Science Education International
35(3),240-249
https://doi.org/10.33828/sei.v35.i3.7

ORIGINAL ARTICLE

Empirical Analysis of Physics Test Instruments to Measure Graphical Representation Abilities in "Temperature and Heat" Topics

Ananda Aprilia[1], Wipsar Sunu Brams Dwandaru[2]

[1]Department of Science Education, Faculty of Mathematics and Natural Sciences, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia,
[2]Department of Physics Education, Faculty of Mathematics and Natural Sciences, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia

***Corresponding Author:** anandaaprilia.2024@student.uny.ac.id\wipsarian@uny.ac.id

ABSTRACT

This study focused on developing physics test instruments for senior high school students on the topics of temperature and heat. The study aimed to determine (i) the quality of the test instrument content, (ii) the feasibility of the test instrument, and (iii) students' graphic representation abilities on "Temperature and Heat" topics. The test instrument development went through three stages, namely test design, testing, and test assembly. The test instrument was tested in the Science Class XII of the Public Senior High School 5 Yogyakarta with a total of 195 students as research subjects. The item analyzes in this study included analysis of model validity, reliability, level of difficulty, and level of students' ability. The research results showed that the test instrument for the graphical representation of "Temperature and Heat" topics had good content quality but still needed to be improved. The feasibility of the graphic representation test instrument showed that there were seven items that were declared valid and reliable even though the items tested did not conform to the *Rasch* model. However, these seven items were still suitable for use on a wide scale to measure graphical representations of "Temperature and Heat" topics. Finally, in the analysis of students' abilities regarding graphical representation individually, the largest percentage was in the low category, while, overall, the students' abilities level was classified as medium. Thus, it was necessary to improve question items that involve graphical representation that meet the criteria for good, valid, and reliable question items, and can improve students' graphic representation abilities.

KEY WORDS: "Temperature and Heat" topics; graphical representation; physics test instruments

INTRODUCTION

Education in the 21st century is characterized by globalization, which brings changes in the learning paradigm. The term "21st-century abilities" refers to the reconstruction of educational goals and learning outcomes according to the needs of the 21st century (Wang et al., 2018). 21st-century education aims to prepare students to face complex educational challenges and be able to adapt (González-Pérez and Ramirez-Montoya, 2022). Thus, there is an urgency for competency-based abilities that prioritize holistic and student-centered learning (Riveros, 2019). This is to ensure that students are able to face challenges in the era of globalization (Kennedy and Sundberg, 2020; Sawitri et al., 2021). One of the 21st-century competency-based abilities that students must have is the ability to represent a concept (Kurniawan, 2024). In the present study, physics learning is a learning activity that has the potential to grow the ability to represent concepts. The ability to represent concepts is important in physics learning in the 21st century (Munfaridah et al., 2021; Pradana et al., 2023). The important role of the ability to represent concepts in physics learning are (i) to make communicating abstract concepts in a concrete way easier (Munfaridah et al., 2021), (ii) to communicate scientific concepts (Treagust et al., 2018), (iii) to improve student learning outcomes (Maries and Singh, 2018), and (iv) to establish a correlation with critical thinking abilities (Abdurrahman et al., 2019).

Physics consists of abstract concepts that require high-level thinking processes. Concepts in physics learning should be understood, comprehended, and mastered by students. Physics concepts can be expressed in various representations or multiple representations, namely: Verbal, graphical, diagrammatic, mathematical, and table (Pebriana et al., 2022; Opfermann et al., 2022). Multiple representations can be used effectively in learning to help students develop creativity (Mutia and Prasetyo, 2018; Bicer, 2021) and understand facts and concepts accepted in classroom learning (Won et al., 2014; Lumbangaol and Tambunan, 2022). One of the problems related to multiple representations in physics learning is that physics concepts are often taught through mathematical formulas or mathematical representations (Harti, 2022). Hence, teaching concerning the physical meaning of these formulas through other representations is still lacking (Redish, 2021). Based on literature studies, teachers in schools only use mathematical and verbal representations by explaining physics formulas using classical learning in class (Yuwono et al., 2017). This causes students to become uninterested or bored with the explanations given by the teacher.

Studying physics does not involve only mastering mathematical skills but also mastering graphical representation. Graphical representation is one of the basic abilities that students must have. However, in reality, students still experience difficulties in connecting and integrating multiple representations and concluding information (Bollen *et al*., 2017; Maries et al., 2017). In addition, it is also known that students have difficulty reading, creating, and interpreting graphs (Stefanel, 2019). This is supported by the results of a literature review showing that graphical representation abilities in physics learning are still low (Ekawati et al., 2019; Febrianti et al., 2019). Based on the findings of Lallo et al., (2020), it is shown that only 3.1% of students have high representation abilities. In line with Fatimah's (2023) findings, it is indicated that as many as 78.2% of students have not been able to solve graphical representation problems. Therefore, in-depth research is needed in graphical representation so that it can be used more often in physics learning. Graphical representation skills are important to apply in physics learning to explain information, connect ideas, interpret data, process data, and draw conclusions. Efforts to overcome the low ability of graphical representation in physics learning require students' participation (Amelia et al., 2023) and the increase of test questions involving graphical representation.

