ISSN : 2549-2764 (ONLINE)
ISSN : 2337-604X (PRINT)
BERKALA ILMIAH PENDIDIKAN FISIKA
https://ppjp.ulm.ac.id/journal/index.php/bipf

Construction of The Diagnostic Test Instrument Heat and Temperature Conception Four-Tier Test: An Exploratory Study

Rinta Destiani*, Taufik Ramlan Ramalis, and Ika Mustika Sari
Department of Physics Education Study Program, Universitas Pendidikan Indonesia
Bandung, Indonesia
*rinta.destiani@gmail.com

DOI:10.20527/bipf.v12i3.18655

Received: 28 January 2024 Accepted: 20 August 2024 Published: 16 October 2024

Abstract
Students regard physics as challenging due to the difficulty of integrating knowledge learned before and after the lesson, which causes misconceptions. One such misconception pertains to the subject of temperature and heat. The objective of this study is to develop a four-tier diagnostic test instrument that can be used to identify concepts regarding heat and temperature. This study employed Rasch analysis as a mixed method. The unidimensionality of the instrument is very good at 50.3%, with a person measure of -0.39 and an Alpha Cronbach of 0.82, which suggests a strong interaction between the person and the details of the question as a whole. The consistency of the responses from students is adequate, as indicated by a person reliability of 0.80 and an item reliability of 0.99. However, the items' quality and the instrument's reliability are exceptional. The Outfit MNSQ Person and Item of 0.99 and the Outfit ZSTD with a value of 0.02 suggest that the quality is improving. The instrument includes ten questions that can be employed to identify students' conceptions. The utilization of this instrument has the potential to yield more comprehensive measurement results regarding students' misconceptions.
Keywords: four-tier diagnostic test; misconceptions; reliability; validity

How to cite: Destiani, R., Ramalis, T. R., & Sari, I. M. (2024). Construction of the diagnostic test instrument heat and temperature conception four-tier test: An exploratory study. *Berkala Ilmiah Pendidikan Fisika, 12*(3), 383-396.

INTRODUCTION
Physics is a natural science with a collection of facts, concepts, and principles, and it is a learning process that provides students with experiences to understand the surrounding environment scientifically (Wea et al., 2021; Zafitri et al., 2018). In the learning process, students' understanding of physics is highly prioritized. Students will be able to understand and engage when they can connect pre-learning and post-learning knowledge. However, establishing these connections can be difficult due to various obstacles, including the unattractiveness of teaching methods and the learning difficulties of students (Wulandari et al., 2018). Physics is among the subjects that necessitate students to engage in deeper thinking to comprehend and acquire knowledge (Ady & Warliani, 2022). Therefore,

many students perceive physics as challenging (Ornek et al., 2008).

Students regard physics as a challenging subject due to the necessity of simultaneously presenting it in various formats, including formulas and calculations, graphs, conceptual explanations, and experiments (Angell et al., 2004). The environment influences the development of students' attitudes, knowledge, and skills in everyday life. When learning physics concepts, students construct newly learned knowledge with knowledge, attitudes, skills, and experiences brought from life (Kiray & Simsek, 2021). Misconceptions result from the inconsistency between students' conceptions and scientific concepts, which can be caused by the fact that each student has a unique knowledge construction of physics (Saputri & Kurniawan, 2021).

Misconceptions can occur because students' conceptions are based on their understanding, which hinders their ability to construct comprehensive concepts (Zafitri et al., 2018). Researchers conducted a diagnostic test with open-ended questions in one of the high schools in Bandung Regency and found that the highest misconception was in the concept of thermal expansion. Students believe that when two rods with different initial lengths but the same coefficient of thermal expansion and temperature change are placed, the length of the rods is inversely proportional to the initial length. Meanwhile, the scientific concept of the statement is that the change in length is directly proportional to the initial length.

The percentage of student misconceptions in the school is very high, with a percentage of 57.85%, according to the research conducted by Febrianti et al. (2019). The highest misconception among students is in the concept of thermal equilibrium, where students believe that when water is in thermal equilibrium and then poured into several containers, the temperature of the water will be evenly distributed. Furthermore, Nusryamsi et al. (2018) stated that the highest misconception in the concept of change of state is 47.03%. The low interest of students in physics subjects and the fact that students merely memorize concepts without comprehending them are the root causes of student misconceptions.

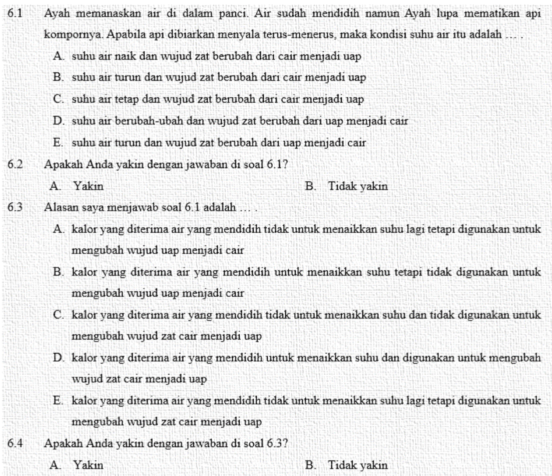
Various measurement tools have been developed to identify student misconceptions (Kiray & Simsek, 2021). Misconceptions can be identified through various methods, including concept maps, multiple-choice tests with open-ended questions, written essays, diagnostic interviews, classroom discussions, and question-and-answer laboratories. A diagnostic test is a type of multiple-choice test employed to identify misconceptions (Utari & Ermawati, 2018). The diagnostic test has been the most effective method for identifying student misconceptions (Putri & Ermawati, 2021).

A diagnostic test is used to determine students' weaknesses, difficulties, levels of achievement, and basic abilities to administer the most suitable treatments (Hidayati et al., 2013). Teachers implement diagnostic tests to update the learning process, while students utilize them to enhance their learning process (Zaleha et al., 2017). The four-tier diagnostic test is a diagnostic test that was developed from the three-tier diagnostic test by incorporating the level of students' confidence in selecting answers and providing reasons. This test is a valuable addition to one-tier to three-tier diagnostic tests (Gurel, 2015). The four-tier diagnostic test allows for a comprehensive diagnosis of student misconceptions by analyzing the answers and reasons provided by students. Furthermore, the confidence level of students' answers and reasons can provide a more comprehensive understanding of the depth of their conceptual comprehension. Teachers can identify concepts that require additional emphasis and assist in developing learning materials to mitigate misconceptions that may arise (Jumilah et al., 2022).

