Semarak International Journal of Machine Learning 6, Issue 1 (2025) 25-33

Semarak International Journal of Machine Learning

Journal homepage: https://semarakilmu.my/index.php/sijml/index
ISSN: 3030-5241

Applying Rasch Model Technique on Stacking and Racking Analysis the Measuring the Understanding of Chemistry Teachers on Computational Thinking Concept

Omay Sumarna[1], Septian Karyana[1,2,*], Agus Setiawan[1], Sjaeful Anwar[1]

[1] Program Studi Pendidikan Ilmu Pengetahuan Alam, Universitas Pendidikan Indonesia, Jl. Dr. Setiabidhi No. 299, Bandung, Indonesia
[2] SEAMEO Regional Centre for QITEP in Science, BBGTK JABAR, Gedung B, Jl. Dr. Cipto No.9, Bandung, Indonesia

ARTICLE INFO

Article history:
Received 28 June 2025
Received in revised form 25 July 2025
Accepted 10 August 2025
Available online 23 September 2025

Keywords:
Computational thinking; stacking; racking; Rasch model

ABSTRACT

This research aimed to measure teachers' ability to understand computational thinking concepts by using stacking and racking analysis techniques in the Rasch model. The participants of this study were chemistry teachers who had been selected to attend training for 6 days. Data collection was carried out before and after participating in the training using the same instrument. The data obtained was then analyzed using WINSTEPS 5.4.1 software. The results showed that teachers who had attended the training experienced changes in their knowledge of computational thinking concepts. Statistically, the change in the ability of teachers' computational thinking concept knowledge moves from -0.33 logit to 2.44 logit. Along with the intervention effect, in certain cases, positive conceptual changes were found due to teachers' lucky guesses and cheating. In other cases, teachers were found to be careless in answering questions. Stacking and racking analysis is essential in detailing any changes in teacher ability, item difficulty, and learning progress.

1. Introduction

The idea of computational thinking (CT) is not new in the education field. CT is a term used since the 1950s that describes the notion of using structured thinking or algorithmic thinking [1]. In the 1960s, Perlis argued that every student from any discipline should learn about programming and the "theory of computation". However, in primary and secondary education, computing first became a concern after Seymour Papert in the 1980s pioneered the idea of developing procedural thinking through programming [2]. Furthermore, Jeannette Wing (2006) published an article entitled "Computational Thinking" which appeared in the March 2006 issue of Viewpoint from ACM Communication. In the article, she defined CT as "solving problems, designing systems, and understanding human behavior, by drawing on the concepts fundamental to computer science. Computational thinking includes a range of mental tools that reflect the breadth of the field of

* *Corresponding author.*
E-mail address: septian@upi.edu

https://doi.org/10.37934/sijml.6.1.2533

computer science" [3]. Based on this, the essence of CT is to think like a computer scientist when facing a problem.

However, the term CT is still not widely recognized by teachers [3]. CT is still often misinterpreted as the ability to use computers [4-6]. In essence, CT is an analytical thinking skill that draws on computer science concepts. However, CT is a fundamental skill that can be used by and is useful for everyone [3]. Some scholars argue that CT is an important skill for all fields of science, not just computer science [7]. The application of computational thinking through systematic problem-solving has already permeated multiple disciplines, establishing CT as a vital competency across various fields [8]. This powerful idea is beginning to have a significant impact in many fields, including medicine and health care, archaeology, traffic engineering, music and law [9], making it important to include CT as a priority in primary and secondary education. Wing (2006) said, "To reading, writing, and arithmetic, we should add computational thinking to every child's analytical ability." In short, CT is a set of problem-solving thought processes that originated in computer science but can be applied in any domain.

Several existing studies have been conducted to measure teachers' perceptions of what constitutes CT using surveys. The survey examined teachers' concepts of CT and how they differed based on teaching experience and subject matter taught. Although teachers agreed with the statement that CT is problem solving, logical thinking, and algorithmic thinking, they tended to view CT as math, using computers, and playing online games. In addition, there was no difference in how teachers viewed CT based on their background in science or grade level, either elementary or secondary [10].