Graphical representation is important for students to understand. Based on research by Volkwyn (2020), it is found that students' understanding of physics can be increased through graphical representation. This is reinforced by the findings of Stefanel (2019) showing that graphical representation has a fundamental role in physics and physics education, which is improving students' understanding. Apart from that, the use of graphical representation can also help students in solving physics problems (Fithrathy and Ariswan, 2019). Various research innovations have been carried out to improve graphical representation abilities, including using multimedia learning (Fithrathy and Ariswan, 2018), inquiry-based laboratory learning (Stefanel, 2019), carrying out computational modeling activities (Araujo et al., 2008), implementing electronic modules (Amelia et al., 2023), use of video simulation software (Rahma and Kurniawan, 2021), and virtual physics laboratory (Rani et al., 2019).

The innovation and improvements of graphical representation abilities can be implemented effectively if there is an assessment of the abilities that are valid, reliable, and can be used on a wide scale. Based on this, it is important to develop graphical representation ability test instruments. Several examples of graphical representation test instruments have been developed on vector physics (Bollen et al., 2017), mechanics (Ekawati et al., 2019), kinematics (Araujo et al., 2008; Volkwyn et al., 2020), waves (Rangkuti and Karam, 2022), thermodynamics (Kim and Nam, 2021), and electricity (Campos *et al*., 2020). However, the development of graphical representation test instruments is still limited. Several instruments are still being developed in the multiple-choice format (Armanto and Fenditasari, 2021) and two-level multiple-choice items (Am and Istiyono, 2022). Based on these limitations, it is necessary to develop test instruments in the form of open-ended questions that can develop students' reasoning abilities of physics concepts. The physics topics that can be developed are "Temperature and Heat". This is because the topics of temperature and heat are quite abstract and can give rise to different interpretations for students when studying them again (Baser, 2006). Students' representation ability in understanding "Temperature and Heat" topics presented in graphical forms reaches a difficulty level of 58.57% (Ma'rifah et al., 2016). Apart from that, "Temperature and Heat" test instruments for measuring graphical representations have not been widely developed. Thus, based on the explanation above, the objectives of the present study are to determine: (i) the quality of the test instrument contents to measure the graphical representation abilities on the "Temperature and Heat" topics, (b) the feasibility of the test instrument to measure graphical representation abilities on "Temperature and Heat" topics, and (c) students' graphical representation abilities on "Temperature and Heat" topics. Through this instrument development, it is hoped that this can become an innovative graphical representation test instrument that is valid and reliable. Thus, the instrument developed can measure students' graphical representations of the topics of "Temperature and Heat".

METHODS

This study used the Research and Development (R&D) method, which focused on developing test instruments. The test instrument developed was used to measure students' graphic representation abilities on the topics of "Temperature and Heat". The development design in this study used the Wilson, Oriondo, and Antonio model, which was modified by Istiyono, Mardapi, and Suparno (2014). The test instrument development went through three stages, namely (1) test design, (2) test testing, and (3) test assembly. The test design part consisted of (a) determining the test objectives, (b) determining the competencies and topics being tested, (c) preparing the test grid, (d) preparing question items and scoring guidelines, (e) validating question items, and (f) improving question items. The flow chart of the instrument development design can be seen in Figure 1.

The respondents participating in the test trials were senior high

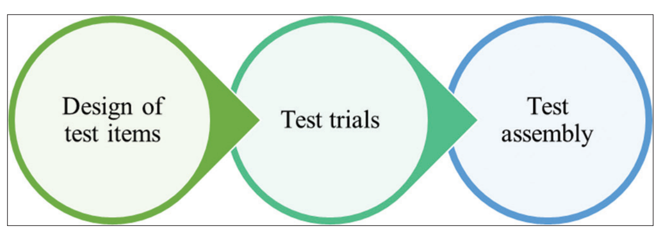


Figure 1: Instrument development design

school students. These respondents had previously studied "Temperature and Heat" topics. The test trials were carried out at the Public Senior High School 5 Yogyakarta, Science Class XII, with a total of 195 students. The test instrument used was a 10-item essay. The indicators for the question items can be seen in Table 1.

The data obtained in the test trials were quantitative data, which were then analyzed using the *Quest* software. The item analysis in this study included analysis of the item model feasibility, reliability, item difficulty level, and student ability level. The item model feasibility analysis can provide empirical validity of the items. An instrument was said to be valid or in accordance with the assessment model if it has an Infit Mean Square (*INFIT MNSQ*) value in the range of 0.77–1.33 (Adam and Kho, 1996; Subali and Suyata, 2011).

The reliability analysis of the question items can be observed in the internal consistency section of the *QUEST* output (Setyawarno, 2021; Yunida and Arthur, 2023). Apart from that, the reliability analysis can also be determined through summary case estimates (Subali and Suyata, 2011). The levels of reliability based on the interpretation of the reliability index can be observed in Table 2.

The difficulty level analysis of the questions can be seen in the difficulty section of the *QUEST* output. An item difficulty level is said to be good if it has an item difficulty index ranging from -2 to +2 (Hambleton et al., 1991; Asysyifa et al., 2019).

Table 1: Indicators for graphical representation question items

Question item number	Graphical representation indicator	Description
1, 2	Collecting data	The ability to collect and examine data presented on a graph.
3, 4	Graph interpretation	The ability to interpret graphs related to relationships among variables.
5, 6	Linking variables	The ability to connect variables on a graph.
7	Comparing graphs	The ability to compare graphs.
8, 9	Graph analysis	The ability to analyze and transform data into a graph according to the concept given.
10	Making a conclusion	The ability to create and review results based on data or information provided.

Table 2: Reliability coefficient (Guilford, 1956)

Reliability coefficient	Reliability level
0.00–0.20	Very low
0.21–0.40	Low
0.41–0.60	Moderate
0.61–0.80	High
0.81–1.00	Very high

The difficulty levels can be observed in Table 3.

