Momentum: Physics Education Journal
7 (2), 2023, 247-258
Available at: ejournal.unikama.ac.id/index.php/momentum
ISSN 2548-9127(print) | 2548-9135 (online)

Rasch modeling: Developing a critical thinking ability test for students on climate change material

Irwan Muhammad Ridwan[a,b], Ida Kaniawati[a]*, Andi Suhandi[a], Taufik Ramlan Ramalis[a], Ifa Rifatul Mahmudah[b], Berry Kurnia Vilmala[c]

[a]Universitas Pendidikan Indonesia, Jl. Dr. Setiabudi No.229, Bandung, Jawa Barat, 40154, Indonesia
[b]Universitas Siliwangi, Jl. Siliwangi No.24, Tasikmalaya, Jawa Barat, 46115, Indonesia
[c]Universitas Muhammadiyah Riau, Jl. KH. Ahmad Dahlan No.88, Pekanbaru, Riau, 28156, Indonesia
e-mail: kaniawati@upi.edu
* Corresponding Author.

Received: 3 March 2023; Revised: 3 April 2023; Accepted: 5 April 2023

Abstract: The goal of this study was to examine critical thinking skills on the topic of climate change. This study is quantitative in nature. The information was gathered by creating a 30-question test to assess the ability to think critically about climate change. Respondents were 52 physics education students from one of Tasikmalaya's campuses. Winstep software version 4.00 was utilized for data analysis. Rasch modeling was employed in this investigation. The results of the Rasch modeling study show a value of 0.38 for person reliability and 0.91 for item reliability, indicating that the instrument items manufactured are dependable. The raw variance value of 23.4% suggests that the standards for dichotomous data construct validity are met. The separation value of 3.15 supports this. It suggests that this item has a good answer distribution and can assess critical thinking skills on the topic of climate change. As a result, it can be used for study.
Keywords: Rasch modelling; critical thinking skills; climate change

How to Cite: Ridwan, I. M., Kaniawati, I., Suhandi, A., Ramalis, T. R., Mahmudah, I. R., & Vilmala, B. K. (2023). Rasch modeling: Developing a critical thinking ability test for students on climate change material. Momentum: Physics Education Journal, 7(2), 247-258. https://doi.org/10.21067/mpej.v7i2.8310

Introduction

Planet Earth is experiencing increasingly apprehensive conditions. The earth begins to experience a process of decreasing quality and indications of running out of natural resources. Industrial growth that produces hazardous waste, ignorance of littering, river water pollution, uncontrolled growth of motorized vehicles, excessive use of energy, excessive fertilization, forest burning, and conversion of natural forests This human activity causes damage to natural resources and builds up the volume of greenhouse gases in the atmosphere (A. Leicht, 2018; Masson-Delmotte et al., 2021).

The industrial revolution influenced worldwide environmental change significantly. These modifications are mostly the result of extensive and damaging human activities. This problem is caused by an unsustainable lifestyle (UNESCO, 2017; von Weizsäcker, E.-U., & Wijkman, 2018). Planet Earth is losing its ability to produce the resources required to meet human needs. This problem cannot be separated from the low level of human awareness of the interrelationship between humans and natural systems, which, if left unchecked, causes current and future generations to experience difficulties in meeting their needs.

To solve these issues, we must be critical of the conditions that exist and alert to the environmental destruction that is occurring around us. Furthermore, it is critical to inculcate a thorough

DOI: 10.21067/ mpej.v7i2.8310

awareness that nature is a friend of life who requires each other and must be properly cared for (Ridwan et al., 2021). Furthermore, problem-solving ability is critical and must be learned by students. The study of earthly phenomena and celestial bodies is known as IPBA. IPBA is a required subject in most physics education majors in the aim that students will grasp the status of the universe by physically studying natural events using physics and mathematics (NSTA, 2020).

One of the learning outcomes of this course is the ability of students to develop crucial solutions to social, economic, and environmental challenges related to the current trend of climate change (UNESCO, 2020). A person with strong critical thinking abilities can participate as a science consumer (National Research Council, 2012). Someone who lacks the ability to think critically, on the other hand, will struggle to compete in a global world (Frijters et al., 2008). As a result, critical thinking skills are essential for students.

Critical thinking abilities have been acknowledged as vital for academic and career success in the twenty-first century (Facione, 2015). With the increasing demands of the workplace and global trends in higher education, educators and employers are placing a premium on critical thinking skills development (Shaw et al., 2020). When faced with an issue, someone with critical thinking skills will be visible in the person's character. When the person speaks, acts, or is fighting about a hard subject, the character will be seen. Several studies have been conducted to investigate the ability to think critically. To properly measure critical thinking skills, the assessment will undoubtedly require a good measurement test. Multiple choice exams, skill tests, and description tests can be used to assess critical thinking skills (R. H. Ennis, 2011). The construction of a multiple-choice exam was chosen in this study to assess physics students' critical thinking ability.

This type of critical thinking ability test research and development has been done before. One of them is to create a broad and material-free critical thinking skills exam (R. Ennis, 2011; Facione, 2015). Many reports have discussed the development of critical thinking tests, such as research that developed a multiple-choice critical thinking ability test to measure HOTs for prospective elementary school teachers (Maryani et al., 2021), analysis of students' critical thinking ability tests using response theory items (Item Response Theory/IRT) on physics material using the Partial Credit Models (PCM) approach (Asysyifa et al., 2019). There are few publications in the literature on the research and creation of test instruments that truly measure critical thinking skills on the topic of climate change. As a result, examinations of critical thinking abilities on climate change are required for physics students.

