JPPIPA 11(1) (2025)

Jurnal Penelitian Pendidikan IPA

Journal of Research in Science Education

http://jppipa.unram.ac.id/index.php/jppipa/index

Development of Understanding Test Instruments for Grade 7 Junior High School Students on Temperature, Heat and Expansion Topics

Fibriyana Safitri[1*], Dadi Rusdiana[2], Achmad Samsudin[2], Arif Widiyatmoko[3]

[1] Departement of Science Education, Universitas Pendidikan Indonesia, Bandung, Indonesia.
[2] Departement of Physics Education, Universitas Pendidikan Indonesia, Bandung, Indonesia.
[3] Departement of Science Education, Universitas Negeri Semarang, Semarang, Indonesia.

Received: September 26, 2024
Revised: December 02, 2024
Accepted: January 25, 2025
Published: January 31, 2025

Corresponding Author:
Fibriyana Safitri
fibriyanasafitri@gmail.com

DOI: 10.29303/jppipa.v11i1.9271

Abstract: The aim of this research is to analyze the validity, reliability, level of difficulty and differentiating power of understanding test items, especially on the temperature, heat and expansion topic that has been developed. The test items consist of 20 multiple choice questions which refer to cognitive process of understanding according to Revised Bloom's Taxonomy. The subjects of this research were 31 students in class VIII of junior high school. Before being tested on students, the understanding test items were evaluated by 5 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 understandingtest items on temperature, heat and expansion can be used to test students' understanding. Validity analysis results using item fit criteria with output in the form of MNSQ and ZSTD values. Meanwhile, the value of reliability item of 0.88 indicates that the quality of the items in the understanding instrument is included in the good category. On the other hand, the distribution of the difficulty level of items is uneven, there are more items in the difficult category than items in the easy category. However, 10 items have a differentiating power of more than 0.40 which is included in the very good category, so that they can differentiate students with high and low understanding very well.

Keywords: Expansion; Heat; Rasch analysis; Temperature; Understanding test

Introduction

Understanding

Understanding is a form of learning outcome that a level higher than knowledge, regarding ability to capture the meaning or essence of what is learned or encountered. Understanding is a process or method that aims to make someone understand or know about something. In learning science, we need to understand concepts. Students' understanding of concepts can be used to complete something problems related to the concept at hand. Conceptual understanding involves applying previously known explanations to new situations (Montfort et al., 2009; Al-Mutawah et al., 2019; Tan et al., 2020). A deeper understanding of a concept is achieved when students apply it in different situations, describe or define it in their own words, create a model of it, or find an appropriate metaphor for it (Slotte & Lonka, 1999; Idris, 2009; Konicheck-Moran et al., 2015). Conceptual understanding is the actual ability achieved by students after experiencing a learning process over a certain period of time. Conceptual

How to Cite:
Safitri, F., Rusdiana, D., Samsudin, A., & Widiyatmoko, A. (2025). Development of Understanding Test Instruments for Grade 7 Junior High School Students on Temperature, Heat and Expansion Topics. *Jurnal Penelitian Pendidikan IPA*, *11*(1), 395–404. https://doi.org/10.29303/jppipa.v11i1.9271

understandingwill be continuous or interconnected between one concept and another, therefore conceptual understandingis very important in learning process.

In the Revised Bloom's Taxonomy, the level of understanding is one level higher than knowledge. Understanding includes the ability to: interpret, give examples, summarize, conclude, compare, and explain (Anderson et al., 2001; Forehand, 2010; Setiani et al., 2019; Bernardo et al., 2015; Cetin-dindar, 2015). Next, the revised bloom's taxonomy dimension indicators (Anderson & Krathwohl, 2001) can be seen on Table 1.

Table 1. Dimension Indicators of the Revised Bloom's Taxonomy

Cognitive Level	Cognitive Process	Indicators
Remember	recognizing	Gaining knowledge of a given phenomenon
	recalling	Express knowledge already possessed from long-term memory
Understand	interpreting	Interpret information from a demonstrated phenomenon
	exemplifying	Find specific examples or illustrations of concepts
	summarizing	Summarize general themes or important points
	inferring	Draw logical conclusions from the information provided
	comparing	Detecting correspondence between two ideas and objects
	explaining	Building a cause and effect system
apply	executing	Apply concepts using mathematical formulations
		Applying concepts in a graph
		Using mathematical formulations to apply a concept
	implementing	Applying concepts to a particular substance
		Determine the application of concepts based on appropriate phenomena
		Apply the concepts you have
analyze	differentiating	Distinguish relevant information to complete a given statement
	organizing	Organizing the concept of a problem or phenomenon
	attributing	Refers to viewpoints, biases, values, and goals
evaluate	checking	Checking the truth of a statement
	critiquing	Detect inconsistencies and compatibility between products and external criteria and procedures
Create	generating	Generate alternative hypotheses based on criteria
	planning	Design procedures to complete tasks
	producing	Creating products

Understanding Test Instrument

One of the tendencies that causes students to fail in doing science questions well is because students have difficulty understanding concepts and do not use good reasoning in solving the questions or tests given. Due to this fact, schools should play a role in helping solve the problems faced by students because understanding concepts is a very important aspect in the principles of science learning (Mutmainna at al., 2018; Suwarto, 2013; Widdiharto, 2008). Students who have a proper understanding of concepts will be able to give examples, compare, explain, draw conclusions, solve scientific problems and be able to see the relationship between science and other fields of science (Radiusman, 2020; Oglivie, 2009). A person is said to understand if he can explain or re-explain the essence of the material or concept he obtained independently (Alan & Afriansyah, 2017; Cheriani et al., 2015; Bayuningsih et al., 2017; Marlina et al., 2018; Oglivie, 2009; Syarif et al., 2019). One way to find out students' understanding of concepts is to use diagnostic tests. Diagnostic tests are tests that can be used to detect the presence or absence of several signs (understanding the concept, misconceptions, not understanding the concept) (Hidayati et al., 2019; Fitrianingrum et al., 2017; Caleon et al., 2010; Wiyono et al., 2016; Syahrul et al., 2015; Nursalam, 2016; Eryilmaz, 2010; Laksono, 2020; Kirburut, 2014).