Students' abilities can be determined through the *QUEST* software in the output summary of the case estimates on the reliability of estimates (Hanna and Retnawati, 2022). Apart from that, students' abilities can also be observed in the estimated values in the *QUEST* software output. Estimated student abilities can be observed in Table 4.

RESULTS

Graphical representation is one of the provisions for students to face the 21st century, especially for the physics subject. This study develops a graphical representation test instrument in "Temperature and Heat" topics. In developing the test instrument, it is necessary to assess the items to measure the feasibility and quality of the items tested on students, which is called empirical validation. Apart from that, through empirical validation, students' abilities in representing graphics can also be determined.

The feasibility and quality of the questions as well as the level of students' abilities are analyzed using the *QUEST* software. The validity and invalidity of the question item instrument are shown in the *MNSQ* infit output (Bond et al., 2021). The *MNSQ* infit output results can be observed in Figure 2.

Based on the results of the distribution of *MNSQ* infit values for each question item in Figure 1, it is found that there are three questions that are invalid. This is because the *MNSQ* infit value of the questions is outside the range of 0.77–1.30. The results of the validity percentage of graphical representation question items can be seen in Figure 3. The question items that are declared invalid can be seen in Table 5.

Apart from determining the validity of the question items, the *QUEST* software can look at the reliability of the instrument. The instrument reliability for the graphical representation of items can be observed from the output results of the item estimate summary, case estimate summary, and internal consistency. The results of the instrument reliability can be observed in Table 6.

Table 3: Levels for difficulty values

Difficulty value	Information
$b>2$	Very difficult
$1<b\leq 2$	Difficult
$-1<b\leq 1$	Moderate
$-2<b\leq -1$	Easy
$b\leq -2$	Very easy

Table 4: Estimated student abilities

Estimated value	Ability levels
>1.00	High
−1.00–+1.00	Moderate
<−1.00	Low

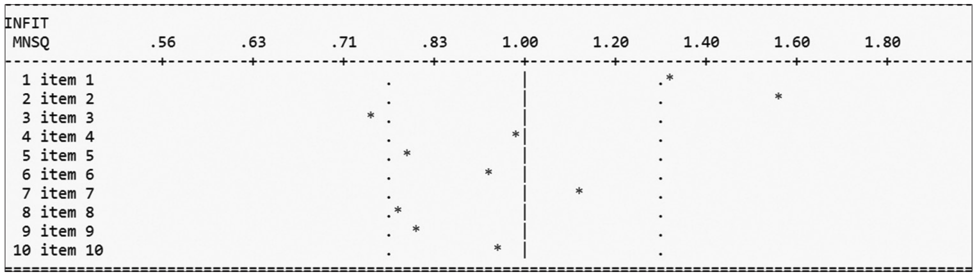


Figure 2: Distribution of the *MNSQ* infit values for each question item

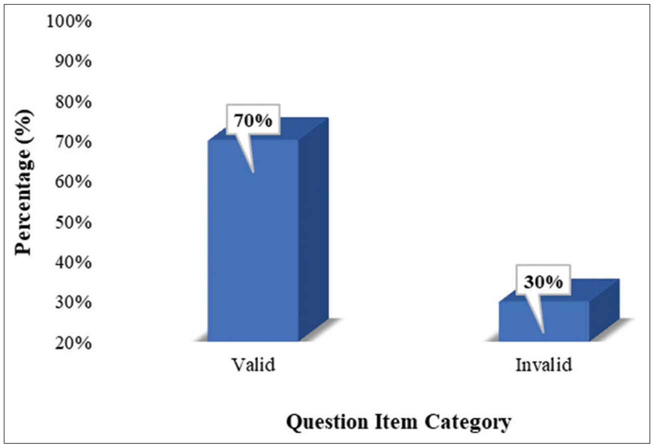


Figure 3: Percentage of question item categories

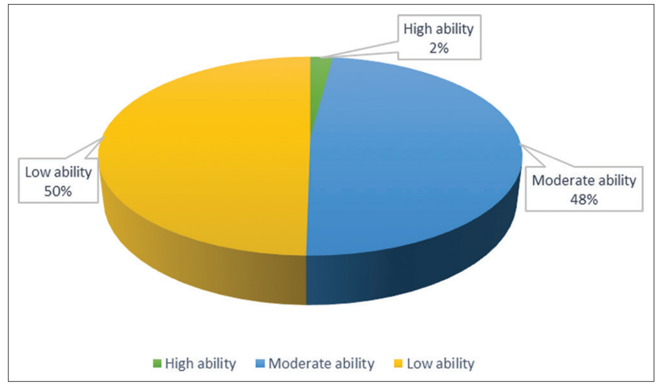


Figure 4: Students' ability levels

The difficulty level of the instrument can be observed from the difficulty output of the *QUEST* software. The results of the difficulty level of the question items can be seen in Table 7.

Furthermore, the output results of the *QUEST* software can display the level of students' abilities in answering graphical representation questions. The level of students' abilities is categorized into three, namely high, moderate, or low. The analysis results of students' ability levels can be observed in Figure 4.

Based on the results of the validity, reliability, and level of difficulty, a test instrument has been produced that is feasible. The results of the analysis show that of the ten question items, there are seven items that are declared feasible and meet the criteria of good quality items. The results of the final characteristics of the instrument can be observed in Table 8.

