

Developing the HETMOFT : A Five-Tier Test for Identifying Student Misconceptions in Heat and Thermodynamics

Aulia Ayuni Novita^{1*}, Agus Danawan¹, Lina Aviyanti¹, and Abu Nawas²

¹Program of Physics Education, Universitas Pendidikan Indonesia, Bandung, Indonesia

²School of Education, University of Adelaide, Adelaide, South Australia

*auliaayuninovita24020@upi.edu

Article Info	Abstract
Article history: Received: January 8, 2025 Accepted: March 28, 2025 Published: June 10, 2025 Keywords: Heat and thermodynamics Misconception Physics learning Rasch model This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License © 2025 Berkala ilmiah Pendidikan Fisika	In the implementation of learning, there are often errors or discrepancies in the interpretation of student concepts, with the concepts of experts called misconceptions. In physics learning, one of the materials where misconceptions often occur is thermodynamics and heat. This study aims to analyze high school students' understanding of the concepts of heat and thermodynamics in Karawang Regency. This study used the Research and Development (R&D) model to develop a test instrument in the form of 15 items distributed through Google Forms to 195 students (105 female and 90 male). The data obtained was analyzed using the Rasch model through Winstep software version 5.1.7.0, which provides an overview of the level of student understanding and identifies differences in understanding between male and female students. The five-tier test instrument was validated by three teachers and two lecturers using the Aiken V formula and obtained a result of 0.69 and a probability of 0.010 with 18 "valid" items and 2 "invalid" items. The results of instrument reliability show person reliability is 0.55 with a "weak" interpretation, item reliability is 0.81 with a "good" interpretation, and Cronbach's alpha (KR-20) value is 0.54 with a "low" interpretation. Based on these results, it can be concluded that the instrument can be used in this study. The results showed significant misconceptions among students, especially regarding the concept of temperature change, with the highest percentage being 44.11%. The main cause of this misconception was identified as coming from students' thinking. Male students showed a better understanding of heat transfer and temperature change, while female students showed a slightly better understanding of substance form change. In addition, a five-tier test instrument is recommended as a learning evaluation tool to help educators identify students' misconceptions in-depth and understand the causes of these misconceptions.
To cite this article: Novita, A. A., Danawan, A., Aviyanti, L., and Nawas, A. (2025). Developing the HETMOFT: A five-tier test for identifying student misconception in heat and thermodynamics. <i>Berkala Ilmiah Pendidikan Fisika</i>, 13(2), 230-248.	

INTRODUCTION

Education is a responsibility for every country, and it greatly affects the progress of a nation. The quality of education can be seen from the quality of learning in the classroom. One effort to improve this quality is to make the learning atmosphere more enjoyable so that students are not bored when taught in class and more easily accept the learning

concepts being studied. Especially in the field of education, qualified human resources are expected to be able to provide discoveries ([Puspitasari, 2018](#); [Darmaji et al., 2019](#); [Dina et al., 2024](#)). Physics learning is seen as a thought process to develop the ability to understand the concepts, principles, and laws of physics that are very abstract. At the high school level, physics learning is expected to be carried out by teachers with process accuracy when learning physics in the classroom. Things that need to be done in physics learning must start with mastering concepts, applying them in problem-solving, and doing it scientifically ([Nasution, 2018](#); [Lidiana et al., 2018](#)). Studying physics not only focuses on facts, laws, theories, principles, models, and mastery of formulas, but also on understanding the basic concepts of the subject matter. One indicator of the success of the learning process is concept understanding ([Hadijah & Napitupulu, 2016](#); [Nurhayati, 2012](#)).

Concept understanding is an understanding related to theoretical principles that requires students to master concepts first ([Fitriani & Astuti., 2017](#)). In the implementation of learning, there are often errors or discrepancies in the interpretation of students' concepts with the concepts of experts, called misconceptions ([Khairaty et al., 2018](#)). The unequal definition of concepts in physics and everyday life can cause errors or discrepancies in a concept ([Nasution & Siregar, 2019](#)). Misconceptions are formed from students' knowledge that is not in accordance with the concept. The influence of misconceptions has an impact on hindering the development of students' understanding of learning concepts ([Hasniyah & Muchtar, 2021](#)). Meanwhile, according to [Fariyani & Rusilowati \(2015\)](#) Misconceptions occur when students do not provide correct explanations of physics concepts in accordance with what scientists have agreed upon; students build theories in their minds that are not necessarily true based on their daily experiences.

Concept understanding in physics learning is very important. One of the main objectives of physics education is to understand the basic concepts of physics, which are very important for understanding the world ([Aviyanti, 2020](#)). Given that the achievement of concepts is the basis for forming logical thinking ([Ifrianti et al., 2019](#)). Good concept mastery can be characterized by students' ability to use their knowledge to solve problems ([Puspita et al., 2019](#)). The impact caused by misconceptions can mislead students in understanding and explaining natural phenomena, thus hindering the process of receiving new information related to the concept ([Furoidah, 2017](#)). Thus, misconceptions in students need to be overcome because errors in understanding concepts will cause errors in subsequent concepts and result in low learning outcomes ([Pebriyanti, 2015](#)).

Misconceptions in various fields of science, especially physics, have been long and widely revealed by researchers from various places ([Hidayati et al., 2024](#)). Based on the findings of research conducted by [Sari \(2021\)](#), from various materials in physics learning, the materials where misconceptions often occur are thermodynamics and heat. Thermodynamics studies heat and how it is transferred ([Yaqin et al., 2017](#)). This is because students' creativity and imagination are also very limited, and the subject matter of thermodynamics is abstract, thus making students' understanding of the subject matter of thermodynamics still relatively low ([Merlina, 2021](#)). In a study by [Sari \(2021\)](#), conceptual errors fell into the high-level misconception category, namely 75.5% of students. In addition, other findings relevant to analyzing students' misconceptions of heat and thermodynamic material, namely, research conducted by [Fitriah \(2017\)](#), [Handayani \(2018\)](#), and [Utari & Ermawati \(2018\)](#), stated that students still experience misconceptions about thermodynamics and heat material. Furthermore, relevant research using the Five-tier test to analyze the misconceptions experienced by students is conducted by [Merlina \(2021\)](#).

In a study by [Sari \(2021\)](#), conceptual errors were included in the high-level misconception category, namely 75.5% of students. In addition, other findings relevant to analyzing students' misconceptions of heat and thermodynamic material, namely research conducted by [Fitriah \(2017\)](#), [Handayani \(2018\)](#), and [Utari and Ermawati \(2018\)](#), stated that students still experience misconceptions about thermodynamics and heat material.

Furthermore, relevant research using the Five-tier test to analyze the misconceptions experienced by students is research conducted by [Merlina \(2021\)](#), which shows the results of the analysis of students' misconceptions identified in the subconcepts of effort, heat, energy, and thermodynamic processes of 81.7%; the subconcept of law I of thermodynamics and its application of 70.3%; the subconcept of law II of thermodynamics and its application of 87.5%; and the Carnot cycle subconcept of 92.2%.

Misconceptions that occur in students need to be considered by teachers in anticipation of the occurrence of sustainable misconceptions in physics learning. If the teacher does not consider misconceptions in participants, it will impact students' learning outcomes. Therefore, misconceptions that occur need to be analyzed and corrected so that the objectives of the learning process can be met ([Syahratinnur & Kafrawi, 2023](#)). This phenomenon hinders the acquisition of new knowledge and can also affect students' attitudes toward science learning, reducing their interest and motivation ([Guerra-Reyes et al., 2024](#)). Reducing misconceptions is very important because misconceptions are a common problem that often arises after learning ([Khusaini et al., 2025](#)). Therefore, a measurement tool is needed to determine students' misconceptions. Making a good test instrument needs to be done as a provision for teachers to measure how far students' concept understanding has progressed after carrying out learning activities. An instrument is a tool that meets academic requirements and can be used to measure a measuring object or collect data about a variable ([Sappaile, 2007](#)). While the test is one way to indirectly assess the magnitude of a person's ability, namely through a person's response to a stimulus or question ([Supardi, 2013](#)).

