JIPF (JURNAL ILMU PENDIDIKAN FISIKA)
p-ISSN: 2477-5959 | e-ISSN: 2477-8451 *Vol. 8 No. 2, May 2023, Page 259-268*

Developing Two Tier Test of Critical Thinking Ability using Rasch Model

Novika Lestari [1*)], Chairun Nisa [2], Mutazam [3], Soeharto [4]
STKIP Melawi, Indonesia[1,3], IAIN Takengon, Indonesia[2], University of Szeged, Hungary[4]
[*)]Corresponding E-mail: novika.lestari02@gmail.com[1]

Received: October 16th, 2022. Revised: March 29th, 2023. Accepted: May 2nd, 2023

Keywords **:**
Critical thinking; Partial Credit Model (PCM); Rasch Model; High Order Thinking; Logic Model, Two-tier Test

ABSTRACT
Critical thinking ability refers to cognitive processes to solve problems so that students are able to relate learning to everyday life. This achievement needs to be measured with valid and reliable instruments. Therefore, the research was conducted with the aim of characterizing the critical thinking test type two tier tests using the Rasch model. This study used a construction design and test validation. The construction process resulted in a critical thinking test of the two-tier type in the form of multiple choice on physics material for class X semester 1 which measured three basic aspects of critical thinking. The validation procedure was based on a trial exam taken by 146 students in Melawi Regency and a test evaluation by two experts. The findings demonstrated that the Aiken's V content validity analysis and the analysis utilizing item response theory both supported the validity of the critical thinking test. The analysis's findings demonstrated the validity and reliability of this test of critical thinking. The 23 created questions were found to be valid based on the results of the content validity test utilizing Aiken V. This test's empirical validity demonstrates that every item has a good level of difficulty and meets the Rasch model. The physics material used in the designed critical thinking test for class X semester 1 was valid, dependable, and had a good level of difficulty.

INTRODUCTION

Students are required to be proficient in dealing with and solving problems in the surrounding environment. Students are expected to be responsive to the conditions faced by the surrounding environment. Therefore, it is necessary to have basic skills which are often referred to as higher order thinking skills (HOTS). The term HOTS refers to skills including Communication, Collaborative, Critical thinking and Creativity. These skills will be the spearhead in the readiness of students to face competition in the world of work.

Indonesian students' HOTS skills are low. The Program for International Student Assessment (PISA) results from 2019 was 396; seven points lower than that of 2015 [1]. This demonstrates the need to improve students' HOTS competencies, including critical thinking ability. Problem-solving starts with using critical thinking as a fundamental skill. Decision-making, leadership, and judgment skills are all impacted by critical thinking [2]. Students must be able to gather and describe information, pose queries, and react to their surroundings [3].

Critical thinking skill refers to cognitive processes that provoke students to reason, assess, defend, make alternatives reflectively and honestly in accordance with available evidence and arguments. Students who are accustomed to thinking creatively will construct knowledge based on the initial knowledge they have by paying attention to the truth of the knowledge itself. So students who are critical will not easily accept statements from others without thinking scientifically first [4] [5] [6] [7].

Students who have critical thinking skills will be able to basic classification with further support, interference, further clarification accompanied by strategies and tactics [5]. These three steps show that critical thinking skills can provide a broader view, creative solutions, and ways of initiative [8], are able to see opinions related and related to everyday life [6], and build ideas, structures and actions. built and channeled to the interests of the good. There are three approaches that can evaluate critical thinking skills carefully which include: the belief that critical thinking can only be assessed in certain contexts; critical thinking can be judged by friends as a reflection; critical thinking assessment can only be seen from documentation, demonstration, and the students' own assessment [8]. One of the measuring tools that can be used to determine students' critical thinking skills is a test. The test is a measuring instrument that can be quantified that explains the stimulus and perspective of an answer. Test must be able to measure the ability of the stimulus and perspective objectively and can be presented in the form of numbers. Therefore, test is interpreted as presenting a series of questions that must be answered to obtain a measure of a person's ability.