In another study conducted to examine pre-service teachers' concept of CT, it was found that pre-service teachers' descriptions of CT varied widely and that familiarity with the term did not necessarily result in a more sophisticated view of CT [11]. When discussing what pedagogical strategies pre-service teachers could use to develop learners' CT skills, the use of simple technology emerged as the most popular response. Overall, they found that prospective teachers had a weak understanding of CT and most had low confidence in their ability to teach CT.

Another study was conducted with elementary school teachers on the conception of CT in Italy [12]. Based on a literature review, they categorized CT elements into four categories, namely mental processes (such as algorithmic thinking), methods (such as automation), practices (such as testing and debugging), and transversal skills (such as creating). Analysis of survey responses from 972 teachers showed that most teachers did not conceptualize CT in these four categories. Teachers also reported that they did not feel prepared to develop CT competencies in their learners. One of the positives of the survey was that many teachers viewed computing and IT tools as separate.

Overall, the studies show that teachers tend to equate CT with computing, and they express a lack of confidence in being able to integrate CT into their classrooms in a meaningful way. The studies also show the potential for teachers to make a big difference in what CT means but how teachers learn to think about what integrating CT into learning looks like is not much explained.

Therefore, teachers need to know and understand how CT can be integrated into science learning, so that it can improve students’ CT skills. To do that, Southeast Asian Ministers of Education (SEAMEO) Regional Centre for Quality Improvement of Teacher and Education Personnel in Science (SEAQIS) developed a training program on Integrating Computational Thinking in Science Classroom. This program was organized for science teachers - Physics, Chemistry, Biology - at the high school level. This training aims to make science teachers have CT knowledge and skills and be able to develop learning designs that integrate CT into science learning.

This program was conducted because there are still few resources that directly integrate CT concepts and practices into the science context [13]. Despite efforts to integrate CT at various levels

[14-17] and in teacher preparation [4,5], there is still limited research on how teachers understand CT, especially in terms of how they think about implementing it in their own classrooms. To integrate CT in learning settings, teachers must be directly involved in the learning process [18].

2. Methodology

2.1 Participants

The participants in this program were 20 chemistry teachers in West Java province. The trainees have different educational backgrounds and different teaching experiences. Table 1 shows the demographic data of the trainees.

Table 1
Demographic data of trainees

Teaching Experience			Education Level	
Category	Numbers	Percentage	Category	Percentage
< 5 year	2	10%	S1	40%
5 – 10 year	2	10%		
10 – 15 year	4	20%		
15 – 20 year	3	15%	S2	60%
> 20 year	9	45%		

Teachers attended the training for 6 days face to face. They received materials on Concept of Computational Thinking, Integration of Computational Thinking in Science Learning, CT-STEM Taxonomy in Science Learning, Review of CT Integration Units in Science Learning, Analysis of Chemistry Topics that will be integrated with CT, and material on Assessment in CT Integrated Science Learning.

2.2 Instrument

The instrument is a multiple-choice test consisting of 15 questions. The questions measure the knowledge and understanding of CT concepts and how to integrate CT in chemistry learning. The items developed by the facilitator team and the content were validated by experts in educational assessment in chemistry.

To ensure the correct measurement of teacher achievement is made and meaningful conclusions can be made, it is imperative that the measurement tool is reliable and valid in measuring teacher proficiency in CT concepts [19]. Therefore, the instrument was statistically tested using Rasch modelling. Table 2 shows the result of Rasch measurement.

