PAEDAGOGIA
Jurnal Penelitian Pendidikan
http://jurnal.uns.ac.id/paedagogia

DOI : 10.20961/Paedagogia.V27i3.91259
Vol. 27, No. 3, 2024, pp. 390-400
p-ISSN 0126-4109
e-ISSN 2549-6670

Analysis of Critical Thinking Skills Test Items on The Topic of Temperature, Heat, and Expansion using The Rasch Model

Fibriyana Safitri[1], Dadi Rusdiana[1*], Achmad Samsudin[1], Arif Widiyatmoko[2]
[1] Universitas Pendidikan Indonesia, Indonesia.
[2] Universitas Negeri Semarang, Indonesia.

Keywords:
Critical Thinking Skills Test Items; Rasch Analysis; Temperature, Heat, and Expansion

Article history
Received: 1 August 2024
Revised: 26 October 2024
Accepted: 27 October 2024
Published: 31 October 2024

*Corresponding Author Email:
dadirusdiana@upi.edu

doi: 10.20961/paedagogia.v27i3.91259

Abstract: This research aims to analyze the validity, reliability, level of difficulty and differentiating power of critical thinking skills test items, particularly on the topic of temperature, heat, and expansion. The test items consist of eight essay questions which refer to the eight elements of critical thinking according to Inch. The subjects of this research were 31 students in class VIII of junior high school. Before being tested on students, the critical thinking skills test items were evaluated by five experts. The method used to analyze validity, reliability, level of difficulty, and differentiating power of items is the Rasch model using MINISTEP software version 4.3.1. The results of the analysis show that the test items can be used to test students' critical thinking skills on the topic of temperature, heat, and expansion. For future education practice, the test items can be combined with education technology, like flipbooks, to know the development of students' critical thinking skills.

How to cite: Safitri, F, Rusdiana, D., Samsudin, A. & Widiyatmoko, A. (2024). Analysis of Critical Thinking Skills Test Items on The Topic of Temperature, Heat, and Expansion using The Rasch Model. *PAEDAGOGIA*, *27*(3), 390-400. doi: 10.20961/paedagogia.v27i3.91259

INTRODUCTION

Critical Thinking Skills

Critical thinking is often considered the ability to think clearly and rationally (Mulnix, 2012; Higgins, 2014; Butler, Pentoney, & Bong, 2017). Critical thinking as the use of thinking strategies that increase the likelihood of a desired outcome. Critical thinking skills include synthesizing, analyzing, and evaluating information to make cognitive decisions and transforming them into the affective domain (Simpson & Courtney, 2002; Kavenuke, Kinyota, & Kayombo, 2020; Putra, Sulaeman, & Wahyuni, 2021). Ennis states that critical thinking is reasonable and reflective thinking that is focused on making decisions about what to do or believe (Norris, 1989; Mustika, Nurkamto, & Suparno, 2020). Reasonable means thinking based on facts to produce the best decision, reflective means searching consciously and firmly for the best possible solution. Thus, critical thinking is goal-directed thinking. The goal of critical thinking is to evaluate the best action or belief.

Norris & Ennis focus the critical thinking framework on the thinking process that involves gathering information and applying criteria to consider a different set of actions or views. Norris & Ennis' framework reveals that complex reasoning requires the integrated use of several thought processes. Norris & Ennis reveal the stages in the critical thinking process, namely: (a) clarifying issues by asking critical questions; (b) collecting information about the issue; (c) beginning to reason from various sides or different points of view; (d) collect information and perform further analysis, if necessary; and (e) making and communicating decisions.

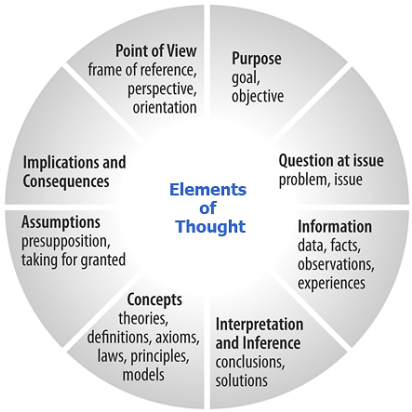

Figure 1. Elements of Critical Thinking Skills

On the other hand, critical thinking skills according to Inch et al. (2006, p. 5-7), consist of eight elements of critical thinking, namely: purpose, question at issue, assumptions, point of view, information, concepts, interpretation and inference, as well as implication and consequences (see Figure 1 above). Each element shows each ability of critical thinking. The explanation of each critical thinking element is presented in Table 1 below.