Item response theory (IRT) was used to evaluate the viability of the critical thinking ability test instrument. The tests employed in the learning process should not only focus on (1) the characteristics of items that depend on the sample group utilized, (2) not demand equality of measurement errors for all test subjects, and (3) not demand parallel classes. The Classical Test Theory (CTT), which pays less attention to the interaction between each student and the item, has a drawback in that it cannot accurately reflect students' abilities. Item Response Theory (IRT) addresses this shortcoming. Some of the weaknesses of the CTT are that the raw score is basically not a measurement result. The raw score is initial information, the raw score has a weak quantitative meaning, the raw score does not indicate a person's ability to a particular task, and the raw score and the percentage of correct answers are not always linear [9] [10]. This is in line with the opinion of Pellham III [11] which states that “the main characteristic of the IRT model is that the reliability (or measurement accuracy) of a score varies as a function of the respondent's score on the construction being assessed. This contrasts with the classical test theory approach, where reliability is determined by a single number (eg, alpha coefficient) that applies to each respondent's score.

The general assumptions of the most widely used IRT model are unidimensionality and local independence [12]. Unidimensional means that only one ability is measured by the items that make up the test, while local independence means that the response to one item is free from the influence of other items. There are three unidimensional IRT models, namely 1 logistic parameter (1 PL), 2 logistical parameters (2 PL) and 3 logistical parameters (3 PL) [13]. The difference is that the 1 PL model only measures the level of difficulty, the 2 PL model measures the level of difficulty and false guesses while the 3 PL model measures the level of difficulty, false guesses and discrimination [14].

The Rasch model is a type IRT with 1 logistic parameter (1PL) with item difficulty parameters. The Rasch model explains that students who have higher abilities are more likely to answer questions correctly than students with lower abilities. This also occurs in the item questions. Items that have high difficulty, the probability that the item can be solved is also low. As a 1PL model, the Rasch model is

DOI: 10.26737/jipf.v8i2.3737

used to solve dichotomous score. The analysis of rasch model which modify the dichotomous score into a polytomy score is called the Partial credit model (PCM). The PCM model is a model for constructing measures using items with two or more response categories [15].

PCM is an extension of Model 1-PL and belongs to the Rasch Model family [16] which has developed operating characteristic functions (OCF) [17]. The presence of the Rasch model as a new measurement system, aims to overcome the limitations of the classical measurement system or Classical Test Theory (CTT) [18] [19]. Estimation of the fit of items and parameters in the Rasch model including the PCM model can be determined by using an iterative proportional fitting algorithm for the loglinear model [20]. This model ensures that one construct is being measured, and the items are systematically and predictably related to one another [21]. If the data deviate greatly from the Rasch model, the cause needs to be considered and the inappropriate person or item may need to be removed [22]

The PCM model from Masters has the characteristics of (1) the polytomus scoring model can give credit to the correct application of the items, (2) the polytomus scoring model will provide an opportunity to increase measurement accuracy compared to the dichotomous model, (3) between stages in solving questions does not always have the same level of difficulty, and (4) the higher category does not always have a higher threshold, nor does the lower category always have a lower threshold [23] [24]. PCM is an analytical model of the IRT form (Item Response Theory) in which students' responses to problems can describe students' abilities [25].

Thus, it is necessary to develop a critical thinking ability test using the Rasch model analysis. This test is intended for high school students of Class X Revised 2013 Curriculum. The specification of the test developed is a two-tier test with closed multiple choice reasons.

METHOD

This is a research to develop a test of students' critical thinking skills in physics. The critical thinking ability indicator was adopted from Ennis' critical thinking aspect (ability) [26]. In detail, the critical thinking indicators applied in the test instrument for students' critical thinking skills are listed in Table 1.

Table 1. Aspect of Critical Thinking Ability

Aspect	HOTS	Indicator	Definition
Basic Clarification	Interpreting (C4)	Explaining	Identifying and state information from a problem
		Categorizing	Distinguishing irrelevant information and relevant information in the problem
	Analyzing (C4)	Connecting	Connecting the variables in the problem
		Interpreting	Making various representations of the problem
	Evaluating (C5)	Examining	Assessing the truth of an explanation or argument
		Criticizing	Checking the explanation of the problem
	Concluding (C5)	Guessing	Finding causal relationships from a problem logically to draw conclusions
		Formulating	Making conclusions from explanations
Further clarifying	Further clarifying (C6)	Clarifying	Evaluating causal relationships with relevant theories
		Proving	Comparing the conclusions drawn with the relevant theory
Strategy and tactics	Making decisions (C6)	Constructing	Designing conclusions from several theories that contain alternative solutions

Two analysis used included content validity and item characteristics. Content validity data were analyzed using Aiken-V, while the analysis of test item characteristics used the Winstep-assisted Rasch model [27]. Test items characteristics test involved 146 students in Melawi Regency, West

Kalimantan. The Rasch model has better accuracy compared to CTT [28] [29]. The grain characteristics were analyzed using the Rasch Model type Partial Credit Model (PCM). This analysis is a development of the dichotomous item analysis which is applied to the polytomy item and included goodness fit items (item match), level of difficulty, reliability with test information function (TIF) curves and standard error measurement (SEM) values, test characteristic curve (TCC) analysis and student ability levels that match the test items (ability).

