfactory model fit indices, factor loadings, and reliability estimates obtained from the CFA support the multidimensional structure of CT and demonstrate that the retained items adequately represent their intended constructs. In addition, CFA provides evidence of convergent and discriminant validity, thereby confirming that the dimensions of abstraction, decomposition, algorithmic thinking, and pattern recognition function as theoretically distinct yet related components of CT (97,

107). Consequently, the CFA findings establish the structural integrity of the instrument and provide support for the theoretical assumptions underlying its development.

Complementing these findings, Rasch analysis offers a more detailed examination of the measurement properties of individual items and the functioning of the scale as a whole. Unlike CFA, which focuses primarily on the covariance structure among items, Rasch modeling evaluates whether each item operates consistently with the expectations of a probabilistic measurement model, thereby providing evidence of item fit, category functioning, person and item reliability, separation indices, local independence, and unidimensionality (105, 108). The Rasch results indicate that the retained items function appropriately across respondents, the rating scale categories operate in an ordered manner, and the instrument demonstrates satisfactory measurement precision. Furthermore, the person–item targeting analysis reveals that the scale is particularly effective for measuring individuals with moderate levels of computational thinking.

The convergence of evidence obtained from CFA and Rasch analysis therefore provides strong support for the validity, reliability of the instrument. Nevertheless, the targeting results suggest that the instrument contains relatively few highly difficult items, limiting its ability to discriminate among respondents with advanced levels of computational thinking.

CONCLUSION

CT is a core 21st-century competency that supports effective, structured, and creative problem-solving across disciplines. This study provides strong evidence for the psychometric robustness of the CT Measurement Tool, which comprises 20 items across four dimensions: abstraction, decomposition, algorithmic thinking, and generalization/pattern recognition. Overall, the findings demonstrate that the instrument possesses satisfactory psychometric properties and can be considered a valid and reliable tool for assessing CT among university students. Evidence from both CFA and Rasch analyses supports the measurement quality, and overall functionality of the scale.

The results from CFA indicate a well-fitting measurement model with satisfactory factor loadings and high internal consistency (ω and α = 0.89), confirming the structural validity of the instrument. Furthermore, Rasch model analysis demonstrated that all items fit the model expectations, satisfy unidimensionality and local independence assumptions, and exhibit good rating scale functioning. High item (0.97) and person reliability (0.88), along with strong separation indices, further confirm the instrument's precision and discriminative capacity.

However, the distribution of item difficulty indicates that the instrument is most sensitive for assessing individuals with moderate levels of computational thinking, while providing relatively limited measurement coverage for respondents with higher levels of computational thinking. This finding suggests that the instrument is effective in differentiating among individuals with average CT skill but may be less capable of distinguishing those with more advanced. Consequently, the development of additional high-difficulty items is necessary to improve the instrument's targeting and enhance its ability

to capture the full range of CT skills. Such refinement would increase measurement precision and strengthen the instrument's capacity to assess individuals across varying levels of ability.

Future studies should focus on expanding item difficulty levels, incorporating broader cognitive dimensions of computational thinking, and conducting additional Rasch-based analyses, such as Differential Item Functioning (DIF), to examine whether items function consistently across demographic groups, including gender, age, and other relevant characteristics. It will provide stronger evidence regarding the absence of item bias and support the broader applicability of the instrument across diverse populations.

Clinical trial number

Not applicable

Human Ethics and Consent to Participate declarations

This study was approved by the Ethics Committee of Universitas Bosowa, Indonesia (Approval No. 31/B.02/DEH/UNIBOS/VI/2026). Informed consent was obtained from all participants prior to data collection.

Consent to Publish

declaration: Not applicable

Declarations

Clinical trial number: Not applicable

Human Ethics and Consent to Participate declarations: This study was approved by the Ethics Committee of Universitas Bosowa, Indonesia (Approval No. 31/B.02/DEH/UNIBOS/VI/2026). Informed consent was obtained from all participants prior to data collection.

Consent to Publish declaration: Not applicable

Data availability: The de-identified datasets generated and analyzed during this study are available from the corresponding author upon reasonable request. Access to the data is subject to ethical considerations, institutional regulations, and the protection of participant confidentiality.

Authors' contribution: SA conceived and designed the study, conducted CFA and Rasch analyses, interpreted the findings, and drafted the original manuscript. SP contributed to the development of the study methodology, provided academic supervision throughout the research process, and critically revised the manuscript for important intellectual content. DN coordinated field activities, managed the research database, and performed data curation. ANAS was responsible for quality assurance procedures, including checking data completeness, and assisted in research administration. SS

coordinated participant recruitment, maintained study documentation, and facilitated communication during the data collection process. All authors collaboratively participated in field data collection, contributed to the interpretation of the findings, reviewed and approved the final version of the manuscript, and agree to be accountable for all aspects of the work, ensuring that questions related to the accuracy or integrity of any part of the study are appropriately investigated and resolved.