The data obtained from this study will be analyzed using the Rasch model. This Rasch model produces more objective measurements and provides detailed diagnostic information about student ability and item difficulty ([Sumintono & Widhiarso, 2015](#)). From the Rasch model analysis, students' ability levels can be determined by looking at their scores and total scores ([Irsalina et al., 2020](#)). The use of the Rasch model in this study is able to produce objective measurements, thus enabling more accurate identification of students' conceptual error patterns, question difficulty levels, and student understanding profiles ([Hamdu et al., 2023](#)).

The use of inappropriate measurement tools will make it difficult for teachers to find more information about concepts that students have not understood ([Nurhayati et al., 2019](#)). Measuring students' concept-understanding abilities and knowing the existence of student difficulties or problems can be done by using diagnostic tests. Diagnostic tests can be used to determine and ensure students' weaknesses and strengths in certain lessons. For teachers, diagnostic tests provide information that can be used to update the learning process, while for students, they can be used to improve the learning process ([Zaleha et al., 2017](#)). There are several forms of diagnostic tests, including one-tier (one level), two-tier (two levels), three-tier (three levels), four-tier (four levels), and five-tier (five levels) diagnostic tests ([Didik et al., 2020](#)). In this study, a five-tier diagnostic test was used, which is a development of a four-tier test.

The Five Tier Test consists of five levels, namely, conceptual questions with one answer key and four exemptions; the second level contains the confidence level of the question-answer; the third level contains the reason for the answer at the first level; the fourth level contains the confidence of the reason for the answer at the third level, and the fifth level contains the source from which students answer questions at the first and third levels ([Inggit et al., 2021](#)). According to [Syahratinnur & Kafrawi \(2023\)](#), the advantage of this five-tier test is that students are given a package of questions with answers accompanied by reasons, equipped with a scale of confidence levels in the answers and reasons chosen in the question items. Students are given several alternative answer choices, reasons, and confidence levels in answering questions.

This research was conducted in Karawang Regency, where the five-tier test is rarely used to evaluate student learning. Students have difficulty answering the first question, so they answer the question arbitrarily. The use of the five-tier test as a learning evaluation tool can also identify the source of the cause if misconceptions occur in students. Therefore, it is hoped that this research can inspire teachers to use the five-tier test as a better learning evaluation tool.

METHODS

This development research uses a development model that uses the Research and Development (R&D) method. R&D is a research method used to produce a certain product and test the effectiveness of the developed product ([Nurhalisa et al., 2021](#)). This research is included in the development research category because it produces a test instrument product. The development model is the basis for developing the product to be produced. Data was collected from students through Google Forms. The data obtained was analyzed using the Rasch model, which provides an overview of students' level of understanding and identifies differences between male and female students. In this study, no intervention is given to students, so data collection on students is effective.

The development model is the basis for developing the product to be produced. The development model used in this research is the 4D form, Define, Design, Development, and Disseminate. According to [Mulyatiningsih \(2016\)](#), the stages in this 4D research development procedure include: Define is a stage for analyzing the curriculum, analyzing the characteristics of students, analyzing materials, and formulating objectives; Design is a stage for collecting material references, making material grids, and making question items in the form of a five-tier test; development is a stage for developing a five-tier test open-ended instrument into a five-tier test close-ended based on the results of a preliminary study of student reasons in the third tier. Furthermore, the instrument is carried out in the validation test stage on teachers, lecturers, and students. The validation test results are then used as revisions so that the items developed are truly suitable for use. Dissemination is a stage where declared feasible instruments would be disseminated and tested on many students to obtain responses.

The participants involved in this study were grade 11 students who had studied heat and thermodynamics at high schools in Karawang Regency. The high school used during the preliminary study data collection differed from those used during the large-scale instrument trial. Participants during the preliminary study data collection consisted of 10 male and 15 female students aged 15-17. Participants were taken from one of the classes that had learned about heat and thermodynamics. Meanwhile, participants during the large-scale instrument trial involved 195 students, including 105 female and 90 male students from three high schools in Karawang Regency.

The sample determination used in this study was carried out using a purposive sampling technique. [Sugiyono \(2019\)](#) states that purposive sampling is a sampling technique with certain considerations. The determination of the sample size used a table with a confidence interval of 90% or an error rate of 10%, as developed by Isaac and Michael in [Sugiyono \(2019\)](#). The population of grade 11 high school students from three high schools in Karawang Regency is 700 students, so the minimum sample size based on the table for determining the sample size developed by Isaac and Michael with an error rate of 10% is $N_T = 195$ students. To determine the minimum sample size required from each school as follows:

$$N_s = \frac{N}{\sum N} N_T \quad (1)$$

Information:

N_s = Number of student samples required,
 N = Number of students in each type of school,
 $\sum N$ = Number of all students,

N_T = Number of samples based on the table

The test used in this study is a five-tier diagnostic test on heat and thermodynamics material with as many as 15 questions. In the Five-Tier Test, the first tier is in the form of multiple-choice questions with several answer choices; the second tier is the level of confidence of students in choosing answers in the first tier; the third tier is the reason for the answer in the first tier; the fourth tier is the level of confidence of students in choosing answers in the third tier; and the fifth tier is the source from which students get answers in the third tier. The five-tier test instrument created has been tested on grade 11 students who have studied heat and thermodynamics. This instrument is given to students online via Google Forms, which contains the final product of the five-tier test instrument development that has passed several validation stages with a specified deadline. The following five-tier diagnostic test instrument design can be seen in Table 1.

Table 1 Five-tier test instrument design

Questions: Diagnose the student's conception
Option Answer (Tier-1) A. First alternative answer B. Second alternative answer C. Third alternative answer D. Fourth alternative answer E. Fifth alternative answer
Confidence Level (Tier-2) A. Alternative if the student's is "confident" of the selected answer B. Alternative if the student's is "not sure" of the selected answer
Reason (Tier-3) A. First alternative answer B. Second alternative answer C. Third alternative answer D. Fourth alternative answer E. Fifth alternative answer
Second Confidence Level (Tier-4) A. Alternative if the student's is "confident" of the selected answer B. Alternative if the student's is "not sure" of the selected answer
Source of answer choices in tier-1 and tier-3 (tier-5) A. First alternative answer B. Second alternative answer C. Third alternative answer D. Fourth alternative answer E. Fifth alternative answer

After the five-tier test instrument on heat and thermodynamics was field tested, the research results need to be processed using Rasch Model analysis with the help of Winstep 5.1.7.0 software. Rasch analysis is part of item response theory that can group persons (subjects) and items (questions) in a distribution map so that the model is able to describe the relationship between the subject's ability level and the question item.

According to [Ünal et al. \(2010\)](#), students' conceptions are grouped into several categories, namely SU (Sound Understanding), which shows that students' conceptions are intact because students can answer questions correctly. The correct reasons are also confident in their answers; PP (Partial Positive) shows that students' understanding of a concept is not intact because of doubts in answering questions, and PN (Partial Negative) shows that the understanding of a concept is not complete because it is only correct in one-tier. The level of confidence is different. MC (Misconception) shows that students' understanding of concepts is not in accordance with scientific concepts, but students are

confident in their understanding, while NU (No Understanding) shows the condition of students who do not understand scientific concepts. In addition, it is also grouped based on the source of the cause of misconception, namely personal thoughts (PT), teacher, book (B), people's explanation (PE), and the internet (I).