RESULTS AND DISCUSSIONS

The physics material on the test instrument consists of Newton's law and Newton's law of gravity and work-energy material. The distribution of critical thinking ability test items for physics class X semester 1 is shown in Table 2.

Table 2. Distribution of Critical Thinking Ability Test Items

Aspect	HOTS		Physics Material		
			Newton Law	**Newton Law of Gravity**	**Work & Energy**
Basic classification	C4	Interpreting	1	13	-
	C4	Analyzing	2,3,9,12	17,18	10,16
	C5	Evaluating	8	14,20	15,22
	C5	Concluding	6,19	11	20
Further clarifying	C6	Further clarifying	4,5,17	-	23
Strategy and tactics	C6	Further clarifying	7,18	-	21

Table 3. Sample Task in Indonesian and English Version

Indonesian Version	19. Kendaraan mudah mengalami slip atau tergelincir pada jalan dengan belokan datar yang basah. Untuk mengatasi permasalahan tersebut, apa yang dapat dilakukan pengelola jalan agar mengurangi jumlah kendaraan yang mengalami slip, kecuali? a. Membuat belokan jalan miring ke dalam b. Mempertebal aspal c. Memberi pagar pembatas jalan d. Memasang rambu-rambu kurangi kecepatan e. Memperkecil belokan di jalan Alasan: a. Karena dipengaruhi oleh gaya sentripetal b. Karena dipengaruhi oleh gaya gesek jalan dan ban mobil c. Karena dipengaruhi oleh gaya sentripetal serta gaya gesek jalan dan ban mobil d. Karena dipengaruhi oleh kecepatan dan gaya sentripetal e. Karena dipengaruhi oleh kecepatan dan percepatan mobil
English Version	19. Vehicles are easy to slip or slip on roads with wet flat turns. To overcome this problem, what can road managers do to reduce the number of vehicles that slip, except? a. Make the road turn inward slope b. Thicken the asphalt c. Provide guardrails d. Installing speed reduction signs e. Minimize turns on the road Reason: a. Because it is influenced by centripetal force b. Because it is influenced by road friction and car tires c. Because it is influenced by centripetal force and road friction force and car tires d. Because it is influenced by speed and centripetal force e. Because it is affected by the speed and acceleration of the car

DOI: 10.26737/jipf.v8i2.3737

The test instrument developed is closed multiple choice with two stages of preparation. The first stage is the dissemination of multiple choice tests that include students' critical thinking skills. This test is called the first level test. After that, the test takers were given an interview to state the reasons for choosing the answer. A test that does not have a rationale for choosing an answer or all students guessing the answer will be rejected. The reasons obtained are used in developing the options at the second level. Students get a score of 4 if they answer the right question and the reason is right, score 3 if they answer the question right the reason is wrong, score 2 if the question is wrong and the reason is right, score 1 if the question is wrong and the reason is wrong.

Content Validity

The content validity carried out refers to Aiken-V equation and coefficient [30]. Aiken-V coefficient value for 4 raters with an error rate of 5% was 0.88. Based on Table 2, it is determined that 21 of the 21 higher order thinking skills (HOTS) test items that reflected physics critical thinking abilities were valid. Based on these results, the composition of the distribution of critical thinking ability test items for class X physics is reconstructed as shown in Table 4.

