

Rasch Analysis for Disposition Levels of Computational Thinking Instrument Among Secondary School Students

Saralah Sovey¹ , Kamisah Osman^{1*} , Mohd Effendi Ewan Mohd Matore¹ 

¹ Faculty of Education, Universiti Kebangsaan Malaysia, Bangi Selangor, MALAYSIA

Received 5 January 2022 • Accepted 10 February 2022

Abstract

Computational thinking is a strategy of thinking to tackle complex problems. There is a paucity of conceptualization and instruments that cogitate on computational thinking disposition and attitudes. This study reacts to these constraints by establishing an instrument to test computational thinking related dispositions and attitudes. The computational thinking disposition Instrument is an indicator of student's disposition towards computational thinking in daily life. The objective of this study is to investigate the psychometric features using Rasch model. Data of 535 form four computer science students in Malaysia were obtained. Instrument consists of 55 core measures in three domains: cognitive, affective and conative. The Rasch analysis indicated good psychometric features of the instrument. In these three domains no items showed disordered thresholds and the reliability was good. As a result, the Rasch analysis provides basis for cautious optimism permitting more detailed and finer level investigation of the instrument.

Keywords: attitudes, computational thinking, disposition, secondary school students

INTRODUCTION

Computational Thinking (CT) is a universal attitude and skill set that should be included in every child's repertoire, making it a critical competency that spans practically all subjects. CT demystify problem solving, designing systems, and not to be missed understanding human behavior by drawing on the concepts fundamental to computer science (Wing, 2006). Thus, computer-based learning is one approach of teaching and learning that has been shown to improve Higher-Order Thinking Skills (HOTS) (Salihuddin et al., 2016); nevertheless, thinking abilities alone are insufficient (Chongo et al., 2020). It necessitates a problem-solving strategy and a tool for problem solving using CT. In general, computer science's objectives include inspiring students to go beyond the screen and explore how computers work and how to solve a variety of problems (Syso & Kwiatkowska, 2015). Olabe et al. (2014) discovered that novel teaching approaches, such as simple Scratch programming, demonstrated a capacity to solve real-world challenges.

In recent times, emphasis has been placed on teaching children to think like computer scientists (i.e.,

computational thinking) and on the importance of prioritizing computer science principles in elementary and secondary schools (Barr & Stephenson, 2011; College Board, 2014; Wing, 2006; Yadav et al. 2014). However, despite the high interest in developing CT among school children and the large investment in CT initiatives, there are a range of issues and challenges to incorporate CT in the school curriculum (Bocconi et al., 2016). Consequently, the education sector faces rising pressure to demonstrate the use of computational thinking in everyday life through the use of computing tools as it is one of the contributing factors before entering the industry. Students who are exposed to CT as part of their education may begin to perceive connections between academic subjects and life in and out of the classroom. Since 2006, the concept of CT as a capability, a set of skills, and a mindset that every child should acquire has gained traction and importance.

However, adaptation of CT concepts in everyday life are not going to be easy and require thorough study (Sondakh et al., 2020a). Most of the attention on embedding CT during the past decade has centered on integration of CT skill in students with only little prominence about their perception, feeling or attitude towards the application of CT in problem solving across

Contribution to the literature

- To raise awareness to the disposition level of computational thinking. We have clarified the concept “computational thinking” in the aspect of disposition framework.
- We have developed an instrument with good psychometric properties to measure students’ disposition level of computational thinking.
- The study revealed the need to develop an instrument to measure students’ disposition to computational thinking. This study has the potential to generate more knowledge and literature on students’ CT disposition. There are very few empirical studies in this regard.

various discipline or specifically in daily life (Sondakh et al., 2020b). Thus, development of an instrument to measure students’ disposition towards CT is required. The new CT disposition tool is presently being developed to assess students’ dispositions. In statistics, empirical evidence is particularly important for building new instruments. The purpose of this study is to determine the validity and reliability of a scale that can be used to assess students’ disposition levels in secondary schools.

LITERATURE REVIEW

CT is a newer curriculum field in this digital culture that has very quickly to be adapted into classrooms. The researchers were unable to anticipate all the issues that arise before implementation (Belanger et al., 2018). Although there has been a broad discussion demystifying pedagogical aspects of CT, the study on assessing CT skills and attitude continues to take place. The attitude which is developed by using CT should be improved to analyses systematic approaches and complex problems (Qiu, 2009). Looking at studies on the past 5 years (2016 to 2020), it became evident that minimal studies were devoted to address the issue of CT disposition between students (Haseski et al., 2018; Jong et al., 2020). CT, likewise, is not only characterized by skills, but also by attitudes (Wing, 2006).

Some researchers view CT as a subset of the critical thinking skills required in today’s society (Tang et al., 2019), the complexity of CT prompts others to dig deeper, implying a more comprehensive understanding of CT as disposition (Wing, 2008). Nurturing students to be self-directed problem solvers in a digital world and arming them with CT skills and knowledge may not be sufficient. While CT makes use of coding knowledge to solve problems, it does not account for the disposition to apply these competencies to pertinent problems. Thus, researchers argue for the importance of CT dispositions as a motivator for persistently distinguishing complex real-world problems and seeking efficient solutions via coding (Abdullah et al., 2012; Denning, 2009). In other words, the CT disposition concept encompasses both the psychological and cognitive dimensions of computational problem solving. The National Research Council (NRC) stated that specific thinking skills are positively correlated with an internal motivation to think

and are constituents of specific thinking dispositions (NRC, 2011). As a result, good thinkers typically possess both thinking abilities and a disposition toward the way of thinking (Barr et al., 2011). Although it has been suggested that CT should be integrated into K-12 classrooms in order to foster students’ dispositions toward CT, a validated measure of CT dispositions appears to be lacking.

Measuring attitude related to CT is required because there does not yet exist any widely adopted standardized assessments (Haseski et al., 2018; Sondakh et al., 2020a; Weese, 2016). As a result, it’s unsurprising that CT evaluation continues to be a significant weakness in this sector. There is no commonly agreed method for assessing CT, making it difficult to assess the impact of interventions correctly and objectively (Grover & Pea, 2013; Kim et al., 2013; Settle et al., 2012; Shute et al., 2017). Nevertheless, there is always an urge to distinguish ways to envision the measurement of CT across all disciplines. Consequently, the issue of assessment in current studies was found lacking compared to the studies investigating approaches to teach CT (Sondakh et al., 2020b). On the other hand, an instrument from the west may not suitable for use in Malaysia due to cross cultural differences. Indeed, the obstacles confronting each individual from a variety of countries, institutions, and levels of education are unique.

Hence, a total of 241 items are created, leaving 143 items after the content validity phase. Then, 87 items underwent a first pilot test incorporating factor analysis and then followed by the second pilot test using Rasch model which remain 55 items finally. This study will determine which final items best fit the requirements of the Rasch model. Georg Rasch founded the Rasch measurement model in 1960. It is a comprehensive statistical method with unique mathematical properties based on a parameter model that combines the difficulty level of items and the respondents’ capabilities, as well as interactions between the two on a similar logit scale (Aziz et al., 2015).