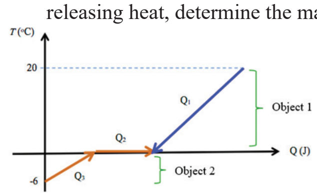
DISCUSSION

The test instrument is developed following a procedure consisting of three steps. The first step is the test design. The graphical representation test instrument on "Temperature and Heat" topics is prepared and designed based on the indicators that can be observed in Table 1. Question items are developed based on six graphical representation indicators. Based on the indicators compiled, there are 10 question items and they are developed in the form of essay questions. This is because essay questions can develop the ability to think openly (Cakir and Cengiz, 2016; Sarwanto et al., 2021), identify assumptions (Larsson, 2021), explore opinions or ideas (Mabruroh and Suhandi, 2017), train thinking skills critically (Marni and Harsiati, 2019), and assess creative thinking abilities (Suyana et al., 2019). Thus, in general, essay questions can train students to think analytically, test students' understanding and reasoning through analyzing, and interpret and explain the information obtained.

The test design step includes several stages, including preparing the grid, question items, and scoring guidelines. Furthermore, the question items are validated by content experts and revised according to suggestions before being tested on students. After fulfilling the content validity, the second step is carried out, namely trial testing. Test trials on instrument development constitute empirical validity. The instrument is tested on students who have received "Temperature and Heat" topics. The empirical validity is carried out to determine the quality and feasibility of the items regarding graphical representation abilities. Apart from that, empirical validity is also used to determine students' abilities in graphical representation. Based on the results of the *MNSQ* infit output in Figure 1, it is found that from the ten question items, seven questions meet the valid criteria (see Appendix 1), whereas the rest of the question items are not valid (invalid) with percentages of 70% versus 30%, respectively. Thus, seven items fit the Rasch model while the other three items do not. According to the study by Planinic et al. (2019), the question items do not match the Rasch model because the students' answer patterns are not suitable for defining the abilities being tested.

The three invalid question items can be observed in Table 5. These invalid question items are included in the indicators

Table 5: Invalid question items

Question item number	Graphical representation indicator	Question Items
1	Collecting data	Lia and Ana are students who are curious about the application of heat expansion of iron. Therefore, they observed a train track at Lempuyangan Station where no trains pass as in the picture below.  (Private Document) The rail is made of iron whose coefficient of linear expansion (α) is $12\times10^{-6}/^{o}C$. Each rail is 2 m long. Lia and Ana observed the train tracks from 07.00 in the morning to 12.00 at noon. The initial temperature of the iron at 07.00 in the morning is 20^{o} C. Lia and Ana observed the increase in temperature of the tracks every hour. Every hour the temperature increases by 3^{o} C consistently until 12.00 noon. Based on the explanation above, determine: a. The minimum gap length between two adjacent rail sections so that they do not collide when expanding! b. Draw a graph of the relationship between the increase in rail length (Y-axis) and the change in temperature (X-axis)! (Use the equation $\Delta L = \alpha L_0 \Delta T$).
2		Lia and Ana are students who are curious about the application of heat expansion of iron. Therefore, they observed a train track at Lempuyangan Station where no trains pass as in the picture below.  (Private Document) The rail is made of iron whose coefficient of linear expansion (α) is $12\times10^{-6}/^{o}C$. Each rail is 10 m long. Lia and Ana observed the train tracks from 07.00 in the morning to 12.00 at noon. The initial temperature of the iron at 07.00 in the morning is 20°C. Lia and Ana observed the increase in temperature of the tracks every hour. Every hour the temperature increases by 5°C consistently until 12.00 noon. Based on the explanation above, determine: a. The minimum gap length between two adjacent rail sections so that they do not collide when expanding! b. Draw a graph of the relationship between the increase in rail length (Y-axis) and the change in temperature (X-axis)! (Use the equation $\Delta L = \alpha L_0 \Delta T$).
3	Chart interpretation	A graph of a relationship between temperature T (°C) and heat Q (J) is observed below. It is known that the specific heats of Object 1 (c_1) and Object 2 (c_2) are 1 cal/g° C and 0.5 cal/g° C, respectively, and the latent heat of Object 2 (L_2) is 80 cal/g. Based on the graph below, determine: a. Which object receives heat? Explain! b. Which object releases heat? Explain! c. Write Black's equation based on the graph! d. If 40% of the object receiving the heat melts into the same phase as the object releasing heat, determine the mass ratio of Objects 1 and 2! 

of collecting data and interpreting graphs. Moreover, these three questions are classified as difficult for students. This is because solving these three question items requires higher-order thinking skills (HOTs); hence, students have difficulty to understand and solve them. Another factor is that students are still not used to working on questions that involve the ability to represent graphics. This is in accordance with the results of the interviews with physics teachers who state that, during

Table 6: Reliability results of the question items

Reliability	Reliability coefficient	Reliability category
Summary of item estimates	0.37	Low
Summary of case estimates	0.75	High
Internal consistency	0.77	High

Table 7: Results of the difficulty levels of the question items

Question item number	Difficulty level	Category
1	−0.49	Moderate
2	−1.01	Easy
3	0.50	Moderate
4	0.39	
5	−0.25	
6	−0.07	
7	0.15	
8	0.30	
9	0.63	
10	−0.16	

Table 8: Characteristic validity of the question item instrument

Question item number	Difficulty level	Categories	Infit *MNSQ*	Status	Categories
4	0.39	Moderate	0.98	Fit	Valid
5	−0.25		0.79		
6	−0.07		0.92		
7	0.15		1.12		
8	0.30		0.79		
9	0.63		0.80		
10	−0.16		0.95		

the physics lesson, graphical representation is not involved in the lesson. Apart from that, time constraints are also one of the causes of students not being able to solve the questions properly and correctly. Hence, students just guess the answer. Thus, students' graphical representation abilities are not optimal.