Instrument validity consists of construct, content, and empirical validity. To test the construct and content validity, the opinions of experts (experts' judgment) can be used (Sugiyono, 2019). The results of the validation data by experts were then analyzed using the Aiken Formula. Aiken's formula is used to determine the V coefficient of rating-scale data statistically (Aiken, 1985). Aiken's V analysis was used in this study to measure expert validation. This method is widely used to measure content validity based on expert judgment. Aiken's V allows quantitative measurement of the level of expert agreement, especially on ordinal data such as Likert scales, with easily interpretable values (0-1), where values close to 1 indicate a high level of validity (Aiken, 1985). This method is also suitable for small expert panels, which are common in validation research and have been widely recognized in scientific research as a credible method. Thus, Aiken's V helps to ensure that the research instruments developed in this study have adequate content validity for further use (Aiken, 1985). Aiken's V formula is formulated as follows:

$$V = \frac{\sum s}{n(c-1)} \quad (2)$$

Information:

V = item validity index
 $s = r - l_0$ = score given by the rater; l_0 = lowest score
 c = number of assessment categories
 n = number of validators

The empirical validity was tested by analyzing the Rasch model using software and entering the student's conception score based on the results of implementing the Five-Tier Test instrument. This validity test is analyzed based on all items, and it is called unidimensionality for each item (Sumintono & Widhiarso, 2015). The validity test based on all items is obtained from the output tables menu in the item dimensionality section. The value of raw variance explained by measures is the validity value based on all items, which can then be interpreted based on Table 2.

Table 2 Instrument unidimensionality interpretation

Raw variance explained by measure	Interpretation
$20\% \leq \text{value} \leq 40\%$	Fulfilled
$40\% \leq \text{value} \leq 60\%$	Retrieved
$\text{value} \geq 60\%$	Special

The quality of each item can be seen based on the outfit mean square (MNSQ) value, which shows the randomness in the form of the number of deviations in the measurement system; the outfit Z-standard (ZSTD) value, which shows how much deviation is likely to occur; and the Pt measure corr value, which shows the relationship between the difficulty of each item and the difficulty of the overall instrument (Sumintono & Widhiarso, 2015). The criteria for checking the quality of each item can be seen in Table 3.

Table 3 Score criteria outfit MNSQ, outfit ZSTD, and Pt Measure Corr

Criteria	Accepted Range	Description
Outfit Mean Square (MNSQ)	$0.5 < \text{MNSQ} < 1.5$	Fit
Outfit Z Standard (ZSTD)	$-2.0 < \text{ZSTD} < + 2.0$	Fit
Point Measure Correlation (Pt Measure Corr)	$0.4 < \text{Pt Measure Corr} < 0.85$	Fit

Instrument reliability is the determination of the instrument in the measurements taken in a study or the respondent's determination in answering the question instrument. This equality shows that the instrument is reliable ([Sumintono & Widhiarso, 2015](#)). The data that was obtained for this reliability includes person reliability (p), item reliability (r), and Cronbach's alpha (KR-20), which shows the interaction that occurs between person reliability and item reliability ([Sumintono & Widhiarso, 2015](#)). The interpretation of the three can be seen in Table 4.

Statistics	Index Value	Interpretation
Person and Item Reliability	$r \leq 0.67$	Weak
	$0.67 < r \leq 0.80$	Simply
	$0.81 < r \leq 0.90$	Good
	$0.91 < r \leq 0.94$	Very Good
	> 0.94	Special
Alpha Cronbach (KR-20)	$KR-20 < 0.50$	Poor
	$0.50 < KR-20 < 0.6$	Bad
	$0.60 < KR-20 < 0.70$	Simply
	$0.70 < KR-20 < 0.80$	Good
	$KR-20 < 0.80$	Very Good

The item difficulty level (item measure) can be sorted from high to low difficulty, from difficult to easy. The Rasch Model can also find each item processed using Winstep software on the output tables menu in the item measure section, with the Measure (M) and Standard Deviation (SD) value data ([Sumintono & Widhiarso, 2015](#)). The item's difficulty level can be determined from its logit value, which can be seen in the measure column. A high logit value indicates the highest level of difficulty of the question. The table following interprets each item's difficulty level based on Table 5.

Criteria	Interpretation
$ME < -1SD$	Easy
$-1SD \leq ME \leq +1SD$	Medium
$ME > +1SD$	Hard

Item Differentiation power shows how far an item can distinguish students who have high abilities from students who have low abilities ([Sumintono & Widhiarso, 2015](#)). Differentiating power can be analyzed using Rasch modeling assisted by Winstep software version 5.1.7.0. This method involves using the Point Measure Correlation (Pt Mean Corr) value to compare the difficulty level of each item with the overall difficulty level of the item. Table 6 shows the interpretation of differentiating power for items.

Pt Mean Corr	Interpretation
$ID \geq 0.40$	Very Good
$0.40 < ID \leq 0.30$	Good
$0.30 < ID \leq 0.20$	less good
$ID < 0.20$	Bad

The students' conceptions that have been analyzed are used to analyze misconception profiles and causes of misconceptions. Students identified with misconceptions need to be analyzed in more depth to discover the misconceptions that occur in them and the sources that cause them to experience misconceptions based on the concepts tested on the five-tier test instrument.

Procedure and development is a research stage carried out for product development. The 4D development model consists of four steps: definition, design, development, and dissemination. Based on the development model, Figure 1 is a flowchart of the learning media development procedure.

The 4D development model consists of 4 stages, which are as follows:

1. Define: at this stage, the analysis was carried out to analyze the curriculum, analyze the characteristics of students, analyze the material, and formulate objectives.
2. At this stage, design was carried out, which included collecting material references, making material grids, and making items in the form of five-tier tests.
3. At this stage, the open-ended five-tier test instrument was developed into a close-ended five-tier test based on the results of the preliminary study of student reasons in the third tier. Furthermore, the instrument was carried out in the validation test stage on teachers, lecturers, and students. The results of the validation test were then used as revisions so that the items developed are truly suitable for use.
4. Disseminate at this stage; the declared feasible instrument would be distributed and tested on many students to obtain a response.

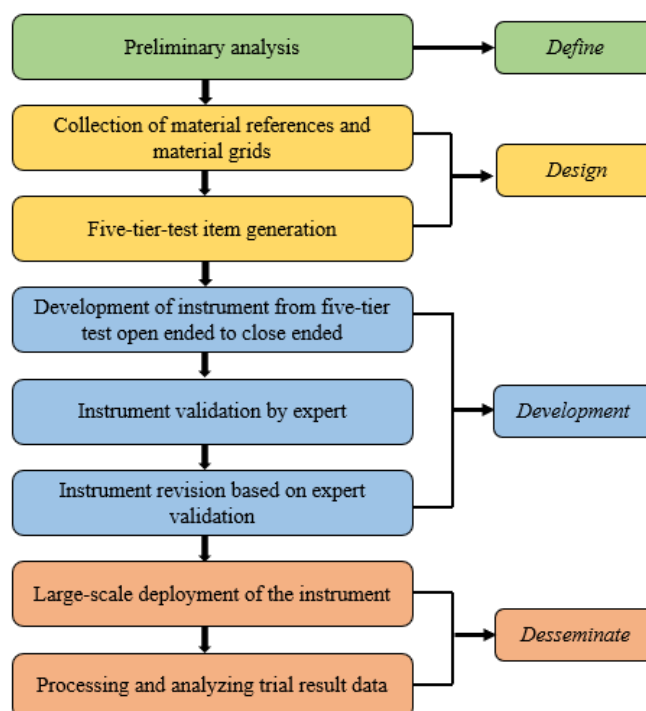


Figure 1 HETMOFT preparation stage with 4D models

RESULTS AND DISCUSSION

Construct Validity and Instrument Content

In this study, the indicators used in making the five-tier test instrument are to measure students' misconceptions and to identify the causes of misconceptions in students. Fives of student conception levels were ident, including SU :rstanding), PP (Partial Positive), PN (Partial Negative), MC (Misconception), and NU (No Understanding). The results of expert validation (expert judgment) on the Five Tier Test instrument on heat and thermodynamics material conducted by two physics education lecturers and three physics teachers were analyzed using the Aiken V Formula in determining Aiken's validity coefficient (Aiken's