Table 2
Summary statistic

		Infit		Outfit		Reliability	Cronbach Alpha (α)
		MNSQ	ZSTD	MNSQ	ZSTD		
Item	Min	0.63	-2.35	0.5	-2.21	0.73	0.81
	Maks	1.44	2.21	2.36	2.68		
Person	Min	0.58	-1.27	0.26	-1.18	0.68	
	Maks	1.39	1.24	2.08	1.66		

In Rasch modelling, the validity of an instrument is determined by analyzing the infit and outfit mean square (MNSQ) values for content validity and the point measure correlation (PTMEA Corr) value for item polarity. Meanwhile, instrument reliability is determined by Cronbach alpha (α), person reliability, and item reliability.

The item fit test is carried out for content validity referring to the MNSQ and ZSTD values. The MNSQ value condition must be in the range of 0.5 - 1.5 and the ZSTD value range is -2.0 - +2.0 [20]. In addition, if the MNSQ value is accepted, the ZSTD value can be ignored [21]. Based on Table 2, the range of MNSQ values for both items and persons is in the predetermined range or appropriate with Rasch modeling. Furthermore, construct validity can be measured by referring to item polarity. The polarity of this item is described by the PTMEA Corr value. Based on the analysis results, the value of PTMEA Corr is positive with a range of 0.22 to 0.75. Therefore, it can be concluded that the items will contribute to the measurement of CT concept understanding ability, and all items developed can be retained as items in the instrument.

The item reliability test is used to assess whether the instrument used measures what should be measured and provides appropriate results. Meanwhile, the person reliability test was conducted to test whether the participants filling out the instrument were the right individuals to answer the instrument [22]. Based on Table 2, it is shown that the Cronbach alpha value is 0.81. This indicates that the instrument is reliable and acceptable with a high level of consistency in measuring teachers' CT comprehension skills. This high value also indicates that the instrument is in good condition with a high level of consistency so that it is acceptable for use in this study.

Furthermore, Table 2 shows that the item reliability value is in the good enough category at 0.73, which indicates that the instrument has good reliability in measuring what needs to be measured and there are enough items to measure accurate reading. In addition, the person reliability value is also in the good enough category, which is 0.68. This indicates that 68% of the results from teacher responses will be repeated if the same respondent answers questions related to alternative instruments [21].

Based on the description, the instruments used in this study were appropriate with Rasch modeling and can be used to measure teachers' CT comprehension skills.

2.3 Data Collecting and Analysis

Before the intervention, teachers took a pre-test to determine their initial knowledge and then they took a post-test using the same questions after the training. Teachers filled in the pre- and post-test answers online so that their answers could be recorded immediately.

The pre- and post-test data measurements are still ordinal data. The Rasch model with WINSTEPS 5.4.1 software was used to convert ordinal data into interval data to have the same scale. The result is calibrated data on the level of ability and difficulty of teacher questions in the same interval.

Data analysis was conducted using stacking and racking techniques [23]. Data analysis with the stacking technique places pre- and post-test data together vertically. Each respondent appeared twice in the data set, while the item appeared once. This allows the researcher to examine each respondent's change after the intervention. In contrast, the racking analysis technique places the pre- and post-test data horizontally. Each pre- and post-test item appears twice in the data set, while the respondent appears once. This allows the researcher to examine the change in difficulty of each item before and after the intervention [23].

3. Results

3.1 Conceptual Changes in Chemistry Teachers' Understanding Ability of Computational Thinking

Changes in the concept understanding ability of chemistry teachers on computational thinking are determined by comparing the value of the pre- and post-test using the stacking technique in the Rasch model [23].

Table 3
The score of chemistry teacher pre- and post-test average ability

Code	Mean (logit)	Diff pre- and post-test (logit)	SD
Pre-test	-0.33	2.77	0.72
Post-test	2.44		1.23

Table 3 presents the results of measuring the mean ability of conceptual understanding change scores before and after the training. Based on Table 3, it is known that the mean size of the post-test item (2.44 logits) is greater than the mean size of the pre-test item (-0.33 logits). This shows that chemistry teachers who attended the training experienced a good understanding of the concept of computational thinking.