Funding Statement: This research was supported by the Ministry of Higher Education, Science, and Technology of the Republic of Indonesia under the Fundamental Research (PFR) scheme in 2026, through contract No. 288/C3/DT.05.00/PL-BARU/2026 and derivative contract No. 365/LL9/PPIP-PL.MONO/2026, PL.BARU 365-010/DIRPM-UNIBOS/IV/2026.

Acknowledgement: The authors would like to express their sincere gratitude to the Ministry of Higher Education, Science, and Technology of the Republic of Indonesia for supporting this research through the Fundamental Research (Penelitian Fundamental Reguler) scheme in 2026. The authors also acknowledge Universitas Bosowa for granting ethical approval and providing institutional support that facilitated the successful completion of this study. Finally, we extend our appreciation to all participants who generously contributed their time and responses to this research.

Statements and Declarations: No potential conflict of interest was reported by the authors.

References

1. Yadav A, Mayfeld C, Zhou N, Hambrusch S, Korb JT. Computational thinking in elementary and secondary teacher education. ACM Trans Comput Educ. 2014;14(1):1–16.
2. Wing J. Computational thinking and thinking about computing. Philos Trans R Soc Math Phys Eng Sci. 2008.
3. Kim Y, Kim J. Application of a software education program developed to improve computational thinking in elementary school girls. Indian J Sci Technol. 2016;9(44):1–9.
4. Grover S, Pea R. Computational thinking: A competency whose time has come. Computer science education: Perspectives on teaching and learning. Bloomsbury Academic; 2018.
5. Amarulloh RR, Iqbal NHM, Aswie V. Enhancing Deep Learning Through Computational Thinking-Based Assessment. Jentre. 2025.
6. Avazova ZS. Teaching critical thinking skills. 2026.
7. Han J. A Systematic Review of Computational Thinking Assessment in the Context of 21st Century Skills. In. 2023. pp. 271–83. Available from: https://www.atlantis-press.com/article/125988879.pdf
8. Angeli C, Giannakos M. Computational thinking education: Issues and challenges. Comput Hum Behav. 2020;(105):106–85.
9. Boom KD, Bower M, Siemon J, Arguel A. Relationships between computational thinking and the quality of computer programs. Educ Inf Technol. 2022;1–22.

10. Zhu Z, Zhang Y, Zhu W, Ma J. Computational Thinking and Academic Achievement: The Mediator Roles of Problem-solving Ability and Learning Anxiety. Int Symp Educ Technol [Internet]. 2023;94–8. Available from: https://ieeexplore.ieee.org/document/10361064/citations
11. Shute VJ, Sun C, Asbell-Clarke J. Demystifying computational thinking. Educ Res Rev. 2017; (22):142–58.
12. Weintrop D, Beheshti E, Horn M, Orton K, Jona K, Trouille L, et al. Defning computational thinking for mathematics and science classrooms. J Sci Educ Technol. 2016;25(1):127–47.
13. Grgurina N, Barendsen E, K V V, Suhre C, Zwaneveld B. Exploring students' computational thinking skills in modeling and simulation projects: A pilot study. Proc Work Prim Second Comput Educ. 2015;65–8.
14. Tang X, Yin Y, Lin Q, Hadad R, Zhai X. Assessing computational thinking: A systematic review of empirical studies. Comput Educ. 2020;(148).
15. Güven I, Gulbahar Y. Integrating computational thinking into social studies. Soc Stud. 2020;111(5):234–48.
16. Tekdal M. Trends and development in research on computational thinking. Educ Inf Technol. 2021;26(5):6499–529.
17. Singh B, Kaunert C. Computational Thinking for Innovative Solutions and Problem-Solving Techniques. In IGI Global; 2024. pp. 60–82.
18. Huang J, Parker MC. Developing computational thinking collaboratively: the nexus of computational practices within small groups. Comput Sci Educ. 2022;1–33.
19. Aho A. Computation and computational thinking. Comput J. 2012;55(7):832–5.
20. Csizmadia A, Curzon P, Dorling M, Humphreys S, Ng T, Selby C et al. Computational thinking - A guide for teachers. Swindon Comput Sch. 2015.
21. Papert S. Situating constructionism. In: Harel I, Papert S, editors. Constructionism. Norwood: Ablex; 1991.
22. Ackermann E. Piaget's constructivism, Papert's constructionism: What's the difference. Futur Learn Gr Publ. 2001;5(3).
23. Gadanidis G. Five affordances of computational thinking to support elementary mathematics education. J Comput Math Sci Teach. 2017;36(2):143–51.
24. Aiken L. Psychological Testing and Assessment. 5 ed. th ed. Boston: Allyn and Bacon, Inc; 1985.
25. Nunnally J. BerstaeinI.H. &. Psychometric Theory. New York: McGraw-Hill; 1994.
26. Azwar S. Dasar-Dasar Psikometrika. Edisi II. Yogyakarta: Pustaka Pelajar; 2016.
27. Lin CY, Pakpour AH. Using Hospital Anxiety and Depression Scale (HADS) on Patients with Epilepsy: Confirmatory Factor Analysis and Rasch Models. Seizure. 2017;45:42–6.
28. Sari CM, Ridwan A, Sastrawijaya Y. Learning Environment of Numeracy: Development and Validation With Confirmatory Factor Analysis and Rasch Model. Pedagogika [Internet]. 2024;153(1):185–98. Available from: https://ejournals.vdu.lt/index.php/Pedagogika/article/download/5483/3259