V) assisted by Microsoft Excel. Since five experts and 20 assessment categories validated the research instrument, the minimum Aiken's V coefficient standard value obtained was 0.69 with a probability of 0.010. Based on expert validation of the Five-Tier Test instrument on heat and thermodynamics material, as many as 18 items, the calculation results obtained are included in the "Valid" category, which has a value greater than the minimum standard value of Aiken's V table, and there are two items with the "Invalid" category, which has a value smaller than the minimum standard value of Aiken's V table. Therefore, it can be concluded that from the results of expert validation, 18 valid items can be used as instruments in this study.

Empirical Validity of Instruments

The Five Tier Diagnostic Test instrument on heat and thermodynamics material was given as many as 18 items using Google Forms media to 25 students consisting of 10 male and 15 female students. At the time of the trial, it was conducted in 1 class, but the number of students available was only 25. The overall validity test using the Rasch model is called unidimensionality. The results of the unidimensionality of the Five-Tier-Test instrument on heat and thermodynamics material obtained from the output of the dimensionality item table in Winstep software version 5.7.1 are shown in Figure 2.

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = ITEM information units				
	Eigenvalue	Observed	Expected	
Total raw variance in observations =	22.0835	100.0%	100.0%	
Raw variance explained by measures =	7.0835	32.1%	32.6%	
Raw variance explained by persons =	1.7416	7.9%	8.0%	
Raw Variance explained by items =	5.3419	24.2%	24.6%	
Raw unexplained variance (total) =	15.0000	67.9%	100.0%	67.4%
Unexplnd variance in 1st contrast =	2.4973	11.3%	16.6%	
Unexplnd variance in 2nd contrast =	2.2198	10.1%	14.8%	
Unexplnd variance in 3rd contrast =	2.0419	9.2%	13.6%	
Unexplnd variance in 4th contrast =	1.5266	6.9%	10.2%	
Unexplnd variance in 5th contrast =	1.4229	6.4%	9.5%	

Figure 2 Output table item dimensionality

Figure 1 shows that the raw variance explained by measures value obtained from the field trial of the Five-Tier-Test instrument on heat and thermodynamics material of as many as 18 items is 32.1%, with the criteria "fulfilled." In addition, according to Fisher ([Aviyanti et al., 2024](#)), the unidimensionality of the instrument can also be seen from the value of unexplained variance in the first contrast; if the value is less than 15%, then the instrument has a good quantity of unidimensionality. It can be seen from the table output above that the value of unexplained variance in the first contrast is 11.3%, so it can be said that the quantity of unidimensionality of the instrument can be said to be good. After testing the overall validity, the next step is to test the validity of each item. Based on the fit-statistics value, the validity test for each item is explained in Table 7.

Table 7 shows that three items do not meet all the value criteria, so they cannot be used. Based on the results of this analysis, as many as 15 items on the Five-Tier-Test instrument on heat and thermodynamics material are valid and can be used in research.

Table 7 Item quality interpretation results

No	MNSQ		ZSTD		Pt Measure Corr.		Value Criteria	Interp.	Desc
	Value	Desc	Value	Desc	Value	Desc			
1	0.89	Fit	-0.01	Fit	0.44	Fit	Three criteria fulfilled	Very Retrieved	Used
2	1.14	Fit	0.51	Misfit	0.34	Fit	Two criteria fulfilled	Retrieved	Used
3	0.74	Fit	-0.63	Fit	0.48	Fit	Three criteria fulfilled	Very Retrieved	Used
4	0.86	Fit	-0.40	Fit	0.55	Fit	Three criteria fulfilled	Very Retrieved	Used
5	0.79	Fit	-0.36	Misfit	0.39	Fit	Two criteria fulfilled	Retrieved	Used
6	1.33	Fit	1.22	Fit	0.29	Fit	Three criteria fulfilled	Very Retrieved	Used
7	1.06	Fit	0.29	Misfit	0.30	Fit	Two criteria fulfilled	Retrieved	Used
8	1.03	Fit	0.21	Misfit	0.31	Fit	Two criteria fulfilled	Retrieved	Used
9	1.63	Misfit	1.61	Fit	0.27	Misfit	One criterion fulfilled	Less Retrieved	Not Used
10	0.85	Fit	-0.44	Fit	0.47	Fit	Three criteria fulfilled	Very Retrieved	Used
11	1.27	Fit	0.65	Fit	0.25	Misfit	Two criteria fulfilled	Retrieved	Used
12	0.47	Fit	-0.32	Misfit	0.34	Fit	One criterion fulfilled	Less Retrieved	Not Used
13	1.02	Fit	0.25	Misfit	0.20	Fit	Two criteria fulfilled	Retrieved	Used
14	0.55	Fit	-0.32	Misfit	0.34	Fit	Two criteria fulfilled	Retrieved	Used
15	0.63	Fit	-0.98	Fit	0.56	Fit	Three criteria fulfilled	Very Retrieved	Used
16	1.72	Misfit	1.73	Fit	0.28	Misfit	One criterion fulfilled	Less Retrieved	Not Used
17	0.62	Fit	0.16	Fit	0.28	Misfit	Two criteria fulfilled	Retrieved	Used
18	1.05	Fit	0.26	Misfit	0.32	Fit	Two criteria fulfilled	Retrieved	Used

Instrument Reliability

The instrument's reliability in this study can be analyzed using Winstep software version 5.1.7.0. On the main menu, Output Tables, select the Summary Statistics menu. The data obtained for this reliability include person reliability (p), item reliability (r), and Cronbach's alpha (KR-20), which shows the interaction between person reliability and item reliability. The results of the reliability test of the five-tier test instrument on heat and thermodynamics are shown in Figure 3.

Figure 3 shows that the person's reliability is 0.55 with a “weak” interpretation. The interpretation results on weak person reliability indicate the possibility of a small number of samples and items. A sample with a large ability range is needed to get a high person reliability coefficient, and the instrument used has many items. Too few items can cause a

low value of person reliability because the more items, the more behavior is measured more precisely (Setiawan et al., 2017).