3.2 Nature of Change Chemistry Teacher Pre-and Post-test Understanding

The underlying reason for this analysis is to find out which teachers experienced positive and negative pre- and post-test changes. A simple description of this can be explained by using a scatter plot graph. Figure 1 illustrates the scatter plot graph of the pre- and post-test measures of teachers who participated in the training.

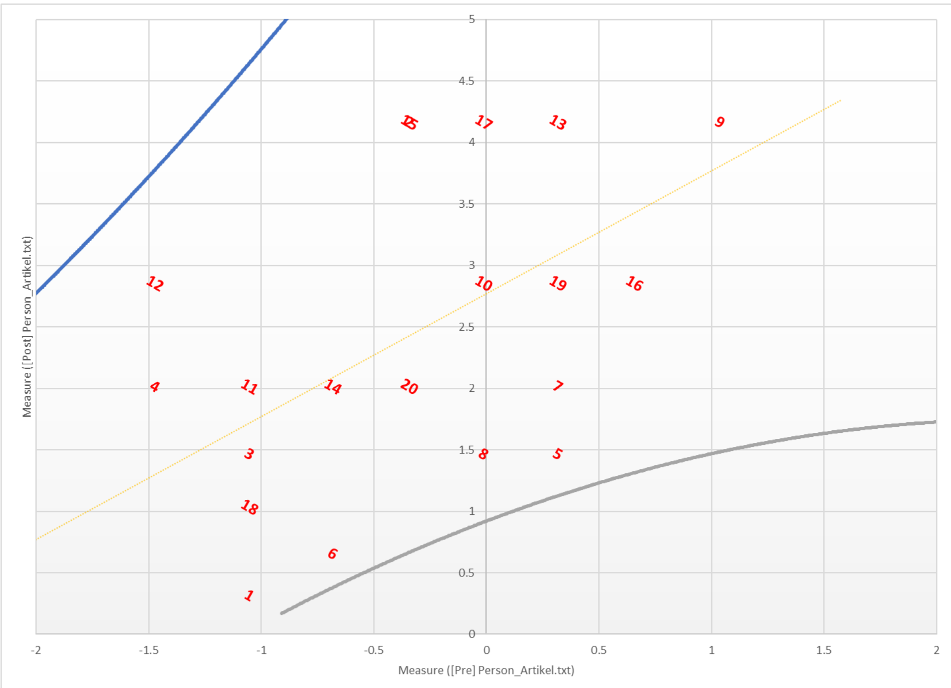

Fig. 1. Scatter plot of chemistry teacher pre- and post-test ability measures

Based on Figure 1, it can be explained that, first, the pretest output range moves from -2 logits to +2 logits, while for post-test output it moves from 0 logits to +5 logits. Based on this information, there is a diversity of teachers' initial knowledge and understanding of the CT concept before they attended the training. Furthermore, teachers get a positive logit value after attending the training. This means that they experienced an increase in knowledge after the training.

Second, there are two lines, a blue line (top) and a grey line (bottom). These two lines represent the areas of change in teachers' knowledge before and after the training. It can be seen from Figure 1 that all teachers are in knowledge change. This means that they experienced a change in knowledge after attending the training. Furthermore, based on Figure 1, there is no significant negative nature of change. This indicates that during the training process, teachers can absorb information about CT concepts and experience changes in knowledge after attending the training.

3.3 The Changes in item difficulty level

Table 4 displays the results of the racking analysis relating to changes in the difficulty level of items in the pre and post-test. Based on Table 4, the average pre-test difficulty level is -0.32 logits, the average post-test is -0.93 logits, and the average pre-post difference is -0.61 logits. This study also found that nine items met a significant change in the level of item difficulty. They were lower than the difference between pre and post-test (-0.61), namely items 1, 4, 5, 6, 7, 8, 9, 10, and 15. Furthermore, there were six items whose difficulty level was greater than the average, namely items 2, 3, 11, 12, 13, and 14. The most difficult item in the pre-test condition was item 1 (2.04 logits), and the easiest item was item 2 (-4.77 logits). While in the post-test condition, the most difficult item was item 11 (1.64 logit), and the easiest items were items 5 and 7 (-1.53 logit). This finding indicates that there is a difference in changes in the level of item difficulty between teachers before and after attending the training.