29. Brauer K, Ranger J, Ziegler M. Confirmatory Factor Analyses in Psychological Test Adaptation and Development. Psychol test Adapt Dev [Internet]. 2023;4(1):4–12. Available from: https://doi.org/10.1027/2698-1866/a000034
30. Kline RB. Principles and practice of structural equation modeling, 4th ed. Principles and practice of structural equation modeling, 4th ed. New York, NY, US: The Guilford Press; 2016. p. xvii, 534–xvii, 534. (Methodology in the social sciences.).
31. Gallagher MW, Brown TA. Introduction to Confirmatory Factor Analysis and Structural Equation Modeling. In SensePublishers, Rotterdam; 2013. pp. 289–314. Available from: https://link.springer.com/chapter/10.1007%2F978-94-6209-404-8_14
32. Bond TG, Fox CM. Applying the Rasch model: Fundamental measurement in the human sciences, 3rd ed. Applying the Rasch model: Fundamental measurement in the human sciences, 3rd ed. New York, NY, US: Routledge/Taylor & Francis Group; 2015. p. xxi, 383–xxi, 383.
33. Avcu A. The Contribution of Rasch Modeling on Measuring Attitudes For Better Classroom Assessment. Int J Psychol Educ Stud. 2025;12(3):169–79.
34. McArthur DL. The Rasch Model for Item Analysis. In Springer, Dordrecht; 1987. pp. 99–127. Available from: https://link.springer.com/chapter/10.1007/978-94-009-3257-9_4
35. Andrich D, Marais I. The Dichotomous Rasch Model—The Simplest Modern Test Theory Model. In Springer, Singapore; 2019. pp. 75–87. Available from: https://link.springer.com/chapter/10.1007/978-981-13-7496-8_6
36. Wright BD, Linacre JM. Observations are Always Ordinal Measurements, however, must be Interval. Arch Phys Med Rehabil. 1989;70(12):857–60.
37. Shin N, Bowers J, Krajcik J, Damelin D. Promoting computational thinking through project-based learning. Discip Interdiscip Sci Educ Res. 2021.
38. Barr V, Stephenson C. Bringing computational thinking to K-12: What is involved and what is the role of the computer science education community? ACM Inroads. 2011;21(1):48–54.
39. Grover S, Pea R. Computational thinking in K–12: A review of the state of the field. Educ Res. 2013;42(1):38–43.
40. Atmatzidou S, Demetriadis S. Advancing students' computational thinking skills through educational robotics: A study on age and gender relevant diferences. Rob Auton Syst. 2016;(75):661–70.
41. Wing J. Computational thinking. Commun ACM. 2006;49(3):33–5.
42. Iste. Computational thinking in K 12 education leadership toolkits [Internet]. 2022. Available from: https://csta.acm.org/Curriculum/sub/CurrFiles/CompThinkingFlyer.pdf
43. Yılmaz FGK, Yılmaz R, Durak HY. Teaching computational thinking in primary education. IGI Glob. 2018.
44. Hooshyar D. Effects of technology-enhanced learning approaches on learners with different prior learning attitudes and knowledge in computational thinking. Comput Appl Eng Educ [Internet]. 2021;1–13. Available from: wileyonlinelibrary.com/journal/cae.

