A rasch model approach to gender-based thermal concept analysis in East Java high schools

Fian Rifqi Irsalina[1], Lina Aviyanti[1]*, Renaldy Fredyan[2]

[1]Department of Physics Education, Universitas Pendidikan Indonesia, Bandung, 40154, Indonesia
[2]Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei, 106335, Taiwan
*Corresponding author, email: lina@upi.edu

Article History
Received: 30 October 2024
Revised: 14 November 2024
Accepted: 15 December 2024

Keywords
Conceptual ability
Physics learning
Rasch model
Thermal concept evaluation

Abstract
This research analyzes high school students' understanding of thermal concepts in East Java, focusing on identifying gender-based differences and misconceptions. A descriptive survey with a non-experimental cross-sectional design was conducted using the Thermal Concept Evaluation (TCE) instrument, distributed via Google Forms to 171 students (94 females and 77 males). Data were analyzed using the Rasch model with Winsteps software, providing a detailed diagnosis of student ability levels and item difficulties. The findings revealed significant misconceptions about thermal conductivity and thermal equilibrium, which were challenging for both genders. Male students demonstrated better comprehension of heat transfer and temperature change, while female students showed a slightly higher understanding of boiling concepts. Practical implications include the need for targeted instructional strategies, such as inquiry-based and problem-based learning, incorporating hands-on experiments and technology-based tools like simulations to visualize abstract thermal phenomena. Policymakers and educators are encouraged to adopt conceptual-based curricula, improve laboratory facilities, and provide teacher training programs to address misconceptions and bridge gender gaps in thermal physics education. These strategies aim to enhance students' understanding and prepare them for advanced applications in science and technology.

Irsalina, F. R., Aviyanti, L., & Fredyan, R. (2025). A rasch model approach to gender-based thermal concept analysis in East Java high schools. *Momentum: Physics Education Journal, 9*(1), 92-108. https://doi.org/10.21067/mpej.v9i1.11007

1. Introduction

Physics education is crucial in developing critical thinking skills, which support problem-solving and a deep understanding of concepts (Salazar et al., 2023). Physics curricula often include activities and strategies to encourage critical thinking, enabling students to engage in open-ended investigations and deepen their understanding of physical concepts (Baker, 2022). To grasp abstract concepts in physics, high-level reasoning abilities are required. Therefore, physics education focuses on the knowledge students acquire and continuously developing their learning skills (Sudirman et al., 2017).

Conceptual understanding is essential for students who have undergone the learning processs (Wiyono et al., 2021). According to Sari et al. (2016, in Safarati & Lubis, 2022), conceptual understanding is a crucial element in the learning process as it represents the stage of understanding abstract information, which requires students to classify objects or phenomena. High-order thinking, including critical thinking skills, is necessary for conceptual understanding. Conceptual understanding can be defined as a student's ability to comprehend a concept related to everyday phenomena based on experience and observation, which is then connected to pre-existing concepts in their thinking. Analyzing students' initial understanding of concepts is critical in preventing conceptual errors. Many conceptual mistakes occur due to several factors, including those related to the students themselves, such as attitudes, interest in learning, or the development of their cognitive structures (Zulrahayu & Mufit, 2024).

Among various concepts in physics, thermal concepts are essential topics that often lead to misconceptions among high school students (Maison et al., 2020; Sofna et al., 2024). Many real-life applications and everyday situations are related to fundamental thermal concepts (Gunawan et al.,

doi: 10.21067/mpej.v9i1.11007

ISSN: 2548-9135

2019). Students tend to form thermal concepts based on daily experiences and intuition, frequently resulting in an understanding that contradicts scientific theory (O'Brien et al., 2021). A deep understanding of thermal concepts is crucial for academic success and introduces new scientific approaches, such as understanding irreversible phenomena, which are important in studying complex systems (Mikielewicz & Mikielewicz, 2024). Additionally, this concept has practical applications in daily life and in developing modern technologies, such as thermal engineering, used to design and manufacture various products (Li, 2022).

Several studies have revealed that students often struggle to understand basic thermal concepts, such as heat transfer, temperature increase, thermal conductivity, and phase changes (GÜLER & ÖZKAYA, 2023; Turgut & Gurbuz, 2012). Maison et al. (2020) found that many students hold incorrect beliefs about the relationship between temperature and objects around them. For instance, there is a misconception that iron in the same room will have a lower temperature than other objects. This misunderstanding likely arises from students' real-life experiences, which do not always align with scientific principles.

A deeper evaluation of students' understanding of thermal concepts is essential to identify and correct these misconceptions. This ensures that students not only understand the theory but can also apply it scientifically in daily life. Such evaluations also help educators design more effective teaching strategies, improving science understanding at the primary and secondary education levels. Chu, Treagust, Yeo, & Zadnik (2012) emphasize the importance of students having a clear and scientific understanding of terms like temperature, heat, and energy (specifically internal energy) so they can explain everyday phenomena related to thermal concepts (Chu et al., 2012).

East Java, one of the provinces with the largest student population in Indonesia, has unique characteristics in the context of science education. According to data from the Central Statistics Agency of East Java for the 2022/2023 academic year, there are over 1,500 senior high schools spread across various regions, both urban and rural, including public and private institutions (Badan Pusat Statistik Provinsi Jawa Timur, 2023). Variations in teaching methods, access to laboratory facilities, and teaching quality significantly influence students' understanding of thermal concepts (Stylos et al., 2023; Villarino, 2023). The region's large student population and diverse socio-economic backgrounds offer a rich context for studying thermal concepts. Furthermore, the research findings in East Java can serve as a reference for addressing similar educational challenges nationwide.

The Rasch model has proven effective in various educational studies (Amiruddin et al., 2024; Asriadi & Hadi, 2021; Handayani et al., 2023). This model can produce objective measurements and provide more detailed diagnostic information about students' abilities and item difficulty levels (Sumintono & Widhiarso, 2015). In evaluating thermal concepts, the Rasch model can help identify patterns of conceptual errors and areas that require special attention in teaching. The unique contribution of using the Rasch model in this research is its ability to produce objective measurements, which allows more accurate identification of students' conceptual error patterns, problem difficulty levels, and students' understanding profiles (Hamdu et al., 2023; Sukarelawan et al., 2024). Unlike approaches that rely on descriptive statistics (Chu et al., 2012; Hadžibegović & Sulejmanović, 2014; Luera et al., 2006), the Rasch model offers evidence-based insights, such as person-item maps and fit statistics, which enhance the evaluation of students' understanding and identify areas requiring instructional intervention. In this study, combining the Rasch model with the TCE instrument addresses research gaps by providing a comprehensive profile of students' understanding of thermal concepts, analyzing different difficulty levels, and identifying gender-based differences.

The Thermal Concept Evaluation (TCE) instrument, developed by Yeo and Zadnik (2001), is reliable for assessing students' understanding of fundamental thermal concepts such as heat, temperature, conduction, and phase changes. Widely used among high school and university students, the TCE has been validated across various educational contexts (Hadžibegović & Sulejmanović, 2014). This instrument has been validated in various educational contexts and demonstrates high reliability in measuring students' understanding of thermal concepts (Stylos et al., 2021). However, most TCE studies remain descriptive and lack diagnostic depth, particularly in analyzing students' understanding based on local characteristics such as gender. This study addresses this gap by integrating the TCE with Rasch model analysis to provide a detailed and diagnostic profile of students' understanding of thermal concepts in East Java.

Combining the TCE instrument with Rasch model analysis provides a comprehensive profile of senior high school student's understanding of thermal concepts, including gender-based differences, in East Java. This approach offers valuable insights for educators and policymakers in designing effective teaching strategies and targeted interventions. The main contribution of this research lies in the Rasch model's ability to identify recurring conceptual errors, assess item difficulty, and analyze students' ability distribution more accurately than traditional methods (Indratno et al., 2023; Sukarelawan et al., 2024). This approach enables a comprehensive understanding of thermal concepts, highlighting gender-based differences and other learning patterns. The findings provide actionable insights for educators and policymakers to develop targeted teaching strategies and improve learning outcomes in physics education by uncovering areas of difficulty and addressing specific weaknesses.

A comprehensive study of senior high school students' understanding of thermal concepts based on gender in East Java, using TCE and Rasch model analysis, has not yet been conducted. This research fills a critical gap in thermal physics education by focusing on East Java's diverse educational context and the gender-based differences in students' understanding. The findings underscore the importance of mastering thermal concepts, which are crucial for real-world applications such as energy systems and advanced technology (Handoyo, 2007; Kruatong et al., 2006). Moreover, it is essential for mastering advanced physics and reducing misconceptions that hinder learning (Stylos et al., 2021). The findings of this research are expected to contribute significantly to improving physics education, particularly in enhancing teaching strategies and addressing misconceptions. Education agencies can use the results to develop conceptual-based curricula, improve laboratory access, and train teachers to identify and correct students' conceptual errors. Furthermore, the insights from East Java, with its diverse educational characteristics, can serve as a model for addressing similar challenges in other regions, ultimately improving the quality of science education at the national level.