This study is significant because students' inadequate ability to think, particularly critical thinking, is inextricably linked to the practices of teachers or lecturers who conduct evaluations by measuring only elements of knowledge (Palavan, 2020; Saprudin, 2019; Suhendi et al., 2018). Students will not have higher-order thinking abilities if they are not given the opportunity to develop them and are not guided to tackle the world's current climate concerns. To address the issue of poor critical thinking abilities, it is vital to create learning that stresses the process of connection between learning materials and real-world circumstances, encouraging students to link and apply learning results in everyday life. Furthermore, educators must create questions that require students to be skeptical of daily phenomena (Andresen et al., 2020; Hudha et al., 2017).

The goal of this project was to create a critical thinking ability exam for physics education students in the IPBA course on climate change. Another goal is to assess the validity and reliability of the tests generated using Rasch modeling. The exam in this research and development is a multiple-choice test of critical thinking skills for physics students on the topic of climate change. The qualities of climate change content, which are more related to stories of real life than calculations, are used to pick multiple-choice examinations. Furthermore, several problems are encountered when examining climate change material (Eggert et al., 2017). The researcher picked the issue of climate change because it is a challenge that humans confront now and in the future, and essential and long-term efforts are required to solve it since climate change has a real influence on human survival.

Method

This research is a sort of R&D that employs the ADDIE model research steps (Alnajdi, 2018). The model has five stages: (a) analyze, (b) design, (c) develop, (d) implement, and (e) evaluate. Only the third stage, development, was investigated in this study. This research will last seven months (from June to December 2022). The information was gathered by creating a 30-question test to assess people's ability to think critically about climate change. The subjects of the study were 52 physics education students from one of Tasikmalaya's campuses. Winstep software version 4.00 was utilized for data analysis. Rasch modeling was employed in this investigation. This modeling was chosen because it used a probabilistic model with the assumption that the chance of the subject successfully answering the question is determined by the subject's competence as well as the qualities of the item. This suggests that test takers with high abilities outperform test takers with poor abilities (Chusni & Suherman, 2021; Eleje et al., 2018; Sharkness & DeAngelo, 2011).

The first step is to do an analysis. A review of existing critical thinking skills assessments is performed at this step. Furthermore, it has been developed. The second stage is preparation. At this point, the indicators for the construction of tests of critical thinking skills are chosen. The five indicators chosen are: (a) interpretation to understand the meaning of a thing; (b) analysis to understand more deeply about a thing through data, information, and others; (c) inference to draw conclusions from collecting data and information; (d) evaluating to assess the credibility of the resulting conclusions; (e) explaining the truth, reasons, and evidence; and (f) self-regulation (Facione, 2015). These six factors are then used as a guideline for constructing the goods. Development is the third stage. The standards for critical thinking ability are used to design the critical thinking ability exam.

Thirty multiple-choice questions were created for this investigation. The thirty questions cover five areas of critical thinking ability. Table 1 shows the five dimensions of critical thinking skills and the characteristics of the items that reflect them.

Table 1. Critical Thinking Indicators and Items representing them

No	Critical Thinking Aspect	Items
1	interpretation	1,2,3,4
2	analysis	5,6,7,8,9,28
3	evaluation	10,11,12,13,14,15
4	inference	16,17,18,19,20,22
5	explanation	21,23,24,25
6	Self-regulation	26,27,29,30

After that, the test is validated. Validation was performed twice: once for content validation and once for empirical validation. Two lecturers from the Department of Physics at the Faculty of Mathematics and Natural Sciences UPI Bandung and one lecturer from the Department of Physics at Siliwangi University, both subject matter experts and experts in physics education, validated the content. This validation takes into account four factors: (a) the items' appropriateness with the indicators; (b) the level of difficulty of the items (the concept of the questions); (c) the usage of language in the items; and (d) the correctness of the concept of the answer key. In addition, content validation was performed to obtain expert input on test items. The average description method was used to analyze data from content validation findings. Furthermore, the test items were updated based on expert recommendations.

Following the revision stage, the empirical validation stage was carried out. Empirical validation was performed on 52 students from one of Tasikmalaya's state universities. This empirical validation is used to verify the validity and reliability of research instruments, as well as to test the suitability of individuals and items at the same time. The instrument will be examined utilizing rapid modeling with Winstep software in this study. An investigation was performed in order to create instruments with high validity and reliability for measuring critical thinking skills. This instrument can then be employed in accordance with the requirements of future research. Rasch modeling has several advantages because it meets the five measurement model principles, which are as follows: first, it can provide a

liner scale with the same intervals; second, it can predict missing data; third, it can provide a more precise estimate; fourth, it can detect model inaccuracies; and fifth, it produces replicable measurements (Sharkness & DeAngelo, 2011; Sumintono & Widhiarso, 2015).

Results and Discussion

Climate change has been a fascinating topic throughout the last decade. This is inextricably linked to the fact that the Earth's temperature continues to rise at an alarming rate, which has a negative impact on life on Earth. This happens naturally, but since the 1800s, human activity has been the primary source of climate change, owing primarily to the use of fossil fuels such as coal, oil, and gas (UNESCO, 2017). Figure 1 depicts the growing temperature of the planet Earth.