45. Gong D, Yang HH, Cai J. Exploring the key infuencing factors on college students' computational thinking skills through fipped-classroom instruction. Int J Educ Technol High Educ. 2020;(17):1.
46. Voogt J, Fisser P, Good J, Mishra P, Yadav A. Computational thinking in compulsory education: Towards an agenda for research and practice. Educ Inf Technol. 2015;(20):715–28.
47. Csizmadia A, Standl B, Waite J. Integrating the constructionist learning theory with computational thinking classroom activities. Inf Educ. 2019;18(1):41–67.
48. McNicholl R. Computational thinking using Code.org. Hello World. 2019;(4):36–7.
49. Hanid MFA, Said MNHM, Yahaya N, Abdullah Z. Effects of augmented reality application integration with computational thinking in geometry topics [Internet]. Education and Information Technologies. Springer US; 2022. Available from: https://doi.org/10.1007/s10639-022-10994-w
50. Jocius R, Byrne WIO, Albert J, Joshi D, Robinson R, Andrews A. Infusing computational thinking into STEM teaching: From professional development to classroom practice. Educ Technol Soc. 2021;24(4):166–79.
51. Liao CH, Chiang CT, Chen IC, Parker KR. Exploring the relationship between computational thinking and learning satisfaction for non – STEM college students. Int J Educ Technol High Educ [Internet]. 2022;19(43). Available from: https://doi.org/10.1186/s41239-022-00347-5
52. Taub R, Armoni M, Bagno E, Ben-ari M. The efect of computer science on physics learning in a computational science environment. Comput Educ. 2015;(87):10–23.
53. Ezeamuzie, Leung JSC. Computational thinking through an empirical lens: A systematic review of literature. J Educ Comput Res. 2021;59.
54. Shivhare, Kumar CA. On the cognitive process of abstraction. Procedia Comput Sci. 2016;(89):243–52.
55. Thalheim B. Abstraction. In: Liu L, Özsu MT, editors. Springer reference Encyclopedia of database systems. New York: Springer; 2009.
56. Anderson JR. Cognitive psychology and its implications. 8th ed. New York: Worth; 2015.
57. Cormen TH, Leiserson CE, Rivest RL, Stein C. Introduction to algorithms. 3rd ed. MIT Press; 2014.
58. Alim S. The Lab Rotation Model Intervention and Psychometric Analysis: Study on Computational and Critical Thinking in Higher Education Students. King Mongkut's Institute of Technology Ladkrabang; 2024.
59. Gaskin J. Model Fit during a Confirmatory Factor Analisys (CFA) in AMOS. Gaskination's Statistics. [Internet]. 2016. Available from: http://youtube.com/Gaskination
60. JASP Team. JASP (Version 0.16.1) Computer software [Internet]. 2022. Available from: https://jasp-stats.org/
61. Hasim MA, Jabar J, Wei VWM. Measuring E-Learning Antecedents in The Context of Higher Education through Exploratory and Confirmatory Factor Analysis. Int J Acad Res Bus Soc Sci. 2024;14(9).

62. Yaakob R, Chik Z. Confirmatory Factor Analysis for Measure the Marketability of Terengganu Graduates Based on Soft Skills and Graduate Awareness. Int Cent Res Resour Dev Qual INDEX Res J [Internet]. 2025;6(3). Available from: https://icrrd.com/public/media/10102025080457-Graduate-Awareness.pdf
63. Sathyanarayana S, Mohanasundaram T. Fit Indices in Structural Equation Modeling and Confirmatory Factor Analysis: Reporting Guidelines. Asian J Econ Bus Acc. 2024;24(7):561–77.
64. Mulaik SA, Goodness of Fit Indices. 2014; Available from: https://onlinelibrary.wiley.com/doi/10.1002/9781118445112.stat06421
65. Cho G, Hwang H, Sarstedt M, Sarstedt M, Ringle CM, Ringle CM. Cutoff criteria for overall model fit indexes in generalized structured component analysis. 2020;8(4):189–202. Available from: https://link.springer.com/content/pdf/10.1057/s41270-020-00089-1.pdf
66. Gao Qhui. Replace or Coexist—Cronbach's Alpha and McDonald's Omega. 2022;4(8):954–61.
67. Kilic S. Cronbach's alpha reliability coefficient. J Mood Disord [Internet]. 2016;6(1):47–8. Available from: https://app.trdizin.gov.tr/makale/TWpBM09ESXpNdz09/cronbach-in-alfa-guvenirlik-katsayisi
68. Taber KS. The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. Res Sci Educ [Internet]. 2018;48(6):1273–96. Available from: https://doi.org/10.1007/s11165-016-9602-2
69. Das Nair R, Moreton BJ, Lincoln NB. Rasch analysis of the Nottingham extended activities of daily living scale. J Rehabil Med. 2011;43(10):944–50.
70. ten Klooster PM, Taal E, van de Laar MAFJ. Rasch analysis of the Dutch health assessment questionnaire disability index and the health assessment questionnaire II in patients with rheumatoid arthritis. Arthritis Rheum. 2008;59(12):1721–8.
71. Boone WJ, Staver JR, Yale MS. Rasch Analysis in the Human Sciences. Dordrecht: Springer; 2014.
72. Meyer JP. Applied Measurement with jMetrik. Routledge; 2014.
73. Fan J, Bond TG. Applying Rasch measurement in language assessment: Unidimensionality and local independence. In Routledge; 2019. pp. 83–102. Available from: https://www.taylorfrancis.com/chapters/edit/10.4324/9781315187815-5/applying-rasch-measurement-language-assessment-jason-fan-trevor-bond
74. Andrich D, Marais I. Violations of the Assumption of Independence I—Multidimensionality and Response Dependence. In Springer, Singapore; 2019. pp. 173–85. Available from: https://link.springer.com/chapter/10.1007/978-981-13-7496-8_14
75. He L, Min S. Item Response Theory. In Springer Nature; 2024. pp. 15–24.
76. Han H. The effectiveness of weighted least squares means and variance adjusted based fit indices in assessing local dependence of the rasch model: Comparison with principal component analysis of residuals. PLoS One [Internet]. 2022;17(9):e0271992. Available from: https://doi.org/10.1371/journal.pone.0271992