Furthermore, the item's reliability, which shows the level of confidence in the items, can be observed in Table 6. Based on the results contained in Table 6, the item estimate summary has a reliability coefficient of 0.37 in the low category. This means that the items tested do not match the model tested (Rasch model). The low value of the item summary is influenced by students who are less thorough and less careful when answering the questions. Based on the findings of Ofianto (2018), several factors that influence reliability include: (a) test complexity, (b) test objectivity, and (c) short test duration that causes low test reliability. This is in line with the three invalid question items in Figure 1. Meanwhile, the summary of case estimates and internal consistency has reliability coefficients of 0.75 and 0.77, respectively, in the high category. This indicates that repeating the test will produce stable and consistent results.

Based on the empirical test results, all questions can be stated as good questions. This can be observed in Table 7. The difficulty level of the questions is in the range of −2 to +2. In general, all questions have a moderate level of difficulty. Based on the research by Adawiah (2020) and Adi et al. (2022), test items are good if the difficulty level of the items is in the moderate category. Even though all the question items are in a good category, there are three question items that are declared invalid in the output of Figure 1. Thus, it can be stated that even though the question items are declared invalid, it does not mean that they are not good. This is of course influenced by various influencing factors.

Based on the findings, it is found that valid question items mean that students could do or solve them, whereas invalid question items indicate that they do not meet the feasibility criteria for the question items. Furthermore, invalid question items are related to the failure to fulfill specified indicators, namely collecting data and interpreting graphs. Meanwhile, the indicators of connecting variables, comparing graphs, analyzing graphs, and providing conclusions meet the feasibility criteria. Furthermore, the findings in the field (physics lesson in class) show that the majority of students are unable to work on and/or solve graphical representation questions on the "Temperature and Heat" topics. Students are of the opinion that the graphical representation is not applied in the "Temperature and Heat" topics. In the "Temperature and Heat" topics, graphs are only found in the relationship between temperature increase (ΔT) and heat (Q). It is very rare for students to encounter graphs on other "Temperature and Heat" sub-topics, such as: (a) the relationship between temperatures of various thermometers (Celcius, Reamur, Fahrenheit, and Kelvin); (b) heat transfer; (c) heat capacity and specific heat; (d) latent heat; and (e) heat expansion. Students feel that the graphical representation questions on "Temperature and Heat" topics are difficult to solve considering the time constraints. This finding is in accordance with Figure 3 in that students have graphical representation ability levels in the high, moderate, and low categories with percentages of 50%, 48%, and 2%, respectively, in the *QUEST* software estimate output. However, in the case of the estimate summary, the reliability of the estimate has a value of 0.75, which states that the level of student ability is classified as moderate. The *QUEST* software of estimate output analyzes students' ability levels one by one, while the case estimate output summary on the estimated reliability analyzes students' ability levels as a whole. Based on these observations, it can be stated that students' abilities in graphical representation of "Temperature and Heat" topics still need to be improved, especially in the ability to collect data and interpret graphics. Thus, based on the results of the test instrument development, it is concluded that validity, reliability, and students' abilities influence each other. However, the validity of the test items does not affect the qualifications of good test items.

After analyzing the questions in the test trial, the next step is assembling the test. At this stage, the test instrument is prepared based on the criteria of validity, reliability, and level of difficulty of the questions so that the test instrument is feasible to measure the graphic representation ability. Based on the results of the analysis of graphical representation items in Table 8, there are seven items that meet the criteria for a good instrument. Meanwhile, the other three questions are not included or have to be revised first if they are going to be used. The seven questions can be tested on a wide scale to measure graphical representation abilities in "Temperature and Heat" topics. In this case, the test instrument still needs to be improved to train students to have one of the competencies that must be possessed by students in the 21st century.

From these findings, it can be seen that several research implications can be sought and proposed. First, in the environment of a learning-teaching process, a test instrument can be used as a teaching tool in addition to its primary function as a testing instrument. In particular, this measurement tool can be used as instructional media that can help students develop their data-collection and data-organization competencies, strengthen their skills in number-space relationships, practice their use of technological devices, and develop their problem-solving skills. Second, in relation to the use of essay-type assessment, the measurement tool has the feature of eliminating the problem of evaluator bias in the scoring of students' answers to open-ended questions. Finally, one mention needs to be made of the linguistic quality of students' essays. One of the major flaws of the student's writing is that the language is not clear. In this case, linguistic support in the form of a language expert's review of students' written responses is needed.

CONCLUSION

This study focuses on developing a test instrument carried out in Science Class XII of the Public Senior High School 5 Yogyakarta on "Temperature and Heat" topics. Based on the analysis results obtained, it can be concluded that the test instrument for measuring the graphical representation of the "Temperature and Heat" topics has good content quality but still needs to be improved. The quality of the content is seen from the difficulty level of the question items. The results obtained show that, overall, the question items have a level of difficulty that is in the range of −2–+2 in the moderate category. Meanwhile, the feasibility of the test instrument can be observed from the validity and reliability of the question items. The results show that there are seven items that are declared valid and reliable even though the items tested do not conform to the Rasch model. However, these seven items are still feasible for use on a wide scale to measure graphical representation of the "Temperature and Heat" topics. The other three questions are not feasible for use on a wide scale and still need to be improved if they are to be used. Finally, in the analysis of students' abilities regarding graphical representation individually, the largest percentage is in a low category, while, overall, it is stated that the student's ability level is classified as in the moderate category. Thus, it is still necessary to improve question items that involve graphical representations that meet the criteria for good, valid, and reliable question items, and can improve students' graphic representation abilities. The study in the development of this test instrument uses a limited number of respondents so that it is advisable that more respondents can be used for further studies. Moreover, it is equally necessary to develop a test instrument with different question items and on different physics topics. Finally, it is also a case that the measurement instrument can also be used as an instructional tool for the learning-teaching process in accordance with the learning objectives.