Table 4. Content Validity

Item	**Aiken-V Value**	**Ket**
1	0,90	valid
2	0,92	valid
3	0,89	valid
4	0,95	valid
5	0,93	valid
6	0,90	valid
7	0,90	valid
8	0,89	valid
9	0,89	valid
10	0,92	valid
11	0,90	valid
12	0,93	valid
13	0,89	valid
14	0,90	valid
15	0,91	valid
16	0,91	valid
17	0,90	valid
18	0,90	valid
19	0,89	valid
20	0,91	valid
21	0,88	valid
22	0,90	valid
23	0,94	valid

Item Estimate

The average value of the INFIT Mean of Square (INFITMNSQ) and its standard deviation can be used as a starting point for total item estimation. The standard deviation is close to 0.00 while the size of the INFITMNSQ is nearly 1.00, or vice versa. When the standard deviation is close to 1.00 and INFITMNSQ is close to 0.00, the test item passes the overall Rasch model fit test using the PCM 1 PL model [31]. Therefore, it can be said that item estimates traditionally suit the Rasch model.

Table 5. Item Estimate

Data	INFIT MNSQ	OUTFIT MNSQ	Infit t	Outfit t
Mean	1,00	1,02	-0,07	0,07
SD	0,10	0,14	1,09	0,96

To find out the criteria for item suitability, the results of the MNSQ outfit data analysis, Outfit ZSTD and PT Measure Corr [32] were used. Since it fell within the range of 0.50 to 1.50, the data from the MNSQ outfit, which ranged from 0.80 to 1.31, was accepted. The range of ZSTD outfit data was -0.9 to 1.9. The Point Measure Correlation demonstrated the range of Rasch's discrimination power, which is between -0.05 and 0.36. The data did not fit the Rasch model according to the Point Measure Correlation data. The MNSQ outfit data and the ZSTD outfit matched the requirements. Therefore, the questions can be used even though the point measure correlation data does not fit. It can be said that every two-tier question on a test of critical thinking skills corresponds to the Rasch model.

Table 6. Goodness of fit test

Item	Analysis			Information			Conclution
	Outfit MNSQ	Outfit ZSTD	PT Measure Corr	Outfit MNSQ	Outfit ZSTD	PT Measure Corr	
S9	1,31	1,9	0,17	Fit	Fit	Not Fit	Fits the model
S11	1,25	2,3	0,34	Fit	Fit	Not Fit	Fits the model
S6	1,16	0,9	0,32	Fit	Fit	Not Fit	Fits the model
S7	1,14	0,8	0,1	Fit	Fit	Not Fit	Fits the model
S15	1,11	0,7	0,07	Fit	Fit	Not Fit	Fits the model
S23	1	0,1	0,28	Fit	Fit	Not Fit	Fits the model
S8	1	0,1	0,3	Fit	Fit	Not Fit	Fits the model
S17	1,06	0,5	0,17	Fit	Fit	Not Fit	Fits the model
S14	1,02	0,2	0,12	Fit	Fit	Not Fit	Fits the model
S2	1,03	0,3	-0,05	Fit	Fit	Not Fit	Fits the model
S18	0,98	0	0,16	Fit	Fit	Not Fit	Fits the model
S4	0,95	-0,2	0,07	Fit	Fit	Not Fit	Fits the model
S13	0,94	-0,4	0,2	Fit	Fit	Not Fit	Fits the model
S16	0,92	-0,6	0,15	Fit	Fit	Not Fit	Fits the model
S5	0,89	-0,9	0,21	Fit	Fit	Not Fit	Fits the model
S1	0,86	-0,8	0,24	Fit	Fit	Not Fit	Fits the model
S21	0,88	-0,8	0,16	Fit	Fit	Not Fit	Fits the model
S19	0,87	-0,6	0,05	Fit	Fit	Not Fit	Fits the model
S20	0,86	-0,9	0,18	Fit	Fit	Not Fit	Fits the model
S12	0,8	1,5	0,36	Fit	Fit	Not Fit	Fits the model
S10	0,81	1,2	0,24	Fit	Fit	Not Fit	Fits the model

Difficulty Index

If the difficulty index of the test item is greater than -2.0 or less than 2.0, it falls into good category [31]. The difficulty index of the test items ranged from -0.36 to 0.24 based on the information in Table 7. Due to the fact that all items' difficulty indexes fell between -1 and 1 (moderate), they all match the required criteria. In PCM, the only item attributes that influences student performance is the item difficulty parameter.