SUMMARY OF 25 MEASURED (EXTREME AND NON-EXTREME) PERSON								
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	4.8	18.0	-1.23	.74				
SEM	.5	.0	.25	.05				
P.SD	2.3	.0	1.21	.26				
S.SD	2.4	.0	1.23	.26				
MAX.	9.0	18.0	.55	1.86				
MIN.	.0	18.0	-4.80	.60				
REAL RMSE	.81	TRUE SD	.89	SEPARATION	1.10	PERSON RELIABILITY .55		
MODEL RMSE	.78	TRUE SD	.92	SEPARATION	1.18	PERSON RELIABILITY .58		
S.E. OF PERSON MEAN = .25								
PERSON RAW SCORE TO MEASURE CORRELATION = .97								
CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .54 SEM = 1.56								
STANDARDIZED (50 ITEM) RELIABILITY = .79								
SUMMARY OF 15 MEASURED (NON-EXTREME) ITEM								
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	7.9	25.0	.00	.58	1.01	.12	.92	.03
SEM	1.4	.0	.38	.05	.04	.17	.06	.14
P.SD	5.3	.0	1.44	.19	.14	.64	.24	.54
S.SD	5.5	.0	1.49	.20	.14	.67	.24	.56
MAX.	20.0	25.0	2.44	1.04	1.25	1.25	1.33	1.22
MIN.	1.0	25.0	-2.97	.46	.73	-1.34	.47	-.98
REAL RMSE	.63	TRUE SD	1.29	SEPARATION	2.04	ITEM RELIABILITY .81		
MODEL RMSE	.62	TRUE SD	1.30	SEPARATION	2.11	ITEM RELIABILITY .82		
S.E. OF ITEM MEAN = .38								

Figure 3 Output table summary statistic

Figure 3 shows that the item reliability value is 0.81 with a "good" interpretation. However, the reliability measure for the construct is VR: 0.84, which is an acceptable value of > 0.8 and is a strong value, so the instrument can still be used in research. The results of this interpretation indicate that the Five-Tier-Test instrument on heat and thermodynamics material has good quality in terms of consistency of the weight of the questions with the answer choices. Then, for the Cronbach's alpha (KR-20) value of 0.54, with a "low" interpretation. Based on the results of the analysis, it can be concluded that the Five-Tier-Test instrument on heat and thermodynamics is reliable enough to be used in this study.

Level of Item Difficulty

The difficulty level of the five-tier test instrument on heat and thermodynamics material can be analyzed using Winstep software version 5.1.7.0 on the item menu: measure. Based on the output of the item measure table, the standard deviation (SD) value of the results of the Five-Tier-Test instrument trial on heat and thermodynamics material is 1.89. Further difficulty level analysis can be done by calculating the frequency and percentage for each interpretation of the items presented in Table 8.

Table 8 Frequency and percentage of item difficulty level

Interpretation	Frequency	Percentage (%)
Hard	2	13
Medium	12	80
Easy	1	7

Based on Table 10, the largest frequency is at the level of difficulty with interpreting "moderate," namely 12 items with a percentage of 80%. For the interpretation of "easy," there is as many as 1 item with a percentage of 7%. While the interpretation of "difficult"

is 2 items with a percentage of 13%. This shows that the level of difficulty of the Five-Tier-Test instrument on heat and thermodynamics material is quite well distributed.

Differentiating Power of the Instrument

Differentiating power analysis can be done using Rasch modeling assisted by Winstep software version 5.1.7.0 to distinguish students with high abilities from students with low abilities in answering question items. The results of the Five Tier Test instrument differentiator test on heat and thermodynamics material are shown in Table 9.

Table 9 Frequency and percentage of item differentiating power

Interpretation	Frequency	Percentage (%)
Very Good	5	28
Good	6	33
Less Good	7	39

Table 9 shows that the largest frequency is in the interpretation of "not good," namely seven items with a percentage of 39%. For the interpretation of "good," there are six items with a percentage of 33%. As for the interpretation of "very good," namely, as many as five items with a percentage of 28%. This shows that the distinguishing power of the Five-Tier Test instrument on heat and thermodynamics material is quite well distributed.

Student Conception Level

Students' conception levels are identified to determine the profile of students' conceptions of heat and thermodynamic materials. In this study, five categories of student conception levels will be identified, including SU (Sound Understanding), PP (Partial Positive), PN (Partial Negative), MC (Misconception), and NU (No Understanding). Based on the analysis results, the rubric of combining answers from 195 students as a sample is used, as identified in Figure 4.

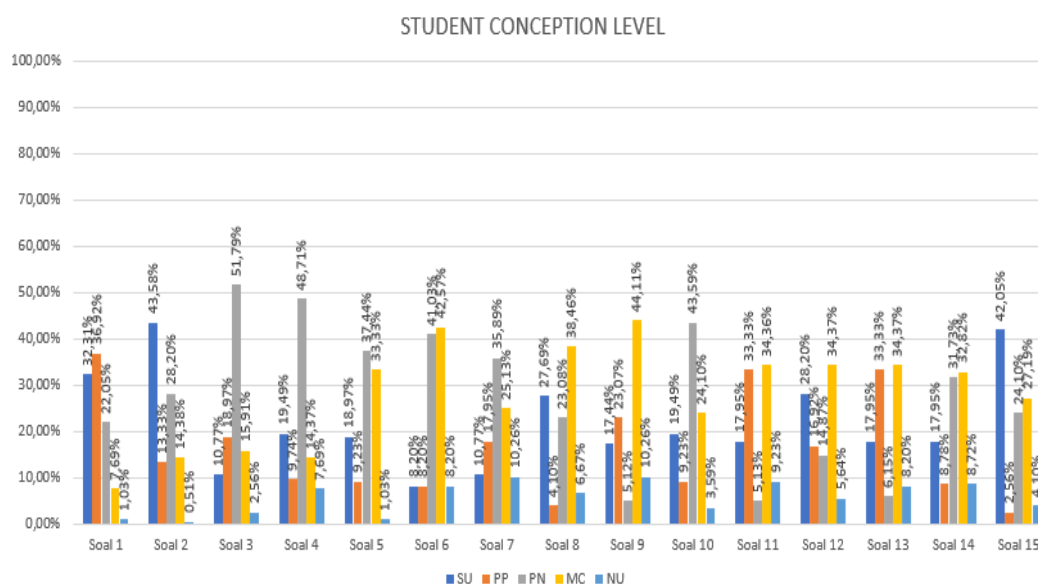


Figure 4 Student conception level

The analysis of student answers presented in Figure 4 shows that the most misconceptions (MC) regarding the concept of temperature change were identified in item 9 at 44.11%. Understanding the concept (SU) was mostly identified in item 2 at 43.58% regarding heat transfer. Understanding the concept but not yet complete (PP) was mostly identified in item number 1 by 36.92% regarding the concept of specific heat. Partially understanding the concept (PN) was mostly identified in item number 3 at 51.79% regarding heat transfer. While the category of not understanding the concept (NU) was

mostly identified by 10.26% in item number 7 regarding the concept of changes in the form of substances and item number 9 regarding the concept of temperature changes. Students' conception level is based on the results of the answers to the five-tier test instrument on heat and thermodynamics material, which can be presented through Wright maps. Based on the results of the analysis of the Wright map to identify the conception profile, it is shown in Figure 5.