Table 1. Expected Capabilities of Each Critical Thinking Skills Element

No	Critical Thinking Element	Ability
1	Purpose	Analyze the objectives of the actions taken
2	Question at Issue	Ask questions that identify the problem
3	Assumptions	Formulate assumptions about a phenomenon that occurs
4	Point of View	Interpret and understand other people's points of view
5	Information	Provide information found and researched
6	Concepts	Explains concepts related to the material and/or phenomena discussed from one or several scientific disciplines
7	Interpretation and Inference	Presenting the results of data processing from observed phenomena in the form of interpretation and drawing conclusions
8	Implication and Concequences	Make implications of observed phenomena by linking the concepts that apply therein and predicting the consequences that occur from a phenomenon based on the concepts studied

Critical Thinking Skills Test Items

There is currently quite a bit of research examining critical thinking skills. In the assessment process, an instrument is definitely needed to measure critical thinking skills accurately. According to Ennis in (Mas'ula, 2020), by using a test a person's critical thinking skill can be measured. This test can be in the form of a reasoned multiple-choice test, a skills test, or an essay test. The test is a set of questions that must be answered, must be responded to, or tasks that must be carried out by the students being tested. Tests are used to measure the extent to which students have mastered the learning delivered, especially in terms of critical thinking skills (Suarjana, 2020).

In previous research, Pradana (2017) has developed an essay test to measure physics students' critical thinking abilities on geometric optics material. Apart from that, Syahbana (2012) has also developed a contextual-based learning tool on the material of prisms and pyramids to measure the mathematical critical thinking skills of class VIII junior high school students. Furthermore, Amalia & Susilaningsih (2014) have developed an instrument for assessing the critical thinking skills of high school students on acid-base material. Hartini & Sukardjo (2015) also developed a higher-order thinking multiple choice test (HOT-MCT) for measuring science critical thinking skills for class VII SMP/MTs. Meanwhile,

the analysis of critical thinking skills tests items on the topic of temperature, heat, and expansion with the Rasch model has not yet been done. Based on this description, research was carried out in the form of an analysis of critical thinking skills test items on the topic of temperature, heat, and expansion using the Rasch model.

Temperature, Heat, and Expansion

One of the learning outcomes that must be achieved by students in the "Kurikulum Merdeka" is that students are able to measure the amount of temperature caused by the heat energy provided, as well as being able to differentiate heat insulators and conductors. The temperature, heat, and expansion chapters include coverage of energy content and its changes, focusing on the Integrated Science element, namely process skills. The curriculum objectives include students being able to:

1) Understand the concept and measure the temperature difference of an object.
2) Describe the difference between temperature and heat
3) Name objects that have high specific heat
4) Calculate the heat needed for an object to increase its temperature.
5) Explain the meaning of expansion.
6) Mention examples of expansion that occur in the surrounding environment.

Related to the curriculum objectives, students need critical thinking skills in mastering these concepts because several previous studies emphasized that temperature and heat are considered quite difficult topic by students, so they need to be taught using innovative models or methods (Güneş et al., 2010; Serevina, Astra, & Sari, 2018; Gürses, Şahin, & Güneş, 2022). Thus, the development of critical thinking skills test items in this topic also aims to make students accustomed to solving complex problems using rational and systematic thinking. Critical thinking is a complex ability, so the use of essay tests is more recommended than limited response tests (Hidayati et al., 2024). Actually, research on the development of temperature, heat, and expansion test items has been widely conducted (Nursa'adah & Bunawan, 2021; Sukarelawan et al., 2021). However, research that develops critical thinking skills test items on that topic and then analyzes them using the Rasch model has not been widely conducted. The Rasch model offers a method for psychometric analysis that helps create test questions and offers important insights about evaluating student learning (Hamdu et al., 2020). Therefore, this study focuses on developing critical thinking skills test items on the topic of temperature, heat, and expansion and analyzing them using the Rasch model.

METHOD

To test students' critical thinking skills, the written test items were developed in the form of eight essay questions. The critical thinking skills test item developed is based on the framework of critical thinking skills according to Inch (2006). The developed written test contained questions about temperature, heat, and expansion. After the critical thinking skills test item has been developed, the next stage is expert validation by five evaluators (lecturers and teachers) to determine whether the items are suitable for use or ideal for use after several revisions. Based on the feedback from the evaluator, the researcher made improvements to the test items that had been created. These improvements include content that is in accordance with the product target, grammar, writing punctuation, and the use of images/graphs/tables. Then, after revisions have been made according to the evaluator's suggestions, the items can be tried out on students who have learned about temperature, heat, and expansion.