Diagnostic tests function to determine students' strengths and weaknesses when learning something, so that the results can be used as a basis for providing follow-up. This test can consist of a number of questions or requests to perform something. The purpose of diagnostic tests is to see student learning progress related to the process of finding student weaknesses in certain material. The approach taken by teachers in diagnosing student learning difficulties varies, depending on the learning difficulties faced by students (Rusilowati, 2015).

There are several diagnostic test instruments that can be used, namely interviews, open questions, concept maps, and description or multiple choice tests (1). Of the four diagnostic tools, multiple choice tests are very suitable for measuring students' level of understanding of concepts. Multiple choice tests are an appropriate choice to measure students' understanding of the material being studied. Multiple choice tests have several advantages, namely they are time efficient, can be used to measure large numbers of test takers, and are

easy to score because they are objective (Murti et al., 2018).

Temperature, Heat, and Expansion

Temperature, heat, and expansion are materials studied by students from elementary school to university level. In everyday life, temperature is a measure of how hot or cold a substance or object is. A hot oven is said to have a high temperature, while frozen ice is said to have a low temperature (Idawati et al., 2016). Heat is a form of energy that can move from a higher temperature object to a lower temperature object (Yanti et al., 2014). Expansion is a physical change in an object due to the object receiving heat or warmth. One of these physical changes is a change in the volume of an object. In general, objects will increase in volume if they receive heat. Changes in the volume of an object are directly proportional to changes in temperature (Endaryono et al., 2023). The material on temperature, heat and expansion is given specifically to junior high students with the application of the concepts in everyday life. In general, the subject of temperature, heat and expansion is divided into four parts, namely temperature, expansion, heat which can affect the shape of objects, Black's principle and heat transfer (Sofianto et al., 2020).

The material on temperature, expansion and heat is one of the difficult materials in science learning, but if this material is prepared in the form of interesting media and learning methods, it can improve the skills that students must have in the 21st century (Supriyadi et al., 2021). The topic of temperature, expansion and heat is also one of the topics where there are many misconceptions. Students' misconceptions also occur a lot in everyday life and students don't realize it, such as students equating heat with heat, heat with energy, cold is not part of heat and there are many physics misconceptions that students don't realize (Sofianto et al., 2020). Based on the results of research (Lestari et al., 2017), on the concept of temperature, the average student who experienced misconceptions was 11.53%, where students thought the temperature of an object was proportional to the mass of the object. In the concept of heat, the average student who experiences misconceptions is 16.44%, where students think that heat is energy that flows from one energy to another. Also, regarding the concept of expansion, the average student who experienced misconceptions was 8.61%, where students thought that the diameter of an object decreases when it expands.

Method

To test students' understanding, the written test items were developed in the form of 20 multiple choice questions. The test item developed is based on Revised Bloom's Taxonomy. The written test developed contains questions related to temperature, heat and expansion. After understanding test items has been developed, the next stage is expert validation by 5 evaluators (lecturers and teachers) to find out whether the items are suitable for use or not, or suitable for use after several revisions. 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. The subjects of this research were 31 junior high school students. The results were then analyzed for validity, reliability, level of difficulty and differentiating 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.

Table 2. MNSQ and ZSTD Criteria (Sumintono et al., 2015)

Outfit	Value received
MNSQ	$0.5 < MNSQ < 1.5$
ZSTD	$-2.0 < ZSTD < +2.0$

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. 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 3 (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

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 item 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. 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 ≤ 0.40	Good
0.20 < ID ≤ 0.30	Not Good
ID ≤ 0.19	Poor

Meanwhile, the level of difficulty is seen from the JMLE MEASURE score, where the interpretation can be seen in Table 6.

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

In the independent curriculum, the expected learning outcomes on the topic of temperature, heat and expansion are that students can measure the amount of temperature caused by the heat energy provided, as well as being able to differentiate between insulators and heat conductors. The learning objectives for the topic of temperature, heat and expansion, which are included in the scope of energy content and its changes, include: (1) Understand the concept and measure the difference in temperature of an object, (2) Describe the difference in temperature and heat, (3) Mention objects that have a high specific heat, (4) Calculate the heat needed for an object to increase its temperature, (5) Explain the meaning of expansion, and (6) Mention examples of expansion that occur in the surrounding environment. The understanding test instrument developed consists of 20 multiple choice questions. In the initial stage of instrument development, a question construction design is created, as in Table 7.

Table 7. Construction of Understanding Test Questions

Question Construction Descriptor	Option	Question Number
Statement	Symbol	5,11
Statement	Statement	6,10,14
Statement	Figure	2,13
Statement	Graph	18
Statement-Table	Graph	15
Statement-Table	Statement	1,3,16
Statement-Table	Symbol	4,8,12
Statement-Figure	Symbol	7
Figure statements	Statement	9,17,19,20

The understanding referred to in this article is a cognitive level that is one level higher than knowledge in accordance with Bloom's Taxonomy. Thus, cognitive abilities at the level of understanding consist of several aspects as in Table 8.