Andayani et al. (2019) conducted a field study in Bandung City, interviewing physics teachers, and discovered that Classical Test Theory (CTT) is still employed in test instruments that implement the analysis. The classical test theory has been extensively employed, but its group-dependent and item-dependent characteristics have resulted in its limitations. Consequently, modern test theories such as Item Response Theory (IRT) have been developed to overcome the constraints of classical theory by providing a more comprehensive understanding of the test's characteristics for each item. Sumintono and Widhiarso (2015) asserted in their book that the Rasch model and IRT 1PL are similar in emphasizing difficulty levels. However, due to their distinctions, the IRT and Rasch models are perceived as distinct measurement models. The main goal of the Rasch model is to create a measurement scale with the same interval. The Rasch model uses score data based on persons and items simultaneously. Both scores form the basis for obtaining pure scores to indicate the individual's ability level and the difficulty level of the questions. The purpose of this study is to determine the validity of the HTC4TT instrument in identifying student misconceptions regarding temperature and heat materials.

METHOD

This research is a mixed method using an exploratory sequential design consisting of two main phases: qualitative followed by quantitative (Creswell, & Guettermen. (2019). Qualitative data was obtained based on diagnostic tests with open-ended questions at one of the high schools in Bandung Regency. The test results were used to collect student response choices, which were used in developing the four-tier diagnostic test. After experts validated the four-tier diagnostic test instrument, the HTC4TT instrument was piloted with 65 students at one of the high schools in Bandung Regency. After the revision stage, the instrument was piloted extensively with 242 students (130 and 112 students). An example of test instrument development is shown in Figure 1.



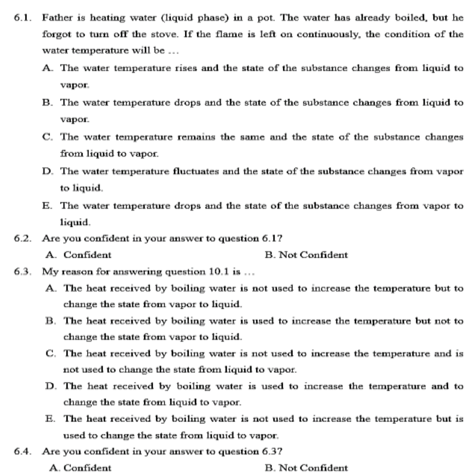
6.1 Ayah memanaskan air di dalam panci. Air sudah mendidih namun Ayah lupa mematikan api kompornya. Apabila api dibiarkan menyala terus-menerus, maka kondisi suhu air itu adalah … .
A. suhu air naik dan wujud zat berubah dari cair menjadi uap
B. suhu air turun dan wujud zat berubah dari cair menjadi uap
C. suhu air tetap dan wujud zat berubah dari cair menjadi uap
D. suhu air berubah-ubah dan wujud zat berubah dari uap menjadi cair
E. suhu air turun dan wujud zat berubah dari uap menjadi cair

6.2 Apakah Anda yakin dengan jawaban di soal 6.1?
A. Yakin B. Tidak yakin

6.3 Alasan saya menjawab soal 6.1 adalah … .
A. kalor yang diterima air yang mendidih tidak untuk menaikkan suhu lagi tetapi digunakan untuk mengubah wujud uap menjadi cair
B. kalor yang diterima air yang mendidih untuk menaikkan suhu tetapi tidak digunakan untuk mengubah wujud uap menjadi cair
C. kalor yang diterima air yang mendidih tidak untuk menaikkan suhu dan tidak digunakan untuk mengubah wujud zat cair menjadi uap
D. kalor yang diterima air yang mendidih untuk menaikkan suhu dan digunakan untuk mengubah wujud zat cair menjadi uap
E. kalor yang diterima air yang mendidih tidak untuk menaikkan suhu lagi tetapi digunakan untuk mengubah wujud zat cair menjadi uap

6.4 Apakah Anda yakin dengan jawaban di soal 6.3?
A. Yakin B. Tidak yakin

(a)



6.1. Father is heating water (liquid phase) in a pot. The water has already boiled, but he forgot to turn off the stove. If the flame is left on continuously, the condition of the water temperature will be …
A. The water temperature rises and the state of the substance changes from liquid to vapor.
B. The water temperature drops and the state of the substance changes from liquid to vapor.
C. The water temperature remains the same and the state of the substance changes from liquid to vapor.
D. The water temperature fluctuates and the state of the substance changes from vapor to liquid.
E. The water temperature drops and the state of the substance changes from vapor to liquid.

6.2. Are you confident in your answer to question 6.1?
A. Confident B. Not Confident

6.3. My reason for answering question 10.1 is …
A. The heat received by boiling water is not used to increase the temperature but to change the state from vapor to liquid.
B. The heat received by boiling water is used to increase the temperature but not to change the state from vapor to liquid.
C. The heat received by boiling water is not used to increase the temperature and is not used to change the state from liquid to vapor.
D. The heat received by boiling water is used to increase the temperature and to change the state from liquid to vapor.
E. The heat received by boiling water is not used to increase the temperature but is used to change the state from liquid to vapor.

6.4. Are you confident in your answer to question 6.3?
A. Confident B. Not Confident

(b)

Figure 1 (a) Example of test instrument development and (b) Example of test instrument development in english

After data cleaning, outlier and misfit data were removed, resulting in 172 students (95 from Bandung Regency and 77 from Bandung City). Validity, reliability, and instrument difficulty level analyses were conducted using Rasch analysis. The developed instrument is illustrated in Figure 1.

The data analysis in the research utilized the Rasch model with the rating scale model, an extension of the dichotomous model. Scoring for all items consists of more than two categories. Each item (k) threshold has a specific estimated difficulty (F), which predicts a 50:50 chance of choosing one type over the other. The criteria that must be met to fulfil the requirements for good unidimensionality (Sumintono & Widhiarso, 2015) are shown in Table 1.