ACKNOWLEDGMENTS

The authors would like to thank the Faculty of Mathematics and Natural Sciences, Yogyakarta State University, for supporting this study. The authors would also like to acknowledge their gratitudes to the Directorate of Research, Technology, and Community Service, Directorate General of Higher Education, Research, and Technology, Ministry of Education, Culture, Research, and Technology of Indonesia for granting the research project under the contract No. 146/E5/PG.02.00.PL/2023.

REFERENCES

Abdurrahman, A., Setyaningsih, C.A., & Jalmo, T. (2019). Implementing multiple representation-based worksheets to develop critical thinking skills. *Journal of Turkish Science Education*, 16(1), 138-155.

Adam, R.J., & Kho, S.T. (1996). *Acer Quest Version 2.1*. Victoria, Camberwell: The Australian Council for Education Research.

Adawiah, R. (2020). The application-based analysis of questions item quality in junior high school. *Journal of Indonesian Student Assessment and Evaluation*, 6(2), 140-148.

Adi, N.R.M., Amaruddin, H., Adi, H.M.M., & A'yun, I.L.Q. (2022). Validity and reliability analysis using the Rasch model to measure the quality of mathematics test items of vocational high schools. *Journal of Research and Educational Research Evaluation*, 11(2), 103-113.

Am, M.A., & Istiyono, E. (2022). Multiple representation abilities of high school students in physics: A study of modern response theory. *Thabiea: Journal of Natural Science Teaching*, 5(1), 85-97.

Amelia, F., Ekawati, E.Y., & Fauzi, A. (2023). Application of the electronic sigil module to improve students' knowledge competency and graphic representation ability. *Journal of Physics Materials and Learning*, 13(2), 53-61.

Araujo, I.S., Veit, E.A., & Moreira, M.A. (2008). Physics students' performance using computational modeling activities to improve kinematics graphs interpretation. *Computers and Education*, 50(4), 1128-1140.

Armanto, I.E., & Fenditasari, K. (2021). *Physics Graphical Representation Test of Straight Motion Kinematics Based On Boti Boat Local Wisdom: Development and Validity*. In: 6th International Seminar on Science Education, pp. 350-357.

Asysyifa, D.S., Wilujeng, I., & Kuswanto, H. (2019). Analysis of students critical thinking skills using partial credit models (PCM) in physics learning. *International Journal of Educational Research Review*, 4(2), 245-253.

Baser, M. (2006). Effect of conceptual change oriented instruction on students' understanding of Heat and Temperature concepts. *Journal of Maltese Educational Research*, 4(1), 64-79.

Bicer, A. (2021). Multiple representations and mathematical creativity. *Thinking Skills and Creativity,* 42, 100960.

Bollen, L., Van Kampen, P., Baily, C., Kelly, M., & De Cock, M. (2017). Student difficulties regarding symbolic and graphical representations of vector fields. P*hysical Review Physics Education Research*, 13(2), 020109.

Bond, T.G., Yan, Z., & Heene, M. (2021). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. 4th ed. London: Routledge.

Cakir, H., & Cengiz, O. (2016). The use of open-ended versus closed-ended questions in Turkish classrooms. *Open Journal of Modern Linguistics*, 6(2), 60-70.

Campos, E., Zavala, G., Zuza, K., & Guisasola, J. (2020). Students' understanding of the concept of the electric field through conversions of multiple representations. *Physical Review Physics Education Research*, 16(1), 1-19.

Ekawati, R., Setiawan, A., Wulan, A.R., & Rusdiana, D. (2019). The use of classroom assessment based on multi-representation ability in Mechanics concept. *Journal of Physics: Conference Series*, 1157(3), 1-6.

Fatimah, F. (2023). Analysis of students' graphic interpretation abilities on rectilinear motion material. *Journal of Education, Science and Technology*, 10(2), 554-566.

Febrianti, J., Akhsan, H., & Muslim, M. (2019). Analysis of misconceptions about temperature and heat among students at SMA Negeri 3 tanjung raja. Journal of Physics Innovation and Learning, 6(1), 90-102.

Fithrathy, A., & Ariswan. (2019). Developing physics learning multimedia to improve graphic and verbal representation of high school students. *Journal of Physics: Conference Series*, 1233(1), 1-10.

González-Pérez, L.I., & Ramírez-Montoya, M.S. (2022). Components of education 4.0 in 21st century skills frameworks: Systematic review. *Sustainability*, 14(3), 1493.

Guilford, J.P. (1956). *Fundamental Statistics in Psychology and Education*. 3rd ed. New York: McGraw-Hill Book Company.

Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of Item Test Response Theory*. Newbury Park: Sage Publications.

Harti, N. (2022). Development of multi-representation-based physics teaching materials to improve students' analytical thinking skills at MAN 2 Makassar. *Educational Scientific Journal*, 1(7), 719-730.

Kennedy, T.J., & Sundberg, C.W. (2020). 21st century skills. In: *Science Education in Theory and Practice: An Introductory Guide to Learning Theory.* Cham: Springer International Publishing, pp. 479-496.

Kim, J.H., & Nam, J. (2021). Thermodynamic identities with sunray diagrams. *European Journal of Physics*, 42(3), 035101.