Tabel 7. Difficulty

Item	Difficulty	Information
S19	0,24	Good
S7	0,23	Good
S4	0,17	Good
S6	0,12	Good
S8	0,12	Good
S15	0,11	Good
S18	0,1	Good
S1	0,09	Good
S10	0,09	Good
S23	0,07	Good
S21	0,01	Good
S9	-0,01	Good
S13	-0,03	Good
S20	-0,04	Good
S12	-0,08	Good
S14	-0,11	Good
S17	-0,15	Good
S16	-0,15	Good
S2	-0,2	Good
S5	-0,22	Good
S11	-0,36	Good

Information Function and Standard error measurement (SEM)
The reliability of the test has a total peak of information at a value of 27.45 on the ability or theta scale of -0.16 with an error of interpretation of 0.19. These outcomes suggested that the test's degree of accuracy was average. According to Hambleton, the estimation is more accurate the higher the information function [12]. In other words, a test that is useful for focusing on a certain level of aptitude can be described as having a concealed nature by the information function [33].

Figure 1 depicts the results of the critical thinking ability test, which consists of 23 questions and was administered to 146 high school students in the Melawi Regency. It reveals that the test items will be valid for students with ability levels between -2.74 and 3.31, or those who fall between the categories of very low and very high abilities.

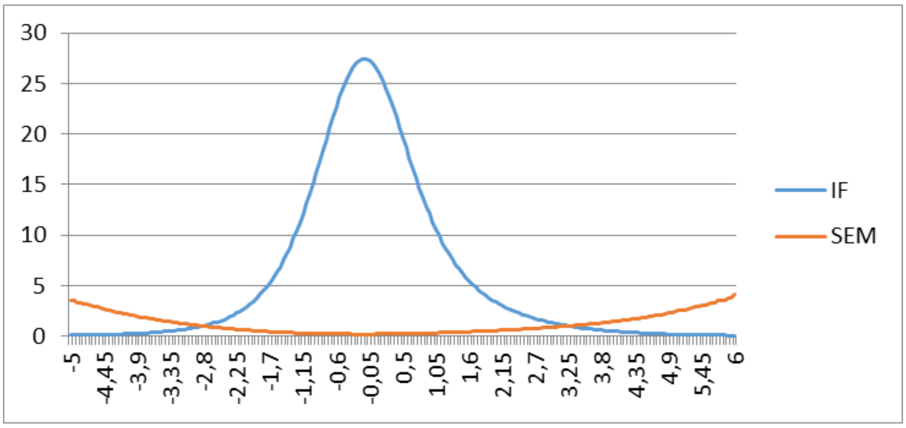

Fig 1. *IF & SEM*

Test characteristic curve (TCC)

According to Figure 2, students with an ability level of -5 (extremely poor) received a score of 21 out of a possible 92, or 22.82% of the possible score. Students with an ability of 6 (extremely high) will receive a score of 84 out of a possible 92, or 91.30% of the possible score. The estimations of proficiency under the item pattern scores, however, tend to differ from those under the grade-correction scores for students with the same correct-item scores but different item responses [34].

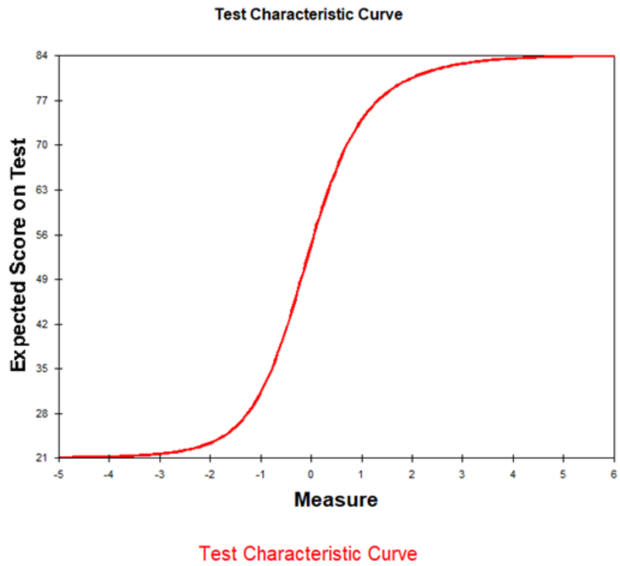

Fig 2. *TCC*

The level of student ability that matches the test item (ability)

According to Table 7, which categorizes students' critical thinking abilities into five levels, 19.18% of students had very high critical thinking abilities, 19.86% had high critical thinking abilities, 37.67% had moderate critical thinking abilities, 17.12% had low critical thinking abilities, and 6.16% had very low critical thinking abilities. This demonstrates that Melawi Regency kids' critical thinking skills were still rated as moderate. The fact that the teacher did not encourage students to develop their critical thinking abilities and that they exclusively measured learning outcomes with multiple choice questions is an indication of the reason of this incident [35].