Additionally, the Rasch measurement model can convert four Likert scale alternatives for each ordinal-scale data item to a scale proportional to the size of the logits unit. As a result, the pupils would have no difficulty making judgments that are most

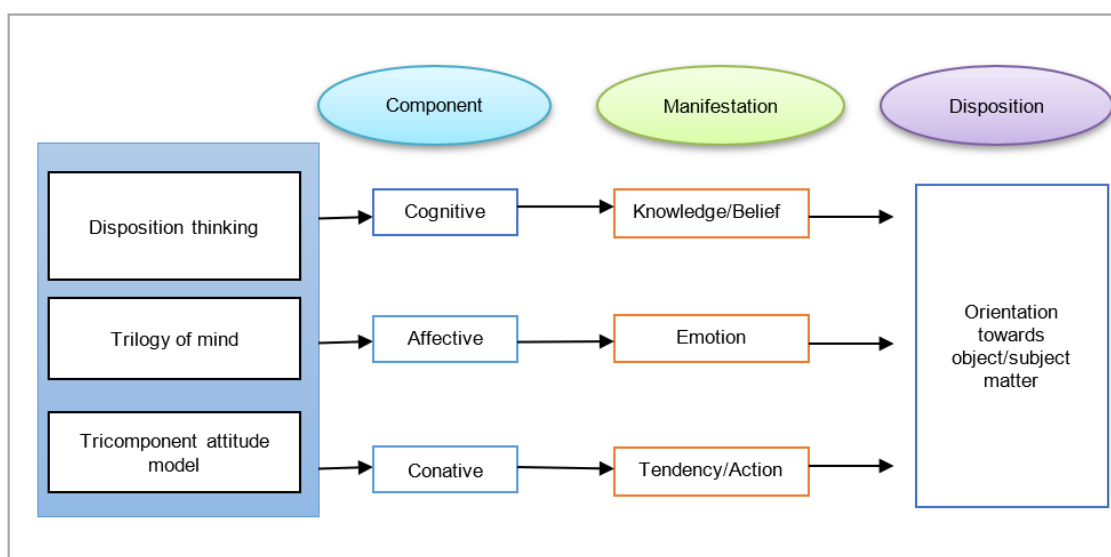


Figure 1. Theoretical framework

representative of themselves. At the Rasch analytics stage, calibration between students and their responses to the items can be used to verify the compatibility of each item produced in the model, hence avoiding an item recurrence on the same measure (Wright & MOK, 2004). Only high-quality items are preserved for subsequent testing by precise testing based on the MNSQ value in the range of 0.60 to 1.40 (Bond & Fox, 2007), the PTMEA CORR value, and local independence. Additionally, local independence can ensure that each produced item measures a distinct latent construct and does not overlap with other constructed items (Baghaei, 2008).

As a result, the research gap can be resolved by conducting psychometric feature testing on the development of instrument items and applying rigorous empirical analytic techniques such as the Rasch model. The Rasch model has undoubtedly attracted the attention of numerous researchers' both domestically and internationally in order to validate the item on their instrument's development (Balsamo et al., 2014; Othman et al., 2014). This article examines the validity and reliability of measuring instruments using the Rasch model's three core assumptions, namely item fit, unidimensionality, and local independence. The primary objective of this research is to ascertain the dispositions that secondary school students exhibit when practicing CT in their daily lives. Thus, it is thought that endorsing an item for each CT disposition construct using the Rasch model is capable of improving item quality measurement.

CT DISPOSITION FRAMEWORK

Long-term involvement in computational practices with an emphasis on the CT process, as well as ample learning opportunities in a motivating environment, are required to cultivate CT dispositions (Brennan & Resnick, 2012). CT disposition recognized as the values,

motivations, feelings, stereotypes, and attitudes applicable to CT (Barr & Stephenson, 2011). Furthermore, disposition is a person's consistent internal motivation to act toward CT, or to respond to, persons, events, or circumstances in habitual, yet potentially malleable ways (CSTA, 2017).

Since dispositions have been recognized as a psychological construct, social psychologists classify disposition as "an attitudinal tendency" (Facione, 2000; Facione et al., 1994; Sands et al., 2018). Despite this, dispositions are often being defined as a "cast or habit of the mind" or "frame of mind" which is necessary for exercising critical thinking (Beyer, 1995). In addition, theorists argue that thinking requires something more fundamental than knowledge or skill i.e., a set of dispositions (Beyer, 1988; Norris & Ennis, 1989). Therefore, in this study, disposition is defined as internal motivation and a combination of attitudes, values and beliefs which comprises, dispositional thinking theory (Beyer, 1995), Tripartite classification of mind (Hilgard, 1980) and Tricomponent attitude model (Schiffman et al., 2012) to form a theoretical framework for this study. Figure 1 describes the integration of cognitive, affective and conative components which accomplish neither the three modes of mental functioning nor modes of attitudes to distinguish the disposition construct towards CT.

METHODOLOGY

Research Design

The study took a quantitative approach, focusing on a cross-sectional quantitative survey. The quantitative technique was used for this study because it enables the collection and analysis of data in a numerical framework to explain the phenomena being studied (Gay & Mills, 2018). The data was collected via a self-administered internet survey since it is less expensive, requires no

copying of surveys, and requires no coding. Additionally, it is simple to administer and capable of collecting detailed and ordered data (Creswell, 2012; Creswell & Creswell, 2018). Thus, the data are nearly immediately ready for statistical analysis (Hair et al., 2017). The data collection method was an online survey in which participants were required to respond to all items before submitting their responses; this eliminated the potential of missing data.

Study Sample

In this study, the researchers have employed probability sampling. Sampling is intended to be applied to selected individuals because they have experiences at the center of the phenomenon (Creswell, 2009). Probability sampling techniques employ some type of random selection and allow for the calculation of sampling error, hence reducing selection bias. Thus, 535 secondary school students with a background in computer science were surveyed for this study, with 247 males (46%) and 288 females (54%). They were selected using simple probability sampling from four zones: north (80; 15%), east (80; 15%), west (252; 47%), and south (123; 23%). To proceed with the data collection procedure, authorization from the Ministry of Education is required. Henceforth, the researcher must obtain permission from the principal before meeting with the respondents. Respondents were invited to participate in research until the required sample size was attained. The respondents were required to meet the following criteria:

1. have a background in computer science,
2. be willing to complete questionnaires, and
3. be able to complete items online.

The number of respondents in the field study is sufficient in accordance with Linacre's recommendation (Linacre, 1994), who specified a minimum requirement of 108 respondents for polytomous data with a 99 percent confidence interval and a calibration value of 0.5 logits in order to implement the Rasch measurement model analysis. The Rasch measurement model was used to evaluate the data in this study to determine item fit, polarity, local independence, unidimensionality, item-individual map, reliability, and separation index for both items and respondents. The overall response rate for surveys distributed is 92 percent. This response rate is deemed sufficient because it exceeds 90% and is consistent with previous research (Marret & Choo, 2017; Masek & Nasaruddin, 2016).

Instrument Development

To develop the instrument, we adopted Miller and Lovler's (2019) scale development guideline, which divided the test development process into 11 steps. The first step in the instrument development is to define the testing universe, the target audience and the purpose of

the test. The testing universe is the body of knowledge that the test represents, the target audience is the group of individuals who will take the test, and the purpose of the test is the information that the test will provide to the test user. This stage provides the foundation for all other development activities. Accordingly, the test plan specifies the characteristics of the test, including an operational definition of the construct and content to be measured, the format of the questions, and the administration and scoring of the test. Then, on the third step we choose the item format whether objective or subjective based on information in the test plan. After writing the test items, we administer them as a test, with appropriate instructions for the test administrator and test taker, to a sample of the target audience. This test provides objective data to help determine whether the items yield the desired information in order to choose the effective items only. Consequently, administration instructions will be designed as a guidance for the person administering the test, another for the person taking the test and a third for the person scoring the test and interpreting its results.