Table 8. Aspects of Understanding

Aspects of Understanding	Question Numbers
Interpreting	6,15,18
Exemplifying	2, 11,13
Classifying	7,8,12
Summarizing	9,10,20
Inferring	3,16,19
Comparing	1,4,5
Explaining	14,17

After ensuring that the distribution and construction of the questions are balanced and represent all aspects of understanding, the next step is to create indicators of item based on each aspect of understanding. This can be seen in Table 9.

Table 9. Indicator of Item

Aspect of Understanding	Indicator	Item Number
interpreting	Translating the meaning of the variables in the equation $Q = m.c.\Delta t$	6
	Interpret the results of temperature observations over time	15
	Representing a phenomenon in the form of a graph of the relationship between temperature and volume and density of water	18
exemplifying	Give examples of objects that are conductors	2
	Illustrate an example of a change in the state of an object from liquid to solid (freezing)	13
	Detect events that change the state of substances that release heat	11
classifying	Categorize materials that are insulators	7
	Classify examples of heat transfer by conduction and convection	8
	Classify the processes of change of form that absorb heat and those that release heat	12
summarizing	Abstracting the main points of the working principle of a thermos	9
	Summarizes information related to heat transfer in everyday life	10

Aspect of Understanding	Indicator	Item Number
	Provides general ideas regarding the use of land and sea breezes	20
infering	Conclude the results of the experiment on the effect of temperature on the time needed to dissolve granulated sugar	3
	Predict the increase in metal length due to expansion if the metal is heated for a certain time	16
	Draw inferences from experimental results to prove the existence of gas expansion	19
comparing	Map temperature measurement results with the same value based on scale conversion	1
	Matching the results of measuring the temperature of an object using two different types of thermometers	4
	Compare the amount of heat absorbed by two objects given the same treatment	5
explaining	Explain the concepts that apply in the event of heating water until the water turns into steam	14
	Constructing the causes of window glass breaking when the temperature rises significantly	17

The understanding test items consist of 20 multiple choice questions, was tested on 31 students and evaluated by five expert evaluators consisting of lecturers and science teachers. The results of trials on students are then analyzed to identify validity, reliability, level of difficulty 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 1.

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
10	30	31	-3.43	1.03	1.11	.42	9.90	3.22	A-.29	.16	96.8	96.7	Q10
5	29	31	-2.67	.75	1.23	.56	6.29	2.46	B-.29	.22	93.5	93.5	Q5
20	20	31	-.28	.43	1.51	2.52	1.86	2.19	C .06	.46	54.8	72.4	Q20
7	28	31	-2.20	.63	1.07	.31	1.50	.77	D .16	.26	90.3	90.2	Q7
3	24	31	-1.06	.47	1.16	.79	1.39	.85	E .24	.39	74.2	78.1	Q3
9	27	31	-1.84	.56	1.18	.59	1.23	.54	F .15	.30	87.1	87.0	Q9
18	10	31	1.50	.44	1.21	.98	1.20	.68	G .35	.49	71.0	75.8	Q18
15	13	31	.94	.42	1.18	.91	1.16	.67	H .39	.51	64.5	74.2	Q15
19	15	31	.59	.42	1.09	.50	1.04	.23	I .46	.51	67.7	73.2	Q19
4	9	31	1.70	.45	.99	.03	.77	-.53	J .52	.48	71.0	76.3	Q4
12	10	31	1.50	.44	.95	-.19	.73	-.78	i .56	.49	71.0	75.8	Q12
16	29	31	-2.67	.75	.82	-.14	.37	-.27	h .37	.22	93.5	93.5	Q16
11	10	31	1.50	.44	.81	-.91	.68	-.95	g .64	.49	83.9	75.8	Q11
13	18	31	.07	.42	.79	-1.20	.68	-1.23	f .65	.49	83.9	72.8	Q13
17	23	31	-.85	.45	.79	-1.05	.61	-.84	e .57	.41	77.4	75.2	Q17
14	10	31	1.50	.44	.73	-1.31	.58	-1.35	d .69	.49	83.9	75.8	Q14
8	8	31	1.91	.47	.71	-1.32	.51	-1.26	c .68	.46	87.1	77.7	Q8
6	8	31	1.91	.47	.68	-1.50	.57	-1.03	b .68	.46	93.5	77.7	Q6
1	8	31	1.91	.47	.62	-1.87	.42	-1.59	a .74	.46	87.1	77.7	Q1
MEAN	18.0	31.0	-.23	.59	.98	-.10	1.66	.09			80.6	80.0	
P.SD	8.6	.0	2.01	.32	.23	1.09	2.32	1.35			11.3	7.7	

Figure 1. MNSQ and ZSTD Scores

Based on the output in Figure 1, the interpretation of the validity ofunderstanding test items can be seen in Table 10. The items can be said to be fit (suitable) when meeting the "accepted" criteria on the fit order.