77. Santoso A, Afendi FM, Pardede T, Retnawati H, Rafi I, Apino E, et al. Deep-Rasch as an Alternative to Rasch Modeling under Assumption Violations and Small Sample Sizes. Cauchy. 2025;10(2):970–85.
78. Colledani D, Pizzio APG, Devita M, Anselmi P. Investigating the Functioning of Rating Scales With Rasch Models. Assessment. 2024;10731911241245792.
79. Andrich D, Rasch Rating-Scale M. 2016;75–94. Available from: https://www.taylorfrancis.com/chapters/edit/10.1201/9781315119144-11/normal-ogive-multidimensional-models-hariharan-swaminathan-jane-rogers
80. Kang HA, Su YH, Chang HH. A note on monotonicity of item response functions for ordered polytomous item response theory models. Br J Math Stat Psychol [Internet]. 2018;71(3):523–35. Available from: https://pubmed.ncbi.nlm.nih.gov/29516492/
81. Wind SA. Detecting Rating Scale Malfunctioning With the Partial Credit Model and Generalized Partial Credit Model. Educ Psychol Meas. 2022;001316442211162.
82. Arifuddin H, Arismunandar A, Arifuddin A, Nur A, Arifuddin H, Rahman A. The many-facet rasch model in the analysis of instruments measuring teachers' attitudes towards earthquake disaster preparedness in indonesia. Heal Tadulako J (Jurnal Kesehat Tadulako). 2024.
83. Saidi SS, Siew NM. Reliability and Validity Analysis of Statistical Reasoning Test Survey Instrument using the Rasch Measurement Model. J Math Educ [Internet]. 2019;14(3):535–46. Available from: https://files.eric.ed.gov/fulltext/EJ1227027.pdf
84. Sheptian R, Arthur R, Wardhani RMKP. Psychometric analysis of mathematics readiness assessment tool for elementary school students: a rasch model approach. Learning. 2024;4(4):1213–20.
85. Martín ES, Boeck PD. What Do You Mean by a Difficult Item? On the Interpretation of the Difficulty Parameter in a Rasch Model. In Springer, Cham; 2015. pp. 1–14. Available from: https://link.springer.com/chapter/10.1007/978-3-319-07503-7_1
86. Wu M, Tam HP, Jen TH, Two-Parameter IRT. Models. In Springer, Singapore; 2016. pp. 187–205. Available from: https://link.springer.com/chapter/10.1007/978-981-10-3302-5_10
87. Cho S, Kim H, Song Y, Nam E, Jones B, Ndosi M et al. Validation of the Korean Leeds Satisfaction Questionnaire in Rheumatoid Arthritis With Rasch Models. Int J Rheum Dis. 2022.
88. La Porta F, Franceschini M, Caselli S, Cavallini P, Susassi S, Tennant A. Unified balance scale: an activity-based, bed to community, and aetiology-independent measure of balance calibrated with Rasch analysis. J Rehabil Med. 2011;43(5):435–44.
89. González-de Paz L, Kostov B, López-Pina JA, Solans-Julián P, Navarro-Rubio MD, Sisó-Almirall A. A Rasch analysis of patients' opinions of primary health care professionals' ethical behaviour with respect to communication issues. Fam Pract. 2015;32(2):237–43.
90. Røe C, Damsgård E, Fors T, Anke A. Psychometric properties of the pain stages of change questionnaire as evaluated by Rasch analysis in patients with chronic musculoskeletal pain. BMC Musculoskelet Disord. 2014;15:95.
91. Ramp M, Khan F, Misajon RA, Pallant JF. Rasch analysis of the Multiple Sclerosis Impact Scale MSIS-29. Health Qual Life Outcomes. 2009;7:58.