We then follow up the pilot test with other studies that provide the necessary data for validation and norming. Thus, conducting the pilot test and analyzing its data are an integral part of the test development process. Quantitative item analysis examines how well each test item performs. Subsequently, in revision of the test step, items are dropped based on their consistency, difficulty, discrimination and bias until a final form of the test is reached. After, the test has been revised, we conduct the validation study by administering the test to another sample of pupil. Standards for the validation study are similar to those for designing the pilot study. The validation process provides sufficient information on reliability and validity. After validation is complete, we develop norms (distribution of test scores used for interpreting an individual's test score) and cut scores. At the end of the validation process, the test manual is assembled and finalized. [Figure 2](#) depicts the development process.

Each student self-assessed the instrument in Malay. The instrument is composed of 55 items. Three constructs of CT disposition are included in the instrument (cognitive, affective, and conative). The instrument is scored on a four-point Likert scale, with 1 indicating "strongly disagree" and 4 indicating "strongly agree." One week is allotted for completing the questionnaires. The raw scores for scales are calculated using the mean score. To begin constructing the scale, a literature review and interviews with experts (professional and lay) were conducted, and a list of the characteristics that a 'person' must possess was compiled. The list was then transformed into statements describing behaviors that students may evaluate. Students indicate their level of agreement with the items using a four-point Likert scale (1=strongly disagree,

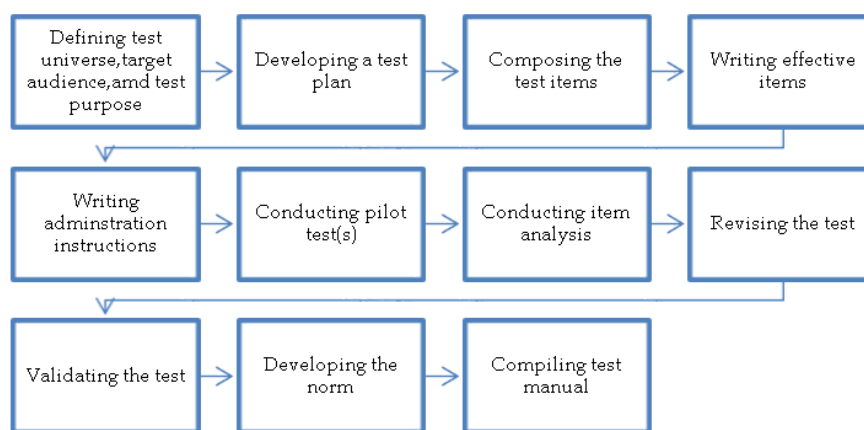


Figure 2. Instrument development and validation process

2=disagree, 3=agree, and 4=strongly agree). A middle point was omitted from this instrument (Sumintono & Widhiarso, 2014) to avoid respondents responding without making a choice (Fisher, 2006). The scale is better appropriate for this study’s usage of the Rasch model than the standard scoring approach.

With 55 items, the scale is intended to assess three primary domains of disposition. A linguist and two educational professionals reviewed the drafted scale for clarity, language, spelling, and punctuation issues. After making the necessary modifications, an instrument consisting of 55 items was constructed. The following are the findings from the validity and reliability evaluations of the data.

Rasch Model

The data were analyzed using the Rasch measurement model and suitable to evaluate and assess an instrument’s psychometric qualities in terms of validity and reliability. Software WINSTEPS version 3.71.0 (Linacre, 2007) was used to analyze the following aspects of item functionality:

1. item fit based on in-fit and outfit values in the range of 0.60 to 1.4 logits (Bond & Fox, 2007),
2. item polarity based on a positive value of point measure correlation (PTMEA CORR),
3. local item dependency analysis,
4. unidimensionality based on the principal component analysis of residuals (PCA), and
5. gender differential item functioning.

The reliability factor can be calculated using a good internal consistency value (Cronbach’s alpha), which is considered acceptable when it surpasses 0.7 (Nunnally & Bernstein, 1994). The item separation index is used to describe a range of item difficulty levels, whereas the individual separation index is used to describe a range of students’ ability levels when responding to questionnaires.

Table 1. Descriptive statistic

	Mean	SD	N
Cognitive	3.2407	0.4885	535
Affective	3.2604	0.5213	535
Conative	3.1409	0.5473	535

Note. SD: Standard deviation

RESULTS AND ANALYSIS

The psychometric properties of the instrument were determined using the Rasch measurement modal. A total of 55 items were evaluated for item fit, polarity, local independence, unidimensionality, reliability index, and separation index, based on three constructs. Items that fit and contribute to the psychometric features of the instrument were kept, while items that did not fit were submitted for revision or elimination (Linacre, 2010). Additionally, the Rasch measurement model may be used to assess the adequacy of the Likert scale employed in this study using Linacre’s six criteria (Linacre, 2002).

Descriptive Statistics

Table 1 contains descriptive statistics on the mean and standard deviation of three CT disposition constructs. The highest degree of CT disposition was discovered to be affective (M=3.2604, SD=0.5213), followed by cognitive (M=3.2407, SD= 0.4885). Conative, on the other hand, exhibits the lowest level with (M=3.1409, SD=0.5473). The highest mean score on the affective dimension reflects respondents’ excitement, interest, awareness, and empowerment to learn and apply CT in daily life. Finally, responders’ cognitive abilities reveal that they are capable of acquiring CT knowledge, creative thinking, value, and perception. The conative level with the lowest mean score reflects persistence, tolerance, collaboration abilities, and still self-confidence in practicing CT in daily life. This score indicates an individual’s eagerness to learn more about CT in depth. Thus, the findings emphasize construct validity, or the degree to which the questions on an instrument correspond to the corresponding theoretical construct (DeVon et al., 2007). Additionally, it contains

conclusions regarding the dimensionality of sub-constructs and validation of the conceptual framework's structure.

Item Fit

Item fit refers to the analysis of the Rasch measurement model's fit for each item in a questionnaire (Ariffin, 2008). Fit statistics is the criteria of mean square (MNSQ) to identify the information weighted (in-fit) and outlier-sensitive (outfit). MNSQ values are ranging from zero to infinity with an expected value of 1. Items outside the range of value of MNSQ are considered over-fit or misfit. Over-fit means the items are too predictive while misfit means the items are erratic (Bond & Fox, 2007). As per Likert-scaled polytomous data used in this investigation, the mean square MNSQ value chosen was between 0.60 and 1.4 (Bond & Fox, 2007). Meanwhile, the productive Zstd value ranged between 2.0 and +2.0 (Bond & Fox, 2007), and this value can be discarded if the MNSQ value is accepted (Linacre, 2005). The MNSQ score is calculated using both in-fit and outfit values. This ensures that the items that fit the model are considered in the subsequent analysis, while misfit items do not contribute to the measurement of constructs and are thus considered weak. All values are in the suggested range; it can be concluded that there is no item that need to be removed. The value varied between 0.81 and 1.31 logits. The standard error value for the data was observed to be between 0.08 and 0.10, which refers to the element of precision in a calculation (Linacre, 2005). Fisher (2007) considered the error value's range to be excellent.

Polarity

Additionally, item fit can be determined based on the polarity of the item by calculating the PTMEA CORR value. This value refers to a collection of items that all measure the same construct, assuming that the items all measure the same construct (Bond & Fox, 2007). The PTMEA CORR value achieved in this study was between 0.43 and 0.68, which was within the minimal value of 0.3 (Wu & Adam, 2007). Item C42 "I am confident of being able to solve difficult problems by thinking computationally", represent the maximum polarity index meanwhile minimum PTMEA CORR index is 0.43 of item C1 "I appreciate the group members' contributions during problem solving" also from the same construct namely, conative. The positive PTMEA CORR score indicated that the retained items could contribute to the instrument's psychometric features, allowing it to distinguish computer science students. In addition, this indicates that all the items used are parallel to the measurement of CT disposition. Table 2 depicts the MNSQ values and polarity values of the items.