A total of 13 items met the MNSQ and ZSTD criteria, while 3 items did not meet just one of the criteria, so these questions did not need to be changed or replaced (they could be used). For item number 5, 10, and 20, although the results of the item analysis show that these items do not meet the criteria, based on the results of expert evaluation, these items are suitable for use to identify students' understanding with several revisions, so that several changes/improvements are made to these items. According to the expert evaluator's advice. Expert assessment as an evaluation strategy offers many advantages such as high quality of responses and the possibility of obtaining extensive information regarding the product being tested (Cabero & Llorente, 2013; Fernández-Gómez et al., 2020;). Expert judgment is sometimes used as the only indicator of the content validity of a research instrument (Escobar-Pérez et al., 2008). 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 very important role.

Table 10. Interpretation of the Validity of Understanding Test Questions

No.	MNSQ score	ZSTD score	MNSQ	ZSTD
P1	0.42	-1.59	Not Accepted	Accepted
P3	1.39	0.85	Accepted	Accepted
P4	0.77	-0.53	Accepted	Accepted
P5	6.29	2.46	Not Accepted	Not Accepted
P6	0.57	-1.03	Accepted	Accepted
P7	1.50	0.77	Not Accepted	Accepted
P8	0.51	-1.26	Accepted	Accepted
P9	1.23	0.54	Accepted	Accepted
P10	9.90	3.22	Not Accepted	Not Accepted
P11	0.68	-0.95	Accepted	Accepted
P12	0.73	-0.78	Accepted	Accepted
P13	0.68	-1.23	Accepted	Accepted
P14	0.58	-1.35	Accepted	Accepted
P15	1.16	0.67	Accepted	Accepted
P16	0.37	-0.27	Not Accepted	Accepted
P17	0.61	-0.84	Accepted	Accepted
P18	1.20	0.68	Accepted	Accepted
P19	1.04	0.23	Accepted	Accepted
P20	1.86	2.19	Not Accepted	Not Accepted

Reliability

To determine the increase in students' understanding, an understanding test instrument was used that had been validated by experts (three lecturers

and two teachers). In this research, reliability analysis with Rasch modeling uses the MINISTEP 4.3.1 software on the menu output 3.1 summary statistics. The analysis results display the values of person reliability and item reliability as in Figure 2.

SUMMARY OF 31 MEASURED Person

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	11.6	20.0	.53	.62	.97	-.01	1.44	.06
SEM	.6	.0	.22	.01	.06	.20	.38	.23
P.SD	3.4	.0	1.22	.04	.31	1.08	2.08	1.27
S.SD	3.4	.0	1.24	.04	.32	1.10	2.11	1.29
MAX.	18.0	20.0	3.01	.78	1.57	2.11	9.90	4.11
MIN.	6.0	20.0	-1.63	.58	.38	-2.17	.27	-1.89
REAL RMSE	.65	TRUE SD	1.03	SEPARATION	1.57	Person RELIABILITY	.71	
MODEL RMSE	.62	TRUE SD	1.05	SEPARATION	1.69	Person RELIABILITY	.74	
S.E. OF Person MEAN = .22								

Person RAW SCORE-TO-MEASURE CORRELATION = 1.00 (approximate due to missing data)
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .75 SEM = 1.71 (approximate due to missing data)
STANDARDIZED (50 ITEM) RELIABILITY = .88

SUMMARY OF 20 MEASURED (EXTREME AND NON-EXTREME) Item

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	18.0	31.0	-.23	.59				
SEM	2.0	.0	.46	.07				
P.SD	8.6	.0	2.01	.32				
S.SD	8.8	.0	2.06	.33				
MAX.	31.0	31.0	1.91	1.83				
MIN.	8.0	31.0	-4.68	.42				
REAL RMSE	.69	TRUE SD	1.89	SEPARATION	2.74	Item	RELIABILITY	.88
MODEL RMSE	.67	TRUE SD	1.90	SEPARATION	2.83	Item	RELIABILITY	.89
S.E. OF Item MEAN = .46								

Figure 2. Person Reliability and Item Reliability

Based on the results of the analysis, cronbach alpha score which shows that students' interaction with the overall understanding test instrument is 0.75, which is included in the good category. Person reliability which shows that the consistency of students' answers is 0.71, which is included in the sufficient category. Meanwhile the item reliability of 0.88 indicates that the quality of the items in the understanding instrument is included in the good category (Sumintono & Widhiarso, 2015).

Level of Difficulty

The items in the understanding test instrument can be divided into very easy, easy, difficult and very difficult categories. The difficulty level of the items is seen from the JMLE MEASURE score.

Mapping the level of difficulty of items can be seen in Figure 3. In Figure 3, the right side of the dotted line shows the question items and the left side shows the students' ability to answer the questions. Positive value of"*measure*"indicates that the question item is classified as difficult, while negative value of "*measure*"indicates that the item is relatively easy. So the questions have larger measure value indicates that the item has a higher level of difficulty. Thus, question items number 1,4,6,8, 11,12,13,14,15,18,19 are classified as difficult or very difficult and the other questions are classified as easy or very easy depending on the JMLE MEASURE value of each item (JMLE MEASURE values can be seen in Figure 4). From this mapping, it can be seen that students with code 01L can only work on easy questions, while students with codes 15P and 23 P can work on questions that are classified as difficult (or very difficult) and get a high score on the understanding test instrument.