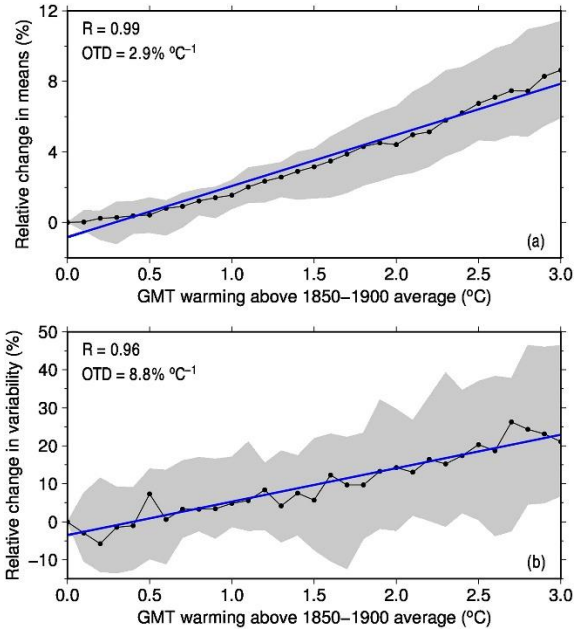

Figure 1. The trend of increasing Earth's temperature (Zhang et al., 2018)

This research and development begins with defining research objectives, which include producing assessments of critical thinking abilities relevant to the issue of climate change as well as knowing the validity, reliability, and applicability of people and products at the same time. Following that, the benchmarks used in the development of test items were chosen. In this research and development, a criterion was chosen, namely the ability to think critically, which consists of the previously reported six critical thinking skills (Facione, 2015).

Students were allowed 100 minutes to work individually on the test items. Following the completion of the test, the students' answers were corrected and analyzed using the Winstep application. There are 30 elements that fit the Rasch model based on data analysis of the Winsteps application. Table 2 shows the whole set of results.

Table 2. Validity dan Reliability of the Instrument

Variable	Logit average	Separation	Reliability	Validity
Person	0,24	0,74	0,36	23.4%
Item	0,00	3,15	0,91	

Table 2 demonstrates that the general quality of student response patterns, the quality of the instruments utilized, and the interactions between people and goods are all important. Furthermore, the person's measure value is 0.24 logit, indicating that students' abilities are likely to be stronger than the average difficulty level of questions with 0.00 logit.

The consistency of student answers is measured using person reliability. The person reliability value of 0.36 indicates that student answers are inconsistent. Item dependability is used to assess the quality of items based on the outcomes of student responses. The item dependability value is 0.91, indicating that the quality of the questions created in the instrument's reliability component is good. Cronbach's alpha (KR 20) is a measure of dependability that takes into account the interaction between the individual and the things as a whole. The Cronbach alpha reliability coefficient of 0.38 derived from sum.stat is in the poor category.

The validity of the instrument is another piece of data that can be seen. The raw variance of the dichotomous questions' instrument validity value must be greater than 20%, and the raw variance of the polytomous data must be greater than 40%. According to Table 2, the raw variance value is 23.4%, indicating that the instrument created fulfills the required construct validity and is certified valid.

The separation value identifies the grouping of people and stuff. Because it identifies groups of respondents and groups of items, the greater the separation value, the higher the quality of the instrument in terms of all respondents and item items. The value of separation can be used to derive the grouping of items in Rasch modeling. Because it can identify the objects that have been created, the higher the value, the higher the quality of the instrument. Another equation, stratum division, is employed more precisely:

$$H=[(4xSEPARATION)+1]/3 \quad (1)$$

With a separation item value of 3.15, H = [(4x3.15)+1]/3 = 4.5 is rounded up to 4, which means that there are four groups of item questions, which can be interpreted as very difficult, difficult, easy, and very easy. In the picture that appears, on the left side there is an L code, which indicates a male, and a P code, which indicates a female, while on the right side there is an S code, which indicates a question.

To see the distribution of student abilities and the distribution of question difficulty maps, see Figure 2.