2. Method

2.1. Research Design

This study uses a descriptive survey method with a non-experimental cross-sectional design to describe and analyze the understanding of thermal concepts among students in East Java based on gender. Data will be collected at one point through the Thermal Concept Evaluation (TCE) instrument, distributed via Google Forms to students. The obtained data will be analyzed using the Rasch model, which provides a detailed overview of students' understanding levels and identifies differences between male and female students. This research design allows for an in-depth analysis without requiring intervention or special treatment, making mapping student understanding profiles effective.

The descriptive survey method with a non-experimental cross-sectional design was chosen because it offers several advantages in describing and analyzing student understanding. This method allows for efficient data collection from many respondents, providing broad coverage of student understanding across various educational contexts, including different levels and institutions (Brown & Singh, 2021). The cross-sectional design facilitates data collection at a single point to identify common patterns and misconceptions, such as the frequent misconceptions about heat and temperature (Budiarti et al., 2017; Sukarelawan et al., 2019). From a practical standpoint, this method is resource-efficient and easy to implement, making it ideal for large-scale studies in resource-limited environments (Msoma, 2020).

2.2. Participants

This study involved 171 senior high school (SMA) students in East Java, consisting of 94 female and 77 male students, with sample distribution from various public and private schools. The sampling technique used was simple random sampling (SRS), which ensures that each student has an equal chance of being selected and produces unbiased estimates (McEwan, 2020). This random sampling is important to ensure that the research results can be generalized to a broader population, especially given the homogeneity of the student population in East Java (Latpate et al., 2021; Singh, 2003).

The number of samples was determined using the Slovin formula, a common approach for large population sample sizes. This formula considers the total population, margin of error, and confidence level (Sevilla et al., 1993). With a population of high school students in East Java of around 500,000 students (Badan Pusat Statistik Provinsi Jawa Timur, 2024). Sample size calculation using a margin of error of 7.5% (0.075) produces the following sample estimate:

$$\boldsymbol{n} = \frac{\boldsymbol{N}}{\boldsymbol{1+Ne^2}} \tag{1}$$

Where:
$\boldsymbol{n}$ = sample size,
$\boldsymbol{N}$ = population size (500.000)
$\boldsymbol{e}$ = margin of error (7,5%)

A margin of error of 7.5% was chosen to balance resource efficiency and the accuracy of research results, with a confidence level of 87% to 99%, which is sufficient for the context of this study (Majdina et al., 2024). The selection of this margin of error also helps manage resources effectively, as a larger sample size would require more cost and time (Rendón-Macías & Villasís-Keever, 2017) while still providing reliable results (Choi & Tran, 2016; McClure & Lee, 2005).

2.3. Test Instrument

The instrument used in this study was the Thermal Concept Evaluation (TCE), developed by Yeo & Zadnik (2001), to assess the extent to which students possess a scientifically accurate understanding of basic thermal concepts. The TCE consists of 26 multiple-choice questions covering various topics, such as heat, temperature, thermal conduction, and phase changes. This instrument was administered to students online via Google Forms, with a specified time limit. The thermal concepts were selected because they are fundamental topics in physics education under the Kurikulum Merdeka for Phase F.

The TCE (Thermal Concept Evaluation) instrument by Yeo & Zadnik (2001) was chosen for this study because of its reliability and validity in measuring high school student's understanding of thermal concepts, as well as its ability to identify misconceptions and differences in understanding between groups, including by gender (Chu et al., 2012; Yeo & Zadnik, 2001). Previous research has shown that male and female students often experience similar misconceptions regarding thermal concepts despite receiving university-level education (Stylos et al., 2021). The TCE effectively identifies common misconceptions in both groups, highlighting the importance of tailored educational interventions to address gaps in understanding thermal physics (Loverude, 2023). Thus, this instrument allows a more in-depth analysis of gender differences in understanding thermal concepts.

2.4. Data Analysis

The data analysis in this study was conducted using the Rasch Model approach with the assistance of Winsteps 3.73 software. The analysis employed a multiple-choice format, with correct answers scoring one and incorrect answers scoring 0. The steps involved interpreting the data to produce a mapping of students' abilities and the difficulty levels of the items, ensuring the fit between students and the items tested. The following key points were analyzed following the guidelines of Boone & Staver (2020) and Sumintono & Widhiarso (2015).

Person Measure: This measures a student's ability on the logit scale. Using this approach, each student is assigned a score that reflects their ability to understand thermal concepts based on their responses to the given items.

Person Fit: This evaluates whether the students' responses conform to the Rasch model using criteria such as the outfit mean square (MNSQ) within the range of $0.5 < MNSQ < 1.5$, outfit z-standard (ZSTD) within $-2.00 < ZSTD < 2.00$, and point measure correlation (PT Mean Corr) within $0.4 < PT\ Measure < 0.85$.

Wright Map (Person – Item Map): The Wright Map displays the distribution of students' abilities and the difficulty levels of the items on the same logit scale. It allows us to see whether the items are appropriately matched in difficulty to the students' abilities. This helps identify whether any items are too easy or too difficult for most students.

Reliability of Person and Item: This measures the consistency of the measurements within the student and item data. A high Person Reliability indicates a consistent measurement of students' abilities, while a high Item Reliability indicates a consistent measurement of item difficulty. The following is a table showing the criteria used in interpreting person and item reliability:

Table 1. Interpretation Person and Item Reliability

PR/IR Range	Interpretation
$PR/IR \geq 0.95$	Excellent
$0.95 > PR/IR \geq 0.91$	Very Good
$0.91 > PR/IR \geq 0.81$	Good
$0.81 > PR/IR \geq 0.68$	Moderate
$0.68 > PR/IR$	Poor

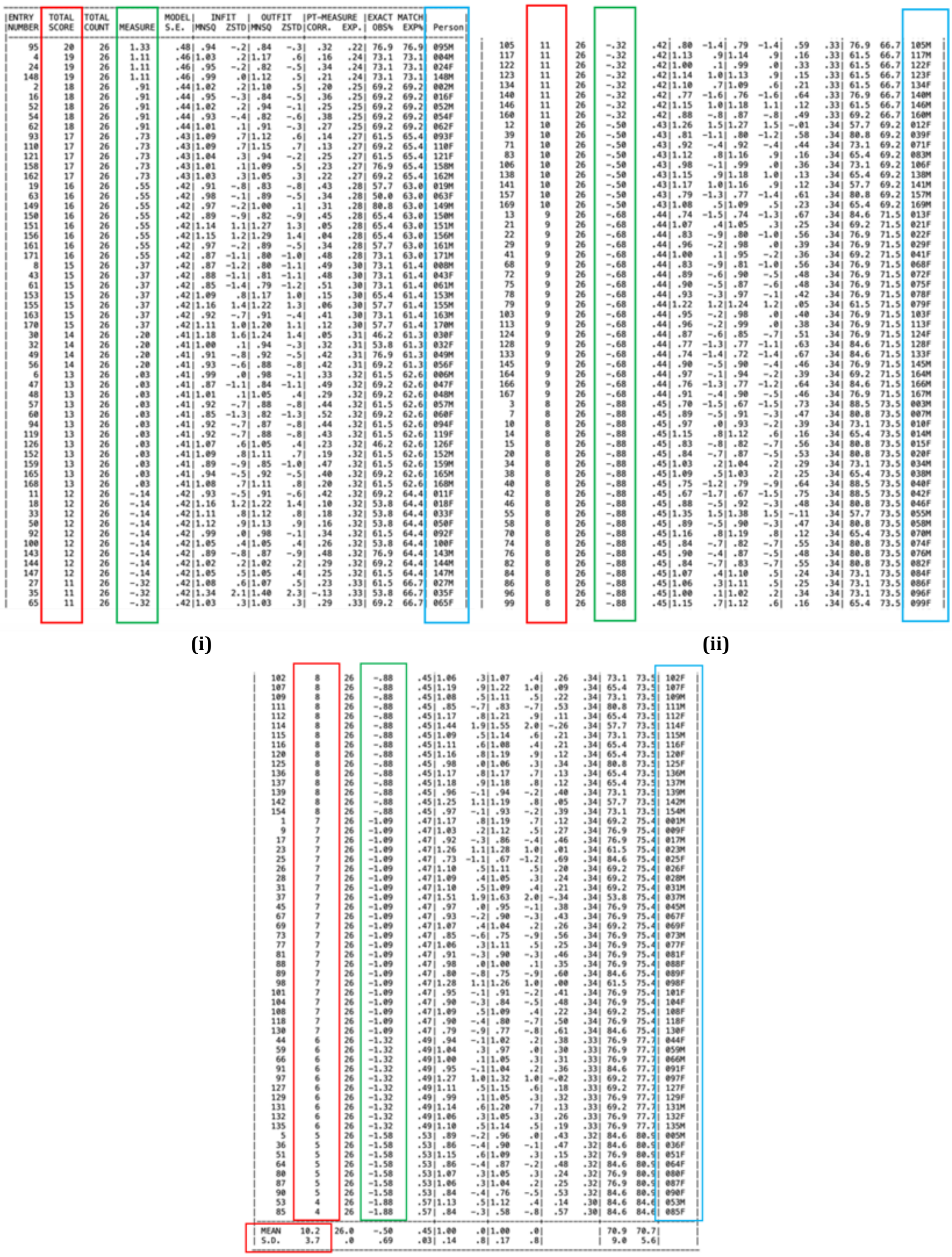
Person Separation and Item Separation: These measures indicate how well the instrument can differentiate students' abilities across multiple levels and how well it can differentiate items based on their difficulty. A high Person Separation value (greater than 1.00) indicates that the respondents' abilities are spread across different levels (Krishnan & Idris, 2014). An Item Separation value greater than 2.00 indicates that the instrument can effectively distinguish at least two groups of ability levels or item difficulty levels (Linacre, 1994).