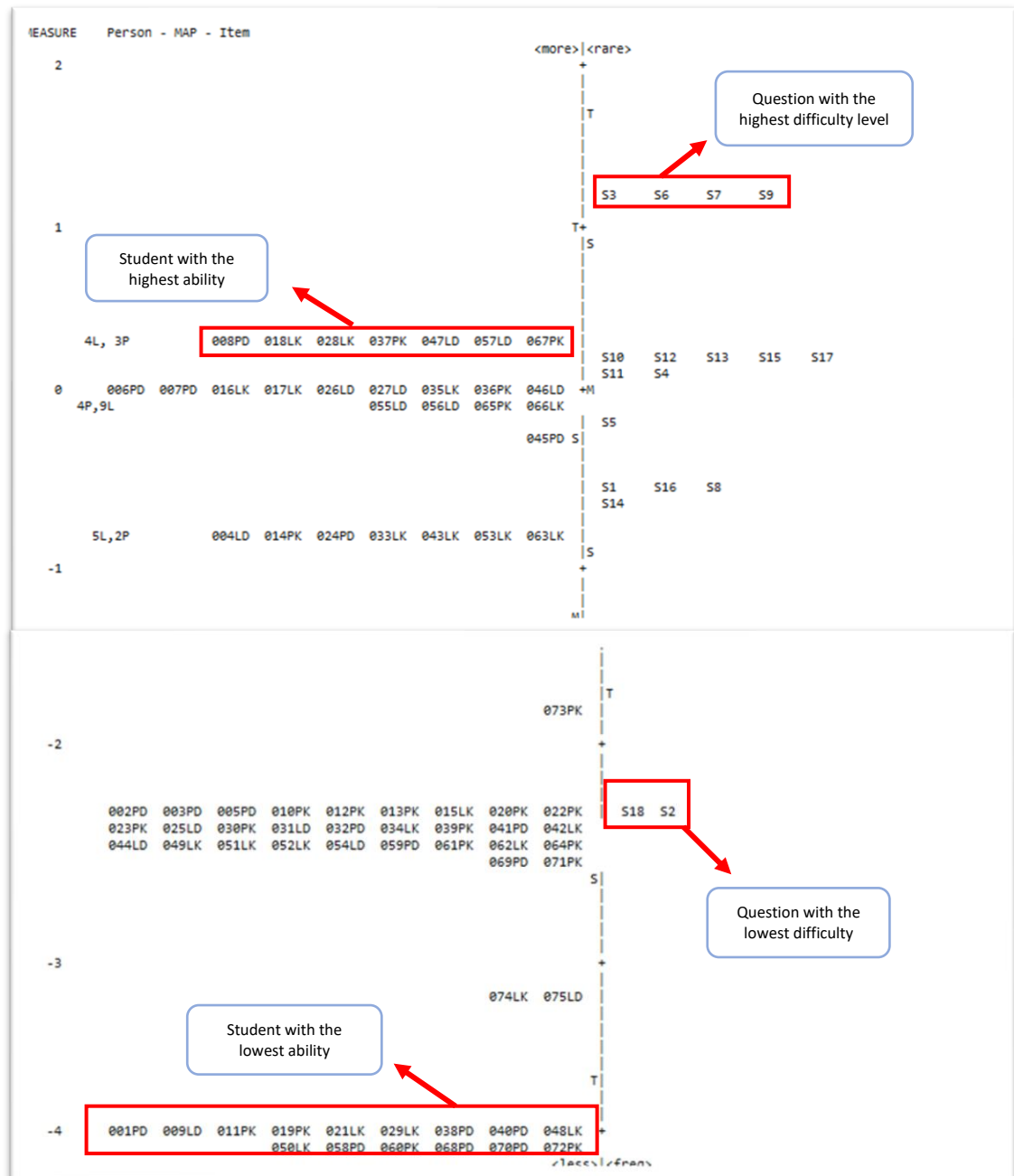


Figure 5 Wright Map the distribution of student conceptions

In Figure 5, it can be seen that the P and L codes on the left are female and male student data, while on the right, the letter S code can be seen, which is a question item. The analysis shows that the items with codes S3, S6, S7 and S9 are the most difficult questions, while the easiest item is item number 2. In addition, Figure 4 also shows that several students

experience misconceptions, namely 1P, 9L, 11P, 19P, 21L, 29L, 38P, 40P, 48L, 50L, 58P, 60P, 68P, 70P, and 72P.

Misconception Profile and Causes of Misconception

The misconception profile and the causes of students' misconceptions about heat and thermodynamics can be seen through the combination of students' answers to the five-tier test instrument on heat and thermodynamics. The results of the recapitulation of students' misconception profiles and causes of misconceptions in heat and thermodynamics are shown in Table 10.

Table 10 Misconception profile and causes of misconception

Question Number	MC-PT		MC-PE		MC-T		MC-B		MC-I	
	N	N (%)	N	N (%)	N	N (%)	N	N (%)	N	N (%)
1	15	7.69	0	0.00	0	0.00	0	0.00	1	0.51
2	28	14.36	0	0.00	0	0.00	0	0.00	0	0.00
3	27	13.84	2	1.03	0	0.00	1	0.51	1	0.51
4	15	12.82	3	1.54	0	0.00	0	0.00	0	0.00
5	49	25.13	8	4.10	0	0.00	3	1.54	5	2.56
6	61	31.28	7	3.59	7	3.59	4	2.05	4	2.05
7	49	25.13	0	0.00	0	0.00	0	0.00	0	0.00
8	61	31.28	9	4.61	2	1.03	3	1.54	0	0.00
9	65	33.33	19	9.74	0	0.00	2	1.03	0	0.00
10	47	24.10	11	5.64	2	1.03	4	2.05	2	1.03
11	50	25.64	10	5.13	3	1.54	1	0.51	3	1.54
12	48	24.61	18	9.23	0	0.00	0	0.00	1	0.51
13	44	22.56	17	8.12	0	0.00	2	1.03	2	1.03
14	36	18.46	14	7.18	0	0.00	1	0.51	2	1.03
15	36	18.46	7	3.59	0	0.00	3	1.54	1	0.51
Average		21.91		4.23		0.48		0.82		0.75

Table 10 presents the recapitulation of misconception profiles and causes of misconceptions in heat and thermodynamics. Overall, it can be seen that the highest percentage of sources of misconception is caused by personal thinking (MC-PT), and the smallest percentage of sources of misconception is caused by the teacher (MC-T). This shows that in this study, the number of students who experienced misconceptions in heat and thermodynamics was mostly sourced from the students' thoughts. According to [Handhika et al. \(2016\)](#), student perceptions affect student conceptions, so misconceptions based on personal thoughts often occur.

Based on the identification of misconception profiles and causes of misconceptions in 195 students using the five-tier test instrument on heat and thermodynamics, several misconceptions were obtained. Most of these misconceptions have been identified by previous research, so this shows that students tend to experience misconceptions in the same content. One of them is on the concept of temperature change, where students assume that the ice cream stick will absorb heat from the ice cream so that the temperature of the stick increases. This misconception is in line with the findings in previous research. Whereas if two objects are in a room and contact at a certain time interval, they will have the same temperature because there is absorption and loss of heat in the same amount, and it does not depend on the material ([Yeo & Zadnik, 2001](#)).

Five sources of misconceptions were identified: personal thoughts, friends, teachers, books, and the internet. The cause of misconceptions originating from personal thinking (MC-PT) can occur when students have wrong thinking, reasoning, or intuition when interpreting a concept. The cause of misconceptions originating from friends (MC-PE) can occur due to explanations conveyed by others who are wrong about a concept. The cause

of teacher-sourced misconceptions (MC-T) can occur when the teacher has misconceptions about a concept or a teacher's lack of ability to convey a concept so that students will misunderstand the teacher's explanation. The cause of misconceptions originating from books (MC-B) can occur when students do not sufficiently understand the meaning of the book's contents or there are translation errors from the original book. Then, the cause of misconceptions sourced from the internet (MC-I) can occur due to the ease of access to find information, but the information is often not filtered for truth, causing misconceptions.

CONCLUSION

Based on the findings obtained from the research on the development of a five-tier test instrument to identify misconceptions in heat and thermodynamics, it can be concluded that the five-tier test instrument is able to identify students' misconceptions and find out the source of the causes of students' misconceptions in heat and thermodynamics. There are five categories of misconceptions consisting of the Sound Understanding (SU) category at 23.19%, Partial Positive (PP) category at 16.38%, Partial Negative (PN) category at 27.93%, Misconception (MC) category at 28.21%, No Understanding (NU) category at 5.85%. This shows that many students still experience misconceptions in heat and thermodynamics. The profile of misconceptions and causes of misconceptions of high school students in heat and thermodynamics is in the Misconception from Personal Thought (MC-PT) category of 21.91%, the Misconception from People Explanation (MC-PE) category of 4.23%, the Misconception from Teacher (MC-T) category of 0.48%, the Misconception from Book (MC-B) category of 0.82%, and the Misconception from Internet (MC-I) category of 0.75%. The limitation of this study is that researchers did not provide direct intervention through learning to students, so they cannot compare students' conceptions before and after learning about heat and thermodynamic material. To overcome this challenge, educators need to provide direct intervention to compare students' conceptions or misconceptions before and after learning. This is so that educators can find out changes in students' conceptions if there were misconceptions before learning. In addition, it is necessary to conduct interviews with students to support the results of identifying misconceptions and causes of misconceptions. Educators can consider the use of the five-tier test instrument as a learning evaluation tool to identify student misconceptions and the causes of student misconceptions.

