Journal of Advanced Sciences and Mathematics Education
Volume 5, Issue 2, 359 - 372
e_ISSN: 2798-2351
DOI: 10.58524/jasme.v5i2.818

Design and psychometric validation of a metacognitive instrument for physics learning: A focus on heat concepts

Hera Novia* Universitas Pendidikan Indonesia, INDONESIA

Siska Dewi Aryani Universitas Pendidikan Indonesia, INDONESIA

Andhy Setiawan Universitas Pendidikan Indonesia, INDONESIA

Muhammad Zahran Universitas Pendidikan Indonesia, INDONESIA

Article Info

Article history:

Received: July 11, 2025
Revised: Aug 13, 2025
Accepted: Sept 05, 2025

Keywords:

Metacognitive Knowledge;
Metacognitive Questionnaire;
Instrument Validation;
Physics Education;
Rasch Model.

Abstract

Background: Metacognitive knowledge, awareness of one's own cognition, tasks, and learning strategies, is critical for independent learning but is often underemphasized in physics education.
Aim: This study aimed to develop and validate a questionnaire to measure high school students' metacognitive knowledge within the specific context of physics, focusing on the topic of heat.
Method: The research involved content validation by six physics education experts and construct validation through empirical testing with 163 high school students. Content validity was established using Aiken's V, while construct validity and reliability were evaluated using Rasch model analysis.
Results: The final instrument consisted of 28 items, of which 26 met the Rasch model fit criteria. The analysis confirmed high person reliability (0.82) and item reliability (0.98), indicating strong internal consistency and measurement stability.
Conclusion: The findings support the questionnaire's validity and reliability as a tool for assessing metacognitive knowledge in physics. This validated instrument provides a foundation for future research and instructional practices aimed at enhancing students' metacognitive skills.

To cite this article: Novia, H., Aryani, S. D., Setiawan, A., & Zahran, M. (2025). Design and psychometric validation of a metacognitive instrument for physics learning: A focus on heat concepts. *Journal of Advanced Sciences and Mathematics Education, 5(*2), 359-372.

INTRODUCTION

An individual's understanding of their cognitive processes, commonly called metacognitive knowledge, plays a critical role in personal and intellectual development (Novia et al., 2019). This self-awareness enables learners to reflect on, evaluate, and regulate their thinking, enhancing learning outcomes and problem-solving capabilities across diverse disciplines. Although half a century has passed since Flavell (1979) introduced the concept, its integration into science education (particularly in physics) remains limited. In contrast, extensive research in language learning consistently demonstrates that students with well-developed metacognitive knowledge tend to perform better in self-directed learning environments (Peskin & Astington, 2004; Teng, 2025; Wenden, 1998).

Metacognitive knowledge encompasses awareness of one's cognitive strengths and limitations, understanding the demands of specific tasks, and knowing which strategies are most effective for learning and problem-solving (Pintrich, 2002). This triadic framework, knowledge of self, tasks, and strategy, is a foundation for independent learning and academic resilience (Efklides & Vlachopoulos, 2012). However, despite the conceptual significance of metacognition, its effective assessment remains a critical challenge for educators seeking to foster these skills, which leads to the next point. Instruments designed to capture metacognitive knowledge in physics are exceedingly scarce (Sari, 2019), highlighting the need for context-specific tools grounded in strong psychometric principles.

* **Corresponding author:**
Hera Novia, Universitas Pendidikan Indonesia, INDONESIA
heranovia@upi.edu

In physics education, heat presents a significant context for assessing metacognitive knowledge (Beck et al., 2019; Hikmah et al., 2021; Sukarelawan et al., 2021). As an abstract and often misunderstood concept, heat requires learners to actively plan and monitor their thought processes to develop scientifically accurate understandings. Misconceptions commonly arise from intuitive beliefs, imprecise everyday language, and oversimplified textbook explanations. For example, students often incorrectly believe that a metal chair is inherently 'colder' than a wooden desk in the same room, rather than understanding differences in thermal conductivity, or they may think of cold as a substance that flows into objects (Zahran et al., 2025). All of these necessitate metacognitive reflection to resolve (Agnezi, 2023; Taherdoost, 2016). The significant mental effort required to overcome these deep-seated misunderstandings makes this an ideal context to assess students' metacognitive abilities, directly aligning with the focus of this research.