3. Results and Discussion

3.1. Person Measure (Student Ability Level)

From the results of the Rasch analysis, we can determine students' ability levels by examining their logit values and total scores. A higher logit value indicates that a student can better answer the TCE questions correctly. This corresponds to the total score column, which shows the number of questions a student answered correctly (Sumintono & Widhiarso, 2015).

Figure 1 shows students' abilities based on the person measure table output. In Figure 1, the Total Score column is marked in red, the Measure column in green, and the Person column in blue. From Figure 1, it can be seen that the student with code 095F has a logit value of +1.33 logit (see Measure column) and was able to answer 20 out of 26 questions correctly (see Total Score column), indicating that student 095F has the highest ability in this test. Conversely, students with codes 053M and 085F have logit values of -1.88 and could answer only 4 out of 26 questions correctly, indicating they have the lowest ability levels in this test. For comparison, students with codes 002M, 016F, 052F, 054F, and 062F all have the same logit value of +0.91 logit, and these five students answered 18 out of 26 TCE questions correctly.



(i) (ii)

(iii)

Figure 1. Person measure (student ability) from highest to lowest logit

Additionally, we can compare the ability levels of three students in understanding thermal concepts: students with codes 019M (logit +0.55), 168M (logit +0.03), and 076M (logit -0.88). Student 019M has the highest logit value, indicating better performance than the other two students. The logit difference between 019M and 168M is 0.52, which suggests that the ability of 019M is approximately 1.68 times higher than that of 168M. Furthermore, the logit difference between 019M and 076M is 1.43, meaning that the ability of 019M is about 4.18 times higher than that of 076M. Finally, when comparing 168M and 076M, the logit difference is 0.91, indicating that the ability of 168M is approximately 2.48 times higher than that of 076M. From this comparison, it is evident that 019M has the highest ability level, followed by 168M, and 076M has the lowest ability level.

3.2. Person Fit (Individual Fit Level)

Based on the person fit analysis results from the Winsteps output, as shown in Figure 2, several students displayed response patterns that deviated from the Rasch model's (misfit) expectations. Figure 2 presents the Person Fit table, which evaluates the consistency of individual students' responses against the model's assumptions. According to the outfit MNSQ and ZSTD criteria, the student identified as 037M exhibited the most significant misfit, with an outfit MNSQ value of 1.91 and a ZSTD value of 1.95, both exceeding the acceptable range (0.5 < MNSQ < 1.5). Similarly, the student with the code 114F showed an MNSQ value of 1.73 and a ZSTD value of 2.08, falling outside the acceptable limits. The student labeled 035F also demonstrated misalignment with the model, with an MNSQ value of 1.63 and a ZSTD value of 2.06, indicating a lack of conformity to the Rasch model's assumptions.

Person STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Person
37	7	26	-1.09	.47	1.51	1.9	1.63	2.0	A-.34	.34	53.8	75.4	037M
114	8	26	-.88	.45	1.44	1.9	1.55	2.0	B-.26	.34	57.7	73.5	114F
35	11	26	-.32	.42	1.34	2.1	1.40	2.3	C-.13	.33	53.8	66.7	035F
55	8	26	-.88	.45	1.35	1.5	1.38	1.5	D-.11	.34	57.7	73.5	055M
97	6	26	-1.32	.49	1.27	1.0	1.32	1.0	E-.02	.33	69.2	77.7	097F
156	16	26	.55	.42	1.15	1.2	1.29	1.4	F .04	.28	65.4	63.0	156M
23	7	26	-1.09	.47	1.26	1.1	1.28	1.0	G .01	.34	61.5	75.4	023M
98	7	26	-1.09	.47	1.28	1.1	1.26	1.0	H .00	.34	61.5	75.4	098F
12	10	26	-.50	.43	1.26	1.5	1.27	1.5	I-.01	.34	57.7	69.2	012F
151	16	26	.55	.42	1.14	1.1	1.27	1.3	J .05	.28	65.4	63.0	151M
142	8	26	-.88	.45	1.25	1.1	1.19	.8	K .05	.34	57.7	73.5	142M
79	9	26	-.68	.44	1.22	1.2	1.24	1.2	L .05	.34	61.5	71.5	079F
30	14	26	.20	.41	1.18	1.6	1.24	1.4	M .05	.31	46.2	61.3	030F
155	15	26	.37	.42	1.16	1.4	1.22	1.3	N .06	.30	57.7	61.4	155M
107	8	26	-.88	.45	1.19	.9	1.22	1.0	O .09	.34	65.4	73.5	107F
18	12	26	-.14	.42	1.16	1.2	1.22	1.4	P .10	.32	53.8	64.4	018F
112	8	26	-.88	.45	1.17	.8	1.21	.9	Q .11	.34	65.4	73.5	112F
131	6	26	-1.32	.49	1.14	.6	1.20	.7	R .13	.33	69.2	77.7	131M
170	15	26	.37	.42	1.11	1.0	1.20	1.1	S .12	.30	57.7	61.4	170M
120	8	26	-.88	.45	1.16	.8	1.19	.9	T .12	.34	65.4	73.5	120F
70	8	26	-.88	.45	1.16	.8	1.19	.8	U .12	.34	65.4	73.5	070M
1	7	26	-1.09	.47	1.17	.8	1.19	.7	V .12	.34	69.2	75.4	001M
146	11	26	-.32	.42	1.15	1.0	1.18	1.1	W .12	.33	61.5	66.7	146M
138	10	26	-.50	.43	1.15	.9	1.18	1.0	X .13	.34	65.4	69.2	138M
137	8	26	-.88	.45	1.18	.9	1.18	.8	Y .12	.34	65.4	73.5	137M
153	15	26	.37	.42	1.09	.8	1.17	1.0	Z .15	.30	65.4	61.4	153M
BETTER FITTING OMITTED													
171	16	26	.55	.42	.87	-1.1	.80	-1.0	.48	.28	73.1	63.0	171M
8	15	26	.37	.42	.87	-1.2	.80	-1.1	z .49	.30	73.1	61.4	008M
60	13	26	.03	.41	.85	-1.3	.82	-1.3	y .52	.32	69.2	62.6	060F
61	15	26	.37	.42	.85	-1.4	.79	-1.2	x .51	.30	73.1	61.4	061M
111	8	26	-.88	.45	.85	-.7	.83	-.7	w .53	.34	80.8	73.5	111M
73	7	26	-1.09	.47	.85	-.6	.75	-.9	v .56	.34	76.9	75.4	073M
90	5	26	-1.58	.53	.84	-.4	.76	-.5	u .53	.32	84.6	80.9	090F
85	4	26	-1.88	.57	.84	-.3	.58	-.8	t .57	.30	84.6	84.6	085F
74	8	26	-.88	.45	.84	-.7	.82	-.7	s .55	.34	80.8	73.5	074F
82	8	26	-.88	.45	.84	-.7	.83	-.7	r .55	.34	80.8	73.5	082F
68	9	26	-.68	.44	.83	-.9	.81	-1.0	q .56	.34	76.9	71.5	068F
22	9	26	-.68	.44	.83	-.9	.80	-1.0	p .56	.34	76.9	71.5	022F
15	8	26	-.88	.45	.83	-.8	.82	-.7	o .56	.34	80.8	73.5	015F
39	10	26	-.50	.43	.81	-1.1	.80	-1.2	n .58	.34	80.8	69.2	039F
105	11	26	-.32	.42	.80	-1.4	.79	-1.4	m .59	.33	76.9	66.7	105M
89	7	26	-1.09	.47	.80	-.8	.75	-.9	l .60	.34	84.6	75.4	089F
157	10	26	-.50	.43	.79	-1.3	.77	-1.4	k .61	.34	80.8	69.2	157M
130	7	26	-1.09	.47	.79	-.9	.77	-.8	j .61	.34	84.6	75.4	130F
40	8	26	-.88	.45	.75	-1.2	.79	-.9	i .64	.34	88.5	73.5	040F
128	9	26	-.68	.44	.77	-1.3	.77	-1.1	h .63	.34	84.6	71.5	128F
140	11	26	-.32	.42	.77	-1.6	.76	-1.6	g .64	.33	76.9	66.7	140M
166	9	26	-.68	.44	.76	-1.3	.77	-1.2	f .64	.34	84.6	71.5	166M
13	9	26	-.68	.44	.74	-1.5	.74	-1.3	e .67	.34	84.6	71.5	013F
133	9	26	-.68	.44	.74	-1.4	.72	-1.4	d .67	.34	84.6	71.5	133F
25	7	26	-1.09	.47	.73	-1.1	.67	-1.2	c .69	.34	84.6	75.4	025F
3	8	26	-.88	.45	.70	-1.5	.67	-1.5	b .73	.34	88.5	73.5	003M
42	8	26	-.88	.45	.67	-1.7	.67	-1.5	a .75	.34	88.5	73.5	042F
MEAN	10.2	26.0	-.50	.45	1.00	.0	1.00	.0			70.9	70.7	
S.D.	3.7	.0	.69	.03	.14	.8	.17	.8			9.0	5.6	