Table 2. Fit statistics of measurement items

Item	Measure	Standard Error	MNSQ		PTMEA	
			In-fit	Outfit	Corr.	Exp.
K49	0.33	0.08	0.88	0.90	0.63	0.61
K3	0.72	0.08	0.87	0.90	0.62	0.63
K34	0.71	0.08	0.98	1.03	0.61	0.63
K52	0.10	0.08	0.94	0.91	0.60	0.60
K32	-0.14	0.08	0.94	0.94	0.59	0.58
K10	-0.13	0.08	0.84	0.83	0.59	0.58
K50	0.50	0.08	0.88	0.91	0.59	0.62
K37	0.34	0.08	0.99	1.06	0.59	0.61
K31	0.05	0.08	0.89	0.96	0.58	0.59
K6	0.00	0.08	0.93	0.98	0.57	0.59
K29	-0.37	0.09	1.07	1.08	0.57	0.57
K38	1.05	0.08	1.31	1.35	0.56	0.64
K35	0.45	0.08	1.25	1.27	0.55	0.61
K46	0.21	0.08	1.16	1.15	0.55	0.6
K26	-0.71	0.09	0.99	1.23	0.54	0.54
K43	0.28	0.08	1.05	1.09	0.54	0.61
K21	-0.05	0.08	1.21	1.32	0.51	0.59
K33	-0.86	0.09	1.00	1.22	0.48	0.53
K5	-0.36	0.09	1.27	1.24	0.47	0.57
A11	0.69	0.08	0.81	0.81	0.67	0.63
A22	0.64	0.08	0.89	0.9	0.65	0.62
A3	0.97	0.08	1.10	1.10	0.65	0.64
A7	0.14	0.08	0.88	0.87	0.62	0.60
A41	0.11	0.08	1.06	1.01	0.61	0.60
A18	-0.12	0.08	0.91	0.88	0.61	0.58
A31	-0.20	0.08	0.88	0.88	0.61	0.58
A27	0.09	0.08	0.99	0.97	0.61	0.59
A14	-0.09	0.08	0.87	0.91	0.60	0.58
A9	-0.12	0.08	0.91	0.88	0.60	0.58
A26	0.07	0.08	1.14	1.09	0.60	0.59
A1	-0.68	0.09	1.02	0.96	0.58	0.54
A40	0.38	0.08	1.13	1.11	0.58	0.61
A4	-0.36	0.09	0.98	0.99	0.57	0.57
A6	-0.72	0.09	1.02	1.01	0.53	0.54
A19	-1.20	0.09	1.03	0.92	0.52	0.50
A10	-0.91	0.09	1.13	1.21	0.50	0.53
C42	0.58	0.08	0.85	0.82	0.68	0.62
C40	0.98	0.08	0.94	0.94	0.67	0.64
C45	0.25	0.08	0.87	0.82	0.66	0.60
C44	0.48	0.08	0.91	0.87	0.66	0.62
C38	0.53	0.08	1.06	1.03	0.66	0.62
C39	0.60	0.08	0.95	0.92	0.66	0.62
C29	0.58	0.08	0.99	0.98	0.66	0.62
C10	0.54	0.08	0.90	0.89	0.65	0.62
C21	-0.35	0.09	0.87	0.84	0.61	0.57
C19	-0.21	0.08	0.98	0.95	0.61	0.58
C28	0.61	0.08	1.03	1.09	0.61	0.62
C31	0.30	0.08	1.02	1.02	0.60	0.61
C22	-0.03	0.08	1.09	1.23	0.58	0.59
C7	-0.54	0.09	0.94	0.94	0.57	0.55
C8	-0.76	0.09	0.93	1.08	0.55	0.54
C17	-0.49	0.09	0.99	1.07	0.54	0.56
C3	-1.27	0.09	1.07	0.97	0.50	0.50
C6	-1.12	0.09	1.11	1.19	0.47	0.51
C1	-1.48	0.10	1.15	1.13	0.43	0.48

Table 3. Standard correlation of residual values

Correlation	Item-construct	Item-construct
0.47	C1-Conative I appreciate the group members' contributions during problem solving	C3-Conative I am willing to tolerate group members during problem solving
0.41	A6-Affective I would like to increase my mastery in computational thinking in the future	A4-Affective I would like to learn programming to apply computational thinking
0.40	C1-Conative I appreciate the group members' contributions during problem solving	C6-Conative I encourage the sharing of ideas to find effective solutions
0.37	C6-Conative I encourage the sharing of ideas to find effective solutions	C8-Conative I try to get more ideas to solve problems in groups
0.35	C3-Conative I am willing to tolerate group members during problem solving	C6-Conative I encourage the sharing of ideas to find effective solutions
0.35	C28-Conative I am willing to take risks to solve a problem	C29-Conative I am ready to step out of my comfort zone to solve complex problems
0.35	A10-Affective I feel happy to be able to find a solution that has never been used by anyone else	A19-Affective I enjoyed my idea becoming a reality in the problem - solving process
0.32	C44-Conative I dare to face problems that will arise in new situations	C31-Conative I can adapt to face uncertainty in solving a problem
0.30	A26-Affective I admit that thinking like a computer expert gives an edge in my learning	A27-Affective I realize that all complex problems can be solved effectively when thinking like a computer expert
0.29	C44-Conative I dare to face problems that will arise in new situations	C39-Conative I dare to accept challenges to solve complex problems

Local Independence

The following feature of item measurement is local item dependence analysis. Local item dependence is often quantified using the standardized residual correlation value between two items, which should not exceed 0.3 (Balsamo et al., 2014). Likewise, if the correlation between two items is greater than 0.7, only one item is kept and the other is excluded from the model (Linacre, 2005). The retained item will be determined using the MNSQ value, which should be close to or equal to 1.0 (Bond & Fox, 2015; Linacre, 2005), as this value represents the predicted value for model fit (Aziz et al., 2015). This procedure is taken to ensure that retained items do not duplicate existing ones (Matore et al., 2020). Ten matching residual correlation coefficients ranging from 0.29 to 0.47 are shown in Table 3. Correlation values greater than 0.3 were preserved because the association remained within the approved range of 0.7 (Aziz et al., 2015) and the pair of items was within the same construct. None of the items breach the 0.70 limit indicating item independence in instrument.

Gender Differential Item Functioning (GDIF)

The purpose of this analysis was to determine the presence of gender differential item functioning (GDIF)

in the instrument used. Winstep uses a two-tailed t-test to determine the significance of the difference between two index difficulties when analyzing GDIF. For all DIF analyses, the confidence level is 95 percent and the level of t critical value is 2.0. Additionally, the GDIF Contrast index is utilized to demonstrate the difference in gap confirmation levels between males and females when males and females are compared. According to Lai and Eton (2002), the Likert scale requires a value of 0.5 logits DIF contrast. Meanwhile, in Pallant and Tenant (2007), Wright and Panchalakesan suggest that GDIFs with a size less than 0.5 logits are regarded inconsequential. A low GDIF Contrast index indicates that the item is more easily affirmed by female respondents. DIF Measurement is the difficulty index of this group while all other variables are kept constant. The DIF contrast results indicate that 11 out of 55 items illustrate the relevance of GDIF in terms of $t \geq \pm 2$ logit value. However, the GDIF contrast (± 0.5 logits) indicates that 11 items do not exhibit significant GDIF, as indicated by the GDIF index being less than 0.5 logit. The GDIF Contrast value ranges between -0.41 and 0.49. As such, it is identified that 55 items remain. Items that passed this GDIF analysis demonstrated that they satisfy the disposition testing element of fairness. The study uses DIF to identify all 55 items that did not exhibit evidence of