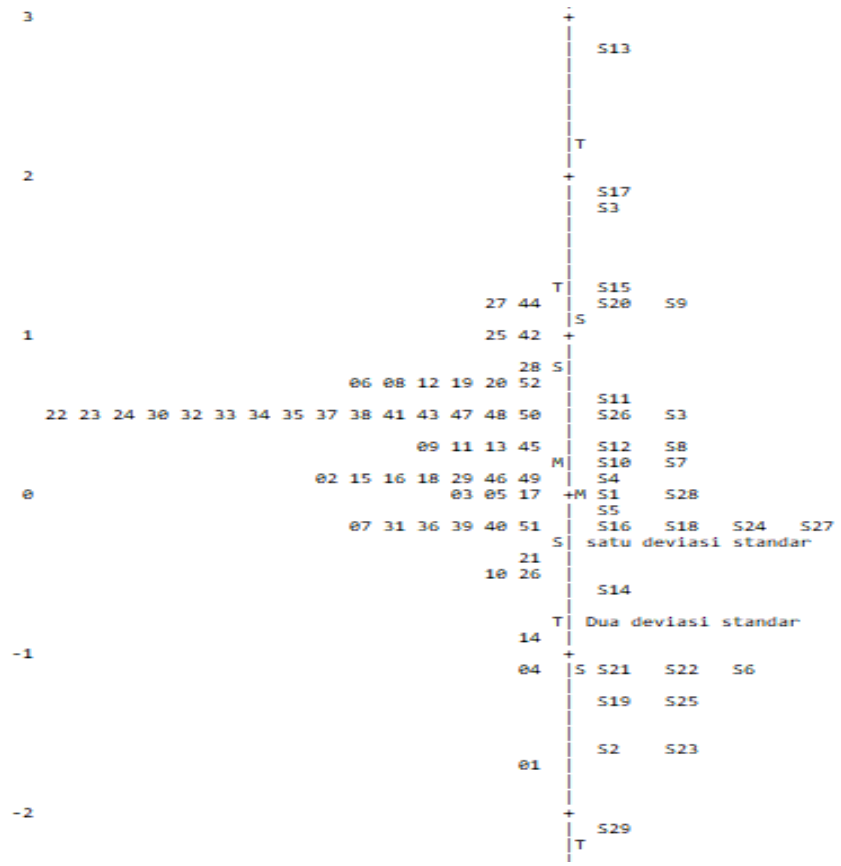

Figure 2. Distribution of student abilities (left side) and item difficulty (right side)

Figure 2 depicts two kids with exceptional ability, 27P and 44L. These two students' logit value is 11.24 (< +2). Students with extremely low skills, 01L, 04P, and 14L, had sequential logit values of -1.71, -1.07, and -0.88 (nearly close to -2). The problem-difficulty map is shown on the right. It can be seen that question S13 is the most difficult (2.83 logit), with only four students answering properly, and question S17 is answered correctly by nine students (1.89 logit), implying that the possibility of all students answering correctly is extremely low. Furthermore, the question with the lowest logit score (-2.12) is S29. In this example, practically all students can solve this problem correctly, as proven by 47 of 52 sample students successfully answering this question. Figure 3 depicts the number of questions successfully answered by students with the question code S17, indicating that the questions are very simple to work on.

See the various human activities below!

Based on the pictures on the side, the attitude you will take to reduce the impact of global warming on Earth is....

A. burning waste, raising poultry, and operating manufacturing machines
B. walking, riding a motorcycle, and planting trees
C. walking, tree planting, and raising poultry
D. raising poultry, walking, and operating manufacturing machines

Figure 3. The easiest question (code S17)

The S13 question was one that students answered incorrectly the fewest times. Although questions with code S13 and response choice D can also be correct, most students chose option D with the proper answer key, namely C. As a result, this question with code S13 requires change in terms of the multiple-choice answer choices in the form of a dichotomy or true-false, so that the answer choices offered clearly demonstrate the proper answer choice. Figure 4 depicts the questions that at least half of the students properly answered, indicating that the questions with the question code S13 is relatively difficult.

The increasing number of motor vehicles results in an increase in urban air temperature which can interfere with the breathing process of living things. The exhaust gas from these vehicles is a greenhouse gas and poison to the body.

Based on these problems, the Government has taken various ways to overcome them.

1) Make an anti-pollution exhaust using plasma technology for motor vehicles.
2) Making CO2 detection devices in various public places.
3) Promote greenery in public open area.
4) Provide counseling about the importance of environmental health

According to the information text above, the right technique for the government to deal with motor vehicle exhaust gas is...

A. 1) and 2)
B. 1), 2), and 3)
C. 1), 2), and 4)
D. 1), 2), 3), and 4)

Figure 4. The most difficult question (S13)

Students are within the mean and two standard deviations (T), indicating that they are all in the typical range. Despite the fact that three students, 01L, 04P, and 14L, were beyond the two standard deviation (T) boundaries, they were in the outlier category with very poor abilities. If any items fall outside of the standard deviation limitations, they must be adjusted because the questions may be too difficult or too easy. The average logit item value is always set at 0.0 logit, indicating the scale's beginning reference point; the average logit person is 0.24 logit, which is more than the initial reference

point. This demonstrates that the average person's achievement exceeds the difficulty level of the questions.

There are many students who cannot do S13, S17, and S3 questions because the level of difficulty of the items whose scores are 2.83 logit, 1.89 logit, and 1.76 logit, are higher than the person's ability with an average logit of 0. 24 logs. The logit score of the student's ability which is lower than the logit item item means that the probability of being able to answer questions with a larger logit correctly is less than 50%. Students 27P (1.24 logit) and 44 (1.24 logit) will have no difficulty in doing the S11 (0.58 logit), S26 (0.5 logit), and S3 (0.5 logit) questions well because their level The difficulty is lower than the person's ability.

When a person's ability with the same logit value, namely questions 27P (1.24 logit) and 44L (1.24 logit), is compared to questions with nearly the same logit value, namely questions S20 (1.21 logit) and S9 (1.21 logit), the probability that these two students can correctly answer the S20 and S9 questions is 50%. Students 27P (1.24 logit) and 44L (1.24 logit), students 25P (1.04 logit) and 42L (1.04 logit) can work on questions S11 (0.58 logit), S26 (0.50 logit), and S8 (0.25 logit) well because the level of difficulty is lower than their ability, but they will struggle with questions S17 (1.89 logit) and S30 (1.76 logit) because the level of difficulty is higher than student ability.