92. Liargovas P, Poland R. Encyclopedia of Quality of Life and Well-Being Research. Encyclopedia of Quality of Life and Well-Being Research. 2014.
93. Embretson SE, Reise SP. Item Response Theory for Psychologists. 1st ed. Psychology; 2000.
94. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equ Model A Multidiscip J [Internet]. 1999;6(1):1–55. Available from: https://doi.org/10.1080/10705519909540118
95. Hair JF, Babin BJ, Anderson RE, Black WC. Multivariate Data Analysis [Internet]. 8th ed. Cengage Learning; 2019. Available from: https://www.cengage.uk/c/multivariate-data-analysis-8e-hair-babin-anderson-black/9781473756540/?searchIsbn=9781473756540
96. DeVellis RF, Thorpe CT. Scale Development: Theory and Applications. 5th ed. Thousand Oaks, CA: Sage; 2022.
97. Hair JF, Hult GTM, Ringle CM, Sarstedt M. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). 3rd ed. Sage; 2022.
98. Gertsen R, Klaassen S, Butter R. Development and preliminary validation of the moral authorship questionnaire. Learn Environ Res [Internet]. 2022;25(3):663–86. Available from: https://doi.org/10.1007/s10984-021-09400-y
99. Dang VH, Nguyen LT, Pham HT. Validity of school – level environment questionnaire (SLEQ): a case of high school teachers in Vietnam. Learn Environ Res [Internet]. 2023;(0123456789). Available from: https://doi.org/10.1007/s10984-023-09468-8
100. Alim S, Purwasetiawatik TF, Saudi ANA, Susanti S. The Short Version of IPIP-BFM Scale Properties Based on Bugis-Makassar Cultural Background: Do the Items Match with Javanese Culture? JP3I (Jurnal Pengukuran Psikol dan Pendidik Indones. 2022;11(2):196–208.
101. Worthington RL, Whittaker TA. Scale Development Research: A Content Analysis and Recommendations for Best Practices. Vol. 34, The Counseling Psychologist. Worthington, Roger L.: Department of Educational, School, and Counseling Psychology, University of Missouri, Columbia, MO, US, 65211, WorthingtonR@missouri.edu: Sage Publications; 2006. pp. 806–38.
102. Dunn TJ, Baguley T, Brunsden V. From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. Vol. 105, British Journal of Psychology. Dunn, Thomas J.: Division of Psychology, Nottingham Trent University, Burton Street, Nottingham, United Kingdom, NG1 4BU, thomas.dunn02@ntu.ac.uk: Wiley-Blackwell Publishing Ltd.; 2014. pp. 399–412.
103. Linacre JM. Optimizing rating scale category effectiveness. J Appl Meas. 2002;3(1):85–106.
104. Reckase MD. Unifactor latent trait models applied to multifactor tests: Results and implications. Vol. 4, Journal of Educational Statistics. US: American Educational Research Assn; 1979. pp. 207–30.
105. Bond T, Yan Z, Heene M. Applying the Rasch Model: Fundamental Measurement in the Human Sciences. 4th ed. Routledge; 2020.
106. Denovan A, Dagnall N, Drinkwater K. The Ego Resiliency Scale-Revised: Confirmatory Factor Analysis and Rasch Models. J Psychoeduc Assess. 2022;40(6):707–21.

107. Kline RB. Principles and Practice of Structural Equation Modeling. 5th ed. Guilford Press; 2023.
108. Linacre JM. Winsteps Rasch Measurement Computer Program User's Guide. Winsteps.com; 2024.