Figure 2. Person Fit

The high ZSTD values indicate a significant mismatch between the student's responses and the Rasch model, which could be attributed to factors such as uncertainty or inaccuracies in answering, misconceptions, overconfidence, or a misalignment between the student's ability level and the difficulty of the items being tested (Edwards & Alcock, 2010; Hermann-Abell & DeBoer, 2016; Nopiah et al., 2012; Paek et al., 2013). The outfit MNSQ, ZSTD, and PT Measure Corr values that fall outside the expected range suggest that the student's response patterns need to be more fully consistent with the assumptions of the Rasch model. In such cases, further analysis of the student's response patterns is required to determine the cause of this misfit—whether it is due to the student's understanding of the items, external factors such as anxiety, or other variables that may influence their performance during the test (Edwards & Alcock, 2010; Planinic et al., 2019).

3.3. Wright Map (Person – Item Map)

Using Rasch analysis, the distribution of students' abilities and the difficulty levels of TCE items can be visualized through a map known as the Wright Map, as shown in Figure 3. According to Salman & Aziz (2015, cited in Asriadi & Hadi, 2021), the Wright Map in Figure 3 illustrates the relationship between test participants' abilities and item difficulty levels on the same scale. This representation also provides insights into students' readiness by aligning item challenges with students' abilities (Asriadi & Hadi, 2021). In Figure 3, the Wright Map demonstrates that students with higher abilities are positioned above the more difficult items. In comparison, those with lower abilities appear below the easier items, effectively highlighting the balance between item difficulty and student performance levels.

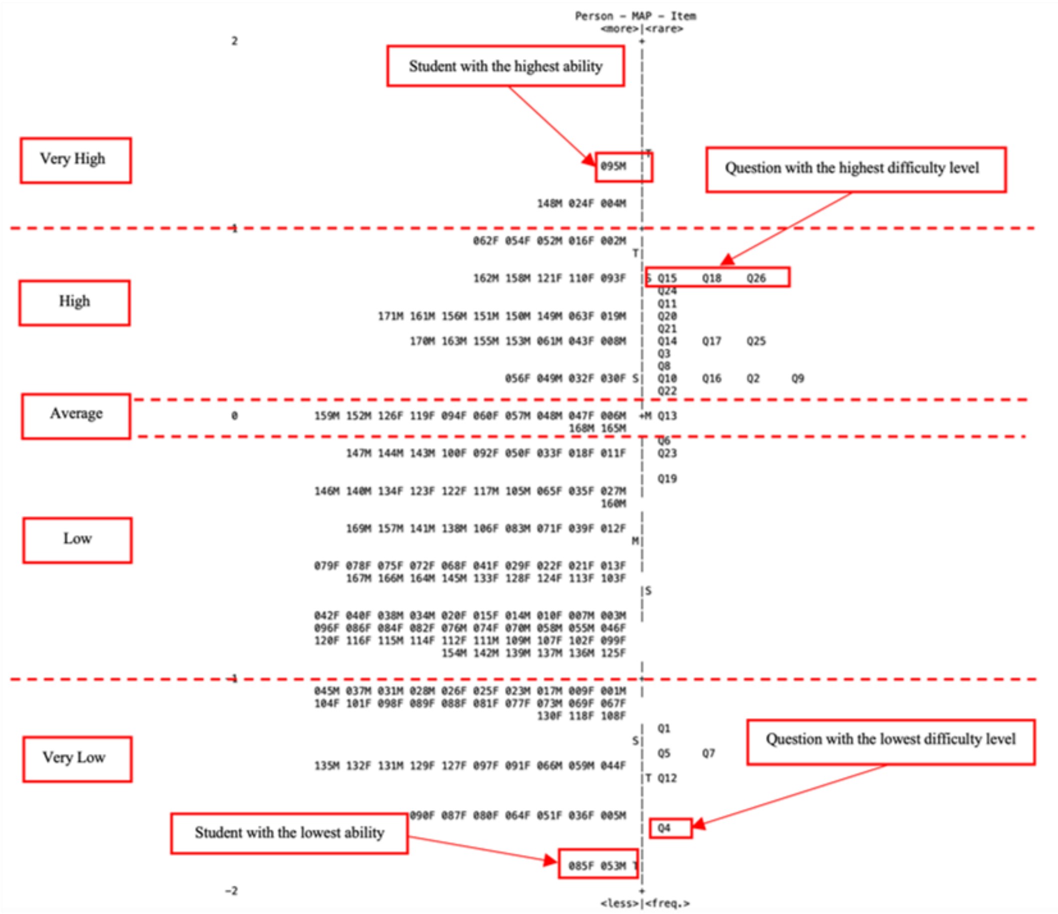


Figure 3. Wright Map (Person - Item Map)

In Figure 3, the map on the left shows nine students (095M, 004M, 024F, 148M, 002M, 016F, 052M, 054F, and 062F) who exhibit high ability in answering questions related to the understanding of thermal concepts. These nine students completed all the items at the given difficulty levels, earning the maximum score. According to the Wright Map, it is evident that some students share the same ability level, such as students with the codes 019M, 063F, 149M, 150M, 151M, 156M, 161M, and 171M, all of whom are positioned at a logit of +0.55. These eight students could correctly answer 16 items but needed help with the more complicated items (logit +0.63). In the lower-left section of the Wright Map, two students (053M and 082F) demonstrated a low ability to understand thermal concepts, as they needed help to answer the most straightforward question with a logit of -1.68 (Q4). This suggests teachers need special attention, as traditional teaching approaches appear less effective for students with low ability in thermal concepts (Sumintono & Widhiarso, 2015).

Additionally, Figure 3 also shows that the items with the highest difficulty levels are Q15, Q18, and Q26, with logits of +0.76, +0.73, and +0.73, respectively, followed by Q24 (+0.66), Q11 (+0.63), Q20 (+0.53), Q21 (+0.43), Q14 (+0.40), Q17 (+0.40), Q25 (+0.37), Q3 (+0.34), Q8 (+0.28), Q9 (+0.22), Q10 (+0.22), Q16 (+0.20), Q2 (+0.17), and Q22 (+0.14). These items are classified as complex and were only completed by the nine high-ability students: 095M, 004M, 024F, 148M, 002M, 016F, 052M, 054F, and 062F. This means that only approximately 5.26% of students could answer these difficult questions. On the other hand, eight items with logits below zero indicate that these items are classified as easy. The easiest item is Q4, with a logit of -1.68, followed by Q12, Q7, Q5, Q1, Q19, Q23, and Q6, with logits of -1.37, -1.28, -1.25, -1.14, -0.30, -0.14, and -0.08, respectively. According to Darmana et al. (2021, cited in Amiruddin et al., 2024), logit values reflect the difficulty level of an item; the higher the logit, the more difficult the item (Amiruddin et al., 2024).

From the analysis of the Wright Map, it is evident that most students fall within the low to very-low ability categories in understanding thermal concepts. This indicates the need for intervention in the learning process to enhance students' understanding of thermal concepts. Evaluating the teaching methods is also recommended to deepen the overall understanding of thermal concepts.

The percentage of student ability based on gender can be seen in Table 2, which provides a detailed overview of the distribution of ability levels within each gender group. This table categorizes students' abilities into five levels: Very High, High, Average, Low, and Very Low, offering a comprehensive breakdown of how male and female students differ in their understanding of thermal concepts. The data presented in Table 2 serves as a valuable reference for analyzing gender-based patterns in student performance and identifying specific areas that require targeted educational interventions.

Table 2. Percentage of Students' Ability Levels by Gender

Gender	Level of Ability (%)					Total (%)	Freq.
	Very High	High	Average	Low	Very Low		
Male	3.90	23.38	9.09	45.45	18.18	100	77
Female	1.06	11.70	5.32	52.13	29.79	100	94

Based on Table 2, there is a significant difference in the ability of male and female students to understand thermal concepts, with most students falling into the low to very low ability categories. A higher percentage of female students are in these categories, and only a tiny proportion achieve very high ability levels (3.90% of males and 1.06% of females). This uneven distribution highlights students' challenges in grasping thermal concepts and underscores the need for innovative teaching strategies to enhance understanding, particularly for female students.