Item Fit Level

Item fit is used to assess the items' suitability (validity). The item's compatibility level can clearly observe the quality of the item's suitability with the model. The acceptability of these objects can be determined by whether they perform normally when measuring. Figure 5 depicts the applicability of the things from highest to lowest order.

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
2	44	52	-1.56	.39	1.07	.35	1.21	.73	A .07	.22	82.7	84.7	S2
3	23	52	.50	.29	1.10	1.49	1.19	2.00	B .02	.24	50.0	59.4	S3
4	28	52	.09	.29	1.10	1.41	1.15	1.77	C .05	.25	59.6	60.6	S4
12	26	52	.25	.29	1.10	1.51	1.12	1.47	D .06	.24	53.8	59.6	S12
18	31	52	-.16	.29	1.09	1.09	1.12	1.21	E .07	.25	59.6	63.4	S18
26	23	52	.50	.29	1.08	1.21	1.11	1.17	F .08	.24	65.4	59.4	S26
30	10	52	1.76	.36	1.02	.15	1.10	.45	G .11	.17	80.8	80.7	S3
5	30	52	-.08	.29	1.03	.42	1.09	1.03	H .16	.25	69.2	62.4	S5
10	27	52	.17	.29	1.07	1.03	1.05	.69	I .13	.25	48.1	59.9	S10
11	22	52	.58	.29	1.05	.72	1.07	.70	J .14	.23	59.6	60.5	S11
20	15	52	1.21	.31	1.03	.27	1.07	.44	K .13	.21	71.2	71.2	S20
22	41	52	-1.15	.35	1.04	.28	1.07	.38	L .15	.24	80.8	79.3	S22
16	32	52	-.25	.29	1.02	.25	1.05	.54	M .20	.25	69.2	64.7	S16
8	26	52	.25	.29	.99	-.20	1.04	.49	N .25	.24	61.5	59.6	S8
9	15	52	1.21	.31	1.01	.15	1.00	.05	O .19	.21	71.2	71.2	S9
15	14	52	1.30	.32	1.01	.12	.98	-.02	o .19	.20	73.1	73.0	S15
21	41	52	-1.15	.35	.97	-.06	1.01	.12	n .26	.24	80.8	79.3	S21
17	9	52	1.89	.37	.96	-.08	1.00	.11	m .21	.17	82.7	82.7	S17
13	4	52	2.83	.52	.98	.10	.93	.04	l .16	.12	92.3	92.3	S13
1	29	52	.01	.29	.95	-.63	.94	-.70	k .34	.25	61.5	61.4	S1
14	36	52	-.61	.31	.93	-.50	.95	-.30	j .36	.25	75.0	70.9	S14
23	44	52	-1.56	.39	.95	-.11	.94	-.08	i .29	.22	86.5	84.7	S23
28	29	52	.01	.29	.95	-.71	.93	-.79	h .35	.25	61.5	61.4	S28
19	42	52	-1.27	.36	.94	-.21	.88	-.38	g .36	.23	82.7	81.1	S19
6	41	52	-1.15	.35	.91	-.37	.85	-.58	f .40	.24	80.8	79.3	S6
24	31	52	-.16	.29	.91	-1.07	.91	-.92	e .41	.25	75.0	63.4	S24
27	32	52	-.25	.29	.90	-1.08	.88	-1.18	d .44	.25	69.2	64.7	S27
29	47	52	-2.12	.48	.90	-.15	.72	-.57	c .40	.19	90.4	90.4	S29
7	27	52	.17	.29	.87	-2.04	.85	-1.94	b .48	.25	75.0	59.9	S7
25	42	52	-1.27	.36	.82	-.82	.71	-1.16	a .58	.23	82.7	81.1	S25
MEAN	28.7	52.0	.00	.33	.99	.1	1.00	.2			71.7	70.7	
P.SD	11.2	.0	1.12	.06	.07	.8	.12	.9			11.5	10.5	

Figure 5. Item Conformance Level

Based on the fit order output items, in general, it was obtained that all items met the MNSQ score criteria, 0.5 < MNSQ < 1.5. Likewise for Z-STD scores, all items fulfill, -2.0 < Z-STD < +2.0. However, there are several items that do not meet the Pt Measure Correlation criteria. Even so, because all of the items on the MNSQ and Z-STD scores met the criteria even though the items on the Pt Measure corr did not fulfill them, all of these items could be considered fit, meaning that the items were maintained.

Person Fit Level

Rasch modeling can also find individuals whose response patterns differ/are incompatible. The pattern of different/inappropriate responses indicates that there is a disparity between the answers supplied based on their ability and the ideal model. This can be used to determine the consistency of students' thinking and to determine whether fraud has occurred. Figure 6depicts the order of the level of disagreement with the model; one student's response pattern is declared unfit, namely 09P. The mean-square outfit value (1.52 logit) of 09P students is beyond the permitted limit of 1.5 logit, whereas all students meet the standards in the z-std outfit aspect.

Pattern of Student Response

This scalogram will reveal the direct causes of student response patterns and why the response patterns do not match the model. The scalogram also has the ability to identify fraud (students cheating on each other). The first indicator is to see if the same person logit value is detected, followed by checking questions that are correctly and mistakenly answered. The scalogram will plainly show this.

Figure 7 depicts students 41P, 43P, 47P, and 48P who allegedly cheated on each other by claiming a similar pattern of responses to answers. This is why the four students' response patterns aren't perfect. Furthermore, 35P students were careless and stated that there was a lucky guess where the third and fourth easiest items in S23 and S29 could not be answered correctly, even though the more difficult questions (e.g., questions S9 and S13) could be answered correctly.