Table 2. MNSQ and ZSTD Criteria

Outfit	Value received
MNSQ	$0,5 < \text{MNSQ} < 1,5$
ZSTD	$-2,0 < \text{ZSTD} < +2,0$

The academic background of the subjects of this research is 31 students of grade VIII junior high school. They are studying the material of temperature, heat, and expansion according to the curriculum syllabus. The results were then analyzed for validity, reliability, difficulty level, and distinguishing power.

The test results were then calculated using the Rasch model using MINISTEP 4.3.1 software. In this research, the validity of the items is seen from the score outfit mean square (MNSQ) and Z-standard outfit (ZSTD). The values received can be seen in Table 2 (Sumintono & Widhiarso, 2015). The test item can be said to be fit (suitable) if it meets the "acceptable" criteria in the fit order.

Test reliability is related to the problem of the certainty of test results, where a test can be said to have a high level of confidence if the test can provide constant results. Reliability analysis with Rasch modeling was conducted using the ministep 4.3.1 application on the output menu 3.1 summary statistics. The results of the analysis display the person reliability and item reliability values. Rasch analysis can display values, personal reliability and Cronbach alpha. The interpretation of person reliability score can be seen in Table 3 and the interpretation of Cronbach alpha can be seen in Table 4 (Sumintono & Widhiarso, 2015).

Table 3. Interpretation of Person Reliability Score

r	Interpretation
$r \geq 0,94$	Excellent
$0,90 \leq r < 0,94$	Very good
$0,80 \leq r < 0,90$	Good
$0,67 \leq r < 0,80$	Enough
$r < 0,67$	Weak

Based on Table 3 above, when the value of a person's reliability score (r) is more than or equal to 0,94, it means excellent. Then, when the value of r is $0,90 \leq r < 0,94$, it means very good. Interpretation of r is good, when the value of r is $0,80 \leq r < 0,90$. Interpretation of r is enough, when the value of r is $0,67 \leq r < 0,80$. And, when the value of r is less than 0,67, it means weak.

Table 4. Interpretation of Cronbach Alpha Score

α	Interpretation
$\alpha \geq 0,80$	Very good
$0,7 \leq \alpha < 0,8$	Good
$0,6 \leq \alpha < 0,7$	Sufficient
$0,5 \leq \alpha < 0,6$	Bad
$\alpha < 0,5$	Very Bad

Differentiating power shows the ability of items to differentiate students with high ability and students with low ability. Rasch analysis provides information point-measure correlation (PTMEASURE-AL COOR) to identify the differentiating power of items. Test items can be categorized as having very good, good, not good, and poor differentiating power. The interpretation can be seen in Table 5.

Table 5. Interpretation of Ptmeasure-Al Coor

Ptmeasure-Al Coor (ID)	**Interpretation**
$ID > 0,40$	Very good
$0,30 < ID \leq 0,40$	Good
$0,20 < ID \leq 0,30$	Not Good
$ID \leq 0,19$	Poor

Meanwhile, the level of difficulty is seen from the JMLE MEASURE score. The JMLE MEASURE value obtained is then interpreted to determine its meaning, where the interpretation can be seen in Table 6. When the value of Level of Difficulty (TK) is more than 2,01, it means very difficult. When the value of TK is $0,00 < TK \leq 2,01$, it is interpreted as difficult. Interpretation of easy is shown by the value of TK in -2,01

< TK ≤ 0,00. And, very easy is an interpretation of TK with the value lower than or equal to -2,01.

Table 6. Interpretation of The Level of Difficulty

Level of Difficulty (TK)	Interpretation
TK > 2,01	Very Difficult
0,00 < TK ≤ 2,01	Difficult
-2,01 < TK ≤ 0,00	Easy
TK ≤ -2,01	Very Easy

RESULT AND DISCUSSION

The critical thinking skills test item consists of eight essay questions, was tested on 31 students, and was evaluated by five expert evaluators consisting of lecturers and science teachers. The results of trials on students are then analyzed to identify validity, reliability, difficulty level, and differentiating power.

Validity

Validity analysis results using criteria fit item with output in the form of MNSQ and ZSTD values can be seen in Figure 2. Expert judgment is sometimes used as the only indicator of the content validity of a research instrument ((Escobar-Pérez & Cuervo-Martínez, 2008); (Fernández-Gómez et al., 2020)). This is because the MNSQ and ZSTD scores are obtained based on students' answers which are of course influenced by other factors, so the results of expert evaluations also have a crucial role.