Table 1 Unidimensionality test requirements

Parameter	**Accepted Value**
Raw Variance explained by measures	>20%
Unexplained Variance in 1st contrast	<15%
Unexplained Variance in 2st contrast	<15%
Unexplained Variance in 1st contrast Eigen Value	<3
Unexplained Variance in 2st contrast Eigen Value	<3

Table 2 shows the reliability criteria using Cronbach's Alpha.

Table 2 Reliability criteria using cronbach's alpha

Value	**Reliability Criteria**
$< 0{,}5$	Poor
$0{,}5 - 0{,}6$	Bad
$0{,}6 - 0{,}7$	Fair
$0{,}7 - 0{,}8$	Good
$> 0{,}8$	Very Good

(Sumintono & Widhiarso, 2015)

Table 3 shows the person and item reliability criteria.

Table 3 Person and item reliability criteria

Value	**Criteria**
$< 0{,}67$	Poor
$0{,}67 - 0{,}80$	Bad
$0{,}80 - 0{,}90$	Fair
$0{,}91 - 0{,}94$	Good
$> 0{,}94$	Very Good

(Sumintono & Widhiarso, 2015)

The four-tier test instrument's validity is obtained using Outfit MNSQ and Outfit ZSTD data, as shown in Table 4.

Table 4 Outfit MNSQ and Outfit ZSTD data

Criteria	**Accepted Value**
Outfit Mean Square (*Outfit* MNSQ)	$0{,}5 < Outfit\ \text{MNSQ} < 1{,}5$
Outfit Z Standard (*Outfit* ZSTD)	$-2{,}0 < Outfit\ \text{ZSTD} < 2{,}0$
Pt Measure Corr	$0{,}4 < Pt\ Measure\ Corr < 0{,}85$

(Sumintono & Widhiarso, 2015)

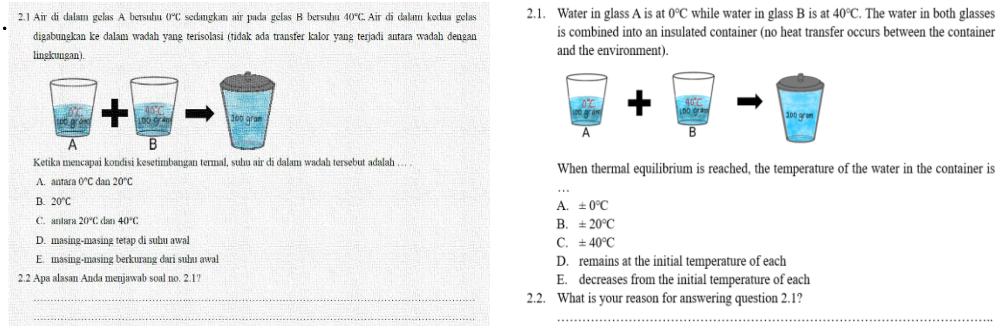
RESULT AND DISCUSSION

Interviews with high school physics teachers in Bandung City, Bandung Regency, Tasikmalaya City, and Cianjur Regency showed that students' comprehension of physics varies in each school. However, most students struggle to comprehend the material because they only remember it, rather than comprehend it. Students encounter numerous abstract physics concepts during the physics learning process, making connecting with related concepts challenging. For instance, experimental results must be represented through graphs, conceptual explanations, and mathematical calculations. Some physics materials that are regarded as challenging include heat and temperature, thermodynamics, mechanics, and gas kinetic theory. Based on the research of Febrianti, Akhsan & Muslim (2019), Nursyamsi, Sujiono & Yani (2018), Utari & Liliawati (2019), Ni'mah, Kusairi & Supriana (2019), Sartika, Mansyur & Kendek (2021), the findings of frequent misconceptions are shown in Table 4.

Table 4 Misconception findings on heat and temperature materials based on literature study

Subject Matter	**Misconceptions**
Temperature	When poured into several glasses with the same volume of water The temperature of the water in one container will be divided equally. When a number of waters in a container with different temperatures are combined in one container, the temperature does not change.
Expansion	The length of a solid is inversely proportional to the change in temperature. The change in length of the two sticks is the same because they are made of the same material, even though their initial lengths are different. Hot water poured into a glass expands evenly to all surfaces so that the glass can break. Heat is in the form of particles. In the expansion process, the length of the heat particles pushes the metal particles so that they expand. Water with temperatures below 0°C is usually solid and returns to liquid when the temperature is above 0°C.
The Effect of Heat on The Form of Substances	Water that reaches 100°C will shrink in volume. 0°C 4°C Water increases in volume between 0 °C and 4 °C. Above 4°C, the volume decreases as the temperature increases. When boiling, the temperature of the water will continue to increase if the fire continues to burn. Water turns into gas (evaporation) because water releases heat.
The Effect of Heat on The Temperature of Substances	Heat capacity is directly proportional to temperature. The same mass of two objects causes both objects to experience the same temperature increase even though their specific heat is different. Water is a better heat absorber than land, making it easier for water temperatures to rise.
Heat Transfer	The heating of a metal rod includes convection. Seawater is a better heat sink than land. Heat transfer from the coffee to the spoon is an example of radiation. The thickness of the object affects how much heat is absorbed.

Following a preliminary study, the researcher developed a two-tier diagnostic test instrument consisting of open-ended questions with 18 items. The instrument was based on the outcomes of the analysis and literature review. It consisted of open-ended queries and was implemented to ascertain the reasons behind students' responses to the questions. Figure 2 illustrates the test instrument in the form of open-ended questions. The researcher obtained the students' reasons for responding to questions on the open-ended question test instrument by analyzing the test results. The reasons for the answers provided by students as a whole demonstrate that students need help to explain scientific concepts, and a significant number of students have an incomplete and in-depth understanding of heat and temperature material. The reasons provided by students are used by researchers to develop four-tier diagnostic test instruments.



2.1 Air di dalam gelas A bersuhu 0°C sedangkan air pada gelas B bersuhu 40°C. Air di dalam kedua gelas digabungkan ke dalam wadah yang terisolasi (tidak ada transfer kalor yang terjadi antara wadah dengan lingkungan).