CONFLICTS OF INTEREST

The author declares that there is no conflict of interest.

AUTHOR CONTRIBUTIONS STATEMENT

AD : Academic supervisor, reviewer of this research paper and provided guidance in this research, LA : provided guidance and supervision to refine this research project, AN : provide guidance and supervision in writing this research paper. AAN : Concept, wrote, methods, data collection, data analysis, and draft manuscript.

REFERENCES

- Aiken. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychology Measurement*, 131–142.
- Aviyanti, L. (2020). *An investigation into indonesian pre-service physics teachers' scientific thinking and conceptual understanding of physics*. (Doctoral dissertation, Flinders University, College of Education, Psychology and Social Work.).
- Aviyanti, L., Gani, A. W., Febriyanti, T., Ribie, S., & Nawas, A. (2024). Promoting higher-order thinking skills and learning motivation through the teams games tournaments learning model in physics education: A rasch model analysis. *Indonesian Journal Of*

- Science And Mathematics Education*, 07(3), 560–576. <https://doi.org/10.24042/ij sme.v5i1.23648>
- Darmaji, D., Agus Kurniawan, D., & Irdianti, I. (2019). Physics education students' science process skills. *International Journal of Evaluation and Research in Education (IJERE)*, 8(2), 293–298. <https://doi.org/10.11591/ijere.v8i2.28646>
- Didik, L. A., Wahyudi, M., & Kafrawi, M. (2020). Identifikasi miskonsepsi dan tingkat pemahaman mahasiswa tadris fisika pada materi listrik dinamis menggunakan 3-tier diagnostic test. *JNSI: Journal of Natural Science and Integration*, 3(2). <https://doi.org/10.24014/jnsi.v3i2.9911>
- Dina, T. F., Prahani, B. K., Marianus, M., Wibowo, F. C., & Sanjaya, L. A. (2024). Critical thinking skills student profile and PBL needs assisted by android physics module. *Momentum: Physics Education Journal*, 8(1), 11–22. <https://doi.org/10.21067/mpej.v8i1.9059>
- Fariyani, Q., & Rusilowati, A. (2015). Pengembangan four-tier diagnostic test untuk mengungkap miskonsepsi fisika siswa sma kelas x. *JISE*, 4(2).
- Fitriani-grum, A., & Astuti, B. (2017). Penerapan instrumen three-tier test untuk mengidentifikasi miskonsepsi siswa sma pada materi keseimbangan benda tagar. *Phenomenon: Jurnal Pendidikan MIPA*, 7(2), 88-98.
- Fitriah, L. (2017). Diagnosis miskonsepsi siswa pada materi kalor dengan menggunakan three-tier essay dan open-ended test items. *Berkala Ilmiah Pendidikan Fisika*, 5(2).
- Furoidah, A. (2017). *Identifikasi Miskonsepsi Siswa SMA Negeri 3 Jember Konsep Dinamika Rotasi Dengan Metode Four Tier*. Skripsi. Universitas Jember.
- Guerra-Reyes, F., Guerra-Dávila, E., Naranjo-Toro, M., Basantes-Andrade, A., & Guevara-Betancourt, S. (2024). Misconceptions in the Learning of Natural Sciences: A Systematic Review. *Education Sciences*, 14(5), 1–17. <https://doi.org/10.3390/educsci14050497>
- Hadijah, S., & Napitupulu, E. E. (2016). Pengaruh pembelajaran kooperatif tipe jigsaw terhadap kemampuan pemahaman konsep dan komunikasi matematik siswa smp negeri 4 percut sei tuan. *TABULARASA*, 13(3), 285–299. <https://doi.org/10.24114/jt.v13i3.4584>
- Hamdu, G., Hadiana, D., Sylvia, N., Apipatunnisa, I., & Yulianto, A. (2023). Measuring changes of students conceptual understanding of literacy and numeracy in natural science by using rasch model. *Jurnal Ilmiah Sekolah Dasar*, 7(3), 489–497. <https://doi.org/10.23887/jisd.v7i3.59943>
- Handayani, N. D. (2018). *Identifikasi miskonsepsi siswa menggunakan four-tier diagnostic test pada pokok bahasan hukum termodinamika di sma kabupaten bondowoso*. Skripsi. Universitas Jember.
- Handhika, J., Cari, C., Soeparmi, A., & Sunarno, W. (2016). Student conception and perception of newton's law. *AIP Conference Proceedings*, 1708, 1–5. <https://doi.org/10.1063/1.4941178>
- Hasniah, F., & Muchtar, Z. (2021). Pengembangan uji instrumen tiga tingkat dengan cri untuk mendeteksi miskonsepsi dalam pembelajaran reaksi redoks. *Jurnal Inovasi Pembelajaran Kimia*, 03(02), 123–135. <https://doi.org/10.24114/jipk.v3i2.26503>
- Hidayati, F. N., Akhsan, H., & Syuhendri. (2024). Identifikasi miskonsepsi siswa kelas x pada materi elastisitas dan hukum hooke di sma negeri 1 indralaya. *Jurnal Inovasi dan Pembelajaran Fisika*, 03(2), 1–9.
- Ifrianti, S., Istihana, I., Amaliyah, E., Komikesari, H., Sodikin, S., & Jamilah, S. (2019). The influence of concept attainment models on science processes skills. *Journal of Physics: Conference Series*, 1155(1). <https://doi.org/10.1088/1742-6596/1155/1/012035>
- Inggit, S. M., Liliawati, W., & Suryana, I. (2021). Identifikasi miskonsepsi dan penyebabnya menggunakan instrumen five-tier fluid static test (5tfst) pada peserta