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE 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
3	74	31	-.58	.23	2.32	3.76	2.23	3.64	A .66	.57	22.6	51.5	Q3
5	45	31	.74	.21	1.38	1.63	1.35	1.49	B .49	.63	12.9	36.3	Q5
8	23	31	1.86	.25	.86	-.44	1.23	.74	C .62	.65	74.2	57.4	Q9
6	58	31	.17	.21	.84	-.67	.87	-.52	D .68	.60	41.9	41.1	Q6
4	94	31	-1.74	.26	.78	-.84	.76	-.89	d .56	.52	74.2	50.0	Q4
2	48	31	.61	.21	.65	-1.81	.67	-1.62	c .70	.62	61.3	36.0	Q2
7	55	31	.38	.21	.64	-1.85	.63	-1.83	b .48	.60	45.2	39.4	Q7
1	88	31	-1.36	.25	.54	-2.03	.52	-2.14	a .70	.55	58.1	48.5	Q1
MEAN	60.6	31.0	.00	.23	1.00	-.28	1.03	-.14			48.8	45.0	
P.SD	22.1	.0	1.10	.02	.55	1.88	.53	1.85			21.1	7.4	

Figure 2. MNSQ And ZSTD Scores on The Critical Thinking Skills Test Items

Based on the output in Figure 2, the interpretation of the validity of the critical thinking skills test items can be seen in Table 7. The items can be considered fit (suitable) when meeting the "accepted" criteria on the fit order.

A total of six items meet the MNSQ and ZSTD criteria, while item K1 does not meet just one of the criteria, so this item does not need to be changed or replaced (can be used). For item K3, even though the results of the analysis of the item show that the item does not meet the criteria, based on the results of the expert evaluation, the item is suitable for use to identify students' critical thinking skills in the "assumption" critical thinking element with several revisions. So, several changes/improvements were made to the item according to the expert evaluator's suggestions. Item K3 is not used to identify students' critical thinking skills in all elements of critical thinking, only in the "assumption" element.

Table 7. Interpretation of The Validity of Critical Thinking Skills Test Items

No.	MNSQ Score	ZSTD Score	MNSQ Description	ZSTD Description
K1	0,52	-2,14	Accepted	Not Accepted
K2	0,67	-1,62	Accepted	Accepted
K3	2,23	3,64	Not Accepted	Not Accepted
K4	0,76	-0,89	Accepted	Accepted
K5	1,35	1,49	Accepted	Accepted
K6	0,87	-0,52	Accepted	Accepted
K7	0,63	-1,83	Accepted	Accepted
K8	1,23	0,74	Accepted	Accepted

Reliability

To determine the increase in students' critical thinking skills, the critical thinking skills test items are used, which have been validated by experts (lecturers and science teachers). In this research, reliability analysis with the Rasch model uses the MINISTEP software on the output menu 3.1 summary statistics. The analysis results display the values of person reliability and item reliability, as in Figure 3.

SUMMARY OF 31 MEASURED Person

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	15.6	8.0	-.14	.44	.99	-.02	1.03	.02
SEM	.9	.0	.17	.01	.09	.19	.11	.20
P.SD	4.8	.0	.94	.03	.50	1.04	.62	1.10
S.SD	4.8	.0	.96	.03	.51	1.06	.63	1.11
MAX.	26.0	8.0	2.10	.54	2.10	1.96	3.09	2.68
MIN.	10.0	8.0	-1.22	.42	.32	-1.84	.33	-1.70
REAL RMSE	.49	TRUE SD	.81	SEPARATION	1.66	Person RELIABILITY	.73	
MODEL RMSE	.45	TRUE SD	.83	SEPARATION	1.86	Person RELIABILITY	.78	
S.E. OF Person MEAN = .17								

Person RAW SCORE-TO-MEASURE CORRELATION = 1.00 (approximate due to missing data)
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .74 SEM = 2.42 (approximate due to missing data)
STANDARDIZED (50 ITEM) RELIABILITY = .96

SUMMARY OF 8 MEASURED Item

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	60.6	31.0	.00	.23	1.00	-.28	1.03	-.14
SEM	8.3	.0	.42	.01	.21	.71	.20	.70
P.SD	22.1	.0	1.10	.02	.55	1.88	.53	1.85
S.SD	23.6	.0	1.18	.02	.59	2.01	.57	1.98
MAX.	94.0	31.0	1.86	.26	2.32	3.76	2.23	3.64
MIN.	23.0	31.0	-1.74	.21	.54	-2.03	.52	-2.14
REAL RMSE	.25	TRUE SD	1.07	SEPARATION	4.29	Item RELIABILITY	.95	
MODEL RMSE	.23	TRUE SD	1.08	SEPARATION	4.73	Item RELIABILITY	.96	
S.E. OF Item MEAN = .42								

Figure 3. Point-Measure Correlation Value in The Critical Thinking Skill Test Items

Based on the results of the analysis, the value of Cronbach alpha amounts to 0.74, which shows that students' interaction with the overall critical thinking skills test items is in the good category. The value of person reliability shows that the consistency of students' answers is 0.73, which is included in the sufficient category. Meanwhile, the value of item reliability of 0.95 indicates that the quality of the items in the critical thinking skills test items is included in the special category. To determine the quality of items, the analysis uses item reliability because the items are validated, not the subject (as in person

reliability). If a test is reliable, then the test can be measured consistently; the higher the reliability of the test, the better the quality of the test (Sari, 2020; Kholifah, Suhendarti, & Liberna, 2022).