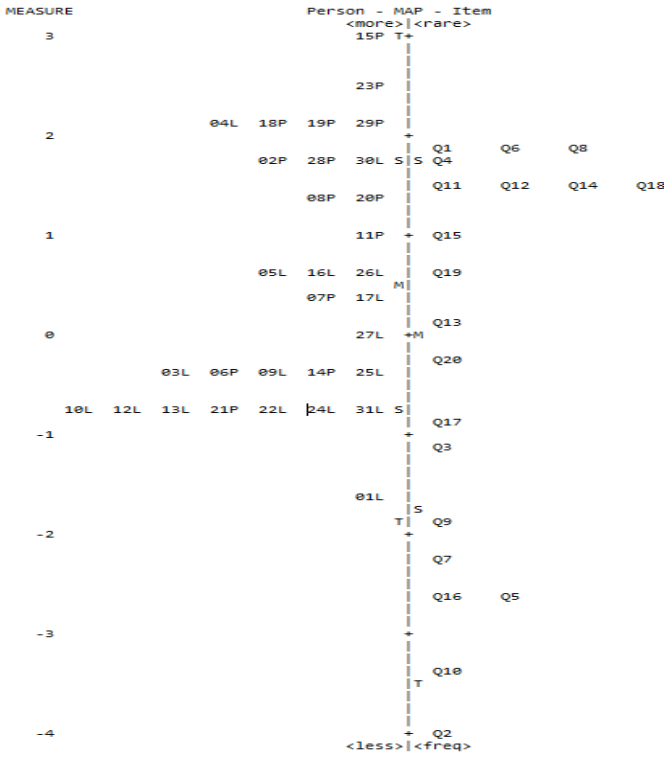

Figure 3. Mapping of the level of difficulty of the understanding test instrument

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
1	8	31	1.91	.47	.62	-1.87	.42	-1.59	.74	.46	87.1	77.7	Q1
6	8	31	1.91	.47	.68	-1.50	.57	-1.03	.68	.46	93.5	77.7	Q6
8	8	31	1.91	.47	.71	-1.32	.51	-1.26	.68	.46	87.1	77.7	Q8
4	9	31	1.70	.45	.99	.03	.77	-.53	.52	.48	71.0	76.3	Q4
11	10	31	1.50	.44	.81	-.91	.68	-.95	.64	.49	83.9	75.8	Q11
12	10	31	1.50	.44	.95	-.19	.73	-.78	.56	.49	71.0	75.8	Q12
14	10	31	1.50	.44	.73	-1.31	.58	-1.35	.69	.49	83.9	75.8	Q14
18	10	31	1.50	.44	1.21	.98	1.20	.68	.35	.49	71.0	75.8	Q18
15	13	31	.94	.42	1.18	.91	1.16	.67	.39	.51	64.5	74.2	Q15
19	15	31	.59	.42	1.09	.50	1.04	.23	.46	.51	67.7	73.2	Q19
13	18	31	.07	.42	.79	-1.20	.68	-1.23	.65	.49	83.9	72.8	Q13
20	20	31	-.28	.43	1.51	2.52	1.86	2.19	.06	.46	54.8	72.4	Q20
17	23	31	-.85	.45	.79	-1.05	.61	-.84	.57	.41	77.4	75.2	Q17
3	24	31	-1.06	.47	1.16	.79	1.39	.85	.24	.39	74.2	78.1	Q3
9	27	31	-1.84	.56	1.18	.59	1.23	.54	.15	.30	87.1	87.0	Q9
7	28	31	-2.20	.63	1.07	.31	1.50	.77	.16	.26	90.3	90.2	Q7
5	29	31	-2.67	.75	1.23	.56	6.29	2.46	-.29	.22	93.5	93.5	Q5
16	29	31	-2.67	.75	.82	-.14	.37	-.27	.37	.22	93.5	93.5	Q16
10	30	31	-3.43	1.03	1.11	.42	9.90	3.22	-.29	.16	96.8	96.7	Q10
2	31	31	-4.68	1.83	MINIMUM MEASURE				.00	.00	100.0	100.0	Q2
MEAN	18.0	31.0	-.23	.59	.98	-.10	1.66	.09			80.6	80.0	
P.SD	8.6	.0	2.01	.32	.23	1.09	2.32	1.35			11.3	7.7	

Figure 4. JMLE MEASURE values on understanding test instruments

The JMLE MEASURE values obtained are then interpreted to determine their meaning. Interpretation of the level of difficulty of the items on the understanding test instrument can be seen in Table 11.

Table 11. Interpretation of the Level of Difficulty of Understanding Test Items

Number	Level of Difficulty	Interpretation
P1	1.91	Difficult
P2	-4.68	Very Easy
P3	-1.06	Easy
P4	1.70	Difficult
P5	-2.67	Very Easy
P6	1.91	Difficult
P7	-2.20	Very Easy
P8	1.91	Difficult
P9	-1.84	Easy
P10	-3.43	Very Easy
P11	1.50	Difficult
P12	1.50	Difficult
P13	0.07	Difficult
P14	1.50	Difficult
P15	0.94	Difficult
P16	-2.67	Very Easy
P17	-0.85	Easy
P18	1.50	Difficult
P19	0.59	Difficult
P20	-0.28	Easy

Based on Table 11, a trial of the understanding test instrument shows that in the test instrument there are 5 items that are included in the very easy category, 4 items that are included in the easy category, and 11 items that are included in the difficult category. There are no items that are very difficult. On very easy items, most students can answer correctly. If you look at the results of instrument testing, the distribution of difficulty levels in the understanding test items is not evenly distributed (the most items are in the "difficult" category). Instrument trials were carried out on students who had received material on temperature, heat and expansion. Apart from the quality of the questions, other factors that can influence the results of this trial include the condition and characteristics of the students. The trial was carried out in the last hour, so some students lacked focus.