One practical approach is inquiry-based learning, which encourages students to explore concepts through hands-on experiments and in-depth questioning (Wartono et al., 2019). This strategy helps students build their understanding through direct experience, making it particularly engaging for female students in the lower ability categories. Additionally, project-based learning allows students to connect thermal concepts to real-world situations, fostering motivation and emphasizing the relevance of physics in everyday life (Otoluwa et al., 2024). Problem-based learning (PBL) can also develop students' analytical and problem-solving skills by challenging them to solve real-world thermal issues, such as reducing heat loss in household systems (Muthaharoh & Sukarelawan, 2023; Seibert, 2021). Teachers should tailor their strategies to minimize gender gaps by incorporating real-life contexts, such as cooking or weather, to engage female students. Meanwhile, technical and engineering-focused applications, like designing thermal insulation, could appeal more to male students. Advanced educational technologies, such as interactive simulations and augmented reality, can also support these strategies by visualizing abstract concepts like heat transfer and equilibrium, making them easier to understand. These tools enhance students' engagement and help bridge the gap between theoretical and practical learning (Malik et al., 2022; Masrifah & Amiroh, 2023; Nugraheni & Mundilarto, 2022; Nuryantini et al., 2022).

By integrating student-centered approaches with technology-based media, it is possible to address gender gaps and improve overall comprehension of thermal concepts. Combining inquiry-based, project-based, and problem-based learning with tools like simulations, animations, and e-modules enables more interactive and engaging learning experiences. These strategies provide a multifaceted framework to enhance students' understanding of thermal physics, reduce misconceptions, and foster critical thinking, particularly for lower-ability categories (Nusir et al., 2013; Perdana et al., 2022).

3.4. Reliability and Separation of Person and Item

The reliability and separation of students and test items are displayed in summary statistics for the overall instrument analysis. These summary statistics are shown in Figure 4, which provides a detailed overview of the reliability and separation values for both students (person) and test items, highlighting the consistency and effectiveness of the instrument in measuring student abilities and item difficulties. The figure shows that the reliability value for students (person) is 0.56, indicating a low level of reliability in measuring students' consistency in the test. This value suggests considerable variation in the students' response patterns, possibly due to uncertainty or difficulty in answering the given items. In contrast, the reliability for test items is 0.94, indicating very good reliability. This suggests that the items consistently measure the intended ability, namely understanding thermal concepts. In other words, the items in this test have a stable level of difficulty and can be relied upon to measure student ability consistently.

SUMMARY OF 171 MEASURED Person

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	10.2	26.0	-.50	.45	1.00	.0	1.00	.0
S.D.	3.7	.0	.69	.03	.14	.8	.17	.8
MAX.	20.0	26.0	1.33	.57	1.51	2.1	1.63	2.3
MIN.	4.0	26.0	-1.88	.41	.67	-1.7	.58	-1.6

REAL RMSE .46 TRUE SD .52 SEPARATION 1.13 Person RELIABILITY .56
MODEL RMSE .45 TRUE SD .53 SEPARATION 1.18 Person RELIABILITY .58
S.E. OF Person MEAN = .05

Person RAW SCORE-TO-MEASURE CORRELATION = 1.00
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .62

SUMMARY OF 26 MEASURED Item

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	66.9	171.0	.00	.17	1.00	.0	1.00	.0
S.D.	25.7	.0	.71	.01	.11	1.4	.14	1.4
MAX.	127.0	171.0	.76	.19	1.24	2.9	1.28	2.7
MIN.	41.0	171.0	-1.68	.16	.83	-2.6	.78	-2.6

REAL RMSE .18 TRUE SD .69 SEPARATION 3.87 Item RELIABILITY .94
MODEL RMSE .17 TRUE SD .69 SEPARATION 3.96 Item RELIABILITY .94
S.E. OF Item MEAN = .14

Figure 4. Reliability and Separation of Person and Item

Although the Cronbach's Alpha for the entire test is 0.62, which indicates moderate reliability, it falls within an acceptable range for exploratory studies (Hayat, 2024; Malapane & Ndlovu, 2024). The low reliability for students points to potential gaps in their understanding, which may arise from instructional shortcomings. It is essential to address these gaps by improving teaching methods to ensure that students have a clearer comprehension of the material being tested. Studies have shown that inconsistencies in student responses often reflect a lack of alignment between teaching approaches and assessment goals (Clayson, 2018).

The disparity between person reliability and item reliability is a common occurrence in educational assessments. While item reliability tends to be high due to the consistent nature of the questions, person reliability can be low due to individual differences in response patterns. Factors such as students' education level, awareness, or fluctuating confidence can contribute to inconsistent responses (Fekken et al., 1999; Tourangeau et al., 2020; Zou & Bolt, 2023). This highlights the need to address individual differences through tailored instructional strategies that consider diverse learning styles and cognitive processes.

Person reliability and item reliability often diverge because of the different factors influencing them. Person reliability is affected by individual response patterns, which may vary due to differences in education, awareness, or personal characteristics, such as response style or self-perception. Conversely, item reliability tends to remain high because of the stability of fact-based or limited-response questions, which yield more consistent results across populations (Hartig et al., 2007; Tourangeau et al., 2020).

The discrepancies in reliability highlight the need for strategies to enhance assessment validity. Educators should employ diverse assessment methods and ensure that test items align with desired learning outcomes. Additionally, analyzing student problem-solving patterns, including response times, can provide insights into cognitive processes and help develop adaptive learning strategies tailored to students' needs. These steps can improve students' comprehension of the material and reduce measurement errors (Park et al., 2023; Pipia, 2014).

Subsequently, the separation analysis results from the Wright Maps indicate a person separation value of 1.13 and an item separation value of 3.87. A person separation value of 1.13 suggests that the test effectively differentiates various student ability levels, categorizing them into at least two ability groups. However, there is still room for further differentiation. This value indicates that the instrument can effectively separate respondents with varying abilities (greater than 1.00) (Krishnan & Idris, 2014). On the other hand, the item separation value of 3.87 demonstrates the instrument's excellent ability to differentiate levels of item difficulty, surpassing the reliability criteria and indicating that the test items can effectively distinguish more than two levels of difficulty (greater than 2.00) (Linacre, 1994). These values reflect the reliability and effectiveness of the instrument in measuring student ability and item difficulty, with person separation showing the test's capacity to differentiate student abilities and item separation indicating that the test items are well-structured to assess the mastery of the concepts being tested (Bintang & Suprananto, 2024; Ummah et al., 2022).

3.5. Analysis of Understanding of Thermal Physics Concepts Based on Gender

In the TCE instrument developed by Yeo & Zadnik (2001), students' alternative conceptions are categorized into four groups: 1) Heat transfer and temperature changes, 2) Boiling, 3) Heat conductivity and equilibrium, and 4) Freezing and melting. Table 3 below presents the percentage of correct student responses for the items within these four conceptual groups, categorized by gender.

Table 3. The Correct Responses in the Four Thermal Concept Groups Based on Gender

Item	No.	Percentage of correct response (%)			
		M	F	Overall	Group Mean
Heat transfer and temperature changes					
Q3	1	48.1	18.1	33.1	36.5
Q7	2	75.3	59.6	67.4	
Q9	3	33.8	34.0	33.9	
Q10	4	45.5	24.5	35.0	
Q13	5	53.2	26.6	39.9	
Q15	6	26.0	22.3	24.2	
Q20	7	35.1	22.3	28.7	
Q21	8	39.0	22.3	30.7	
Q22	9	44.2	28.7	36.4	
Q23	10	40.3	42.6	41.4	
Q25	11	33.8	28.7	31.2	
Boiling					
Q3	12	72.7	75.5	74.1	58.9
Q7	13	58.4	72.3	65.4	
Q9	14	48.1	34.0	41.0	
Q10	15	63.6	72.3	68.0	
Q13	16	54.5	37.2	45.9	
Heat conductivity and equilibrium					
Q14	17	27.3	33.0	30.1	28.3
Q16	18	31.2	37.2	34.2	
Q17	19	32.5	28.7	30.6	
Q18	20	19.5	28.7	24.1	
Q24	21	29.9	22.3	26.1	
Q26	22	27.3	22.3	24.8	
Freezing and melting					
Q1	23	64.9	62.8	63.9	39.6
Q2	24	44.2	27.7	35.9	
Q8	25	35.1	30.9	33.0	
Q11	26	18.2	33.0	25.6	

Based on the data in Table 3, analyzed using the student conceptual understanding levels as outlined by Apridonata et al. (2022), students' understanding of thermal physics concepts can be classified into three levels: low (≤30%), moderate (30% < x ≤ 60%), and high (60% < x ≤ 100%) (Apridonata et al., 2022). Gender-based analysis reveals that male students perform better than female students in most categories, although overall understanding across both genders remains predominantly low to moderate. This highlights significant gaps in thermal concept comprehension that require targeted interventions.

In the Heat Transfer and Temperature Changes concept, the overall understanding level is categorized as low, with an average percentage of 36.5%. Male students demonstrate better comprehension, achieving moderate to high scores on specific questions, such as Q7 (75.9%) and Q9 (39.4%). Female students struggle more, with scores often below 30%, such as Q3 (18.1%) and Q10 (24.5%). This indicates that while male students generally understand this concept better, both genders face challenges in achieving higher comprehension levels.

The Boiling concept shows a moderate overall understanding level, with an average percentage of 58.9%. Female students slightly outperform male students on several questions, such as Q4 (75.5%), Q5 (72.3%), and Q12 (72.3%), which fall into the high understanding category. This suggests that female students may find the Boiling concept more relatable or accessible, although both genders show a reasonable level of comprehension compared to other concepts.