Level of Difficulty

The questions in the critical thinking skills test items can be divided into very easy, easy, difficult, and very difficult categories. The difficulty level of items is seen from the JMLE MEASURE score. Mapping the level of difficulty of items can be seen in Figure 4.

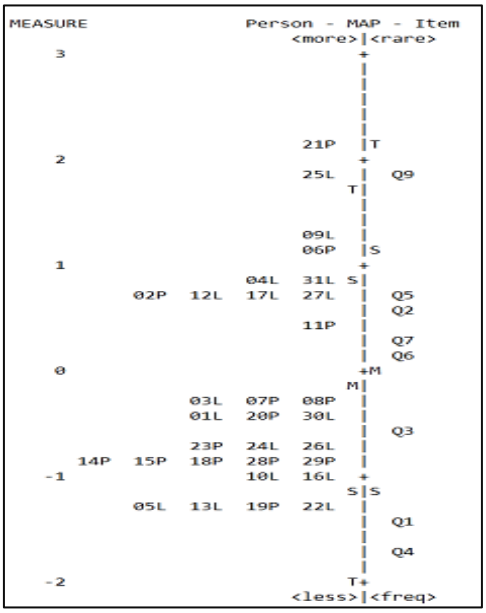

Figure 4. The Level of Difficulty of Critical Thinking Skill Test Items

A positive value indicates that the item is classified as difficult, while a negative value indicates that the item is relatively easy. Items with a larger measure value mean they have a higher difficulty level. Thus, items K1, K3, and K4 are classified as easy or very easy, and the others are classified as difficult or very difficult depending on the JMLE MEASURE value of each item (the JMLE MEASURE value can be seen in Figure 5). From this mapping, it can be seen that students with codes 05L, 13L, 19P, and 22L can only do well on items that are included in the very easy category, while students with codes 21P and 25L can even do items that are classified as very difficult well and get a high score on trials critical thinking skills test items.

Item STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE 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
8	23	31	1.86	.25	.86	-.44	1.23	.74	.62	.65	74.2	57.4	Q9
5	45	31	.74	.21	1.38	1.63	1.35	1.49	.49	.63	12.9	36.3	Q5
2	48	31	.61	.21	.65	-1.81	.67	-1.62	.70	.62	61.3	36.0	Q2
7	55	31	.30	.21	.64	-1.85	.63	-1.83	.48	.60	45.2	39.4	Q7
6	58	31	.17	.21	.84	-.67	.87	-.52	.68	.60	41.9	41.1	Q6
3	74	31	-.58	.23	2.32	3.76	2.23	3.64	.66	.57	22.6	51.5	Q3
1	88	31	-1.36	.25	.54	-2.03	.52	-2.14	.70	.55	58.1	48.5	Q1
4	94	31	-1.74	.26	.78	-.84	.76	-.89	.56	.52	74.2	58.0	Q4
MEAN	60.6	31.0	.00	.23	1.00	-.28	1.03	-.14			48.8	45.0	
P.SD	22.1	.0	1.10	.02	.55	1.88	.53	1.85			21.1	7.4	

Figure 5. JMLE MEASURE Value on The Critical Thinking Skill Test Items

The JMLE MEASURE values obtained are then interpreted to determine their meaning. Interpretation of the difficulty level of the items on the critical thinking skills test instrument can be seen in Table 8. The results of the research show that there are 2 items that are included in the very easy category (K1 and K4), 1 item that is included in the easy category (K3), and 4 items that are included in the difficult category (K2, K5, K6, K7). Meanwhile, item K8 is included in the very difficult category. Not many students can get the maximum score on item K8.

Table 8. Interpretation of The Level of Difficulty

No.	Difficulty Level	Interpretation
K1	-1,36	Very Easy
K2	0,61	Difficult
K3	-0,58	Easy
K4	-1,74	Very Easy
K5	0,74	Difficult
K6	0,17	Difficult
K7	0,30	Difficult
K8	1,86	Very Difficult

In this research, the distribution of the level of difficulty was uneven. There are far more items in the difficult category than items in the easy category. This is because critical thinking involves high-order thinking skills (HOTS), so the items used to test critical thinking skills do not only ask students to simply remember, restate, or refer without processing. Items with difficult categories train students to have HOTS. HOTS is included in critical thinking skills. Thus, the distribution of difficult items with slightly more will support the development of critical thinking skills of students. From the results of the difficulty level test, there are three questions that are categorized as very easy and easy. Then, there are five questions that are included in the difficult category and very difficult. The composition of the difficulty level can be said to be balanced between easy and difficult questions. The questions in the difficult category are slightly increased in accordance with the purpose of the study to assess the critical thinking skills of students. This will support teachers in assessing the critical thinking skills of students as a whole.