Table 4. Standardized residual variance (in eigenvalue units)

	Eigenvalue	Empirical		Modeled
Total raw variance in observations	92.7	100.0%		100.0%
Raw variance explained by measures	37.7	40.7%		40.5%
Raw variance explained by persons	21.1	22.8%		22.7%
Raw Variance explained by items	16.6	17.9%		17.8%
Raw unexplained variance (total)	55.0	59.3%	100.0%	59.5%
Unexplained variance in 1st contrast	4.1	4.4%	7.4%	
Unexplained variance in 2nd contrast	3.2	3.5%	5.9%	
Unexplained variance in 3rd contrast	2.8	3.1%	5.1%	
Unexplained variance in 4th contrast	2.1	2.3%	3.9%	
Unexplained variance in 5th contrast	1.7	1.9%	3.2%	

Table 5. Statistical summary for person

	Raw score	Count	Measure	Model error	In-fit		Outfit	
					MNSQ	ZSTD	MNSQ	ZSTD
Mean	179.8	55.0	2.41	0.32	1.00	-0.30	1.01	-0.3
Standard deviation	22.8	0.0	1.80	0.28	0.47	2.5	0.48	2.5
Max	220.0	55.0	8.17	1.83	3.84	9.1	3.81	9.1
Min	96.0	55.0	-2.07	0.21	0.11	-8.3	0.11	-8.2
Real RMSE	0.44	True S. D.	1.74	Separation	3.97	Person reliability		0.94
Model RMSE	0.43	True S. D.	1.74	Separation	4.09	Person reliability		0.94

Note. Person raw score-to-measure correlation=.94; Cronbach's alpha (KR-20) Person raw score reliability=.97

injustice when a group of students with varying skill levels of the same sex was compared.

Unidimensionality

Compliance with the unidimensionality assumption indicates that a collection of items in the designed instrument measures only a single construct (Wright & Master, 1982). Unidimensionality is a detection of the construct validity in the test that has been developed. Items should test constructs which measure a single dimension only. Furthermore, unidimensionality is very important to measure the internal consistency of the instrument using the principal component analysis (PCA). The PCA was determined to be 40.7 % of variance explained by measures, which was close to model estimates of 40.5 % and was sufficient in comparison to the minimal value given by Linacre (2012), which is 40%. Additionally, unexplained variance in the first contrast was 4.4 %, and values less than 5% are well accepted (Linacre, 2007, 2016). Additionally, the reported variance's Eigenvalue was 4.1 and less than 5.0 (Linacre, 2005), indicating the absence of a second dimension. Additionally, Table 4 indicates that the unexplained variation in first to fifth contrast was between 3% and 5%, which is also considered very good (Linacre, 2007). Likewise, the ratio of variation described by item size (17.9%) to variance explained by the first component (4.4%) was 4.07, exceeding the three-ratio minimum value (Conrad et al., 2012; Embretson & Reise, 2000).

Reliability Index

The instrument's reliability index is provided in Table 4. The interpretation of person reliability is

equivalent to Alpha Cronbach or KR20 (Wright & Master, 1982). Cronbach's alpha is 0.97, and item reliability is 0.98, both of which are considered excellent values (Nunnally & Bernstein, 1994). In this study, the reliability index for respondents is 0.94, which is an acceptable range (Pallant, 2001; Sekaran, 2003) which could expect consistency level personal situation arrangement in the log scale if this sample answers different set item, but to measure that the same construct (Wright & Masters, 1982).

Separation Index

The instrument items' reliability was determined using the index of person separation, which is comparable to Cronbach's alpha. The term "person separation" refers to the process of classifying persons and estimating how well a measure can separate individuals on a construct. The presence of a high degree of person separation or stratification (two distinct levels of performance, i.e., high and low, that can be separated based on test scores, person reliability of 0.7) indicates that the measure may be sensitive to distinguishing between high and low performers. Separation of items is used to validate the item hierarchy. The presence of a high degree of item separation or stratification (three items representing three distinct levels of difficulty, namely high, medium, and low; item reliability of 0.9) indicates that the person sample is sufficiently large to corroborate the item difficulty hierarchy (Linacre, 2017). Individual separation index is recorded at 6.88, as per in Table 5, which means there were 7 ability level of respondent's ability level and is regarded adequate when it exceeds 2 (Fox & Jones, 1998; Linacre, 2012).

Table 6. Statistical summary for item

	Raw score	Count	Measure	Model error	In-fit		Outfit	
					MNSQ	ZSTD	MNSQ	ZSTD
Mean	1748.8	535.0	0.00	0.08	1.00	-0.1	1.01	0.0
Standard deviation	83.9	0.0	0.60	0.00	0.12	1.8	0.14	1.7
Max	1940.0	535.0	1.05	0.10	1.31	4.5	1.35	4.8
Min	1590.0	535.0	-1.48	0.08	0.81	-3.2	0.81	-3.0
Real RMSE	0.09	True S. D	0.59	Separation	6.88	Item Reliability		0.98
Model RMSE	0.08	True S. D	0.59	Separation	7.04	Item Reliability		0.98

Note. UMean=0.000; UScale=1.000; Item raw score-to-measure correlation=-0.99

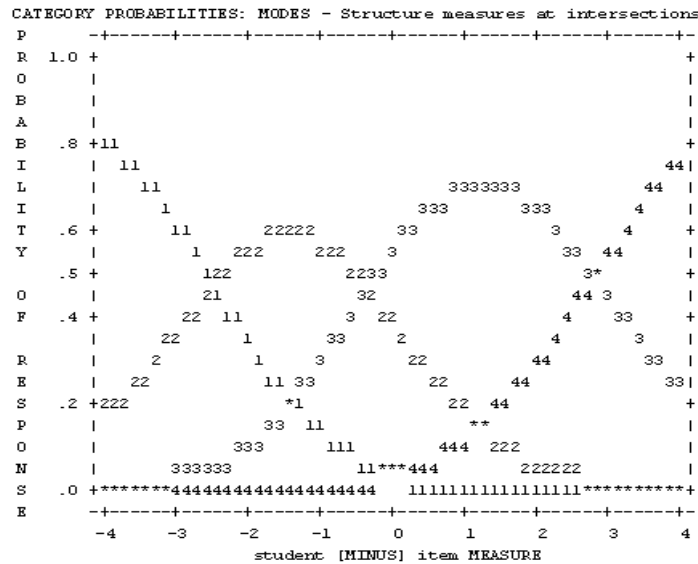


Figure 3. Threshold for scale review

Table 7. Mean category of measurement: Observed average

Category Label	Score	Observed		Observed average	Sample expected	MNSQ		Structure calibration	Category measurement
		Count	%			In-fit	Outfit		
1	1	268	1	-0.35	-0.68	1.20	1.28	None	(-3.65)
2	2	3161	11	0.58	0.61	0.99	1.02	-2.47	-1.42
3	3	14392	49	1.75	1.76	0.97	1.00	-0.34	1.26
4	4	11604	39	3.44	3.43	0.99	0.98	2.81	(3.94)

Meanwhile, Table 6 represents the item separation index was 3.97, which is considered acceptable. This means the scale can be statistically differentiated into 4 difficulty levels. Increased item separation index values imply a more effective separation of items of varied difficulty. Separation is dependent on item reliability (Wright & Masters, 1982). This outcome confirms Linacre’s (2005) assertion that separation indexes of two and above indicate greater reliability. As a result, the instrument has a wide spread when it comes to determining the level of CT disposition. Additionally, it demonstrates that the tool is measuring what it is designed to measure, thereby establishing its validity.