Table 7. The level of students' abilities that match the test items

Category	Frequency	Percentage
Very high	28	19,18
High	29	19,86
Moderate	55	37,67
Low	25	17,12
Very low	9	6,16
Total	**146**	**100,00**

According to research on a variety of physics topics, including momentum and impulse [36], matter and its transformations [37], and vectors [38] Indonesian high school students have poor critical thinking skills. In middle school, the same incident occurred. Using 15 critical thinking tasks with 13 elements, Nuryanti, Zubaidah, and Diantoro discovered that junior high school students' critical thinking skills are still lacking [39]. Therefore, through assessing the learning process and the assessment it generates, the teacher plays a crucial role in developing critical thinking skills.

DOI: 10.26737/jipf.v8i2.3737

CONCLUSION AND SUGGESTION

According to this study, physics critical thinking test created was valid, dependable, and of a good level of difficulty. This critical thinking exam instrument can be used to assess students' abilities in all ability levels, from poor to high.

ACKNOWLEDGMENTS

We would like to thank DP2M (Directorate of Research and Community Service) for funding this research.

REFERENCES

[1] OECD. (2015). Programme for international student assessment (PISA). Re*sults from PISA 2015*.
[2] Kalelioğlu, F., & Gülbahar, Y. (2014). The effect of instructional techniques on critical thinking and critical thinking dispositions in online discussion. *Journal of Educational Technology & Society*, *17*(1): 248-258.
[3] Chukwuyenum, A. N. (2013). Impact of critical thinking on performance in mathematics among senior secondary school students in Lagos State. *IOSR Journal of Research & Method in education*, *3*(5): 18-25.
[4] Hatcher, D. L. (2011). Which test? Whose scores? Comparing standardized critical thinking tests. *New Directions for Institutional Research*, *2011*(149): 29-39.
[5] McMurray, M. A., Beisenherz, P., & Thompson, B. (1991). Reliability and concurrent validity of a measure of critical thinking skills in biology. *Journal of Research in Science Teaching*, *28*(2): 183-191.
[6] Simpson, E., & Courtney, M. (2002). Critical thinking in nursing education: literature review. *International Journal of Nursing Practice*, *8*(2): 89-98.
[7] Barnett, J. E., & Francis, A. L. (2012). Using higher order thinking questions to foster critical thinking: A classroom study. *Educational Psychology*, *32*(2): 201-211.
[8] Kamali, Z., & Fahim, M. (2011). The Relationship between Critical Thinking Ability of Iranian EFL Learners and Their Resilience Level Facing Unfamiliar Vocabulary Items in Reading. *Journal of Language Teaching and Research*, *2*(1): 104-111.
[9] Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan rasch pada assessment pendidikan*. Trim komunikata.
[10] Sumintono, B. (2018, February). Rasch model measurements as tools in assesment for learning. In *1st International Conference on Education Innovation (ICEI 2017)* (pp. 38-42). Atlantis Press.
[11] Brookfield, S. D. (1997). Assessing critical thinking. *New Directions for Adult and Continuing Education*, *1997*(75): 17-29.
[12] Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory* (Vol. 2). Sage.
[13] Retnawati, H. (2016). *Validitas Reliabilitas & Karakteristik Butir*. Parama Publishing.
[14] Parmaningsih, T. J., & Saputro, D. R. S. (2021, February). Rasch analysis on item response theory: Review of model suitability. In *AIP Conference Proceedings* (Vol. 2326, No. 1, p. 020017). AIP Publishing LLC.
[15] van der Linden, W. J., & Hambleton, R. K. (Eds.). (2013). *Handbook of modern item response theory*. Springer Science & Business Media.
[16] Widhiarso, W. (2010). Model politomi dalam teori respons butir. *Available at SSRN 2593459*.
[17] Engelhard, G. (2005). Item Response Theory (IRT) models for rating scale data. *Encyclopedia of statistics in behavioral science*.
[18] Yilmaz, H. B. (2019). A Comparison of IRT Model Combinations for Assessing Fit in a Mixed

Format Elementary School Science Test. *International Electronic Journal of Elementary Education*, *11*(5): 539-545.