Differentiating Power

The differentiating power of items can be seen based on the value of point-measure correlation (PTMEASURE-AL COOR) from each item (as in Figure 6).

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE 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
3	74	31	-.58	.23	2.32	3.76	2.23	3.64	A .66	.57	22.6	51.5	Q3
5	45	31	.74	.21	1.38	1.63	1.35	1.49	B .49	.63	12.9	36.3	Q5
8	23	31	1.86	.25	.86	-.44	1.23	.74	C .62	.65	74.2	57.4	Q9
6	58	31	.17	.21	.84	-.67	.87	-.52	D .68	.60	41.9	41.1	Q6
4	94	31	-1.74	.26	.78	-.84	.76	-.89	d .56	.52	74.2	50.0	Q4
2	48	31	.61	.21	.65	-1.81	.67	-1.62	c .70	.62	61.3	36.0	Q2
7	55	31	.30	.21	.64	-1.85	.63	-1.83	b .48	.60	45.2	39.4	Q7
1	88	31	-1.36	.25	.54	-2.03	.52	-2.14	a .70	.55	58.1	48.5	Q1
MEAN	60.6	31.0	.00	.23	1.00	-.28	1.03	-.14			48.8	45.0	
P.SD	22.1	.0	1.10	.02	.55	1.88	.53	1.85			21.1	7.4	

Figure 6. Value of Point-Measure Correlation in Critical Thinking Skill Test Items

Criteria for the differentiating power of items include very good, good, poor, and not good.

Interpretation of the differentiating power of critical thinking skills test items can be seen in Table 9. Table 9 shows that all items have a differentiating power of more than 0.40, which is included in the "very good" category. This means that the tested critical thinking skill test items can differentiate students with high and low critical thinking skills.

Table 9. Interpretation of The Differentiating PoweroOf Critical Thinking Skill Test Items

No	PTMEASURE-AL COOR	Interpretatior
K1	0,79	Very Good
K2	0,70	Very Good
K3	0,66	Very Good
K4	0,56	Very Good
K5	0,49	Very Good
K6	0,68	Very Good
K7	0,48	Very Good
K8	0,62	Very Good

Critical thinking includes skills in analyzing arguments, making conclusions using inductive or deductive reasoning, judging or evaluating, and making decisions or solving problems. Critical thinking allows students to study problems systematically, face various challenges in an organized manner, formulate innovative questions, and design solutions. People with ideal critical thinking skills are highly curious, have extensive experience, are full of self-confidence, are open-minded, flexible, diligent in searching for relevant information, and reasonable in selecting criteria (Facione, 2000). In the critical thinking element "purpose", students are expected to be able to analyze the purpose of the actions taken. Through question item K1, it is hoped that students can write down the purpose of using insulating material on Teflon handles. In the critical thinking element "question at issue", students are expected to be able to ask questions that identify problems. In question item K2, students are asked to write down research questions that can be asked regarding an experiment to determine which objects can conduct heat well and which are difficult to conduct heat. In the critical thinking element "assumptions", the expected ability is that students are able to formulate assumptions about a phenomenon that occurs. Assumptions are temporary guesses that are accepted as a foundation of thinking because they are considered correct. Assumptions are in the form of guesses, estimates, predictions, and forecasts. In other words, assumptions are the result of individual thoughts whose truth is not yet known and needs proof. The question items K3 are intended for students to formulate assumptions if in an area the duration of sunlight during the year is short.

The next element of critical thinking is "point of view", where the expected ability is that students can interpret and understand other people's points of view. In question item K4, it is hoped that students will be able to interpret other people's points of view about an action, namely immediately pouring very hot water into a glass cup. Students are asked to assess whether the actions taken by other people in the problem are appropriate or not and explain the reasons. In the critical thinking element "information", the expected ability is that students can provide information that is found and researched. Question item K5 is formulated so that students provide information regarding the differences in the Celsius, Reamur, Fahrenheit, and Kelvin thermometer scales based on the images presented. The ability to understand and convey information correctly is very important in everyday life so as not to give rise to misunderstandings and misconceptions. Moreover, nowadays information sources can be accessed easily, anywhere and at any time. Then, in the critical thinking element "concept", it is hoped that students can explain concepts related to the material and/or phenomena discussed from one or several scientific disciplines. The question item K6 is made so that students can explain Azas Black when mixing hot and cold water.