The Heat Conductivity and Equilibrium concept is the most challenging, with an average understanding level categorized as low (28.3%). Both male and female students exhibit difficulty, with low percentages of correct responses, such as Q18 (19.5% for males and 28.7% for females) and Q24 (29.9% for males and 22.3% for females). This highlights the need for improved teaching strategies and deeper conceptual scaffolding to address gaps in this area.

The Freezing and Melting concept has an average understanding level of moderate, at 39.6%. Male students perform slightly better, achieving a high understanding of Q1 (64.9%), while female students score highly on the same question (62.8%). However, for other questions like Q2 and Q11, the understanding levels drop into the low and moderate categories for both genders. While this concept shows better comprehension than Heat Conductivity and Equilibrium, the persistent gender-based gaps warrant further attention.

The findings reveal that male students perform better in most thermal physics concepts, particularly in Heat Transfer and Temperature Changes. In contrast, female students show a slight advantage in the Boiling concept. However, the persistent low levels of understanding in areas like Heat Conductivity and Equilibrium suggest the need for targeted instructional approaches to improve comprehension across genders.

The observed gender differences in understanding thermal concepts align with previous research. Stylos (2021) identified similar gender-related disparities, with male students often performing better in concepts requiring spatial reasoning, such as Heat Transfer. On the other hand, female students excel in tasks requiring meticulous observation and contextual interpretation, as seen in their performance in the Boiling concept. This highlights the role of cognitive and experiential factors in shaping gender-based performance in physics.. Similarly, Al-Khatri et al. (2020) found distinct differences in thermal comfort experiences between genders, which could suggest a cognitive variance in how thermal phenomena are perceived and processed.

The conceptual gaps in Heat Conductivity and Equilibrium reflect more profound challenges in understanding abstract processes such as energy transfer at the molecular level. Traditional teaching methods may fail to address these abstract concepts adequately, as students often rely on macroscopic observations rather than microscopic principles. This aligns with findings by Nottis et al. (2017), who noted that students frequently misunderstand heat as a substance rather than a transfer process. To address these challenges, instructional strategies must cater to diverse learning styles and bridge gender gaps. Incorporating real-world contexts, such as cooking or weather phenomena, can make the material more relatable, particularly for female students. For male students, technical applications like designing thermal insulation systems can provide engaging challenges. These approaches can ensure that both genders find relevance in thermal physics concepts. Visual tools such as bar charts, as referenced in Figure 5, can help educators identify trends in understanding and highlight areas needing improvement. For example, higher scores in the Boiling concept suggest a better grasp of this topic among both genders.

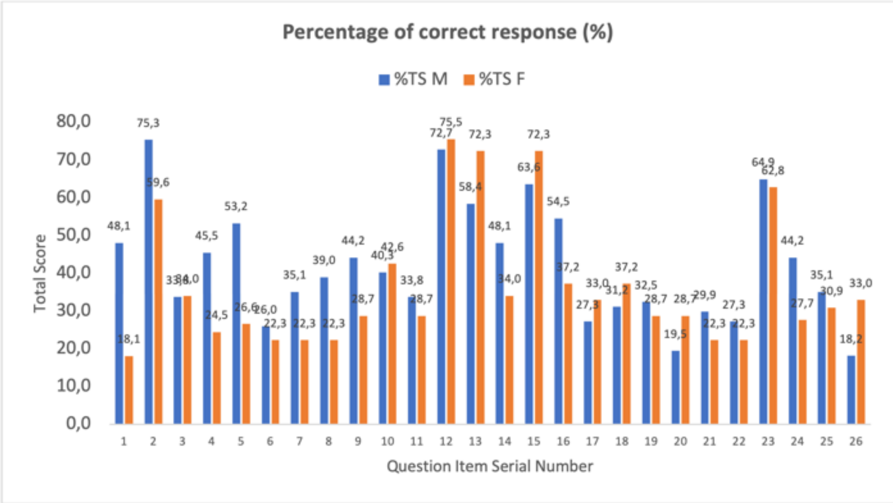


Figure 5. Percentage of Correct Responses in the Four Thermal Concept Groups Based on Gender

4. Conclusion

This study reveals that high school students in East Java exhibit low to moderate levels of understanding of thermal concepts, with significant misconceptions observed in key areas such as heat conductivity and thermal equilibrium, which students find particularly challenging. Gender-based analysis shows that male students perform better in heat transfer and temperature changes, while female students demonstrate a slightly higher understanding of boiling concepts. Using the Rasch model, this research effectively identifies error patterns, item difficulty levels, and conceptual weaknesses, providing valuable insights into students' learning needs and enabling targeted interventions. To address these challenges, educators are encouraged to adopt instructional strategies that focus on correcting misconceptions through visual explanations and hands-on experiments that bridge the gap between abstract concepts and real-world phenomena. Technology-based learning tools like interactive simulations should also be integrated to enhance students' engagement and understanding. Developing conceptual-based curricula, improving laboratory access, and providing teacher training programs are critical to minimizing gender-based gaps and fostering a more comprehensive mastery of thermal concepts. These findings offer a strong foundation for shaping effective teaching strategies and educational policies to improve physics education and equip students with the skills needed for advanced technological applications.

Author Contributions

All authors have equal contributions to the paper. All the authors have read and approved the final manuscript.

Funding

No funding support was received.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgement

The author would like to sincerely thank Lembaga Pengelola Dana Pendidikan (LPDP) for their scholarship support, which made this research possible. Special thanks are also extended to the high school students in East Java who participated in this study and to all those who provided moral and technical support throughout the research process.

References

Al-Khatri, H., Alwetaishi, M., & Gadi, M. B. (2020). Exploring thermal comfort experience and adaptive opportunities of female and male high school students. *Journal of Building Engineering*, *31*, 101365. https://doi.org/10.1016/j.jobe.2020.101365

Amiruddin, M. Z. Bin, Samsudin, A., Coştu, B., Suhandi, A., Kaniawati, I., & Apriliyanti, N. (2024). Impressive Analysis of High School Students' Abilities in Physics Measurement Materials of the New Indonesian-Paradigm Curriculum Via the Rasch Model. *JIPF (Jurnal Ilmu Pendidikan Fisika)*, *9*(2), 253–265. https://doi.org/10.26737/jipf.v9i2.4808

Apridonata, W., Syarkowi, A., & Purwanto, A. (2022). Profil Aspek Interpretasi Pemahaman Konsep Materi Kinematika Gerak Lurus pada Siswa SMAN 6 Kota Bengkulu. *Jurnal Ilmiah Pendidikan Fisika*, *6*(3), 591–600. https://doi.org/10.20527/jipf.v6i3.6730

Asriadi, M., & Hadi, S. (2021). Analysis of the Quality of the Formative Test Items for Physics Learning Using the Rasch Model in the 21st Century Learning. *JIPF (Jurnal Ilmu Pendidikan Fisika)*, *6*(2), 158–166. https://doi.org/10.26737/jipf.v6i2.2030

Badan Pusat Statistik Provinsi Jawa Timur. (2023, July 20). *Jumlah Sekolah Menengah Atas (SMA) di Bawah Kementerian Pendidikan dan Kebudayaan Menurut Kabupaten/Kota di Provinsi Jawa Timur, 2021/2022 dan 2022/2023*. Pendidikan.

Badan Pusat Statistik Provinsi Jawa Timur. (2024, February 21). *Jumlah Sekolah, Guru, dan Murid Sekolah Menengah Atas (SMA) di Bawah Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Menurut Kabupaten/Kota di Provinsi Jawa Timur, 2023/2024*. Pendidikan - Tabel Statistik.

Baker, B. (2022). *Critical Thinking in the Physics Curriculum*. AIP Publishing LLC. https://doi.org/10.1063/9780735424234

Bintang, R. S., & Suprananto, S. (2024). The Impact of Sample Size, Test Length, and Person-Item Targeting on the Separation Reliability in Rasch Model: A Simulation Study. *JP3I (Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia)*, *13*(1), 1–13. https://doi.org/10.15408/jp3i.v13i1.27975

Boone, W. J., & Staver, J. R. (2020). *Advances in Rasch Analyses in the Human Sciences*. Springer International Publishing. https://doi.org/10.1007/978-3-030-43420-5

Brown, B., & Singh, C. (2021). Development and validation of a conceptual survey instrument to evaluate students' understanding of thermodynamics. *Physical Review Physics Education Research*, *17*(1), 010104. https://doi.org/10.1103/PhysRevPhysEducRes.17.010104

Budiarti, I. S., Suparmi, I., Cari, A., Viyanti, V., Winarti, C., & Handhika, J. (2017). Student's understanding consistency of thermal conductivity concept. In A. G. Abdullah, I. Hamidah, S. Aisyah, A. A. Danuwijaya, G. Yuliani, & H. S. H. Munawaroh (Eds.), *Ideas for 21st Century Education* (1st Edition, p. 4). Routledge.