Another example that can be explained with a scalogram is that 10P student answered item S9 correctly. The logit value of 10P student who is included in the slow response student category is -0.53 logit. The logit value of question number 9 is 1.21 logit in the difficult category. The student should not be able to answer question 9 correctly because the student's logit score is far below the question's logit value. It is possible that 10P students will guess (lucky guess).

Conclusion

According to the findings of research and data analysis using Rasch modeling, thirty multiple choice items on the theme of climate change that have been developed fulfill good test characteristics, are evenly distributed on the person item map, and can reach students' abilities. All items meet the MNSQ score criteria, 0.5 < MNSQ < 1.5. Likewise for Z-STD scores, all items fulfill, -2.0 < Z-STD < +2.0. Despite the fact that numerous items do not match the Pt Measure Correlation requirements. Because all of the items on the MNSQ and Z-STD scores satisfied the criterion while the ones on the Pt Measure corr did not, all of these items could be judged fit, which means that they were kept. Furthermore, the raw variance value of 23.4% indicates that the construct validity standards have been met, and the Item Reability value is 0.91, indicating that the quality of the items created in the reliability aspect instrument is good.

Acknowledgment

Thank you to the entire supervisor team for their direction and assistance, which aided in the research, as well as constructive and helpful talks that allowed the research to be completed and this article to be produced.

References

A. Leicht, J. H. & W. J. B. (eds). (2018). *Issues and trends in Education for Sustainable Development Issues and trends in Education*. UNESCO Publishing.

Alnajdi, S. M. (2018). The Effectiveness of Designing and Using a Practical Interactive Lesson Based on ADDIE Model to Enhance Students' Learning Performances in University of Tabuk. *Journal of Education and Learning*, *7*(6), 212–221.

Andresen, L., Boud, D., & Cohen, R. (2020). Experience-based learning. In *Understanding Adult Education and Training* (pp. 225–239). Allen&Unwin. https://doi.org/10.4324/9781003118299-22

Asysyifa, D. S., WİLUJENG, 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.

Chusni, M. M., & Suherman, S. (2021). Developing authentic assessment instrument based on multiple representations to measure students ' critical thinking skills. *Momentum: Physics Education Journal*, *5*(23), 194–208. https://doi.org/https://doi.org/10.21067/mpej.v5i2.5790

Eggert, S., Nitsch, A., Boone, W. J., Nückles, M., & Bögeholz, S. (2017). Supporting students' learning and socioscientific reasoning about climate change-the effect of computer-based concept mapping scaffolds. *Research in Science Education*, *47*, 137–159.

Eleje, L. I., Onah, F. E., & Abanobi, C. C. (2018). Comparative Study of Classical Test Theory and Item Response Theory Using Diagnostic Quantitative Economics Skill Test Item Analysis Results. *European Journal of Educational & Social Sciences*, *3*(1).

Ennis, R. (2011). *Inquiry: Critical Thinking Across the Disciplines* (Vol. 26, Issue 1). Philosophy Documentation Center.

Ennis, R. H. (2011). Critical Thinking: Reflection and Perspective Part I. *Inquiry: Critical Thinking Across the Disciplines*, *26*, 4–18.

Facione, P. (2015). Critical Thinking: What It Is and Why It Counts. *Insight Assessment*.

Frijters, S., ten Dam, G., & Rijlaarsdam, G. (2008). Effects of dialogic learning on value-loaded critical thinking. In *Learning and Instruction* (Vol. 18, pp. 66–82). Elsevier Science. https://doi.org/10.1016/j.learninstruc.2006.11.001

Hudha, M. N., Batlolona, J. R., & others. (2017). How are the physics critical thinking skills of the students taught by using inquiry-discovery through empirical and theorethical overview? *Eurasia Journal of Mathematics, Science and Technology Education*, *14*(2), 691–697.

Maryani, I., Prasetyo, Z. K., Wilujeng, I., Purwanti, S., & Fitrianawati, M. (2021). HOTs Multiple Choice and Essay Questions: A Validated Instrument to Measure Higher-Order Thinking Skills of Prospective Teachers. *Journal of Turkish Science Education*, *18*(4), 674–690.

Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., & Gomis, M. I. (2021). Climate change 2021: the physical science basis. *Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, *2*.

NSTA. (2020). *Standards for Science Teacher Preparation*. 3–6.

Palavan, Ö. (2020). The Effect of Critical Thinking Education on the Critical Thinking Skills and the Critical Thinking Dispositions of Preservice Teachers. *Educational Research and Reviews*, *15*(10), 606–627.

Rachmadtullah, R. (2020). Critical Thinking Instrument Test (CTIT): Developing and analyzing Sundanese students' critical thinking skills on physics concepts using Rasch analysis. *International Journal of Psychosocial Rehabilitation*, *24*(08).