Kurniawan, F.A. (2024). Collaborative science learning model based on multiple representatives in elementary schools. *Tambusai Education Journal*, 8(1), 8232-8247.

Lallo, A.T.J., Sinaga, P., & Utari, S. (2020). Inventory of graphic understanding and the ability to represent junior high school students' physics concepts on the subject of kinematics. *Physics Education Forum*, 5(2), 1-5.

Larsson, K. (2021). Using essay responses as a basis for teaching critical thinking-a variation theory approach. *Scandinavian Journal of Educational Research*, 65(1), 21-35.

Lumbangaol, S.T.P., & Tambunan, L.O. (2022). Development of a multi-representation based HOTS test on the learning outcomes of Physics education students on the Condensed Matter Physics course. *Journal of Economics, Social and Humanities*, 3(6), 98-112.

Ma'rifah, E., Parno, & Mufti, N. (2016). Identify student difficulties in temperature and heat material. *Journal of Physics Learning*, 4(5), 124-133.

Mabruroh, F., & Suhandi, A. (2017). Construction of critical thinking skills test instrument related to the concept on soundwave. *Journal of Physics: Conference Series*, 812(1), 012056.

Maries, A., & Singh, C. (2018). Case of two electrostatics problems: Can providing a diagram adversely impact introductory physics students' problem solving performance? *Physical Review Physics Education Research*, 14(1), 1-14.

Maries, A., Lin, S.Y., & Singh, C. (2017). Challenges in designing appropriate scaffolding to improve students' representational consistency: The case of a Gauss's law problem. *Physical Review Physics Education Research*, 13(2), 020103.

Marni, S., & Harsiati, T. (2019). Critical thinking patterns of first-year students in argumentative essays. *Journal for the Education of Gifted Young Scientists*, 7(3), 683-697.

Munfaridah, N., Avraamidou, L., & Goedhart, M. (2021). The use of multiple representations in undergraduate physics education: What do we know and where do we go from here? *Eurasian Journal of Mathematics, Science and Technology Education*, 17(1), 1-19.

Mutia, N.B., & Prasetyo, Z.K. (2018). The effectiveness of students' worksheet based on multiple representations to increase creative thinking skills. *Journal of Education and Learning*, 12(4), 631-637.

Ofianto, O. (2018). Analysis of instrument test of historical thinking skills in senior high school history learning with Quest programs. *Indonesian Journal of History Education*, 6(2), 184-192.

Opfermann, M., Schmeck, A., & Fischer, H.E. (2021). *Multiple Representations and Learning Physics. Physics Education*. Cham: Springer, pp. 175-199.

Pebriana, I.N., Supahar, S., Pradana, P.W., & Mundilarto, M. (2022). Investigating Multiple Representations Abilities of High School Students on Linear Motion. In: *5th International Conference on Current Issues in Education (ICCIE 2021)*. Vol. 640. Amsterdam: Atlantis Press, pp. 232-237.

Planinic, M., Boone, W.J., Susac, A., & Ivanjek, L. (2019). *Rasch* analysis in physics education research: Why measurement matters. *Physical Review Physics Education Research*, 15(2), 020111.

Pradana, P.W., Febriani, F., Ibnusaputra, M., & Jumadi, J. (2023). Development of physics test instrument to measure verbal representation of high school students on optical instrument topic. *Journal of Science Education Research*, 9(10), 7963-7968.

Rahma, F.N., & Kurniawan, E.S. (2021). Assessment of students' graphic representation skills on the concept of parabolic motion assisted by video simulation software modellus. *SPECTRA: Journal of Science Education Studies*, 7(2), 134-140.

Rangkuti, M.A., & Karam, R. (2022). Conceptual challenges with the graphical representation of the propagation of a pulse in a string. *Physical Review Physics Education Research*, 18(2), 1-14.

Rani, S.A., Mundilarto, W., & Dwandaru, W.S.B. (2019). Physics virtual laboratory: An innovative medium in 21st century learning. *Journal of Physics: Conference Series*, 1321(2), 1-8.

Redish, E.F. (2021). Using mathematics in physics: Overview. *The Physics Teacher*, 59(5), 314-318.

Riveros, H.G. (2019). Physics teaching in the 21st century. *Journal of European Education*, 8(1), 17-29.

Sarwanto, Fajari, L.E.W., & Chumdari. (2021). Open-ended questions to assess critical-thinking skills in Indonesian elementary school. *International Journal of Instruction*, 14(1), 615-630.

Sawitri, Y., Asrizal, A., & Mufit, F. (2021). Analysis of physics e-books assisted by application of learning house using quantum learning models to improve the 21st century skills of high school students. *Journal of Physics: Conference Series*, 1876(1), 012041.

Setyawarno, D. (2021). *Practical Guide to Item Response Theory: Item Analysis with Classical and Modern Approaches*. Yogyakarta: Science Education Department.

Stefanel, A. (2019). Graphs in physics education: From representation to conceptual understanding. In: *Mathematics in Physics Education*. Germany: Springer, pp. 195-231.

Subali, B., & Suyata, P. (2011). *Guide to Analyzing Educational Measurement Data to Obtain Empirical Evidence of Validity using the Quest Program*. Indonesia: LPPM UNY.

Suyana, I., Nadaipah, S., Sinaga, P., & Feranie, S. (2019). Constructing essay questions to assess scientific creative and critical thinking simultaneously related to collision problems based on students' responses. *Journal of Physics: Conference Series*, 1204(1), 1-9.