Choi, S. W., & Tran, D. H. D. (2016). Calculating sample size. Alpha, Beta, Mu and Sigma – it's all Greek. *Anaesthesia*, *71*(4), 462–464. https://doi.org/10.1111/anae.13369

Chu, H.-E., Treagust, D. F., Yeo, S., & Zadnik, M. (2012). Evaluation of Students' Understanding of Thermal Concepts in Everyday Contexts. *International Journal of Science Education*, *34*(10), 1509–1534. https://doi.org/10.1080/09500693.2012.657714

Clayson, D. E. (2018). Student evaluation of teaching and matters of reliability. *Assessment & Evaluation in Higher Education*, *43*(4), 666–681. https://doi.org/10.1080/02602938.2017.1393495

Edwards, A., & Alcock, L. (2010). Using Rasch analysis to identify uncharacteristic responses to undergraduate assessments. *Teaching Mathematics and Its Applications*, *29*(4), 165–175. https://doi.org/10.1093/teamat/hrq008

Fekken, G. C., Holden, R. R., & Cotton, D. H. G. (1999). Person reliability of psychiatric patients' responses to a psychopathology inventory. *Journal of Clinical Psychology*, *55*(10), 1299–1306. https://doi.org/10.1002/(SICI)1097-4679(199910)55:10<1299::AID-JCLP10>3.0.CO;2-Z

GÜLER, M., & ÖZKAYA, A. (2023). Content Analysis of Studies Conducted Between 2007- 2022 on Misconceptions about the Heat and Temperature. *International Journal of Education Technology and Scientific Researches*, *8*(24), 2134–2151. https://doi.org/10.35826/ijetsar.666

Gunawan, G., Harjono, A., Hermansyah, H., & Herayanti, L. (2019). Guided Inquiry Model Through Virtual Laboratory to Enhance Students' Science Process Skills on Heat Concept. *Jurnal Cakrawala Pendidikan*, *38*(2), 259–268. https://doi.org/10.21831/cp.v38i2.23345

Hadžibegović, Z., & Sulejmanović, S. (2014). Fundamental thermal concepts understanding: the first-year chemistry student questionnaire results. *Bulletin of the Chemists and Technologists of Bosnia and Herzegovina*, *42*, 21–30.

Hamdu, G., Hadiana, D., Sylvia, N., Apipatunnisa, I., & Yulianto, A. (2023). Measuring Changes of Students Conceptual Understanding of Literacy and Numeracy in Natural Science by Using Rasch Model. *Jurnal Ilmiah Sekolah Dasar*, *7*(3), 489–497. https://doi.org/10.23887/jisd.v7i3.59943

Handayani, Y., Rahmawati, R., & Widiasih, W. (2023). Using Rasch Model to Analyze Reliability and Validity of Concept Mastery Test on Electricity and Magnetism Topic. *JIPF (Jurnal Ilmu Pendidikan Fisika)*, *8*(2), 226–239. https://doi.org/10.26737/jipf.v8i2.3877

Handoyo, E. A. (2007). The Interesting of Learning Thermodynamics Through Daily Life. *Maranatha Teaching and Learning International Conference 2007 "Teaching and Learning in Higher Education for Developing Countries,"* 151–158.

Hartig, J., Hölzel, B., & Moosbrugger, H. (2007). A Confirmatory Analysis of Item Reliability Trends (CAIRT): Differentiating True Score and Error Variance in the Analysis of Item Context Effects. *Multivariate Behavioral Research*, *42*(1), 157–183. https://doi.org/10.1080/00273170701341266

Hayat, B. (2024). Meta-Analysis of Coefficient Alpha: Empirical Demonstration Using English Language Teaching Reflection Inventory. *Journal of Curriculum and Teaching*, *13*(2), 231–243. https://doi.org/10.5430/jct.v13n2p231

Hermann-Abell, C. F., & DeBoer, G. E. (2016). Using Rasch Modeling and Option Probability Curves to Diagnose Students' Misconceptions. *American Association for the Advancement of Science (AAAS)*, 1–12.

Indratno, T. K., Sukarelawan, Moh. I., Puspitasari, A. D., & Soeharto, S. (2023). Students are Not Sure about Their Conceptual Understanding: A Comparative Study of the Level of Conceptual Understanding and the Level of Confidence Using Rasch Modeling. *Indonesian Review of Physics*, *6*(2), 75–81. https://doi.org/10.12928/irip.v6i2.6901

Krishnan, S., & Idris, N. (2014). Investigating Reliability and Validity for the Construct of Inferential Statistics. *International Journal of Learning, Teaching and Educational Research*, *4*(1), 51–60.

Kruatong, T., Sung-ong, S., Singh, P., & Jones, A. (2006). Thai High School Students' Understanding of Heat and Thermodynamics. *Kasetsart Journal of Social Sciences*, *27*(2), 321–330.

Latpate, R., Kshirsagar, J., Gupta, V. K., & Chandra, G. (2021). Simple Random Sampling. In R. Latpate, J. Kshirsagar, V. K. Gupta, & G. Chandra (Eds.), *Advanced Sampling Methods* (pp. 11–35). Springer Singapore. https://doi.org/10.1007/978-981-16-0622-9_2

Li, X. (2022). Field grand challenge for thermal engineering. *Frontiers in Thermal Engineering*, *2*, 1–7. https://doi.org/10.3389/fther.2022.1045838

Linacre, J. M. (1994). *Many-Facet Rasch Measurement (MFRM)* (Second Edition). MESAPRESS.

Loverude, M. E. (2023). Student Understanding of Thermal Physics. In *The International Handbook of Physics Education Research: Learning Physics* (pp. 3-1-3–38). AIP Publishing LLCMelville, New York. https://doi.org/10.1063/9780735425477_003

Luera, G. R., Otto, C. A., & Zitzewitz, P. W. (2006). Use of the Thermal Concept Evaluation to Focus Instruction. *The Physics Teacher*, *44*(3), 162–166. https://doi.org/10.1119/1.2173324

Maison, M., Safitri, I. C., & Wardana, R. W. (2020). Identification of Misconception of High School Students on Temperature and Calor Topic using Four-Tier Diagnostic Instrument. *EDUSAINS*, *11*(2), 195–202. https://doi.org/10.15408/es.v11i2.11465

Majdina, N. I., Pratikno, B., & Tripena, A. (2024). Penentuan Ukuran Sampel Menggunakan Rumus Bernoulli dan Slovin: Konsep dan Aplikasinya. *Jurnal Ilmiah Matematika Dan Pendidikan Matematika*, *16*(1), 73. https://doi.org/10.20884/1.jmp.2024.16.1.11230

Malapane, T. A., & Ndlovu, N. K. (2024). Assessing the Reliability of Likert Scale Statements in an E-Commerce Quantitative Study: A Cronbach Alpha Analysis Using SPSS Statistics. *2024 Systems and Information Engineering Design Symposium (SIEDS)*, 90–95. https://doi.org/10.1109/SIEDS61124.2024.10534753

Malik, A., Wandira, A., Kuntadi, D., & Nugraha, A. R. (2022). How can interactive multimedia direct instruction model improve student cognitive learning outcomes? *Momentum: Physics Education Journal*, *6*(2), 104–118. https://doi.org/10.21067/mpej.v6i2.6686

Masrifah, M., & Amiroh, D. (2023). The multi-representational ability profile of physics students in the interactive multimedia assisted problem-based learning during the Covid-19 pandemic. *Momentum: Physics Education Journal*, *7*(2), 188–200. https://doi.org/10.21067/mpej.v7i2.7501

McClure, F. D., & Lee, J. K. (2005). Sample Sizes Needed for Specified Margins of Relative Error in the Estimates of the Repeatability and Reproducibility Standard Deviations. *Journal of AOAC INTERNATIONAL*, *88*(5), 1503–1510. https://doi.org/10.1093/jaoac/88.5.1503

McEwan, B. (2020). Sampling and validity. *Annals of the International Communication Association*, *44*(3), 235–247. https://doi.org/10.1080/23808985.2020.1792793

Mikielewicz, J., & Mikielewicz, D. (2024). Influence of thermodynamics on the development of technology and science. *Archives of Thermodynamics*, *45*(2), 51–61. https://doi.org/10.24425/ather.2024.150851

Msoma, A. (2020). Baseline, Cross-sectional Survey of Knowledge, Attitudes, and Practices Related to Efficient Use of Government Resources in the Health Sector in Malawi. *Malawi Medical Journal*, *32*(3), 146–152. https://doi.org/10.4314/mmj.v32i3.7

Muthaharoh, S., & Sukarelawan, Moh. I. (2023). Improving high school students' research skills through light diffraction experiments using the problem-based learning model. *Jurnal Riset Dan Kajian Pendidikan Fisika*, *10*(2), 64–71. https://doi.org/10.12928/jrkpf.v10i2.566

Nopiah, Z. M., Jamalluddin, M. H., Othman, N. A. I. H., Asshaari, I. A., & Osman, M. H. (2012). Reliability Analisis of Examination Questions in a Mathematics Course Using Rasch Measurement Model. *Sains Malaysiana* , *41*(9), 1171–1176.