Table 4
Data of item measures of pre- and post-test

Item	Pre-test	Post-test	Diff Pre and Post
Item 1	2.04	1.32	-0.72
Item 2	-4.77	-2.8	1.97
Item 3	-1.05	-0.14	0.91
Item 4	-0.58	-2.8	-2.22
Item 5	0.59	-1.53	-2.12
Item 6	1.56	-2.8	-4.36
Item 7	-0.13	-1.53	-1.4
Item 8	0.1	-0.69	-0.79
Item 9	-0.81	0.67	1.48
Item 10	-0.13	0.29	0.42
Item 11	0.59	1.64	1.05
Item 12	0.59	-0.69	-1.28
Item 13	-1.58	-2.8	-1.22
Item 14	-1.3	0.67	1.97
Item 15	0.1	-2.8	-2.9
Mean	-0.32	-0.93	-0.61

3.4 The Changes in Teacher Understanding and Item Difficulty Level

In addition to the effect of the training intervention, there are three other factors that tend to affect changes in teacher ability and the level of difficulty of the questions, namely 1) guessing the correct answer or lucky guess, 2) cheating, and 3) carelessness. These three factors can be identified from teachers' item response patterns using a scalogram. Table 5 displays the teachers' item response patterns. For example, the response patterns for teachers 1 and 18. Both teachers did not understand items 12 (-0.69 logit), 3 (-0.14 logit), and 10 (0.29 logit). Meanwhile, teachers 1 and 18 were able to answer correctly the more complicated item, item 14 (0.67 logit). This condition implies the existence of a lucky guess. This is also confirmed by the Rasch output shown in Figure 2. Both teachers had knowledge thresholds below the difficulty level of the questions, but both teachers were able to answer the questions correctly.

Furthermore, there is an indication of cheating in the response patterns of teachers 10 and 19. This indication can be seen from the Rasch measurement output that all indicators have the same value. Based on the response patterns in Table 5, teacher 4 could not answer item 5 (-1.53 logit) which was easy but could answer item 8 (-0.69 logit) which was more difficult than item 5. In addition, they had a high post-test proficiency.

Table 5
Scalogram

```
GUTTMAN SCALOGRAM OF RESPONSES:
PERSON |ITEM
       |   11   1 1 1 1    Pre-      Post-      Pre & Post     Item
       |246355782309411    Test      Test       Difference     Response Pattern
       |---------------
    18 +111111110000111    -1.11     .51        1.62           Lucky Guess
     1 +111111010001100    -1.11     -.5        .61            Lucky Guess

    10 +111111111111110    -.02      2.58       4.13           Cheating
    19 +111111111111110    .33       2.58       2.25           Cheating

    16 +111111111111011    0.69      2.58       1.89           Careless
     4 +111110111111101    -1.55     1.67       3.22           Careless
    11 +111111101111011    -1.11     1.67       2.78           Careless
```