[19] Amelia, R. N. (2021). Identifikasi Item Fit Dan Person Fit Dalam Pengukuran Hasil Belajar Kimia. *Jurnal Ilmiah WUNY, 3*(1).

[20] Mellenbergh, G. J., & Vijn, P. (1981). The Rasch model as a loglinear model. *Applied Psychological Measurement*, *5*(3): 369-376.

[21] Aoyama, K. (2007). Investigating a hierarchy of students' interpretations of graphs. *International Electronic Journal of Mathematics Education*, *2*(3): 298-318.

[22] Boone, W. J., & Noltemeyer, A. (2017). Rasch analysis: A primer for school psychology researchers and practitioners. *Cogent Education*, *4*(1): 1416898.

[23] Isgiyanto, A. (2013). Perbandingan Penyekoran Model Rasch dan Model Partial Credit pada Matematika. *Jurnal Kependidikan*, *43*(1).

[24] Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*(2): 149-174.

[25] Asysyifa, D. S., Wilujeng, İ., & Kuswanto, H. (2019). Analysis of students critical thinking skills using partial credit models (PCM) in physics learning. *International Journal of Educational Research Review*, *4*(2): 245-253.

[26] Ennis, R. (2011). Critical Thinking: Reflection and Perspective Part I. *Inquiry: Critical Thinking Across the Disciplines*, *26*(1): 4-18.

[27] Winsteps. (2022). WINSTEPS Rasch Software - Winsteps Facets. https://www.winsteps.com/winsteps.htm (accessed Oct. 12, 2022).

[28] Tarigan, E. F., Nilmarito, S., Islamiyah, K., Darmana, A., & Suyanti, R. D. (2022). Analisis Instrumen Tes Menggunakan Rasch Model dan Software SPSS 22.0. *Jurnal Inovasi Pendidikan Kimia*, *16*(2): 92-96.

[29] Hardianti, H. Karakteristik Tes Kemampuan Berpikir Kritis Siswa SMA pada Materi Momentum dan Impuls: Perbandingan Classical Theory Test (CTT) dan Model Rasch. *WaPFi (Wahana Pendidikan Fisika)*, *6*(2): 167-173.

[30] DeMars, C. (2010). *Item response theory*. Oxford University Press.

[31] Adams, R., & Khoo, S. T. (1996). ACER Quest: the interactive test analysis system. Version 2.1.

[32] Boone, W. J., Staver, J. R., & Yale, M. S. (2013). *Rasch analysis in the human sciences*. Springer Science & Business Media.

[33] Meyer, J. P., & Zhu, S. (2013). Fair and equitable measurement of student learning in MOOCs: An introduction to item response theory, scale linking, and score equating. *Research & Practice in Assessment*, *8*: 26-39.

[34] Kim, S., Moses, T., & Yoo, H. H. (2015). Effectiveness of item response theory (IRT) proficiency estimation methods under adaptive multistage testing. *ETS Research Report Series, 2015*(1): 1-19.

[35] Asysyifa, D. S., Wilujeng, İ., & Kuswanto, H. (2019). Analysis of students critical thinking skills using partial credit models (PCM) in physics learning. *International Journal of Educational Research Review*, *4*(2): 245-253.

[36] Permata, A. R., Muslim, M., & Suyana, I. (2019, December). Analisis kemampuan berpikir kritis siswa SMA pada materi momentum dan impuls. In *Prosiding Seminar Nasional Fisika (E-Journal)* (Vol. 8, pp. SNF2019-PE).

[37] Ridho, S., Ruwiyatun, R., Subali, B., & Marwoto, P. (2020). Analisis kemampuan berpikir kritis siswa pokok bahasan klasifikasi materi dan perubahannya. *Jurnal Penelitian Pendidikan IPA, 6*(1): 10-15.

[38] Arini, W., & Juliadi, F. (2018). Analisis kemampuan berpikir kritis pada mata pelajaran fisika untuk pokok bahasan Vektor siswa kelas X SMA Negeri 4 Lubuklinggau, Sumatera Selatan. *Berkala Fisika Indonesia*, *10*(1): 1-11.

[39] Nuryanti, L., Zubaidah, S., & Diantoro, M. (2018). Analisis kemampuan berpikir kritis siswa SMP. *Jurnal Pendidikan: Teori, Penelitian, dan Pengembangan*, *3*(2), 155-158.