Ketika mencapai kondisi kesetimbangan termal, suhu air di dalam wadah tersebut adalah … .

A. antara 0°C dan 20°C
B. 20°C
C. antara 20°C dan 40°C
D. masing-masing tetap di suhu awal
E. masing-masing berkurang dari suhu awal

2.2 Apa alasan Anda menjawab soal no. 2.1?

……………………………………………………………………………………

……………………………………………………………………………………

2.1. Water in glass A is at 0°C while water in glass B is at 40°C. The water in both glasses is combined into an insulated container (no heat transfer occurs between the container and the environment).

When thermal equilibrium is reached, the temperature of the water in the container is …

A. ± 0°C
B. ± 20°C
C. ± 40°C
D. remains at the initial temperature of each
E. decreases from the initial temperature of each

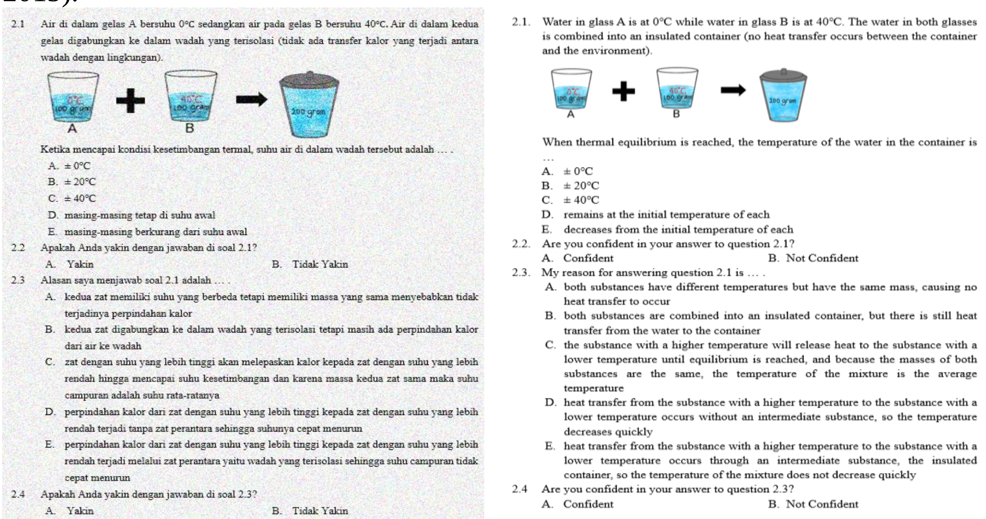
2.2. What is your reason for answering question 2.1?

……………………………………………………………………………………

Figure 2 (a) Test instruments with open-ended question and (b) Test instruments with open-ended questions in english

The rules for writing question items, which include substance/matter, construction, and language, are adjusted again by researchers (Directorate General of Primary and Secondary Education Ministry of Education and Culture, 2015).

The two-tier diagnostic test, which consisted of open-ended questions, was subsequently developed into a four-tier diagnostic test, which was multiple-choice, as shown in Figure 3.



2.1 Air di dalam gelas A bersuhu 0°C sedangkan air pada gelas B bersuhu 40°C. Air di dalam kedua gelas digabungkan ke dalam wadah yang terisolasi (tidak ada transfer kalor yang terjadi antara wadah dengan lingkungan).

Ketika mencapai kondisi kesetimbangan termal, suhu air di dalam wadah tersebut adalah … .

A. ± 0°C
B. ± 20°C
C. ± 40°C
D. masing-masing tetap di suhu awal
E. masing-masing berkurang dari suhu awal

2.2 Apakah Anda yakin dengan jawaban di soal 2.1?

A. Yakin   B. Tidak Yakin

2.3 Alasan saya menjawab soal 2.1 adalah … .

A. kedua zat memiliki suhu yang berbeda tetapi memiliki massa yang sama menyebabkan tidak terjadinya perpindahan kalor
B. kedua zat digabungkan ke dalam wadah yang terisolasi tetapi masih ada perpindahan kalor dari air ke wadah
C. zat dengan suhu yang lebih tinggi akan melepaskan kalor kepada zat dengan suhu yang lebih rendah hingga mencapai suhu kesetimbangan dan karena massa kedua zat sama maka suhu campuran adalah suhu rata-ratanya
D. perpindahan kalor dari zat dengan suhu yang lebih tinggi kepada zat dengan suhu yang lebih rendah terjadi tanpa zat perantara sehingga suhunya cepat menurun
E. perpindahan kalor dari zat dengan suhu yang lebih tinggi kepada zat dengan suhu yang lebih rendah terjadi melalui zat perantara yaitu wadah yang terisolasi sehingga suhu campuran tidak cepat menurun

2.4 Apakah Anda yakin dengan jawaban di soal 2.3?

A. Yakin   B. Tidak Yakin

2.1. Water in glass A is at 0°C while water in glass B is at 40°C. The water in both glasses is combined into an insulated container (no heat transfer occurs between the container and the environment).

When thermal equilibrium is reached, the temperature of the water in the container is …

A. ± 0°C
B. ± 20°C
C. ± 40°C
D. remains at the initial temperature of each
E. decreases from the initial temperature of each

2.2. Are you confident in your answer to question 2.1?

A. Confident   B. Not Confident

2.3. My reason for answering question 2.1 is … .

A. both substances have different temperatures but have the same mass, causing no heat transfer to occur
B. both substances are combined into an insulated container, but there is still heat transfer from the water to the container
C. the substance with a higher temperature will release heat to the substance with a lower temperature until equilibrium is reached, and because the masses of both substances are the same, the temperature of the mixture is the average temperature
D. heat transfer from the substance with a higher temperature to the substance with a lower temperature occurs without an intermediate substance, so the temperature decreases quickly
E. heat transfer from the substance with a higher temperature to the substance with a lower temperature occurs through an intermediate substance, the insulated container, so the temperature of the mixture does not decrease quickly

2.4 Are you confident in your answer to question 2.3?

A. Confident   B. Not Confident

Figure 3 (a) Form of four-tier diagnostic test questions and (b) Form of four-tier diagnostic test questions in english

Figure 3 is a form of four-tier diagnostic test questions in English. After the design stage, experts validated to get good instrument results. Question items are valid if they meet at least two criteria: MNSQ Outfit, ZSTD Outfit, and Pt. Measure Corr, namely (Sumintono & Widhiarso, 2015).