Differentiating Power

The differentiating power of items shows the ability of the items to differentiate between students with high ability and students with low ability. Mark point-measure correlation (PTMEASURE-AL COOR) obtained from the analysis results using MINISTEP software was used to identify the differentiating power of items. Figure 5 displays the values of point-measure correlation of each item.

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
10	30	31	-3.43	1.03	1.11	.42	9.90	3.22	A -.29	.16	96.8	96.7	Q10
5	29	31	-2.67	.75	1.23	.56	6.29	2.46	B -.29	.22	93.5	93.5	Q5
20	20	31	-.28	.43	1.51	2.52	1.86	2.19	C .06	.46	54.8	72.4	Q20
7	28	31	-2.20	.63	1.07	.31	1.50	.77	D .16	.26	90.3	90.2	Q7
3	24	31	-1.06	.47	1.16	.79	1.39	.85	E .24	.39	74.2	78.1	Q3
9	27	31	-1.84	.56	1.18	.59	1.23	.54	F .15	.30	87.1	87.0	Q9
18	10	31	1.50	.44	1.21	.98	1.20	.68	G .35	.49	71.0	75.8	Q18
15	13	31	.94	.42	1.18	.91	1.16	.67	H .39	.51	64.5	74.2	Q15
19	15	31	.59	.42	1.09	.50	1.04	.23	I .46	.51	67.7	73.2	Q19
4	9	31	1.70	.45	.99	.03	.77	-.53	J .52	.48	71.0	76.3	Q4
12	10	31	1.50	.44	.95	-.19	.73	-.78	i .56	.49	71.0	75.8	Q12
16	29	31	-2.67	.75	.82	-.14	.37	-.27	h .37	.22	93.5	93.5	Q16
11	10	31	1.50	.44	.81	-.91	.68	-.95	g .64	.49	83.9	75.8	Q11
13	18	31	.07	.42	.79	-1.20	.68	-1.23	f .65	.49	83.9	72.8	Q13
17	23	31	-.85	.45	.79	-1.05	.61	-.84	e .57	.41	77.4	75.2	Q17
14	10	31	1.50	.44	.73	-1.31	.58	-1.35	d .69	.49	83.9	75.8	Q14
8	8	31	1.91	.47	.71	-1.32	.51	-1.26	c .68	.46	87.1	77.7	Q8
6	8	31	1.91	.47	.68	-1.50	.57	-1.03	b .68	.46	93.5	77.7	Q6
1	8	31	1.91	.47	.62	-1.87	.42	-1.59	a .74	.46	87.1	77.7	Q1
MEAN	18.0	31.0	-.23	.59	.98	-.10	1.66	.09			80.6	80.0	
P.SD	8.6	.0	2.01	.32	.23	1.09	2.32	1.35			11.3	7.7	

Figure 5. Point-measure correlation value on the understanding test instrument

Question items can be categorized as having excellent, good, poor and not good differentiating power. Interpretation of the differentiating power of the items can be seen in Table 12.

Table 12. Interpretation of the Differentiating Power of Understanding Test Items

Number	PTMEASURE-AL COOR	Interpretation
P1	0.74	Very Good
P3	0.24	Less Good
P4	0.52	Very Good
P5	-0.29	Not Good
P6	0.68	Very Good
P7	0.16	Not Good
P8	0.68	Very Good
P9	0.15	Not Good
P10	-0.29	Not Good
P11	0.64	Very Good
P12	0.56	Very Good
P13	0.65	Very Good
P14	0.69	Very Good
P15	0.39	Good
P16	0.37	Good
P17	0.57	Very Good
P18	0.35	Good
P19	0.46	Very Good
P20	0.06	Not Good

Based on the interpretation in Table 12, there are 10 items that have very good differentiating power, meaning that these items can differentiate students with high and low abilities very well. For items with negative differentiating power, the items are still used after being revised according to the advice of the evaluator (expert).

Conclusion

The results of the analysis show that the understanding test items on the topic of temperature, heat and expansion can be used to test students' understanding. Validity analysis results using criteria fit items with output in the form of MNSQ and ZSTD values. Meanwhile, the value of item reliability of 0.88 shows that the quality of the items in the understanding instrument is included in the good category. On the other hand, the distribution of the level of difficulty is uneven, the most questions in the "difficult" category are 11 questions. However, there are 10 questions that have very good differentiating power, meaning that these items can differentiate students with high and low abilities very well. For items with negative differentiating power, these questions can be used after being revised according to the advice of the evaluator (expert). Instrument trials were carried out on students who had received learning abouttemperature, heat and expansion. Apart from the quality of the items, other factors that can influence the results of this trial include the condition and characteristics of the students. The trial was carried out in the last hour, so some students lacked focus.

Acknowledgments
Thank you to all parties who have helped in this research so that this article can be published

Author Contributions
All authors contributed to writing this article.

Funding
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.

Conflicts of Interest
No conflict interest.

References

Alan, U. F., & Afriansyah, E. A. (2017). Kemampuan Pemahaman Matematis Siswa Melalui Model Pembelajaran Auditory Intellectualy Repetition Dan Problem Based Learning. *Jurnal Pendidikan Matematika*, *11*(1). https://doi.org/10.22342/jpm.11.1.3890.67-78

Al-Mutawah, M. A., Thomas, R., Eid, A., Mahmoud, E. Y., & Fateel, M. J. (2019). Conceptual Understanding, Procedural Knowledge and Problem-Solving Skills in Mathematics: High School Graduates Work Analysis and Standpoints. *International journal of education and practice*, *7*(3), 258-273. https://doi.org/10.18488/journal.61.2019.73.258.273

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., & Wittrock, M. C. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. United States: Pearson Education. Inc.