Ridwan, I. M., Kaniawati, I., Suhandi, A., Samsudin, A., & Rizal, R. (2021). Level of sustainability awareness: where are the students' positions? *Journal of Physics: Conference Series*, *1806*(1), 12135. https://doi.org/10.1088/1742-6596/1806/1/012135

Saprudin. (2019). *ω omega The Potential of Gamification in Developing Pre-Service Physics Teachers ' Critical and Creative Thinking Skills*. *2318*. https://doi.org/10.31758/OmegaJPhysPhysEduc.v5i1.7

Sharkness, J., & DeAngelo, L. (2011). Measuring Student Involvement: A Comparison of Classical Test Theory and Item Response Theory in the Construction of Scales from Student Surveys. *Research in Higher Education*, *52*(5), 480–507. https://doi.org/10.1007/s11162-010-9202-3

Shaw, A., Liu, O. L., Gu, L., Kardonova, E., Chirikov, I., Li, G., Hu, S., Yu, N., Ma, L., Guo, F., Su, Q., Shi, J., Shi, H., & Loyalka, P. (2020). Thinking critically about critical thinking: validating the Russian HElghten® critical thinking assessment. *Studies in Higher Education*, *45*(9), 1933–1948. https://doi.org/10.1080/03075079.2019.1672640

Suhendi, H. Y., Ramdhani, M. A., & Irwansyah, F. S. (2018). Verification concept of assesment for physics education student learning outcome. *Int. J. Eng. Technol.*, *7*(3.21), 321–325.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch: pada Assessment Pendidikan* (I). Trim Komunikata.

UNESCO. (2017). *UNESCO moving forward the 2030 Agenda for Sustainable Development*.

UNESCO. (2020). Education for Sustainable Development A Roadmap. In *Impact* (Vol. 2009, Issue 18). UNESCO Publishing. https://doi.org/10.1111/j.2048-416x.2009.tb00140.x

von Weizsäcker, E.-U., & Wijkman, A. (2018). *Come On! Capitalism, short-termism, population and the destruction of the planet: A report to the club of Rome*. Springer Science+Business Media LLC.

Zhang, X., Tang, Q., Liu, X., Leng, G., & Di, C. (2018). Nonlinearity of runoff response to global mean temperature change over major global river basins. *Geophysical Research Letters*, *45*(12), 6109–6116.

Person STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Person
9	17	30	.31	.41	1.22	1.35	1.52	1.92	A .22	.45	73.3	69.1	09P
26	12	30	-.53	.41	1.12	.73	1.44	1.46	B .29	.43	70.0	70.9	26L
39	14	30	-.19	.41	1.35	2.07	1.43	1.61	C .16	.44	56.7	68.4	39L
27	22	30	1.24	.46	1.14	.66	1.41	1.08	D .28	.43	73.3	77.5	27P
35	18	30	.48	.42	1.11	.70	1.40	1.45	E .31	.45	76.7	70.4	35P
21	13	30	-.36	.41	1.29	1.72	1.36	1.34	F .19	.44	66.7	69.6	21L
1	6	30	-1.71	.49	1.34	1.24	1.35	.75	G .08	.35	76.7	80.6	01L
31	14	30	-.19	.41	.90	-.62	1.34	1.32	H .47	.44	76.7	68.4	31P
10	12	30	-.53	.41	1.16	.97	1.31	1.10	I .28	.43	70.0	70.9	10P
13	17	30	.31	.41	1.29	1.68	1.30	1.21	J .22	.45	53.3	69.1	13L
33	18	30	.48	.42	1.08	.52	1.22	.88	K .36	.45	76.7	70.4	33P
34	18	30	.48	.42	1.08	.52	1.22	.88	L .36	.45	76.7	70.4	34P
51	14	30	-.19	.41	1.17	1.07	1.20	.85	M .30	.44	56.7	68.4	51L
32	18	30	.48	.42	1.18	1.06	1.10	.46	N .32	.45	56.7	70.4	32P
37	18	30	.48	.42	1.16	.95	1.10	.47	O .33	.45	56.7	70.4	37L
44	22	30	1.24	.46	1.16	.71	.98	.09	P .34	.43	73.3	77.5	44L
42	21	30	1.04	.44	1.12	.62	1.01	.15	Q .36	.44	70.0	75.9	42L
4	9	30	-1.07	.44	1.11	.59	1.04	.24	R .32	.40	70.0	74.3	04P
20	19	30	.66	.42	1.06	.38	1.10	.42	S .39	.44	70.0	72.3	20L
2	16	30	.14	.41	1.08	.57	1.09	.45	T .38	.45	63.3	68.1	02P
11	17	30	.31	.41	1.08	.52	1.04	.24	U .39	.45	66.7	69.1	11L
14	10	30	-.88	.43	1.06	.38	1.06	.29	V .36	.41	73.3	73.3	14L
18	16	30	.14	.41	1.02	.16	1.06	.32	W .42	.45	70.0	68.1	18L
15	16	30	.14	.41	1.05	.35	.98	-.02	X .42	.45	63.3	68.1	15P
16	16	30	.14	.41	1.03	.23	.98	-.02	Y .43	.45	70.0	68.1	16L
28	20	30	.84	.43	.90	-.47	1.01	.15	Z .49	.44	80.0	74.1	28P
7	14	30	-.19	.41	.99	.00	.94	-.16	z .45	.44	70.0	68.4	07P
3	15	30	-.02	.41	.98	-.06	.90	-.37	y .47	.44	63.3	67.9	03P
5	15	30	-.02	.41	.97	-.16	.89	-.39	x .48	.44	70.0	67.9	05P
52	19	30	.66	.42	.97	-.13	.95	-.08	w .47	.44	76.7	72.3	52P
22	18	30	.48	.42	.96	-.18	.89	-.37	v .49	.45	70.0	70.4	22P
23	18	30	.48	.42	.96	-.18	.89	-.37	u .49	.45	70.0	70.4	23P
24	18	30	.48	.42	.96	-.18	.89	-.37	t .49	.45	70.0	70.4	24P
12	19	30	.66	.42	.94	-.28	.89	-.32	s .50	.44	76.7	72.3	12P
19	19	30	.66	.42	.94	-.28	.89	-.32	r .50	.44	76.7	72.3	19P
25	21	30	1.04	.44	.93	-.27	.75	-.67	q .52	.44	76.7	75.9	25P
36	14	30	-.19	.41	.92	-.48	.86	-.53	p .51	.44	76.7	68.4	36P
8	19	30	.66	.42	.89	-.54	.84	-.51	o .53	.44	76.7	72.3	08P
38	18	30	.48	.42	.89	-.61	.83	-.63	n .54	.45	76.7	70.4	38P
6	19	30	.66	.42	.88	-.63	.83	-.56	m .54	.44	76.7	72.3	06P
30	18	30	.48	.42	.88	-.65	.82	-.66	l .55	.45	76.7	70.4	30P
50	18	30	.48	.42	.88	-.67	.77	-.86	k .56	.45	70.0	70.4	50P
40	14	30	-.19	.41	.85	-.94	.86	-.49	j .55	.44	76.7	68.4	40L
45	17	30	.31	.41	.85	-.93	.78	-.89	i .57	.45	73.3	69.1	45P
29	16	30	.14	.41	.83	-1.11	.76	-1.03	h .59	.45	76.7	68.1	29P
46	16	30	.14	.41	.82	-1.21	.75	-1.09	g .60	.45	83.3	68.1	46L
41	18	30	.48	.42	.76	-1.49	.65	-1.48	f .65	.45	76.7	70.4	41P
43	18	30	.48	.42	.76	-1.49	.65	-1.48	e .65	.45	76.7	70.4	43P
47	18	30	.48	.42	.76	-1.49	.65	-1.48	d .65	.45	76.7	70.4	47P
48	18	30	.48	.42	.76	-1.49	.65	-1.48	c .65	.45	76.7	70.4	48P
17	15	30	-.02	.41	.74	-1.81	.64	-1.64	b .66	.44	76.7	67.9	17P
49	16	30	.14	.41	.74	-1.79	.64	-1.68	a .67	.45	76.7	68.1	49P
MEAN	16.6	30.0	.24	.42	1.00	.0	1.00	.0			71.7	70.7	
P.SD	3.1	.0	.54	.02	.16	.9	.24	.9			6.7	2.8	