Self-report questionnaires remain one of the most commonly employed tools in educational research to measure these nuanced cognitive processes. However, the effectiveness of such instruments depends heavily on their design and psychometric quality. A review by Radhakrishna (2007) found that while 64% of educational studies utilized questionnaires, a significant proportion failed to report key procedures related to validity (31%) and reliability (33%). Such omissions compromise the accuracy of findings by increasing the risk of measurement error, the gap between a respondent's accurate cognitive attributes and recorded responses (Jalil et al., 2018).

Developing robust instruments for latent constructs such as metacognition requires rigorous validation procedures. Validity, which determines whether an instrument accurately captures the intended construct, must be complemented by reliability, which assesses the consistency of its results (Duncan et al., 2015; Zahran et al., 2025). Among the available psychometric models, Rasch analysis offers a sophisticated approach for educational measurement, enabling the transformation of ordinal data into interval-level measures, detailed examination of item functioning and respondent ability, and confirmation of construct coherence, even in the presence of missing data (Boone & Noltemeyer, 2017; Linacre, 2002; Widhiarso & Sumintono, 2016). Its application in science education has yielded promising outcomes across disciplines such as physics, chemistry, and biology, demonstrating its ability to produce precise, interpretable, and pedagogically meaningful results (Chan et al., 2014; Samsudin et al., 2021; Suryana et al., 2020). Nevertheless, instruments measuring metacognition often remain generic and rarely address the nuanced cognitive demands of specific physics topics, despite evidence that misconceptions in areas such as heat are both persistent and resistant to conventional instruction (Hikmah et al., 2021; Sukarelawan et al., 2021). Building on these methodological advances, the present study designs and validates a physics-specific metacognition questionnaire contextualized to the concept of heat, employing Rasch analysis to produce a statistically robust and instructionally relevant instrument that can enhance both research and classroom practice in advancing students' metacognitive development.

METHOD

Research Design

This study aimed to develop and validate a metacognitive knowledge questionnaire designed to assess students' awareness and regulation of cognition in physics, explicitly emphasizing the concept of heat. A Design-Based Research (DBR) approach was adopted, utilizing the ADDIE model (Analysis, Design, Development, Implementation, Evaluation) as elaborated by Molenda (2003). While the ADDIE model is traditionally applied to instructional design, its systematic, phased approach provides a robust framework for the iterative process of creating and validating a non-instructional tool like a questionnaire. This structured process is particularly suitable because it ensures that each stage of development is deliberate and builds logically on the previous one, which is essential for establishing the instrument's psychometric soundness. This model facilitated a systematic and iterative process to ensure the resulting educational instrument was valid, reliable, and applicable in authentic classroom settings. The overall research design is illustrated in Figure 1.

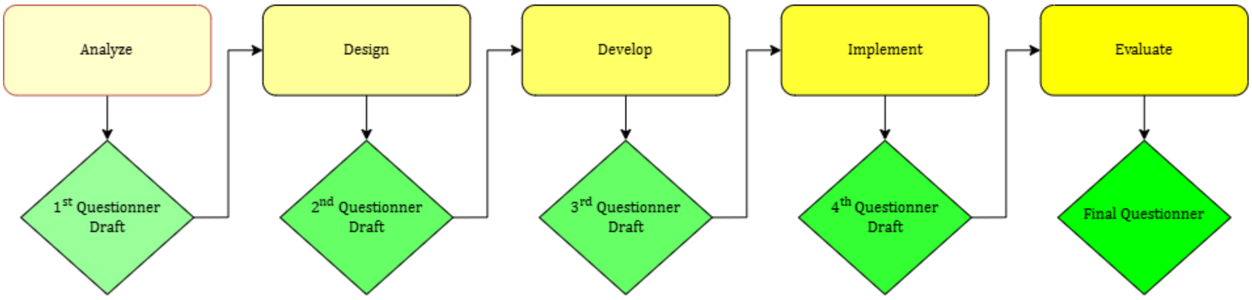

Figure 1. ADDIE Research Design

The development process integrated both qualitative (expert judgment and pilot feedback) and quantitative (Rasch model and reliability statistics) techniques to ensure content relevance, construct validity, and internal consistency. While in non-experimental educational research only the early phases of instrument development may suffice, the validation of diagnostic tools for empirical research demands a complete cycle of content validation, construct validation, and reliability testing to meet recognized psychometric standards (Carver, 1974; Ziegler & Brunner, 2016).