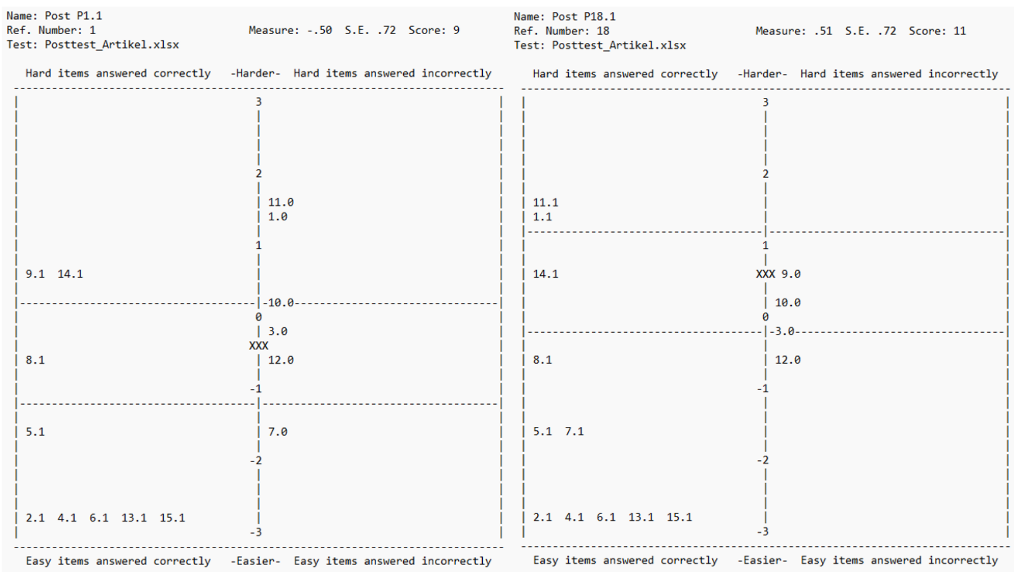

Fig. 2. Person diagnostic maps

4. Conclusions

The findings of this study indicate the conceptual change in teachers' understanding of computational thinking and items after attending the training. This is indicated by the average change from -0.33 logit (pre-test) to 2.44 logit (post-test). In addition, based on the results of the stacking analysis, it was found that all teachers experienced a positive change in ability. However, some special conditions reflect an inappropriate response pattern. After further analysis, it turns out that there are three types of inappropriate response patterns with the three patterns being lucky guess, cheating, and careless. Two teachers might do lucky guesses, namely teachers 1 and 18, and two teachers might do cheating, namely teachers 10 and 19. Finally, three teachers might have done carelessness, namely teachers 4, 11, and 16.

Acknowledgement
The authors would like to thank all parties for their assistance, especially the SEAMEO Regional Center for QITEP in Science for organizing the training for teachers and accommodating all the necessary needs.

Acknowledgement
This research was not funded by any grant.

References
[1] Angeli, Charoula, and Michail Giannakos. "Computational thinking education: Issues and challenges." *Computers in human behavior* 105 (2020): 106185. https://doi.org/10.1016/j.chb.2019.106185
[2] Grover, Shuchi, and Roy Pea. "Computational thinking in K–12: A review of the state of the field." *Educational researcher* 42, no. 1 (2013): 38-43. https://doi.org/10.3102/0013189X12463051
[3] Wing, Jeannette M. "Computational thinking." *Communications of the ACM* 49, no. 3 (2006): 33-35. https://doi.org/10.1145/1118178.1118215
[4] Yadav, Aman, Chris Stephenson, and Hai Hong. "Computational thinking for teacher education." *Communications of the ACM* 60, no. 4 (2017): 55-62. https://doi.org/10.1145/2994591
[5] Czerkawski, Betul C., and Eugene W. Lyman. "Exploring issues about computational thinking in higher education." *TechTrends* 59 (2015): 57-65. https://doi.org/10.1007/s11528-015-0840-3
[6] Yadav, Aman, Chris Mayfield, Ninger Zhou, Susanne Hambrusch, and John T. Korb. "Computational thinking in elementary and secondary teacher education." *ACM Transactions on Computing Education (TOCE)* 14, no. 1 (2014): 1-16. https://doi.org/10.1145/2576872
[7] Denning, Peter J. "The profession of IT Beyond computational thinking." *Communications of the ACM* 52, no. 6 (2009): 28-30. https://doi.org/10.1145/1516046.1516054
[8] Bundy, Alan. "Computational thinking is pervasive." *Journal of Scientific and Practical Computing* 1, no. 2 (2007).
[9] National Research Council, Division on Engineering, Physical Sciences, Computer Science, Telecommunications Board, and Committee for the Workshops on Computational Thinking. *Report of a workshop on the scope and nature of computational thinking*. National Academies Press, 2010.
[10] Sands, Phil, Aman Yadav, and Jon Good. "Computational thinking in K-12: In-service teacher perceptions of computational thinking." *Computational thinking in the STEM disciplines: Foundations and research highlights* (2018): 151-164. https://doi.org/10.1007/978-3-319-93566-9_8

[11] Bower, Matt, and Katrina Falkner. "Computational thinking, the notional machine, pre-service teachers, and research opportunities." In *Proceedings of the 17th Australasian computing education conference (ACE 2015)*, vol. 27, p. 30. Australian Computer Society, 2015.