Table 5 shows question items with question codes S01, S02, S04, S05, S06, S09. S10. S12, S13, S14, S15, and S17 are valid because they meet at least two validity criteria. Meanwhile, items with the codes S03, S07, S08, S11, S16, and S17 are invalid because they only meet 1 criterion.

The items were revised based on expert advice after the expert validation, and invalid items were not used in the study. Consequently, the number of test instrument items totalled 12 items. The instrument was tested on 65 students.

Table 5 Question item feasibility based on expert validity

Code	*Outfit* MNSQ	*Outfit* ZSTD	*Pt. Measure Corr*	Desc.	Code	*Outfit* MNSQ	*Outfit* ZSTD	*Pt. Measure Corr*	Desc.
S01	0.40	-0.9	0.67	Valid	S10	0.57	-1.0	0.77	Valid
S02	0.64	0.4	0.16	Valid	S11	5.18	1.7	0.01	Invalid
S03	1.82	0.9	0.10	Invalid	S12	0.95	0.6	0.13	Valid
S04	0.79	0.8	0.10	Valid	S13	0.74	-0.2	0.65	Valid
S05	0.51	-0.3	0.43	Valid	S14	0.59	0.0	0.34	Valid
S06	0.90	0.0	0.73	Valid	S15	0.38	-1.2	0.74	Valid
S07	4.92	1.8	0,07	Invalid	S16	2.33	1.2	0.26	Invalid
S08	3.86	1.5	0.06	Invalid	S17	1.99	1.3	0.26	Invalid
S09	0.59	0.0	0.34	Valid	S18	0.65	-0.4	0.69	Valid

Items with question codes S02, S03, S04, S05, S06, S07, S08, S09. S10, S11, and S12 are valid because they meet at least two validity criteria. Meanwhile, the item with the code S01 is invalid because it only meets 1 criterion, so the final instrument of the *four-tier* diagnostic test totals 11 items, as shown in Table 6. An item with question code S01 is invalid because it does not meet two criteria. The ZSTD Outfit value is greater than 2.0, and the Pt. Measure Corr. value is less than 0.4.

Table 6 Feasibility of question items based on trial

Code	Logit	SEM	MNSQ Outfit	ZSTD Outfit	Pt. measure Corr.
S01	-0.33	0.08	1.46	2.22	0.39
S02	0.04	0.09	1.07	0.35	0.36
S03	0.43	0.12	1.01	0.15	0.27
S04	0.04	0.09	0.90	-0.33	0.49
S05	-0.07	0.09	1.22	0.96	0.40
S06	0.38	0.11	1.01	0.13	0.40
S07	-0.29	0.08	0.74	-1.43	0.54
S08	0.11	0.09	0.84	-0.53	0.44
S09	-0.13	0.08	0.79	-0.95	0.60
S10	0.26	0.10	1.02	0.17	0.34
S11	-0.53	0.08	1.05	0.33	0.56
S12	0.09	0.09	0.75	-0.93	0.46

Feasibility of *Heat and Temperature Conception Four-Tier Test* Instrument

Unidimensionality

Unidimensionality testing optimizes measurement, ensuring that the information produced is focused on the measured attributes. The unidimensionality test is a measure to evaluate the ability of existing items to measure the range of respondents' abilities and evaluate the ability to measure one variable comprehensively (Sumintono & Widhiarso, 2015). The *Heat and Temperature Conception Four-Tier Test* (HTC4TT) instrument used in the study has an adequate measure of unidimensionality. The analysis results are presented in Table 7.

Table 7 Unidimensionality test

Parameters	Rows
Raw Variances explained by measures	51.0%
Unexplained Variance in 1st contrast	8.0%
Unexplained Variance in 2st contrast	6.7%
Unexplained Variance in 1st contrast Eigen Value	1.7578%
Unexplained Variance in 2st contrast Eigen Value	1.4706%

The raw variance measurement of the data is 51.0%, indicating that the unidimensionality requirements are very good. Another value is raw unexplained variance, which is the variance that the instrument is unable to account for (Laliyo, 2021). The data shows that the unexplained variance value is less than 15% and the eigenvalue is less than 3, meaning that the unidimensionality requirement is good. This implies that the instrument can accurately measure and analyze students' misconceptions regarding heat and temperature.

Rating Scale Testing

Rating scale testing assesses students' comprehension of the rating scale. In accordance with Table 8, the observed average begins at logit -1.26 for rank 1 (do not understand the concept), which increases to logit -0.85 for rank 2 (guessing understanding), logit -0.28 for rank 3 (misconception), logit 0.17 for rank 4 (partial concept understanding), and logit 0.95 for rank 5 (concept understanding).

Table 8 Rating scale test

Rating	OBSVD AVRGE
1	-1.26
2	-0.85
3	-0.28
4	0.17
5	0.95

Person and Item Reliability

Cronbach's alpha value can explain reliability but only shows the overall interaction between person and item. Without determining how the measurements of person and item are separate. Rasch modelling offers two instrument reliability estimates: person and item reliability. These estimates can provide a comprehensive understanding of the quality of student response patterns, the instrument, and the interaction between the individual and the item (Sumintono & Widhiarso, 2015).

Table 9 Reliability of person and item

Category	Parameter	
	Person	Item
Rata-rata Logit (SD)	-0.39 (0.81)	0.00 (0.69)
Separation	1.99	8.21
Reliability	0.80	0.99
Alpha Cronbach	0.82	
Outfit MNSQ	0.99	0.99
Outfit ZSTD	0.02	0.02

Based on Table 9, the Person measure is -0.39 logit, which shows the average value of all students working on the given items. The average value is less than 0.00, indicating a smaller tendency of student ability than the difficulty level of the question. Cronbach's alpha of 0.82 was used to measure reliability, which is the interaction between the person and the items. This value indicates a very good interaction between the person and the items. Person reliability of 0.80 and item reliability of 0.99 indicates that the consistency of answers from students is moderate, and the quality of the items in the reliability aspect of the instrument is excellent. Outfit MNSQ Person and Item of 0.99, the ideal value is 1.00, indicating that the quality is improving. Outfit ZSTD person and item is 0.02 while the ideal value is 0.00, indicating the quality is improving. In general, this instrument is capable of measuring reliably and consistently, thereby enabling it to be employed to represent the identification of student misconceptions.