Furthermore, in the critical thinking element "interpretation and inferences", the expected ability is that students can present the results of data processing from observed phenomena in the form of interpretation and drawing conclusions. Question item K7 is intended for students to draw conclusions based on a table of volume expansion coefficients for several types of substances that have the same initial volume and temperature change. Students need to interpret things and the important points of a

situation, source, event, scoring, agreement, belief, provision, and stage, which is proven by the ability to categorize, look for important points of meaning and explain the meaning of something. This ability is very important because concepts in science learning are formulated systematically, mainly based on observations and experiments (Bilgin, Karakuyu, & Ay, 2015). Lastly, in the critical thinking element "implications and consequences", students are expected to be able to make implications of the observed phenomena relate to the concepts that apply in it and predict the consequences that occur from a phenomenon based on the concepts studied. Question item K8 is created so students can state the implications of anomalous water phenomena and predict the consequences if there is no such phenomenon.

CONCLUSION

The results of the analysis show that the critical thinking skills test items on the topic of temperature, heat, and expansion can be used to test students' critical thinking skills. Validity analysis results using criteria fit item with output in the form of MNSQ and ZSTD values. Meanwhile, the value of item reliability of 0.95 indicates that the quality of the items is included in the excellent category. On the other hand, the distribution of the difficulty level is uneven, there are far more items in the difficult category than items in the easy category. The distribution of items with more difficult categories turned out to be effective in assessing students' critical thinking skills. Thus, further research can also use a similar composition of difficulty levels to assess students' critical thinking skills on other discussion topics. However, all items have a differentiating power of more than 0.40, which is included in the very good category, so they can differentiate students with high and low critical thinking skills very well.

ACKNOWLEDGMENTS

The authors are grateful to the Ministry of Research, Technology and Higher Education Indonesia through the Directorate General of Higher Education, Research, and Technology for providing the funds of the Pendidikan Magister menuju Doktor untuk Sarjana Unggul (PMDSU) scholarship program.

REFERENCES

Amalia, N. F., & Susilaningsih, E. (2014). Pengembangan instrumen penilaian keterampilan berpikir kritis siswa SMA pada materi asam basa. *Jurnal Inovasi Pendidikan Kimia*, *8*(2).

Bilgin, I., Karakuyu, Y., & Ay, Y. (2015). The effects of project-based learning on undergraduate students' achievement and self-efficacy beliefs towards science teaching. *Eurasia Journal of Mathematics Science and Technology Education*, 11(3).

Butler, H. A., Pentoney, C., & Bong, M. P. (2017). Predicting real-world outcomes: Critical thinking ability is a better predictor of life decisions than intelligence. *Thinking Skills and Creativity*, *25*, 38-46.

Escobar-Pérez, J., & Cuervo-Martínez, Á. (2008). Validez De Contenido Y Juicio De Expertos: Una Aproximación A Su Utilización. 6, 27–36.

Facione, P.A. (2000). The disposition toward critical thinking: Its character, measurement, and relationship to critical thinking skill. *Informal logic*, 20 (1).

Fernández-Gómez, E., Martín-Salvador, A., Luque-Vara, T., Sánchez-Ojeda, M. A., Navarro-Prado, S., & Enrique-Mirón, C. (2020). Content Validation through Expert Judgement of an Instrument on the Nutritional Knowledge, Beliefs, and Habits of Pregnant Women. Nutrients 2020, Vol. 12, Page 1136, 12(4), 1136. https://doi.org/10.3390/NU12041136

Güneş, T., Şener Dilek, N., Demir, E.S., Hoplan, M., & Çelikoğlu, M. (2010). A qualitative research on the efforts of teachers about the concept teaching, determination and elimination of the misconceptions. *International Conference on New Trends in Education and Their Implications*, 11-13 November, Antalya-Turkey.

Gürses, A., Şahin, E., & Güneş, K. (2022). Investigation of the Effectiveness of the Problem-based Learning (PBL) Model in Teaching the Concepts of" heat, Temperature and Pressure" and the Effects of the Activities on the Development of Scientific Process Skills. *Education Quarterly Reviews*, *5*(2).

Hamdu, G., Fuadi, F. N., Yulianto, A., & Akhirani, Y. S. (2020). Items quality analysis using rasch model to measure elementary school students' critical thinking skill on stem learning. JPI (Jurnal Pendidikan Indonesia), 9(1), 61-74.