Figures

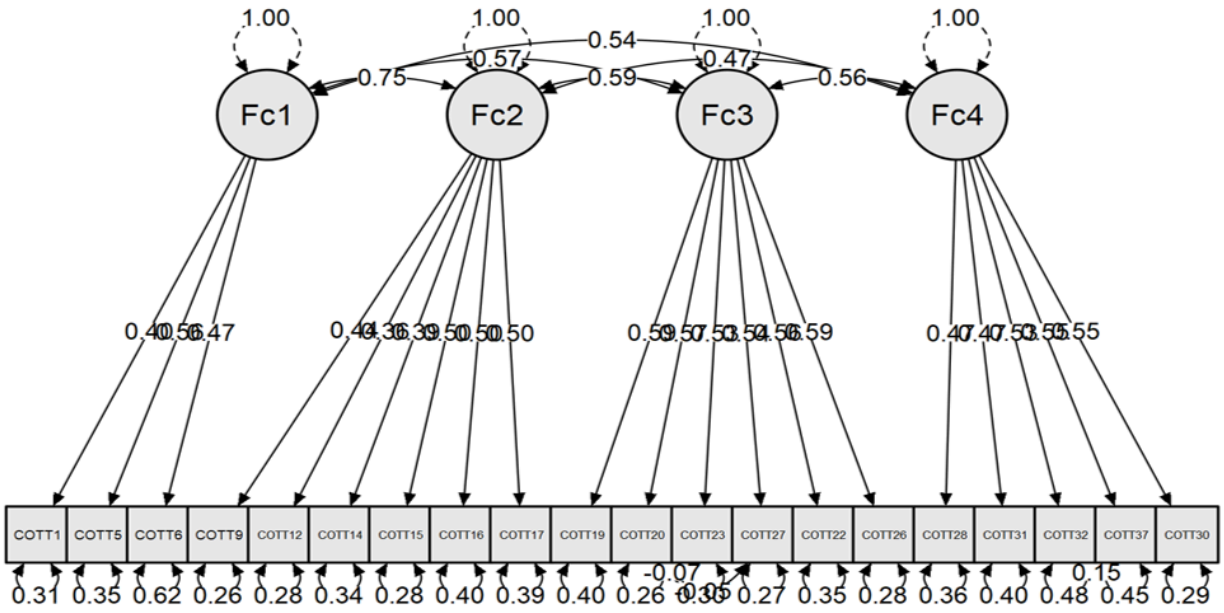


Figure 1

Model plots

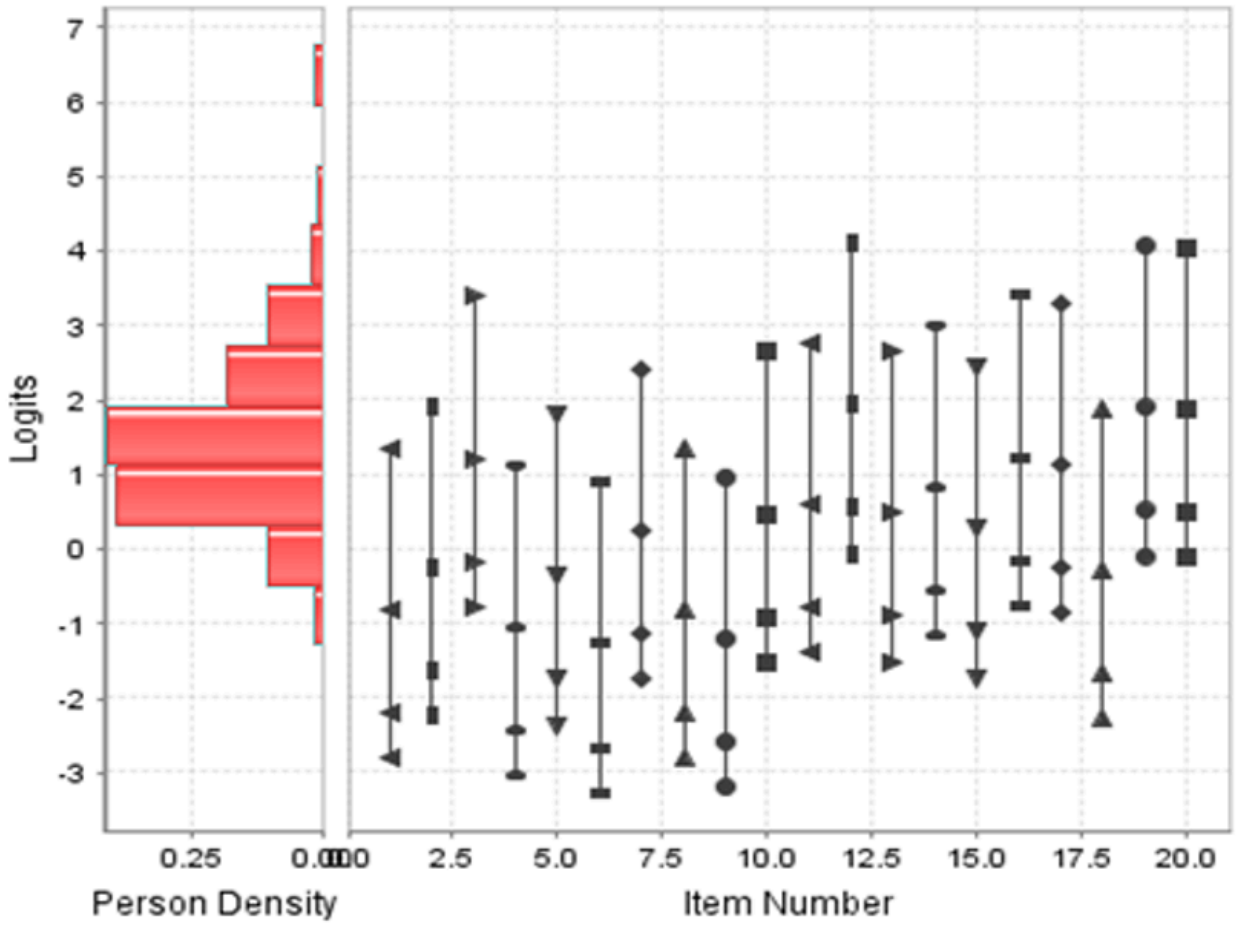