- didik kelas xi sekolah menengah atas. *Journal of Teaching and Learning Physics*, 6(1), 49–68. <https://doi.org/10.15575/jotalp.v6i1.11016>
- Irsalina, F. R., Aviyanti, L., & Fredyan, R. (2020). Pendekatan model rasch untuk analisis konsep termal berbasis gender di sekolah menengah atas di jawa timur. *Momentum : Jurnal Pendidikan Fisika*, 1–22. <https://doi.org/10.21067/mpej.vvix.xxxx>
- Khairaty, N. I., Taiyeb, M., & Hartati. (2018). Identifikasi miskonsepsi siswa pada materi sistem peredaran darah menggunakan three-tier test di kelas xi ipa 1 sma negeri 1 bontonompo. *Jurnal Nalar Pendidikan*, 06(1), 7–13. <https://doi.org/10.26858/jnp.v6i1.6037>
- Khusaini., D. A. I. I., & Rizky Kurniawan, B. (2025). Effect of isomorphic problems with feedback to reduce student misconceptions on simple harmonic motion Article History. *Momentum: Physics Education Journal*, 9(1), 1–13. <https://doi.org/10.21067/mpej.v9i1.10048>
- Lidiana, H., Gunawan., & Taufik, M. (2018). Pengaruh model discovery learning berbantuan media phet terhadap hasil belajar Fisika peserta didik kelas XI SMAN 1 Kediri tahun ajaran 2017/2018. *Jurnal Pendidikan Fisika dan Teknologi*, 4 (1), 33-39. DOI: 10.29303/jpft.v4i1.519.
- Merlina, M. (2021). *Identifikasi miskonsepsi peserta didik menggunakan metode tes diagnostik five-tier pada materi termodinamika*.
- Mulyatiningsih, E. (2016). *Pengembangan model pembelajaran*. <http://staff.uny.ac.id/sites/default/files/pengabdian/dra-endang-mulyatiningsih-mpd/7cpengembangan-model-pembelajaran>
- Nasution, F. H., & Siregar, D. A. (2019). Validitas bahan ajar fisika siswa sma/ma berbasis concepttual chance. *Jurnal Education and Development*, 07(3), 75–78. <https://doi.org/10.37081/ed.v7i3.1169>
- Nasution, S. W. R. (2018). Penerapan model inkuiri terbimbing (guided inquiry) dalam meningkatkan kemampuan berpikir kritis pada pembelajaran fisika. *Jurnal Education and Development*, 3(1). <https://doi.org/10.37081/ed.v3i1.85>
- Nurhalisa, S., Ma'rufi, M., & Baharuddin, M. R. (2021). Pengembangan media pembelajaran berbasis asesmen kompetensi minimum dan pemecahan masalah. *Jurnal Literasi Digital*, 1(3), 192–202. <https://doi.org/10.54065/jld.1.3.2021.63>
- Nurhayati, D. (2012). *Profil miskonsepsi siswa pada konsep suhu dan kalor kelas xi sma*. Skripsi. Universitas Sebelas Maret.
- Nurhayati, S., Lukman, H., & Alsagaf, W. (2019). Pengembangan tes diagnostik three tier multiple choice untuk mengukur konsepsi fisika siswa sma. *Jurnal Pendidikan*, 04(02), 47–54. <https://doi.org/10.26740/jp.v4n2.p47-54>
- Ode, R. W., Takda, A., & Eso, R. (2022). Identifikasi miskonsepsi suhu dan kalor pada peserta didik kelas x ipa sma negeri 5 kendari menggunakan four tier test diagnostic. *Jurnal Penelitian Pendidikan Fisika*, 7. <https://doi.org/10.36709/jipfi.v7i1.23466>
- Pebriyanti, Dwi. , S. Hairunnisyah. , S. (2015). Efektivitas model pembelajaran perubahan konseptual untuk mengatasi miskonsepsi fisika pada siswa kelas x sman 1 praya barat tahun pelajaran 2012/2013. *Jurnal Pendidikan Fisika dan Teknologi*, 01(02), 92–96. <https://doi.org/10.29303/jpft.v1i2.241>
- Puspita, W. I., Sutopo, S., & Yuliati, L. (2019). Identifikasi penguasaan konsep fluida statis pada siswa. *Momentum: Physics Education Journal*, 53–57. <https://doi.org/10.21067/mpej.v3i1.3346>
- Puspitasari, M. D. M. (2018). Investigasi strategi pemecahan masalah dan penguasaan konsep pada materi kalor. *Jurnal Riset Pendidikan Fisika*, 3(2), 1–6. <https://doi.org/10.17977/um058v3i2p1-6>
- Rahayah Ariffin, S., Omar, B., Isa, A., & Sharif, S. (2010). Validity and reliability multiple intelligent item using rasch measurement mode. *Procedia - Social and Behavioral Sciences*, 9, 729–733. <https://doi.org/10.1016/j.sbspro.2010.12.225>

- Sappaile, B. I. (2007). Konsep instrumen penelitian pendidikan. *Jurnal Pendidikan dan Kebudayaan*, 066, 379–391. <https://doi.org/10.24832/jpnk.v13i66.356>
- Sari, R. M. (2021). *Identifikasi miskonsepsi siswa menggunakan tes diagnostik four-tier berbantuan google formulir pada konsep termodinamika di sma negeri 11 tangerang selatan*. (Bachelor's thesis, Jakarta: FITK UIN Syarif Hidayatullah Jakarta).
- Setiawan, D., Cahyono, E., & Kurniawan, C. (2017). Identifikasi dan analisis miskonsepsi pada materi ikatan kimia menggunakan instrumen tes diagnostik three-tier. *Journal of Innovative Science Education*, 06(2), 197–204. <https://doi.org/10.15294/jise.v6i2.15580>
- Sugiyono, S. (2019). *Statistika untuk penelitian*. Alfabeta Bandung.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan rasch pada assesment pendidikan*. Trim Komunikata.
- Supardi, S. (2013). Hasil belajar matematika siswa ditinjau dari interaksi tes formatif uraian dan kecerdasan emosional. *Jurnal Formatif*, 03(02), 78–96. <https://doi.org/10.30998/formatif.v3i2.115>
- Syahratinur, S., Ahmad. A. Z., & Kafrawi, M. (2023). Analisis tingkat pemahaman dan miskonsepsi fisika siswa pada materi suhu dan kalor menggunakan five tier diagnostic test di sman 1 brang rea. *CAHAYA: Journal of Research on Science Education*, 1(1). <https://doi.org/10.70115/cahaya.v1i1.35>
- Ünal, S., Coştu, B., & Ayas, A. (2010). Secondary school students' misconceptions of covalent bonding. *Journal of Turkish Science Education*, 07(2), 3–29. <http://www.tused.org>
- Utari, J. I., & Ermawati, F. U. (2018). Pengembangan instrumen tes diagnostik miskonsepsi berformat four-tier untuk materi suhu, kalor dan perpindahannya. *Inovasi Pendidikan Fisika*, 07(03), 434–439.
- Yaqin, M. K., Prastowo, S. H. B., & Harijanto, A. (2017). Identifikasi pemahaman konsep fisika terhadap pokok bahasan termodinamika pada siswa sma. *Seminar Nasional Pendidikan Fisika*, 2, 1–8.
- Yeo, S., & Zadnik, M. (2001). Introductory thermal concept evaluation: assessing students' understanding. *The Physics Teacher*, 39(8), 496–504. <https://doi.org/10.1119/1.1424603>
- Zaleha, Z., Samsudin, A., & Nugraha, M. G. (2017). Pengembangan instrumen tes diagnostik vcci bentuk four-tier test pada konsep getaran. *Jurnal Pendidikan Fisika Dan Keilmuan (JPfK)*, 3(1), 36. <https://doi.org/10.25273/jpfk.v3i1.980>

AUTHOR PROFILES



Aulia Ayuni Novita
auliaayuninovita24020@upi.edu is a student of the physics education study program at the University of Education Indonesia.
 Universitas Pendidikan Indonesia



Drs. Agus Danawan, M.Si.
ad@upi.edu
 Lecturer at the faculty of Mathematics and Natural Science, Department of Physics Education. Pengembangan Inovasi Pembelajaran Fisika
 Universitas Pendidikan Indonesia
<https://scholar.google.co.id/citations?user=ml7wVNcAAAAJ&hl=en&oi=ao>
<https://www.scopus.com/authid/detail.uri?authorId=57210925998>
<https://sinta.kemdikbud.go.id/authors/profile/6123641>



Lina Aviyanti, S.Pd., M.Si., Ph.D.

lina@upi.edu

Lecturer at the faculty of Mathematics and Natural Science, Department of Physics Education

Universitas Pendidikan Indonesia

<https://scholar.google.com/citations?user=GzsGTqsAAAAJ&hl=en>

<https://sinta.kemdikbud.go.id/authors/profile/6735786>

<https://www.scopus.com/authid/detail.uri?origin=resultslist&authorId=57193869963&zone=>



Dr. Abu Nawas

abu.nawas@adelaide.edu.au

Doctor of Philosophy at the University of Adelaide

The University of Adelaide, South Australia

<https://scholar.google.com/citations?user=fxLR4hcAAAAJ&hl=en>