Every measurement always produces information about the measurement results, which depends on the relationship between the test and the student being measured. Measurement information is presented in Figure 7. The x-axis shows the level of students' ability to do the given diagnostic test. At very low ability levels, the information from the measurement is quite low. At a very high ability level, the information from the measurement is also quite low. At the medium ability level, the information from the measurement is very high. This shows that the diagnostic test produces optimal information when given to students with moderate ability.

Person Separation Index

Person separation index means how well the HTC4TT instrument can analyze student misconceptions. The greater the value of the person separation index means the greater the likelihood of students responding to items with their correct understanding because it can identify respondent groups and item groups (Sumintono & Widhiarso, 2015). In the study, the person separation index (1.99) and item separation index (8.21) mean a fairly good distribution of HTC4TT instruments on students and items. This supports the HTC4TT instrument as a suitable and reliable instrument for analyzing student misconceptions.

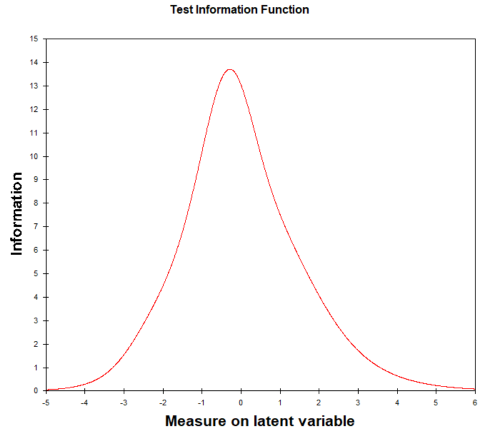


Figure 4 Measurement information function

Validity

Validity is used to test whether the HTC4TT instrument measures what it is intended to measure. By guaranteeing that all HTC4TT items satisfy the Rasch model (item fit), construct validity can be achieved. Item fit explains whether or not the items function normally (Sumintono & Widhiarso, 2015). The Mean-Square Fit Statistic (MNSQ) value shows a measure of randomness or indicates the impact of the item's misfit (not fitting the Rasch model). Items that are too predictable are indicated by MNSQ values that are less than those between 0.5 and 1.5, while larger values indicate items that are not predictable. The Standardized Fit Statistic (ZSTD) value describes the improbability of the data, i.e., its significance if the data fits the model. Outfit, or outlier-sensitive fit, measures the sensitivity of response patterns to items with a certain level of difficulty rather than respondents (person) or vice versa (Sumintono & Widhiarso, 2015). The Point Measure Correlation value is the correlation between the person item score and the person measure. The value must be positive and not near zero (Bond & Fox, 2015).

Based on Table 10, question items with codes S1/KT, S3/P, S4/P, S5/PKW, S6/PKW, S7/PKS, S8/PKS, S9/PKS, S10/PK, and S11/PK S11 are said to be valid because they meet at least two validity criteria. Meanwhile, the item with the code S2/P is invalid because it only meets 1 criterion, so the final instrument of the four-tier diagnostic test includes ten items.

Table 10 Item statistics: Misfit order

Code	Logit	MNSQ Outfit	ZSTD Outfit	Pt. Measure Core	Code	Logit	MNSQ Outfit	ZSTD Outfit	Pt. Measure Core
S1/KT	-1.69	1.15	1.21	0.61	S7/PKS	0.43	0.63	-2.24	0.65
S2/P	0.17	1.57	3.10	0.51	S8/PKS	0.43	0.84	-0.86	0.60
S3/P	0.43	0.78	-1.19	0.54	S9/PKS	0.34	1.08	0.52	0.61
S4/P	-0.35	0.88	-0.91	0.67	S10/PK	-0.98	1.30	2.35	0.56
S5/PKW	0.56	0.98	-0.04	0.58	S11/PK	0.60	0.59	-2.15	0.70
S6/PKW	0.05	1.06	0.42	0.47					

The subsequent step is to examine the Wright Map to generate a map demonstrating the distribution of item difficulty levels and individual abilities in Figure 5.

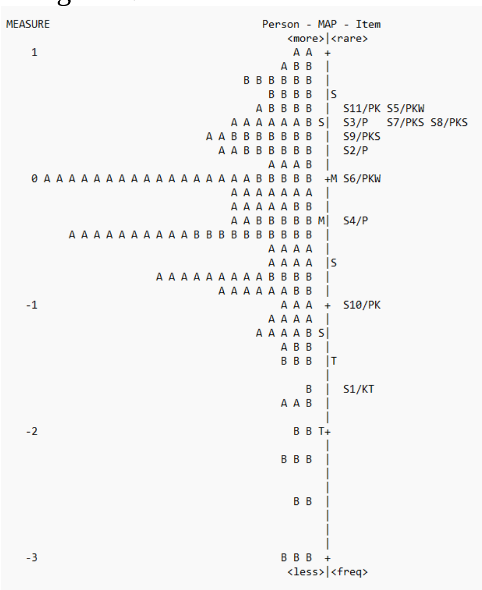


Figure 5 Wright map person (N=172) and item (N=11)

The Wright map demonstrates that the 11 items exhibit a broad range of difficulty levels, from S11/PK, which is the most challenging, to S1/KT, which is the easiest. This demonstrates that the items provided can offer valuable information of the students' abilities. However, the items in the upper right corner are challenging to complete, while the items in the lower right corner are easy to work on.

A comparison of the distance between M-S-T (mean, 1SD, 2SD) reveals that the distribution distance for student ability is greater than the distribution distance for item difficulty. This suggests that the diversity of item difficulty could be more extensive, whereas the diversity of student ability is quite extensive. The logit person is greater (1.03) than the logit item (0.63) when comparing the average logit item and logit person. This suggests that students' overall ability is slightly higher than the difficulty level of the question.