Scale Review

The Rasch measurement model can be used to determine the efficacy of a scale used in an instrument based on six specified criteria (Linacre, 2002). The first

condition is that each concept contain at least ten observations, which was met in this study. With regards to the second condition, each scale must exhibit a probability curve peak, as illustrated in Figure 3.

The study’s use of a four-point Likert scale also met the third condition, as the average measure of each category grew in lockstep with the scale level, increasing by (1) -3.65 logits, (2) -1.42 logits, (3) 1.26 logits, and (4) 3.94 logits. This demonstrates typical, consistent, and steadily rising response patterns (Matore et al., 2020). Following that, all outfit MNSQ values were within the range of 0.97 to 1.20, which meets the fourth condition of outfit MNSQ values being fewer than 2.00 logits. Concerning the fifth criterion, the threshold values of 2.47, 0.34, and +2.81 were ordered in an orderly fashion, indicating that there was no bias in the selection of any category of the scale utilized as shown in Table 7 and Table 8. The sixth requirement specifies that each scale’s restriction should be between 1.00 and 5.00 for a four-

Table 8. Mean category of measurement: Measurement structure

Category label	Structure		Structure to measure			50% Cum. probability	Coherence		RMSR	Estim. Discr.
	Measure	S. E.	At Cat	----Zone----			M->C	C->M		
1	None		(-3.65)	-INF	-2.72		33%	1%	1.5049	
2	-2.47	.07	-1.42	-2.72	-0.21	-2.57	59%	24%	0.8495	0.93
3	-0.34	.02	1.26	-0.21	2.92	-0.28	64%	83%	0.3562	1.00
4	2.81	.02	3.94	2.92	+INF	2.85	75%	60%	0.5407	1.02

Table 9. Revision scale check

Scale	Gaps calculation	Range of acceptance	Decision
S ₁ - S ₂	0.00 - (-2.47)	1.00 < 2.47 < 5.00	Accepted
S ₂ - S ₃	-0.34 - (-2.47)	1.00 < 2.13 < 5.00	Accepted
S ₃ - S ₄	2.81 - (-0.34)	1.00 < 3.15 < 5.00	Accepted

point Likert scale. A check of each restriction as per Table 9, revealed that the difference between each scale category exceeded one and fell within the range of five.

Item-Individual Map

The item-individual map depicts the distribution mapping of the items and respondents in this study on a similar logits scale following the calibration process. In terms of instrument creation or questionnaires, all of the items and respondents depicted in Figure 4 demonstrates varying degrees of difficulty in terms of respondents' agreement with each item (Perera et al., 2018). This mapping is extremely beneficial since it enables researchers to optimize the psychometric qualities of the instrument, they have constructed item (Perera et al., 2018).

The standard deviation values for the items (1.488 to +1.05) indicated that the difficulty level measurement is within the acceptable range of +3.00 to 3.00 (Andrich & Styles, 2004; Hill & Koekemoer, 2013, Linacre, 1994). The student position with the highest arrangement was +8.17 logits, and the student is a male. The student was the easiest to agree on all of the developed instrument's elements. Meanwhile, the lowest rank was held by a female student with a logit value of -2.07, indicating that the student was the most difficult to agree on these items. Additionally, based on the item position hierarchy, item K38 (+1.05 logits) was the most difficult for students to agree on. Item K38 represents "I am capable of suggesting a concept that has never been considered by anyone," equating to the cognitive construct. For the construct of conative, the lowest item position was represented by item C1 (1.488 logits). As a result, this issue was the most easily agreed upon by the students. Item C1 states, "I appreciate the group members' contributions during problem solving." This demonstrates that pupils are capable of expressing their gratitude and respect for their peers. According to Figure 3, the left side of the logits scale represents the order of item difficulty levels, while the right side shows the respondents' position, namely, computer science students. In general, the mean of respondents (2.23

logits) was found to be greater than the mean of items (0.00 logits). This demonstrates that the student can more easily agree on the instrument's items.

DISCUSSION

We detailed the development of a scale to assess secondary school students' disposition toward CT. Item response theory (IRT) method was selected to apply the Rasch measurement model to analyze each test item to determine the validity and reliability of the instrument. To begin, the instrument had good psychometric qualities. The study established strong reliabilities for the construct. During the pilot test, the CT disposition constructs were empirically validated in Malaysia using EFA and Rasch analysis.

Our findings in this investigation substantiated unidimensionality at the scale level. The summation of raw item scores into an interpretable total scale score is acceptable since each component's items all measure the same latent characteristic. Compliance with the unidimensionality assumption is critical in the Rasch measurement model, which is based on the premise that the items in this instrument have a single capacity (Sumintono & Widhiarso, 2014). This is an early indicator of the construct validity at the level of field study.

In general, the estimated abilities of individuals and the difficulty of items spread rather evenly around the logit continuum. However, psychometrically, the items on this scale were insufficient to capture these participants at their high ability levels. In the latter instance, additional "difficult" items at the high ability tiers are required. Given that the CTDI was developed for use in the context of computer science education, there is a strong requirement to clearly differentiate participants at the most self- or other-focused levels.

We explored gender differences in differential item functioning (DIF). To summarise, DIF happens when individuals with the same aptitude level respond differently to an item simply because they belong to distinct groups. In other words, a DIF item is a question that has been skewed by a particular set of people. All items in this study were DIF-free, allowing for meaningful comparisons across groups. These findings established a foundation for further testing of the DIF with more samples, and researchers should proceed cautiously when conducting worldwide comparisons using this instrument.

In addition, Person separation index (PSI) and Cronbach's alpha values were within the margins of reliability. PSI is derived using logit-transformed individual estimations, whereas Cronbach's alpha is calculated using raw scores. PSI is equivalent to Cronbach's alpha when the distribution is normal. PSI and alpha values more than 0.7 are typically regarded adequate (Fisher, 1992; Tavakol & Dennick, 2011). The findings have answered all the possibilities designed to examine the suitability of the items. The item reliability is high and this means the items are stable.

Limitations and Future Directions

The current research had some limitations, which also gave directions for further research. The study's primary limitation is that it was limited to secondary school pupils in a single country, Malaysia. However, we drew our conclusions from highly cited literature on CT in a variety of domains. As a result, the instrument should be applicable to additional domains. Second, caution is advised when applying this instrument to other situations, and more testing with samples from other cultural groups is necessary. Additionally, when extending this instrument to other contexts, it is necessary to investigate differential item functioning in order to draw relevant comparisons. Additionally, replications in various nations would bolster the relevance of the study across diverse countries. Finally, other types of validity, such as convergent and discriminant validity, could be investigated in future studies, however they are beyond the scope of this research. Comparing research across different tests may also provide a more holistic psychometric assessment of the findings from multiple angles. Not only will this analysis influence subsequent analyses, but it may also improve the psychometric qualities of the items. Most critically, the researcher must match appropriate dispositions to pupils in the Malaysian environment. Nonetheless, this questionnaire does not yet cover all of the characteristics listed in the literature, and it is probable that some pertinent variables were omitted. Future study could build on this work by examining additional elements of CT dispositions.

CONCLUSIONS

In summary, the data from each item of the CTDI met the Rasch model's assumptions. Each of the 55 items was maintained. Each item demonstrated a good performance of item fit, polarity, and local independence. This work adds to the body of research about CT teaching and learning by providing a more comprehensive overview of CT dispositions and attitudes, as well as their impact on their readiness to participate in digital workplaces. This is important in order to adhere to a variety of computational thinking concepts across the K-12 curriculum. Rasch analysis

validated the applicability of the CTDI as an instrument for assessing emerging students' attitudes regarding CT in daily life, particularly in the educational context.