Figure 2

Wright map

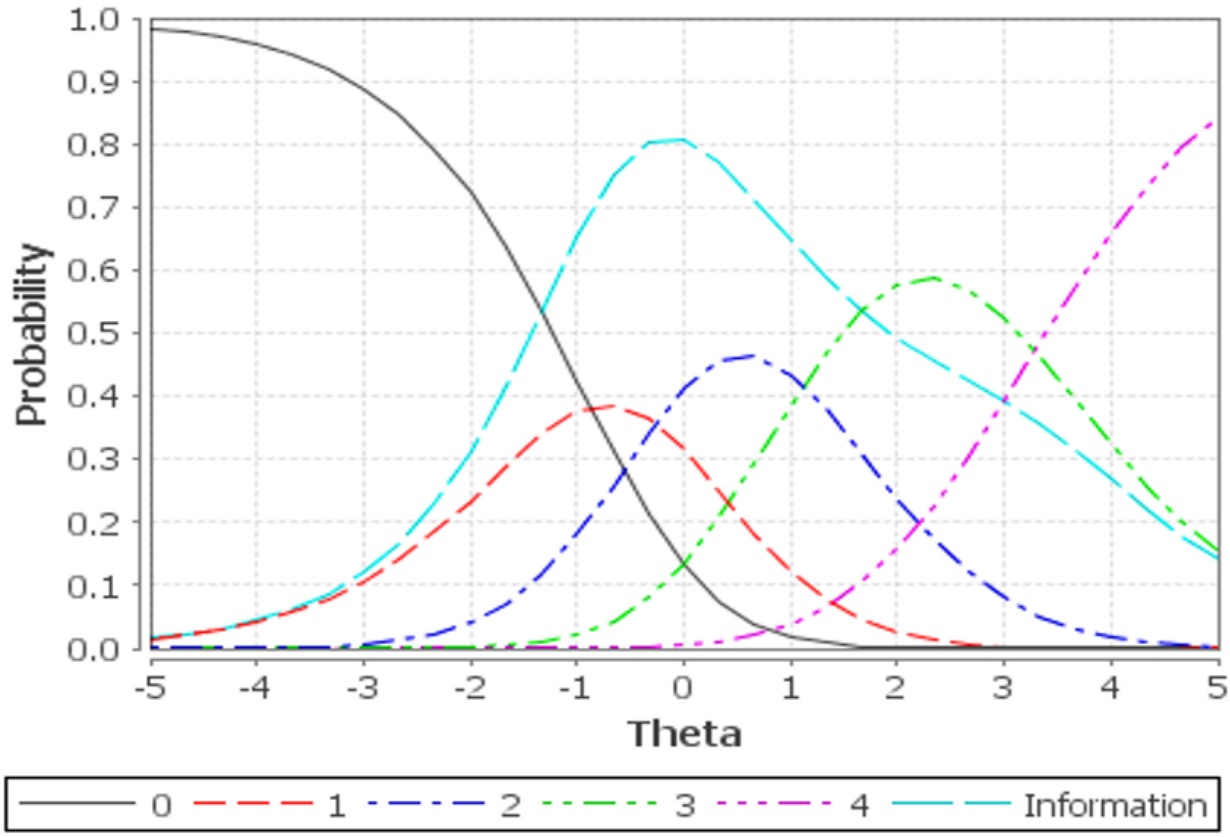


Figure 3

Category probability curves