Table 11 *Item statistics*: Based on problem difficulty level

Code	**Logit**	**MNSQ Outfit**	**ZSTD Outfit**	**Pt. Measure Core**	**Code**	**Logit**	**MNSQ Outfit**	**ZSTD Outfit**	**Pt. Measure Core**
S11/PK	0.60	0.59	-2.15	0.70	S9/PKS	0.34	1.08	0.52	0.61
S5/PKW	0.56	0.98	-0.03	0.58	S6/PKW	0.05	0.63	-2.24	0.65
S3/P	0.43	0.78	-1.19	0.54	S4/P	-0.35	0.88	-0.89	0.67
S7/PKS	0.43	0.88	-0.91	0.67	S10/PK	-0.98	1.30	2.35	0.56
S8/PKS	0.43	0.84	-0.86	0.60	S1/KT	-1.69	1.15	1.21	0.61

Based on Figure 8 and Table 11, the six items with a high difficulty level include S11/PK, S5/PKW, S3/P, S7/PKS, S8/PKS, and S9/PKS. To find out the level of difficulty of the questions described by Table 11, the items were sorted from the highest to lowest difficulty levels. Problem S11/PK is 0.61 *logits*, problem S5/PKW is 0.56 *logits*, problem S3/P and problem S7/PKS have the same difficulty level of 0.43 *logits*, problem S8/PKS is 0.42 *logits*, and problem S9/PKS is 0.33 *logits*. Problem S2/P is a problem that does not fit, so it cannot be used. This is because students cannot understand the problems given properly (problems indicated misconceptions). Problem S6/PKW (0.05 *logit*) is a question with a medium difficulty level. Problems S4/P, S10/PK, and S1/KT are problems with low difficulty. Problem S4/P is -0.35 *logit*, S10/PK is -0.98 *logit* and S1/KT is -1.69 logit.

Disparity in Students' Misconception on Heat and Temperature Materials Based on School Location

Following the identification of students' misconceptions in general, an analysis of the state of items in Table 12 that were biased toward school location (i.e., in the first column of this polytomy file) was conducted.

Table 12 Unidimensionality test

Code	Prob.
S1/KT	0,35
S2/P	0,39
S3/P	0,02
S4/P	0,13
S5/PKW	0,02
S6/PKW	1,00
S7/PKS	0,41
S8/PKS	0,80
S9/PKS	1,00
S10/PK	0,00
S11/PK	0,10

Items that do not contain DIF have a probability above 5%. Based on Table 12, the items coded S3/P, S5/PKW, and S10/PK contain DIF. To find further analysis, the disparity in students' concept understanding of heat and temperature material is shown based on school location with Differential Item Functioning (DIF), as shown in Figure 6.

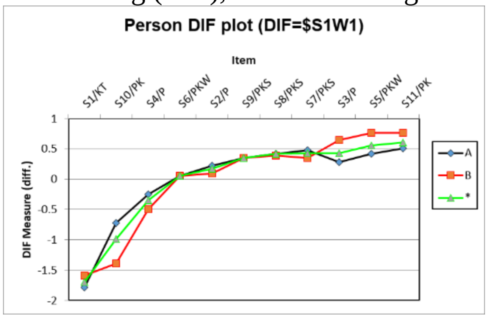


Figure 6 DIF plot of students based on school location

There are three curves based on the student's school location, namely A (Bandung Regency), B (Bandung City), and the * (star) indicates the average. The graph illustrates that the three curves near the upper limit are items with high difficulty (S11/PK, S5/PKW, and S3/P), while the three curves below are items with low difficulty (S1/KT, S10/PK, and S4/P). Items coded S3/P, S5/PKW, and S10/PK contain bias. Based on the graph, questions S3/P (expansion in a glass poured with hot water) and S5/PKW (the impact of the water anomaly), which have a high level of difficulty, are relatively more difficult for students in Bandung City. Meanwhile, question S10/PK (heat transfer), which has a low difficulty level, is more difficult for students in Bandung Regency.

The results of interviews with teachers in several schools conducted by researchers explained that one of the student's difficulties in understanding physics is due to many abstract physics concepts. This is in accordance with the research of Jannah & Rahmi (2020), which shows that students often have misconceptions about abstract material in physics learning because physics is closely related to nature and abstract concepts that are widely used in other related disciplines. The reasons for the answers given by students when conducting open-ended tests as a whole illustrate that many students need help understanding heat and temperature material, cannot explain concepts scientifically, and need completeete understanding. This is consistent with Fitriah's research (2017), which indicates that students may need clarification due to their limited capacity to remember concepts, comprehend material, and apply those concepts to new problems when studying heat and temperature.

The results of expert validation showed that 12 out of 18 items were declared valid. After testing the students, 11 out of 12 questions were declared valid. 2 criteria between MNSQ Outfit, ZSTD Outfit and Pt. Measure Corr, namely (Sumintono & Widhiarso, 2015). After testing the questions, the next instrument distribution was carried out. The Heat and Temperature Conception Four-Tier Test (HTC4TT) diagnostic test instrument is said to be effective in making measurements because of the very good unidimensionality test results of 51.0%. The minimum value of the unidimensionality test is 20% for dichotomous data and 40% for polytomous data (Sumintono &

Widhiarso, 2015). The results of the rating scale test, from rank 1 to rank 5, showed that the average value increases, meaning that the transition of student understanding rating categories occurs consistently (Linacre, 2022).

The Alpha Cronbach value of 0.82 indicates a very good interaction between the person and the items. Alpha Cronbach is used to measure reliability, namely the interaction between a person and an item (Sumintono & Widhiarso, 2015). In addition to using Cronbach's Alpha value, the value shown by person and item reliability indicates that the consistency of answers from students is sufficient, and the quality of the items in the HTC4TT instrument in the reliability aspect is excellent. Outfit MNSQ person and item values show that the instrument's quality is improving. Outfit ZSTD person and item, which is close to the ideal value of 0.00, indicates that the quality is improving. This instrument can generally be measured reliably and consistently, enabling it to identify student misconceptions (Sumintono & Widhiarso, 2015).