Author contributions: All authors have sufficiently contributed to the study, and agreed with the results and conclusions.

Funding: This study is part of transdisciplinary study funded by the Ministry of Higher Education (TRGS/1/2018/UKM/01/6/1).

Acknowledgements: The authors would like to express their gratitude to the Ministry of Education, specialists, teachers, and students who assisted in the data gathering and instrument validation processes. Additionally, the authors would like to thank and express their gratitude to the Faculty of Education at Universiti Kebangsaan Malaysia (UKM).

Declaration of interest: No conflict of interest is declared by authors.

REFERENCES

- Abdullah, N., Zakaria, E., & Halim, L. (2012). The effect of a thinking strategy approach through visual representation on achievement and conceptual understanding in solving mathematical word problems. *Asian Social Science*, 8(16), 30-37. <https://doi.org/10.5539/ass.v8n16p30>
- Andrich, D., & Styles, I. (2004). *Final report on the psychometric analysis of the early development instrument (EDI) using the Rasch model: A technical paper commissioned for the development of the Australian early development instrument (AEDI)*. http://ww2.rch.org.au/emplibrary/australianedi/Final_Rasch_report.pdf
- Ariffin, S. R. (2008). *Inovasi dalam pengukuran dan penilaian [Innovation in measurement and evaluation]*. UKM Press.
- Aziz, A. A., Masodi, M. S., & Zaharim, A. (2015). *Rasch measurement model basis: Scale development and measurement structure*. UKM Press.
- Baghaei, P. (2008). Local dependency and Rasch measures. *Rasch Measurement Transactions*, 21(3), 1105-1106.
- Balsamo, M., Giampaglia, G., & Saggino, A. (2014). Building a new Rasch-based self-report inventory of depression. *Neuropsychiatric Disease and Treatment*, 10, 153-165. <https://doi.org/10.2147/ndt.s53425>
- Barr, D., Harrison, J., & Conery, L. (2011). Computational thinking: A digital age skill for everyone. *Learning and Leading with Technology*, 38(6), 20-23.
- 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>
- Belanger, C., Christenson, H., & Lopac, K. (2018). *Confidence and common challenges: The effects of teaching computational thinking to students ages 10-16*

- [Masters's thesis, St. Catherine University]. St. Catherine University Repository.
- Beyer, B. K. (1988). *Developing a thinking skills programme*. Allyn and Bacon Inc.
- Beyer, B. K. (1995). *Critical thinking*. Phi Delta Kappa Educational Foundations.
- Bocconi, S., Chiocciariello, A., Dettori, G., Ferrari, A., Engelhardt, K., Kamylyis, P., & Punie, Y. (2016). Developing computational thinking in compulsory education. Implications for policy and practice. *EUR-Scientific and Technical Research Reports*. <https://doi.org/10.2791/792158>
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Routledge
- Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge.
- Brennan, K., & Resnick, M. (2012, April). *New frameworks for studying and assessing the development of computational thinking* [Paper presentation]. Annual Meeting of the American Educational Research Association (AERA), Vancouver, BC, Canada. http://scratched.gse.harvard.edu/ct/files/AERA_2012.pdf
- Chongo, S., Osman, K., & Nayan, N. A. (2020). Level of computational thinking level of computational thinking skills among secondary science student: Variation across gender and mathematics achievement skills among secondary science student. *Science Education International*, 31(2), 159-163. <https://doi.org/10.33828/sei.v31.i2.4>
- College Board. (2014). *Advanced placement computer science principles: Curriculum framework*. <http://securemedia.collegeboard.org/digitalServices/pdf/ap/ap-computerscience-principles-curriculum-framework.pdf>
- Computer Science Teachers Association. (2017). *CSTA K-12 Computer Science Standards*. <http://www.csteachers.org/standards>
- Conrad, K. M., Conrad, K. J., Passetti, L. L., Funk, R. R., & Dennis, M. L. (2015). Validation of the full and short-form self-help involvement scale against the Rasch measurement model. *Evaluation Review*, 39(4), 395-427. <https://doi.org/10.1177/0193841x15599645>
- Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson.
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches*. SAGE.
- Denning, P. J. (2009). The profession of IT beyond computational thinking. *Communications of the ACM*, 52(6), 28-30. <https://doi.org/10.1145/1516046.1516054>
- DeVon, H. A., Block, M. E., Moyle-Wright, P., Ernst, D. M., Hayden, S. J., Lazzara, D. J., Savoy, S. M., & Kostas-Polston, E. (2007). A psychometric toolbox for testing validity and reliability. *Journal of Nursing Scholarship*, 39(2), 155-164. <https://doi.org/10.1111/j.1547-5069.2007.00161.x>
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates. <https://doi.org/10.1037/10519-153>
- Facione, N. C., Facione, P. A., & Sanchez, C. A. (1994). Critical thinking disposition as a measure of competent clinical judgment: The development of the California critical thinking disposition inventory. *Journal of Nursing Education*, 33(8), 345-350. <https://doi.org/10.3928/0148-4834-19941001-05>
- Facione, P. A. (2000). The disposition toward critical thinking: Its character, measurement, and relationship to critical thinking skill. *Informal Logic*, 20(1), 61-84. <https://doi.org/10.22329/il.v20i1.2254>
- Fisher, J. W. P. (2006). Survey design recommendations. *Rasch Measurements Transactions*, 20(3), 1072-1074.
- Fisher, J. W. P. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transactions*, 21(1), 1095.
- Fisher, W. (1992). Reliability, separation, strata statistics. *Rasch Measurement Transactions*, 6(3), 238.
- Fox, C. M., & Jones, J. A. (1998). Uses of Rasch modeling in counseling psychology research. *Journal of Counseling Psychology*, 45(1), 30-45. <https://doi.org/10.1037/0022-0167.45.1.30>
- Gay, L. R., & Mills, G. E. (2018). *Educational research: Competencies for analysis and applications*. Merrill Prentice Hall.
- Grover, S., & Pea, R. (2013). Computational thinking in K-12. *Educational Researcher*, 42(1), 38-43. <https://doi.org/10.3102/0013189x12463051>
- Hair, J. F., Celsi, M. W., Oritinau, D. J., & Bush, R. P. (2017). *Essentials of marketing research*. McGraw Hill.
- Haseski, H. I., Ilic, U., & Tugtekin, U. (2018). Defining a new 21st century skill-computational thinking: Concepts and trends. *International Education Studies*, 11(4), 29. <https://doi.org/10.5539/ies.v11n4p29>
- Hilgard, E. R. (1980). The trilogy of mind: Cognition, affection, and conation. *Journal of the History of the Behavioral Sciences*, 16(2), 107-117. [https://doi.org/10.1002/1520-6696\(198004\)16:2<107::AID-JHBS2300160202>3.0.CO;2-Y](https://doi.org/10.1002/1520-6696(198004)16:2<107::AID-JHBS2300160202>3.0.CO;2-Y)
- Hill, C., & Koekemoer, E. (2013). The development of the MACE work-family enrichment instrument. *SA Journal of Industrial Psychology*, 39, 1-16. <https://doi.org/10.4102/sajip.v39i2.1147>