Bayuningsih, A. S., Usodo, B., & Subanti, S. (2017). Analysis of Junior High School Students' Problem-solving Ability Reviewed From Self-regulaterd Learning. *International Journal of Science and Applied Science: Conference Series*, 2(1), 51-56. http//doi.org/10.20961/ijsascs.v2i1.16678

Bernardo, A. B. I., Ganotice, F. A., & King, R. B. (2015). Motivation Gap and Achievement Gap Between Public and Ptivate High School in the Philippines. *Asia Pacific Education Researcher*, 24(4), 657-667. http//doi.org/10.1007/s40299-014-0213-2

Cabero, J., Almenara, J. C., & Cejudo, M. D. C. L. (2013). La aplicación del juicio de experto como técnica de evaluación de las tecnologías de la información y comunicación (TIC). *Revista Eduweb, 7*(2), 11-22. Retrieved from https://revistaeduweb.org/index.php/eduweb/article/download/206/154

Caleon, I., & Subramaniam, R. (2010). Development and Application of a Tree-Tier Diagnostic Test to Assess Secondary Students' Understanding of Waves. *International Journal of Science Education*, 32(7), 939-961. https://doi.org/10.1080/09500690902890130

Cetin-dindar, A. (2015). Studen Motivation in Constructivist Learning Environment. *EURASIA Journal of Mathematics, Science & Technology Education*, 12(2), 233-247. https://doi.org/10.12973/eurasia.2016.1399a

Cheriani., Mahmud, A., Tahmir, S., Manda, D., & Dirawan, G. D. (2015). Problem-Based Learning-Buginese Cultural Knowledge Model-Case Study: Teaching Mathematics at Junior High School. *International Education Studies*, 8(4), 105. https://doi.org/10.5539/ies.v8n4p104

Eryilmaz, A. (2010). Development and Application of Three-tier Heat and Temperature Test: Sample of Bachelor and Graduate Students. *Eurasian Journal of Educational Research*. 40(40), 53-76. https://doi.org/10.1080/282910056

Endaryono, & Dwitiyanti, N. (2023). Perhitungan Korelasi dan Regresi dalam Penentuan Koefisien Muai Volume Beberapa Minyak Nabati. *Jurnal MSA (Matematika Dan Statistika Serta Aplikasinya)*, *11*(1), 81–87. https://doi.org/10.24252/msa.v11i1.23798

Escobar-Pérez, J., & Cuervo-Martínez, Á. (2008). Validez de contenido y juicio de expertos: una aproximación a su utilización. *Avances en medición, 6*(1), 27-36. Retrieved from https://gc.scalahed.com/recursos/files/r161r/w25645w/Juicio_de_expertos_u4.pdf

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*, *12*(4), 1136. https://doi.org/10.3390/nu12041136

Fitrianingrum, A. M., Sarwi, & Astuti, B. (2017). Penerapan Instrumen Three-tier Test untuk Mengidentifikasi Siswa SMA pada materi Keseimbangan Benda Tegar. *Phenomenon*, *7*(2), 88–98. https://doi.org/10.21580/phen.2017.7.2.1118

Forehand, M. (2010). Bloom's taxonomy. *Emerging perspectives on learning, teaching, and technology*, *41*(4), 47-56. Retrieved from https://www.scirp.org/reference/referencespapers?referenceid=2799954

Hidayati, U. N., Sumarti, S. S., & Nuryanto. (2019). Desain instrumen tes three tier multiple choice untuk analisis pemahaman konsep peserta didik. *Jurnal Inovasi Pendidikan Kimia*, *13*(2), 2425–2436. Retrieved from Retrieved from https://journal.unnes.ac.id/nju/index.php/JIPK/article/view/19382

Idawati, S., Baso, U., Selviani, B., & Sunarmi. (2016). Pengaruh Suhu Terhadap Perpindahan Panas Pada Material Yang Berbeda. *Jurnal Dinamika*, *7*(1), 62–73. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/09500799708666915%5

Idris, N. (2009). Enhancing students' understanding in calculus trough writing. *International Electronic Journal of Mathematics Education*, *4*(1), 36-55. https://doi.org/10.29333/iejme/229

Konicek-Moran, R., & Keeley, P. (2015). *Teaching for conceptual understanding in science*. Arlington: NSTA Press, National Science Teachers Association.

Kirbulut, Z. D., & Geban, O. (2014). Using Three-tier Diagnostic Test to Assess Students' Misconceptions of States of Matter. *Eurasia Jurnal of Mathematich, Science and Technology Education*, 10(5), 509-521. https://doi.org/10.12973/eurasia.2014.1128a

Lestari, P. A. S., Rahayu, S., & Hikmawati, H. (2017). Profil Miskonsepsi Siswa Kelas X Smkn 4 Mataram pada Materi Pokok Suhu, Kalor, dan Perpindahan Kalor. *Jurnal Pendidikan Fisika Dan Teknologi*, *1*(3), 146–153. https://doi.org/10.29303/jpft.v1i3.251

Laksono, P. J. (2020). Pengembangan Three Tier Multiple Choice Test pada Materi Kesetimbangan Kimia Mata Kuliah Kimia Dasar Lanjut. *Orbital Jurnal Pendidikan Kimia*, 4(1), 44-63. https://doi.org/10.19109/ojpk.v4i1.5649