CONCLUSION

The study results can be concluded that the Heat and Temperature Conception Four-Tier Test (HTC4TT) instrument can be measured consistently and reliably to identify students' conceptions. The results of the validity test analysis by expert validators showed that 12 items were feasible to use. Furthermore, the instrument was tested on 65 students, showing that as many as 11 of the 12 items were said to be feasible. The last stage, namely distributing the instrument to 242 students, showed that 10 of the 11 items were feasible. Valid question items meet the outfit means-square, outfit z-standard, and point measure correlation criteria.

The results of this study serve as suggestions and solutions for teachers and researchers in physics education. The results indicate that it is necessary to identify students' misconceptions about heat and temperature using the Heat and Temperature Conception Four-Tier Test in formal teaching to optimize learning in areas where misconceptions have been identified. Answer choices for confidence at tier-1 and tier-3 do not adequately reflect students' confidence, so it is recommended to use the percentage of confidence levels. Additionally, future researchers may conduct studies to determine teaching strategies for heat and temperature to reduce students' misconceptions about these topics.

REFERENCES

Ady, W. N. & Warliani, R. (2022). Analisis kesulitan belajar siswa sma terhadap mata pelajaran fisika pada materi gerak lurus beraturan. *Jurnal Pendidikan Islam dan Keguruan, 2*(1), 104-108.

Andayani, A., Purwanto. & Ramalis, T. R. (2019). Kajian implementasi teori respon butir dalam menganalisis instrumen tes materi fisika. *Prosiding Seminar Nasional Fisika*. 37-42.

Angell, C., Guttersrud, Ø., Henriksen, E. K. & Isnes, A. (2004). Physics: Frightful, but fun, Pupils' and teachers' views of physics and physics teaching [Electronic version]. *Science Education*, 88, 683-706.

Bond, T.G., & Fox, C.M. (2015). *Applying the rasch model: Fundamental measurent in the human sciences (3rd ed.).* Routledge Taylor & Francis Group. https://doi.org/10.1088/1751-8113/44/8/085201

Creswell, J. W. & Guettermen, J.C. (2019). *Educational research planning, conducting, and evaluating quantitative and qualitative research (6 ed.).* New York: Pearson.

Febrianti, J., Akhsan, H., & Muslim, M. (2019). Analisis miskonsepsi suhu dan kalor pada siswa SMA Negeri 3

Tanjung Raja. *Jurnal Inovasi dan Pembelajaran Fisika (JIPF)*, 6(1), 90-102.

Fitriah, L. (2017). Diagnosis miskonsepsi siswa pada materi kalor dengan menggunakan three-tier essay dan open-ended test items. *Berkah Ilmiah Pendidikan Fisika*, 5(2), 168-181.

Gurel, D. E. (2015). A Review and comparison of diagnostic instruments to identify students'misconception in science. *Eurasia Journal of Mathematics, Science & Technology Education*, 11(5), 989-1008.

Hidayati, T. Nugroho, S. E. & Sudarmin, S. (2013). Pengembangan tes diagnostik untuk mengidentifikasi keterampilan proses sains dengan tema energi pada pembelajaran ipa terpadu. *Unnes Science Education Journal*, 2(2), 311-319.

Jannag, R. & Rahmi, I. (2020). Pengembangan e-diagnostic four-tier test untuk mengidentifikasi miskonsepsi peserta didik. *Natural Science: Jurnal Penelitian Bidang IPA dan Pendidikan IPA,* 6(2), 151-160.

Jumilah, J., Lestari, E.P., & Wasis, W. (2022). Introduksi miskonsepsi dan penyebab miskonsepsi peserta didik pada sub-materi asas bernoulli memakai four-tier diagnostic test. *Jurnal Literasi Pendidikan Fisika*, 3(1), 20-27.

Kiray, S. A. & Simsek, S. (2021). Determination and evaluation of the science teacher candidates'misconceptions about density by using four-tier diagnostic test. *International Journal of Science and Mathematics Education*, 19, 935-955. https://doi.org/10.1007/s10763-020-10087-5

Laliyo, L.A.R. (2021). *Mendiagnosis sifat perubahan konseptual siswa: Penerapan teknik analisis stacking dan racking rasch model.* Yogyakarta: Deepublish.

Linacre, J.M. (2022). Winstep (Version 5.4.3) [Computer Software]. Portland, Oregon: Winsteps.com.

Ornek, F., Robinson, W. R. & Haugan, M. P. (2008). What makes physics difficult?. *International Journal of Environmental & Science Education*, 3(1), 30-34.

Putri, W. K. & Ermawati, F. U. (2021). Pengembangan, uji validitas dan reliabilitas tes diagnostik four-tier untuk materi getaran harmonis sederhana beserta hasil uji coba terbatasnya. *Journal of Science Education*, 5(1), 92-101.

Saputri, L., Maison. & Kurniawan, W. (2021). Pengembangan four-tier test berbasis website untuk mengidentifikasi miskonsepsi pada materi suhu dan kalor. *Jurnal Ilmiah Teknologi Informasi Asia*, 15(1), 61-68.

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

Suparno, P. (2013). *Miskonsepsi dan perubahan konsep dalam pendidikan fisika.* PT Grasindo: Jakarta.

Utari, J. I. & Ermawati, F. U. (2018). Pengembangan instrumen tes diagnostik miskonsepsi berformat four-tier untuk materi suhu, kalor dan perpindahannya. *Inovasi Pendidikan Fisika*, 7(3). 434-439.

Wea, K. N., Hau, R. R., & Kleruk, E. D. (2021). Penerapan metode pembelajaran inkuiri terbimbing dengan mind mapping untuk meningkatkan pemahaman konsep fisika siswa. *Jurnal Ilmiah Wahana Pendidikan*, 7(8), 770-774.

Wulandari, T. A., Prihandono, T. & Maryani. (2018). Analisis miskonsepsi siswa pada materi suhu dan kalor di kelas xi sma jember. *Seminar Nasional Pendidikan Fisika*, 3(1), 135-139.

Zafitri, R. E., Fitriyanto, S. & Yahya, F. (2018). Pengembangan tes diagnostik untuk miskonsepsi pada materi usaha dan energi berbasis adobe flash kelas xi di ma nw samawa sumbawa besar tahun ajaran 2017/2018. *Jurnal Kependidikan*, 2(2), 19-34.

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), 36-4