[12] Corradini, Isabella, Michael Lodi, and Enrico Nardelli. "Conceptions and misconceptions about computational thinking among Italian primary school teachers." In *Proceedings of the 2017 ACM conference on international computing education research*, pp. 136-144. 2017. https://doi.org/10.1145/3105726.3106194

[13] Ahamed, Sheikh Iqbal, Dennis Brylow, Rong Ge, Praveen Madiraju, Stephen J. Merrill, Craig A. Struble, and James P. Early. "Computational thinking for the sciences: a three day workshop for high school science teachers." In *Proceedings of the 41st ACM technical symposium on Computer science education*, pp. 42-46. 2010. https://doi.org/10.1145/1734263.1734277

[14] Rich, Kathryn M., Aman Yadav, and Christina V. Schwarz. "Computational thinking, mathematics, and science: Elementary teachers' perspectives on integration." *Journal of Technology and Teacher Education* 27, no. 2 (2019): 165-205. https://doi.org/10.70725/303733vqleui

[15] Waterman, Kevin P., Lynn Goldsmith, and Marian Pasquale. "Integrating computational thinking into elementary science curriculum: An examination of activities that support students' computational thinking in the service of disciplinary learning." *Journal of Science Education and Technology* 29, no. 1 (2020): 53-64. https://doi.org/10.1007/s10956-019-09801-y

[16] Ketelhut, Diane Jass, Kelly Mills, Emily Hestness, Lautaro Cabrera, Jandelyn Plane, and J. Randy McGinnis. "Teacher change following a professional development experience in integrating computational thinking into elementary science." *Journal of science education and technology* 29 (2020): 174-188. https://doi.org/10.1007/s10956-019-09798-4

[17] Arık, Merve, and Mustafa Sami Topçu. "Computational thinking integration into science classrooms: Example of digestive system." *Journal of Science Education and Technology* 31, no. 1 (2022): 99-115. https://doi.org/10.1007/s10956-021-09934-z

[18] Stephenson, Chris, Judith Gal-Ezer, Bruria Haberman, and Anita Verno. "The New Educational Imperative: Improving High School Computer Science Education: Final report of the CSTA Curriculum Improvement Task Force." (2005).

[19] Abdaziz, Azrilah, Muhammad Shahar Jusoh, Harith Amlus, and T. S. Salleh. "Construct validity: A Rasch measurement model approaches." *Journal of Applied Science and Agriculture* 9, no. 12 (2014): 7-12.

[20] Linacre, John M. "A user's guide to FACETS: Rasch-model computer programs." *Retrieved December* 18 (2014): 2018.

[21] KAMARUZAMAN, FATHIYAH MOHD, ROSZILAH HAMID, MOHAMAD SATTAR RASUL, MARLISSA OMAR, and MOHD FAKHRUL AZIZIE MOHD ZAID. "VALIDITY AND RELIABILITY OF GS4IR INSTRUMENT FOR ENTRY-LEVEL CIVIL ENGINEERS USING RASCH MODEL." *Journal of Engineering Science and Technology* 18, no. 3 (2023): 1570-1580.

[22] Bond, Trevor G., and Christine M. Fox. *Applying the Rasch model: Fundamental measurement in the human sciences*. Psychology Press, 2013.

[23] Wright, Benjamin D. "Rack and stack: time 1 vs. time 2." *Rasch measurement transactions* 17, no. 1 (2003): 905-906.