Figure 6. Individual Suitability Level

```
TMAN SCALOGRAM OF RESPONSES:
son |Item
    |2 212 2211212  2  1 1 21 21311
    |923956124678451847082361905073
    |------------------------------
 27 +10111111011010111111111000110  27P
 44 +11111111101010110111011110110  44L
 25 +11111111111101111100010000110  25P
 42 +11111111111100101101001101001  42L
 28 +11101111111111110111000000001  28P
  6 +11111110111111010010101101000  06P
  8 +11111110111110011010101101000  08P
 12 +11111101100111011011101100000  12P
 19 +11111101100111011011101100000  19P
 20 +10111111101001111011010010100  20L
 52 +11111110011011110111010000010  52P
 22 +11111101100111001101110100000  22P
 23 +11111101100111001101110100000  23P
 24 +11111101100111001101110100000  24P
 30 +11111110111010110101101010000  30P
 32 +11111011100110101001001101100  32P
 33 +10011111011111111011000010100  33P
 34 +10011111011111110111000010100  34P
 35 +11001111111101111001000100001  35P
 37 +11111011011111010000010101101  37L
 38 +11111110111100011011101010000  38P
 41 +11111111111011110100010010000  41P
 43 +11111111111011110100010010000  43P
 47 +11111111111011110100010010000  47P
 48 +11111111111011110100010010000  48P
 50 +11111111111100001010101100100  50P
  9 +01010111101011111111000010100  09P
 11 +11111110011011000110100101010  11L
 13 +11110001101101001110011001100  13L
 45 +11111011111011110100010010000  45P
  2 +11101110111100001111000000100  02P
 15 +11111001100111001001101100000  15P
 16 +11110101010111101011010010000  16L
 18 +11011111111010010100010000101  18L
 29 +11101111110100111010100000000  29P
 46 +11110111110111101000001000000  46L
 49 +11111111110100011010100000000  49P
  3 +11111110010010100110011010000  03P
  5 +11111011001011100010011001000  05P
 17 +11111111001101011001100000000  17P
  7 +11101110101010010010101001000  07P
 31 +11111111010000110001001000001  31P
 36 +10101111111000110011001000000  36P
 39 +10001011101100000110111100010  39L
 40 +01111111110100011000100100000  40L
 51 +01110011010100111110000001000  51L
 21 +00101111100001001010010110000  21L
 10 +11110010100011000000011010100  10P
 26 +11100111001100000000101000110  26L
 14 +10110101000100001010100010000  14L
  4 +11010001000100100000101100000  04P
  1 +01000000010001001001010000000  01L
    |------------------------------
    |2 212 2211212  2  1 1 21 21311
    |923956124678451847082361905073
```

Figure 7. Students Answer Pattern