Hartini, H., & Sukardjo, S. (2015). Pengembangan higher order thinking multiple choice test untuk mengukur keterampilan berpikir kritis ipa kelas VII SMP/MTs. *Jurnal Inovasi Pendidikan IPA*, *1*(1), 86-101.

Hidayati, N., Suryanti, S., Rahmayumita, R., & Aisya, S. (2024). Development of Critical Thinking Skills Instruments: Cases for Essay Tests. *Jurnal Kependidikan: Jurnal Hasil Penelitian dan Kajian Kepustakaan di Bidang Pendidikan, Pengajaran dan Pembelajaran*, *10*(1), 77-88.

Higgins, S. (2014). Critical thinking for 21st-century education: A cyber-tooth curriculum? *Prospects*, *44*(4), 559-574. DOI: 10.1007/S11125-014-9323-0/TABLES/1.

Inch, E.S., et.al. (2006). Critical Thinking & Communication, The Use of Reason in Argument. United States of America: Pearson Education.

Kavenuke, P. S., Kinyota, M., & Kayombo, J. J. (2020). The critical thinking skills of prospective teachers: Investigating their systematicity, self-confidence and scepticism. *Thinking Skills and Creativity*, *37*, 100677. DOI: 10.1016/J.TSC.2020.100677.

Kholifah, N., Suhendarti, M., & Liberna, H. (2022). Jurnal Pendidikan MIPA. *Jurnal Pendidikan*, *23*(2), 633-641.

Mas'ula, N., & Rokhis, T.A. (2020). Pengembangan Instrumen Tes Kemampuan Berpikir Kritis Mahasiswa Pada Pokok Bahasan Kinematika. *SAP (Susunan Artikel Pendidikan)*, *4*(3).

Mulnix, J. W. (2012). Thinking critically about critical thinking. *Educational Philosophy and theory*, *44*(5), 464-479.

Mustika, N., Nurkamto, J., & Suparno, S. (2020). Influence of Questioning Techniques in EFL Classes on Developing Students' Critical Thinking Skills. *International Online Journal of Education and Teaching*, *7*(1), 278-287.

Nursa'adah, R., & Bunawan, W. (2021, November). Development of Objective Test for High-Level Knowledge of High School Physics Materials on Temperature and Heat. In 6th Annual International Seminar on Transformative Education and Educational Leadership (AISTEEL 2021) (pp. 72-78). Atlantis Press.

Norris, S.P. (1989). Can we test validly for critical thinking?. *Educational researcher*, *18*(9), 21-26.

Pradana, S.D.S., Parno, P., & Handayanto, S.K. (2017). Pengembangan tes kemampuan berpikir kritis pada materi Optik Geometri untuk mahasiswa Fisika. *Jurnal Penelitian dan Evaluasi Pendidikan*, *21*(1), 51-64.

Putra, P.D.A., Sulaeman, N.F., & Wahyuni, S. (2021). Exploring Students' Critical Thinking Skills Using the Engineering Design Process in a Physics Classroom. *The Asia-Pacific Education Researcher*, 1-9.

Sari, L.Y. (2017). Analisis Soal Ulangan Harian Pada Materi Ekologi Kelas X SMA Negeri 2 Lembang Jaya Kabupaten Solok. Saintifik, 1(2).

Serevina, V., Astra, I., & Sari, I.J. (2018). Development of E-Module Based on Problem Based Learning (PBL) on Heat and Temperature to Improve Student's Science Process Skill. *Turkish Online Journal of Educational Technology-TOJET*, *17*(3), 26-36.

Simpson, E., & Courtney, M. (2002). Critical thinking in nursing education: Literature review. *International journal of nursing practice*, *8*, 89-98.

Suarjana, I. (2020). *Pengembangan Instrumen Kemampuan Berpikir Kritis Dan Sikap Peduli Lingkungan Tema 8 Pada Peserta Didik Kelas IV SD* (Doctoral dissertation, Universitas Pendidikan Ganesha).

Sukarelawan, M. I., Sriyanto, S., Puspitasari, A. D., Sulisworo, D., & Hikmah, U. N. (2021). Four-tier heat and temperature diagnostic test (4T-HTDT) to identify student misconceptions. *JIPFRI (Jurnal Inovasi Pendidikan Fisika dan Riset Ilmiah)*, *5*(1), 1-8.

Sumintono, B., & Widhiarso, W. (2015). Aplikasi pemodelan rasch pada assessment pendidikan. Trim komunikata.

Syahbana, A. (2012). Pengembangan perangkat pembelajaran berbasis kontekstual untuk mengukur kemampuan berpikir kritis matematis siswa SMP. *Edumatica: Jurnal Pendidikan Matematika*, *2*(02).