Marlina, R., Nurjahidah, S., Sugandi, A. I., & Setiawan, W. (1018). Penerapan Pendekatan Problem Based Learning untuk Meningkatkan Kemampuan Pemecahan Masalah Matematis Siswa Kelas VII MTs pada Materi Perbandingan dan Skala. *Jurnal Pembelajaran Matematika Inovativ*, 1(2), https://doi.org/10.22460/jpmi.v1i2.p113-122

Murti, Wiyanto, & Hartono. (2018). Studi Komparasi antara Tes Testlet dan Uraian dalam Mengukur Hasil Belajar Kognitif Siswa Kelas XI SMA Negeri 1 Gombong. *Unnes Physics Education Journal*, *7*(1), 23–31. https://doi.org/10.15294/upej.v7i1.22469

Mutmainna, D., Mania, S., & Sriyanti, A. (2018). Pengembangan Instrumen Tes Diagnostik Pilihan Ganda Dua Tingkat Untuk Mengidentifikasi Pemahaman Konsep Matematika. *MaPan*, *6*(1), 56–69. https://doi.org/10.24252/mapan.2018v6n1a6

Montfort, D., Brown, S., & Pollock, D. (2009). An investigation of students' conceptual understanding in related sophomore to graduate-level engineering and mechanics courses. *Journal of Engineering Education*, *98*(2), 111-129. https://doi.org/10.1002/j.2168-9830.2009.tb01011.x

Nursalam, N. (2016). Diagnostik Kesulitan Belajar Matematika: Studi pada Siswa SD/MI di Kota Makasar. *Lentera Pendidikan: Jurnal Ilmu Tarbiyah dan Keguruan*, 19(1), 1-15. https://doi.org/10.1103/10.24252/ip.2016v19n1a1

Ogilvie, C. A. (2009). Changes in students' problem-solving strategies in a course that includes context-rich, multifaceted problems. *Physical Review Special Topics – Physics Education Research, 5*(2), 020102. https://doi.org/10.1103/PhysRevSTPER.5.020102

Radiusman, R. (2020). Studi Literasi: Pemahaman Konsep Anak Pada Pembelajaran Matematika. *FIBONACCI: Jurnal Pendidikan Matematika Dan Matematika*, *6*(1), 1. https://doi.org/10.24853/fbc.6.1.1-8

Rusilowati, A. (2015). Pengembangan Tes Diagnostik Sebagai Alat Evaluasi Kesulitan Belajar Fisika.

Prosiding Seminar Nasional Fisika Dan Pendidikan Fisika, *6*(1), 1–10. Retrieved from http://www.jurnal.fkip.uns.ac.id/index.php/prosfis1/article/view/7684

Syarif, D. F. T., Fathurrahman, M., & Karyani. (2019). *Tehnik Creative Problem Solving*. Yogyakarta: K-Media.

Setiani, R., Sanjaya, I., & Jatmiko, B. (2019). ARICESA as an Alternative Learning Model to Improve Learning Motivation and Understanding of Student Concepts. *International Journal of Instruction*, *12*(2), 383-398. https://doi.org/10.29333/iji.2019.12225a

Slotte, V., & Lonka, K. (1999). Spontaneous concept maps aiding the understanding of scientific concepts. *International Journal of Science Education*, *21*(5), 515-531. https://doi.org/10.1080/095006999290552

Sofianto, E. W. N., & Irawati, R. K. (2020). Efforts to mediate the concept of physics in temperature and heat matter. *Southeast Asian Journal of Islamic Education*, *2*(2), 107–120. https://doi.org/10.21093/sajie.v2i2.2188

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

Syahrul, D. A., & Setyarsih, W. (2015). Identifikasi miskonsepsi dan penyebab miskonsepsi siswa dengan three-tier diagnostic test pada materi dinamika rotasi. *Jurnal Inovasi Pendidikan Fisika (JIPF),* *4*(03), 67-70. http//doi.org/10.26740/ipf.v4n3.p%25p

Supriyadi, U., Marlina, L., & Sriyanti, I. (2021). E-Module Pada Materi Suhu, Pemuaian, dan Kalor: Valid dan Praktis. *Jurnal Ilmiah Pendidikan Fisika*, *5*(2), 201. https://doi.org/10.20527/jipf.v5i2.3787

Suwarto. (2013). *Pengembangan Tes Diagnostik dalam Pembelajaran*. Yogyakarta: Pustaka Pelajar.

Tan, R. M., Yangco, R. T., & Que, E. N. (2020). Students' Conceptual Understanding and Science Process Skills in an Inquiry-Based Flipped Classroom Environment. *Malaysian Journal of Learning and Instruction*, *17*(1), 159-184. Retrieved from https://files.eric.ed.gov/fulltext/EJ1248985.pdf

Widdiharto, R. (2008). *Diagnosis Kesulitan Belajar Matematika SMP dan Alternatif Proses Remidinya.* Yogyakarta: Pusat Pengembangan dan Pemberdayaan Pendidik dan Tenaga Kependidikan Matematika.

Yanti, R. P., & Ihsan, I. (2014). Studi Penentuan Nilai Kalori Pada Buah Durian (Durio Zibethinus). *Teknosains: Media Informasi Sains dan Teknologi, 8(2)*, 161-174. https://doi.org/10.24252/teknosains.v8i2.1900