Participants

A total of 163 senior high school students enrolled in science programs participated in the validation phase of this study. Participants were selected using purposive sampling from selected schools in Bandung, West Java. The key inclusion criterion was that all students had previously received instruction on the concept of heat in their physics curriculum, ensuring they possessed the necessary background knowledge to engage with the questionnaire content. All participants had prior exposure to the concept of heat, and the sample included male and female students. Ethical standards were strictly adhered to, with informed consent obtained from all participants. In addition to the student respondents, six experts in physics education were involved to support the validation process. These experts played a pivotal role by evaluating the questionnaire using a structured validation sheet designed to assess its content, clarity, and overall quality. Their evaluations systematically examined each indicator, focusing on how effectively the questionnaire's content and structure support the assessment of students' metacognitive knowledge. The experts' feedback provided valuable insights into the questionnaire's appropriateness, ensuring that the instrument met educational standards and was suitable for its intended purpose.

Instrument

The evaluation criteria encompassed several key dimensions. Content indicators focused on aligning the questionnaire with the original test instrument—specifically, the concept of heat—and its reflection of metacognitive knowledge components. Construct indicators assessed item formulation's clarity and response instructions' comprehensibility. Language indicators evaluated whether the questionnaire employed communicative, unambiguous, age-appropriate language for students, was grammatically correct according to standard Indonesian, and was free from any elements related to ethnicity, religion, race, or inter-group bias. Each item was designed using a Likert-type scale to capture the degree of agreement with statements related to metacognitive engagement in physics learning (Ziegler & Brunner, 2016).

Data Analysis

Data analysis was conducted using Rasch Model Analysis, supported by Winsteps software and Microsoft Excel. For expert judgment, Aiken's V was employed to assess the overall quality of the questionnaire (Aiken, 1980; Retnawati, 2016). The Rasch Model Analysis was used to evaluate the construct validity and reliability of the questionnaire during implementation. This analytical method enabled the examination of item difficulty, personability, fit statistics (infit and outfit), and the dimensionality of the constructs. The Rasch model was selected for its strengths in developing and validating psychometric instruments to measure a single latent trait, such as metacognitive knowledge (Sumintono, 2018). Unlike classical test theory, the Rasch model transforms ordinal data from the Likert scale into interval-level measures, which allows for a more precise and linear assessment of both item difficulty and person ability on the same scale. This capacity, along with its ability to assess item fit and confirm the instrument's unidimensionality, provides a more rigorous validation than other available methods.

Research Procedure

The research began with the Analysis stage. During this stage, a literature review was conducted to identify sub-concepts of heat capable of measuring aspects relevant to metacognitive knowledge. This phase laid the theoretical groundwork for the instrument by examining metacognitive theory, particularly Pintrich's model, and prior diagnostic tools. Additionally, the instrument's objectives were formulated, and the target respondents, senior high school students enrolled in science programs, were defined to ensure the tool's contextual relevance and educational applicability.

During the Design stage, insights from the analysis informed the conceptual framework of the questionnaire. The instrument was structured to assess three core dimensions of metacognitive knowledge: knowledge of self, knowledge of task, and knowledge of strategies, as outlined by (Pintrich, 2002). Items were composed as declarative statements utilizing a 5-point Likert-type scale, ranging from "strongly disagree" to "strongly agree," to accommodate varying degrees of metacognitive engagement (Joshi et al., 2015). In this phase, the questionnaire's format, structure, and administration procedures were finalized to ensure coherence and ease of use.