Nottis, K. E. K., Prince, M. J., & Vigeant, M. A. (2017). Undergraduate Engineering Students' Understanding of Heat, Temperature, and Energy: An Examination by Gender and Major. *US-China Education Review A*, *7*(3). https://doi.org/10.17265/2161-623X/2017.03.001

Nugraheni, V. Y. S. T., & Mundilarto, M. (2022). Development of 2-D augmented reality integrated physics e-book to improve students' problem-solving skills. *Momentum: Physics Education Journal*, *6*(2), 171–180. https://doi.org/10.21067/mpej.v6i2.6623

Nuryantini, A. Y., Zakwandi, R., & Ariayuda, M. A. (2022). Smartphone for physics practicum in momentum and impulse. *Momentum: Physics Education Journal*, *7*(1), 48–56. https://doi.org/10.21067/mpej.v7i1.6687

Nusir, S., Alsmadi, I., Al-Kabi, M., & Sharadgah, F. (2013). Studying the Impact of Using Multimedia Interactive Programs on Children's Ability to Learn Basic Math Skills. *E-Learning and Digital Media*, *10*(3), 305–319. https://doi.org/10.2304/elea.2013.10.3.305

O'Brien, T. C., Palmer, R., & Albarracin, D. (2021). Misplaced trust: When trust in science fosters belief in pseudoscience and the benefits of critical evaluation. *Journal of Experimental Social Psychology*, *96*, 104184. https://doi.org/10.1016/j.jesp.2021.104184

Otoluwa, S. M., Uloli, R., Abdjul, T., Mursalin, Odja, A. H., & Ntobuo, N. E. (2024). The Influence of the PBL Model to Student Learning Outcomes on Heat and its Transfer Concepts. *Jurnal Penelitian Pendidikan IPA*, *10*(7), 4357–4363. https://doi.org/10.29303/jppipa.v10i7.8003

Paek, I., Lee, J., Stankov, L., & Wilson, M. (2013). Rasch modeling of accuracy and confidence measures from cognitive tests. *J Appl Meas*, *14*(3), 232–248.

Park, H.-J., Lee, D., & Park, H. (2023). Understanding students' problem-solving patterns: Evidence from an allotted response time in a PISA 2012 item. *Frontiers in Psychology*, *13:1050435*, 1–14. https://doi.org/10.3389/fpsyg.2022.1050435

Perdana, R., Asrial, & Maison. (2022). The Impact of Student Motivation on Students' Science Process Skills. *Journal of Education Research and Evaluation*, *6*(4), 729–738. https://doi.org/10.23887/jere.v6i4.52350

Pipia, E. (2014). Practical strategies for teachers to enhance the degree of reliability and validity in assessment. *Journal of Education*, *3*(1), 47–52.

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. https://doi.org/10.1103/PhysRevPhysEducRes.15.020111

Rendón-Macías, M. E., & Villasís-Keever, M. Á. (2017). El protocolo de investigación V: el cálculo del tamaño de muestra. *Revista Alergia México*, *64*(2), 220–227. https://doi.org/10.29262/ram.v64i2.267

Safarati, N., & Lubis, R. H. (2022). Students' Conceptual Understanding and Critical Thinking Skills Through Online Learning Using a Virtual Laboratory. *JIPF (Jurnal Ilmu Pendidikan Fisika)*, *7*(1), 42–49. https://doi.org/10.26737/jipf.v7i1.2221

Salazar, L. M., Díaz, M. H. R., & Slisko, J. (2023). Critical thinking development in physics courses by PBL in virtual collaboration environments. *International Journal of Innovation in Science and Mathematics Education*, *31*(4), 27–39. https://doi.org/10.30722/IJISME.31.04.003

Seibert, S. A. (2021). Problem-based learning: A strategy to foster generation Z's critical thinking and perseverance. *Teaching and Learning in Nursing*, *16*(1), 85–88. https://doi.org/10.1016/j.teln.2020.09.002

Sevilla, C. G., Ochave, J. A., Punsalan, T. G., Regala, B. P., & Uriarte, G. G. (1993). *Pengantar Metode Penelitian: Vol. Cet. 1* (A. Tuwu, Ed.). UI Press.

Singh, S. (2003). Simple Random Sampling. In *Advanced Sampling Theory with Applications: Vols. I–II* (1st ed., pp. 71–136). Springer Netherlands. https://doi.org/10.1007/978-94-007-0789-4_2

Sofna, A., Zakwandi, R., & Desy Purwasih. (2024). In-depth Analysis of The Learning Process in Schools: Reviewing The Quality of Students' Actual Mastery of Concepts on Temperature and Heat Material. *International Journal of Education and Teaching Zone*, *3*(1), 82–93. https://doi.org/10.57092/ijetz.v3i1.193

Stylos, G., Christonasis, A., Georgopoulos, K., & Kotsis, K. T. (2023). The impact of an online physics lab on university students' self-efficacy and understanding of thermal concepts during COVID-19 pandemic. *Journal of Mathematics and Science Teacher*, *3*(2), em049. https://doi.org/10.29333/mathsciteacher/13701

Stylos, G., Sargioti, A., Mavridis, D., & Kotsis, K. T. (2021). Validation of the thermal concept evaluation test for Greek university students' misconceptions of thermal concepts. *International Journal of Science Education*, *43*(2), 247–273. https://doi.org/10.1080/09500693.2020.1865587

Sudirman, S., Maknun, J., & Hasanah, L. (2017). Penggunaan Model Problem Based Learning Berbantuan Media Kit Eksperimen Inkuiri Terhadap Peningkatan Penguasaan Konsep Siswa Kelas XI SMK Pada Materi Fluida Statis. *JIPF (Jurnal Ilmu Pendidikan Fisika)*, *2*(2), 38–41. https://doi.org/10.26737/jipf.v2i2.232

Sukarelawan, Moh. I., Jumadi, J., & Rahman, N. A. (2019). An Analysis of Graduate Students' Conceptual Understanding in Heat and Temperature (H&T) Using Three-Tier Diagnostic Test. *Indonesian Review of Physics*, *2*(1), 9. https://doi.org/10.12928/irip.v2i1.910

Sukarelawan, Moh. I., Sriyanto, Sa'adah, M., Jupri, J., Hikmah, F. N., & Eko Nursulistiyo. (2024). Is it True That My Students don't Understand the Static Fluid Concepts? Rasch Modeling Perspective. *Jurnal Pendidikan Dan Pengajaran*, *57*(1), 127–138. https://doi.org/10.23887/jpp.v57i1.57883

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan* (1st ed.). Trim Komunikata Publishing House.

Tourangeau, R., Yan, T., & Sun, H. (2020). Who Can You Count On? Understanding The Determinants of Reliability. *Journal of Survey Statistics and Methodology*, *8*(5), 903–931. https://doi.org/10.1093/jssam/smz034

Turgut, U., & Gurbuz, F. (2012). Effect of conceptual change text approach on removal of students' misconceptions about heat and temperature. *International Journal of Innovation and Learning*, *11*(4), 386. https://doi.org/10.1504/IJIL.2012.047139

Ummah, K., Mardhiya, J., & Mulyanti, S. (2022). Pengembangan Instrumen Tes Penguasaan Konsep Representasi Kimia pada Lima Indikator Asam Basa dari Alam: Analisis dengan Rasch Model. *JURNAL TARBIYAH*, *29*(2), 212–225. https://doi.org/10.30829/tar.v29i2.1706

Villarino, G. N. B. (2023). Fundamental Thermal Concepts: An Assessment of University College Students' Conceptual Understanding of Everyday Perspectives. *Indonesian Journal of Science and Mathematics Education*, *6*(1), 96–108. https://doi.org/10.24042/ijsme.v6i1.11662

Wartono, Alfroni, Y. F., Batlolona, J. R., & Mahapoonyanont, N. (2019). Inquiry-Scaffolding Learning Model: Its Effect on Critical Thinking Skills and Conceptual Understanding. *Jurnal Ilmiah Pendidikan Fisika Al-Biruni*, *8*(2), 245–255.

Wiyono, K., Sudirman, S., Sury, K., Saparini, S., & Ariska, M. (2021). An Overview of Students' Conceptual Understanding in Kinematics Using Computer Based Test (CBT) Class XI SMA Negeri 19 Palembang. *JIPF (Jurnal Ilmu Pendidikan Fisika)*, *6*(2), 167–177. https://doi.org/10.26737/jipf.v6i2.2037

Yeo, S., & Zadnik, M. (2001). Introductory thermal concept evaluation: assessing students' understanding. *The Physics Teacher*, *39*(8), 496–504. https://doi.org/10.1119/1.1424603

Zou, T., & Bolt, D. M. (2023). Person Misfit and Person Reliability in Rating Scale Measures: The Role of Response Styles. *Measurement: Interdisciplinary Research and Perspectives*, *21*(3), 167–180. https://doi.org/10.1080/15366367.2022.2114243

Zulrahayu, V., & Mufit, F. (2024). Correlation of Concept Understanding and Students' Attitudes towards Learning Physics on the Material of Temperature and Heat in Class XI. *JIPF (Jurnal Ilmu Pendidikan Fisika)*, *9*(2), 181–191. https://doi.org/10.26737/jipf.v9i2.4654