Treagust, D.F., Won, M., & McLure, F. (2017). *Multiple Representations and Students' Conceptual Change in Science. In: Converging Perspectives on Conceptual Change*. England, UK: Routledge, pp. 121-8.

Volkwyn, T.S., Airey, J., Gregorcic, B., & Linder, C. (2020). Developing representational competence: Linking real-world motion to physics concepts through graphs. *Learning: Research and Practice*, 6(1), 88-107.

Wang, Y., Lavonen, J., & Tirri, K. (2018). Aims for learning 21st century competencies in National primary science curricula in China and Finland. *Eurasian Journal of Mathematics, Science and Technology Education*, 14(6), 2081-2095.

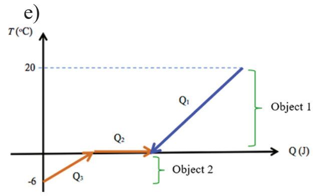
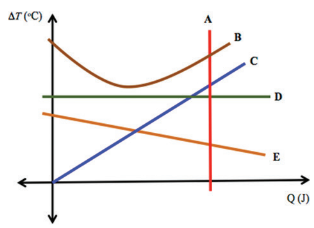
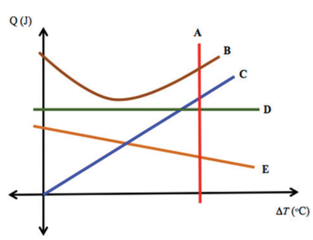
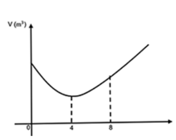
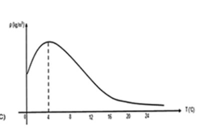
Won, M., Yoon, H., & Treagust, D.F. (2014). Students' learning strategies with multiple representations: Explanations of the human breathing mechanism. *Science Education*, 98(5), 840-866.

Yunida, H., & Arthur, R. (2023). Bloom's taxonomy approach to cognitive space using classic test theory and modern theory. *East Asian Journal of Multidisciplinary Research*, 2(1), 95-108.

Yuwono, G.R., Mahardika, I.K., & Gani, A.A. (2017). The influence of the guided inquiry learning model on students' physics learning outcomes (verbal, pictorial, mathematical and graphic representation abilities) in high school. *Journal of Physics Learning*, 5(1), 60-65.

APPENDIX

Appendix 1: Valid question items

Question Item Number	Graphical Representation Indicator	Question Items
4	Graph interpretation	A graph between temperature – T (°C) vs heat - Q (J) can be observed below. It is known that the specific heats of water and ice are 1 cal/g°C and 0.5 cal/g°C, respectively, and the melting heat of ice is 80 cal/g. Answer the following questions based on the graph: a) Which object receives heat? Explain! b) Which object releases heat? Explain! c) Write Black's equation based on the graph! (Use the equation $Q_{receive} = Q_{release}$) d) If it is known that the mass ratio of Object 2 and Object 1 is 4 : 7, then determine the percentage of the mass of the object (heat receiver) that has melted! e) 
5	Linking variables	A relationship between temperature difference - Δ*T* (°C) vs heat - Q (J) is given in a graph, which can be observed below. For a constant heat capacity, C, determine: a) Which graph shows the relationship between the variables Δ*T* and Q most accurately? Explain! (Q = C Δ*T*) b) Using the relationship between the variables Δ*T* and Q, explain how the value of C can be obtained from the graph below? Explain! 
6		A graph of heat - Q (J) vs temperature difference - Δ*T* (°C) is given as follows. For a constant heat capacity, C, determine: a) Which graph shows the relationship between the variables Q and Δ*T* most accurately? Explain! (Q = C Δ*T*) b) Using the relationship between the variables Q and Δ*T*, explain how the value of C can be obtained from the graph above? Explain! 
7	Comparing graphs	Two graphs regarding the expansion of a liquid is given below. Based on the two graphs, answer the following questions: a) What phenomena are shown by the two graphs above? Explain! b) Describe the differences and similarities between the two graphs above! Explain! 

(Contd...)

Appendix 1: ***(Continued)***

Question Item Number	Graphical Representation Indicator	Question Items
8	Graph analysis	An iron rod with a mass of 500 grams is placed in hot water so that its temperature rises by 50°C from room temperature (30°C). Hence, the amount of heat absorbed by the rod is 11.5 kJ. From the data presented, determine: a) The specific heat of iron! (Q = m c Δ T) b) Draw a graph of *T* (°C) vs Q (kJ)! c) What part of the graph shows the specific heat of iron? Explain using the heat formula!
9		An iron rod with a mass of 100 grams is placed in hot water so that its temperature rises by 20°C from room temperature (30°C). Hence, the amount of heat absorbed by the rod is 0.92 kJ. From the data presented, determine: a) Specific heat of iron! (Q = m c Δ T) b) Draw a graph of T (oC) vs Q (kJ)! c) What part of the graph shows the specific heat of iron? Explain using the heat formula!
10	Giving a conclusion	The expansion graph of a gas relates ΔV vs. ΔT where ΔV and ΔT form the vertical and horizontal axes, respectively. ΔV is the difference in gas volume due to expansion and ΔT is the temperature difference. The measurement results show that when ΔT = 20°C, ΔV = 30 m^3 is obtained, and when ΔT = 120°C, ΔV = 31 m^3 is obtained. From this information, determine: a) The equation for the expansion of a gas that relates ΔV and ΔT! b) Draw a graph of the relationship between ΔV and ΔT! c) Gradient value (slope) of the graph! $\left(m = \frac{\Delta y}{\Delta x}\right)$ d) Value of the initial volume of gas!