- Jong, M. S. -Y., Geng, J., Chai, C. S., & Lin, P. Y. (2020). Development and predictive validity of the computational thinking disposition questionnaire. *Sustainability*, 12(11), 4459. <https://doi.org/10.3390/su12114459>
- Kim, B., Kim, T., & Kim, J. (2013). Paper-and-pencil programming strategy toward computational thinking for non-majors: design your solution. *Journal of Educational Computing Research*, 49(4), 437-459. <https://doi.org/10.2190/ec.49.4.b>
- Lai, J. S., & Eton, D. T. (2002). Clinically meaningful gaps. *Rasch Measurement Transactions*, 15(4), 850.
- Linacre, J. M. (1994). Sample size and item calibration (or person measure) stability. *Rasch Measurement Transactions*, 7(4), 328.
- Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions*, 16(2), 878.
- Linacre, J. M. (2003). *Winsteps computer program version 3.48*. Chicago. <http://www.winsteps.com/a/winsteps-tutorial>
- Linacre, J. M. (2005). *Winstep Rasch measurement computer program*. MESA Press.
- Linacre, J. M. (2007). *A user's guide to WINSTEPS Rasch-model computer program*. MESA Press.
- Linacre, J. M. (2010). When to stop removing items and persons in Rasch misfit analysis? *Rasch Measurement Transactions*, 23(4), 1241.
- Linacre, J. M. (2012). *A user's guide to WINSTEPS: Rasch model computer programs*. MESA Press.
- Linacre, J. M. (2017). *Winsteps Rasch measurement computer program user's guide*. *Winsteps.com*.
- Marret, M. J., & Choo, W. Y. (2017). Factors associated with online victimization among Malaysian adolescents who use social networking sites: A cross-sectional study. *BMJ Open*, 7(6), e014959. <https://doi.org/10.1136/bmjopen-2016-014959>
- Masek, A., & Nasaruddin, N. (2016). Students' perception and readiness on school-based assessment. *Mediterranean Journal of Social Sciences*, 7(6), 189. <https://doi.org/10.5901/mjss.2016.v7n6.p189>
- Matore, M. E. E. M., Zainal, M. A., Noh, M. F. M., Khairani, A. Z., Rahman, N. A., & Idris, H. (2020). Pengujian psikometrik item kecerdasan menghadapi cabaran untuk pelajar lelaki kejuruteraan mekanikal menggunakan model pengukuran Rasch [Psychometric properties of an adversity quotient items for mechanical engineering male students using Rasch measurement model]. *Malaysia Education Journal*, 45(1), 87-100. <https://doi.org/10.17576/jpen-2020-45.01si-11>
- Miller, L. A., & Lovler, R. L. (2019). *Foundations of psychological testing: A practical approach*. SAGE.
- National Research Council. (2011). *Report of a workshop on the pedagogical aspects of computational thinking*. National Academies Press.
- Norris, S. P., & Ennis, R. H. (1989). *Evaluating critical thinking*. Midwest Publications.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. McGraw Hill.
- Olabe, J. C., Basogain, X., Olabe, M. Á., Maíz, I., & Castaño, C. (2014). Solving math and science problem in the real world with a computational mind. *New Approaches in Education Research*, 3(2), 75-82. <https://doi.org/10.7821/naer.3.2.75-82>
- Othman, N., Salleh, S., Hussein, H., & Wahid, H. A. (2014). Assessing construct validity and reliability of competitiveness scale using Rasch model approach. In *WEI International Academic Conference Proceedings* (pp. 113-120). Indonesia.
- Pallant, J. (2001). *SPSS survival manual*. Open University Press.
- Pallant, J. F., & Tennant, A. (2007). An introduction to the Rasch measurement model: An example using the hospital anxiety and depression scale (HADS). *British Journal of Clinical Psychology*, 46(1), 1-18. <https://doi.org/10.1348/014466506x96931>
- Perera, C. J., Sumintono, B., Jiang, N. (2018). The psychometric validation of the principal practices questionnaire based on item response theory. *Educational Leadership*, 2(1), 21-38. <https://doi.org/10.22452/ijel.vol2no1.3>
- Qiu, R. G. (2009). Computational thinking of service systems: dynamics and adaptiveness modeling. *Service Science*, 1(1), 42-55. <https://doi.org/10.1287/serv.1.1.42>
- Salihuddin, M. S., Hasnah, M., Zaleha, A., Norasykin, M. Z., Baharuddin, A., & Mageswaran, S. (2016). Enhancing student's higher order thinking skills (HOTS) through the Socratic method approach with technology. In *1St ICRIL-International Conference on Innovation in Science and Technology*. Technology University of Malaysia.
- Sands, P., Yadav, A., Good, J. (2018). Computational thinking in K-12: In-service teacher perceptions of computational thinking. In M. S. Khine (ed.), *Computational Thinking in the STEM Disciplines* (pp. 151-164). Springer. http://dx.doi.org/10.1007/978-3-319-93566-9_8
- Schiffman, L. G., Lazar, L., & Håvard, K. (2012). *Consumer behavior: A European outlook*. Pearson.
- Sekaran, U. (2003). *Research methods for business: A skill-building approach*. John Wiley & Sons.
- Settle, A., Franke, B., Hansen, R., Spaltro, F., Jurisson, C., Rennert-May, C., & Wildeman, B. (2012). Infusing

- computational thinking into the middle and high-school curriculum. In *Proceedings of the 17th ACM Annual Conference on Innovation and Technology in Computer Science Education - ITiCSE '12* (pp. 22-27). Association for Computing Machinery. <https://doi.org/10.1145/2325296.2325306>
- 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>
- Sondakh, D. E., Osman, K., & Zainudin, S. (2020a). A proposal for holistic assessment of computational thinking for undergraduate: Content validity. (2020). *European Journal of Educational Research*, 9(1), 33-50. <https://doi.org/10.12973/eu-jer.9.1.33>
- Sondakh, D. E., Osman, K., & Zainudin, S. (2020b). A pilot study of an instrument to assess undergraduates' computational thinking proficiency. *International Journal of Advanced Computer Science and Applications*, 11(11), 263-273. <https://doi.org/10.14569/ijacsa.2020.0111134>
- Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial [Rasch model application for social disciplines]*. Trim Komunikata Publishing House.
- Sysło, M. M., & Kwiatkowska, A. B. (2015). Introducing a new computer science curriculum for all school levels in Poland. In A. Brodnik, & J. Vahrenhold (Eds.), *ISEEP* (pp. 141-154). Springer. https://doi.org/10.1007/978-3-319-25396-1_13
- Tang, K. Y., Chou, T. L., & Tsai, C. C. (2019). A content analysis of computational thinking research: An international publication trends and research typology. *The Asia-Pacific Education Researcher*, 29(1), 9-19. <https://doi.org/10.1007/s40299-019-00442-8>
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53-55. <https://doi.org/10.5116/ijme.4dfb.8dfd>
- Weese, J. L. (2016). Mixed methods for the assessment and incorporation of computational thinking in K-12 and higher education. In *Proceedings of the 2016 ACM Conference on International Computing Education Research-ICER '16* (pp. 279-280). Association for Computing Machinery. <https://doi.org/10.1145/2960310>.
- 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>
- Wright, B. D., & Master, G. N. (1982). *Rating scale analysis*. Cesa Press.
- Wright, B. D., & Mok, M. M. (2004). An overview of the family of Rasch measurement models. In E. V. Smith, & R. M. Smith (Eds.), *Introduction to Rasch measurement: Theory, models, and applications*. JAM Press.
- Wu, M., & Adams, R. (2007). *Applying the Rasch model to psycho-social measurement: A practical approach*. Educational Measurement Solutions.
- Yadav, A., Mayfield, C., Zhou, N., Hambrusch, S., & Korb, J. T. (2014). Computational thinking in elementary and secondary teacher education. *ACM Transactions on Computing Education*, 14(1), 1-16. <https://doi.org/10.1145/2576872>

<https://www.ejmste.com>