In the Development stage, an initial draft of the instrument (Draft 1) was constructed based on the theoretical framework and defined objectives. To ensure content validity, six experts in physics education assessed each item for clarity, relevance, and alignment with metacognitive constructs. Their evaluations led to a revised version (Draft 2) and Aiken (1980)'s V index was employed to quantitatively determine the validity of each item, applying a minimum acceptable value of 0.78. A pilot test using Draft 2 was then conducted on a small group of students to evaluate item performance. Quantitative item analysis led to further refinements and the creation of Draft 3. Construct validation followed using Rasch model analysis to examine item fit, response consistency, and dimensionality. Items that did not align with model expectations were revised to improve validity.

As part of the Implementation stage, the refined instrument (Draft 4) was administered to a broader sample of 163 senior high school students from selected Bandung, West Java schools. Having prior exposure to the heat concept, these students completed the entire questionnaire. Rasch model analysis was employed again, utilizing Winsteps software, to assess item and person fit, scale unidimensionality, and overall instrument functionality (Linacre, 2002). The feedback from this phase informed additional adjustments

Finally, in the Evaluation stage, the instrument's reliability was analyzed using Rasch-based metrics, including person reliability, item reliability, and separation indices. These indicators provided internal consistency statistics comparable to Cronbach's alpha while offering detailed insights into scale performance (Boone & Noltemeyer, 2017; Linacre, 2002; Sumintono, 2018). Based on the findings from both construct and reliability analyses, the instrument was finalized as Draft 5, a validated and reliable tool suitable for research and educational use. For a better interpretation of the research procedure, see Figure 2.

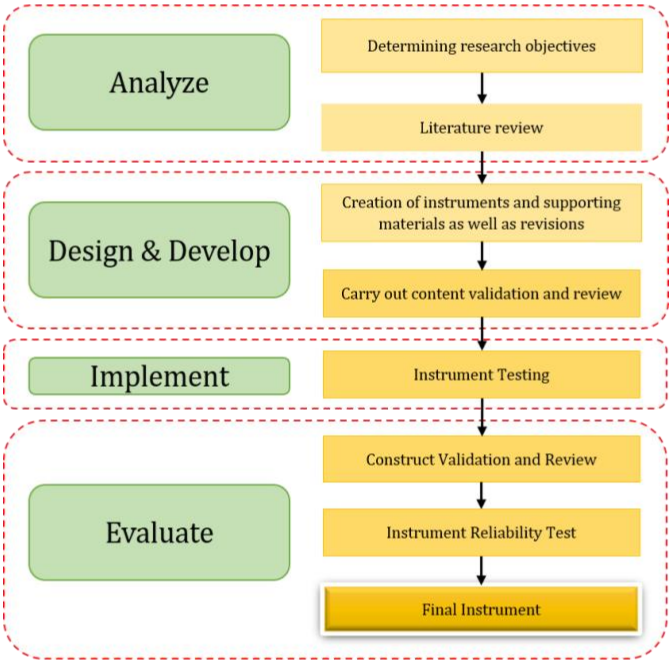

Figure 2. Research Procedure.

Throughout all stages, the development process adhered to rigorous psychometric standards and ethical research guidelines (Carver, 1974; Ziegler & Brunner, 2016), ensuring that the resulting questionnaire was robust instrument for assessing students' metacognitive knowledge in heat.

RESULTS AND DISCUSSION

Analyze

A literature review was conducted using the Scopus database with keywords related to *metacognition*, *metacognitive knowledge*, and *physics learning*. The retrieved data were analyzed using VOSviewer software to map keyword co-occurrences (Figure 3).

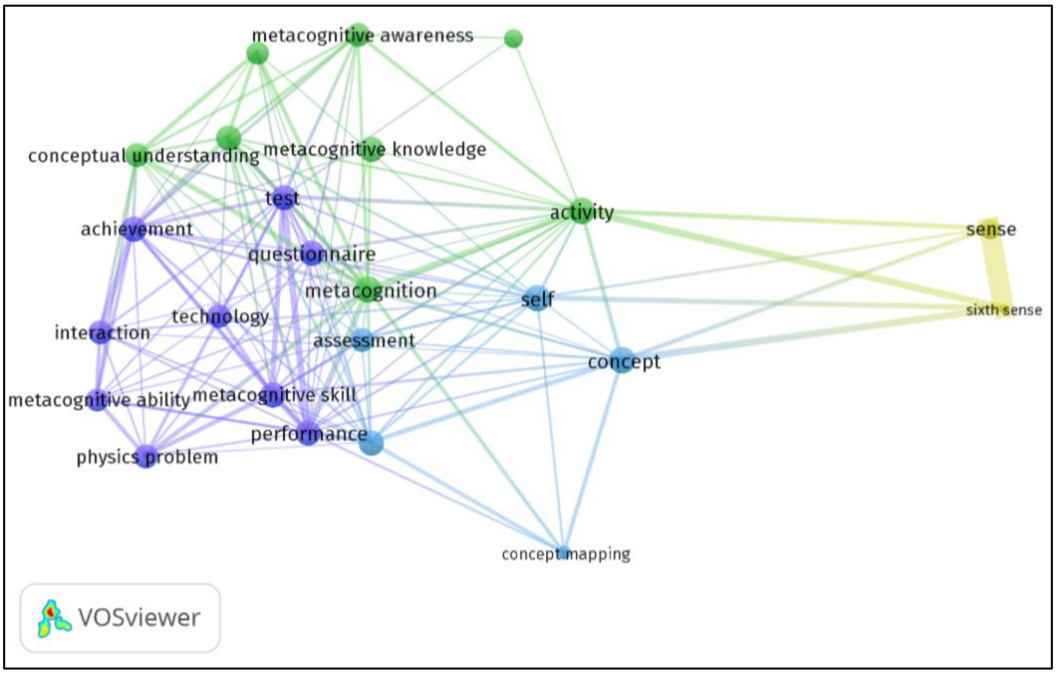

Figure 3. VOSviewer visualization of keyword co-occurrence from Scopus database search on metacognition and physics.

Note. The source of this image was generated by the author using VOSviewer software (van Eck & Waltman, 2010), with data retrieved from the Scopus database.

Figure 3 reveals distinct keyword clusters, including *metacognition*, *metacognitive knowledge*, *metacognitive skill*, *performance*, *assessment*, *test*, and *questionnaire*. These indicate that much of the existing research addresses metacognitive aspects within general educational or psychological frameworks. Additional clusters, such as *conceptual* understanding, achievement, *interaction*, and *physics problems*, reflect the intersection between metacognition and conceptual physics learning.

Notably, the visualization shows no cluster containing heat-related concepts such as *temperature*, *thermal expansion*, or *heat transfer*. This confirms a gap identified during the review: there is a lack of instruments explicitly designed to assess metacognitive knowledge in domain-specific contexts, particularly in thermal physics. While constructs like *activity*, *self*, and *concept mapping* appear, their distance from central terms like *metacognitive awareness* suggests that integrating active learning strategies and conceptual tools with metacognitive assessment in physics is still limited. Peripheral terms like *sense* and *sixth sense*, though interesting, appear tangential to mainstream educational research and likely represent fringe literature.

The bibliometric findings highlight a clear research gap: while general metacognitive assessment is well-represented, domain-specific tools for physics, especially those addressing thermal concepts, are scarce. This aligns with Zohar and Barzilai (2013), who note that most metacognitive instruction in science is generic and not adapted to specific domains, potentially reducing its effectiveness.

Consequently, Figure 3 supports the need for a validated, context-specific instrument focusing on heat-related physics concepts. The present study targets fundamental topics, such as the effect of heat on temperature and shape, and the principle of black objects in heat absorption, which are core to the physics curriculum yet underrepresented in metacognitive research. The instrument's development is grounded in Efklides and Vlachopoulos's (2012) framework, which categorizes metacognitive knowledge into self-related, task-related, and strategy-related knowledge. In this study, these categories were further elaborated into six sub-factors, as outlined in Table 1.

Table 1. Indicators and Sub-Indicators of Metacognitive Knowledge in Physics

Indicator	Sub-Indicator
Metacognitive knowledge about the self	• Perceived ease or fluency in understanding basic physics concepts based on past experience • Perceived difficulty or lack of fluency in understanding basic physics concepts based on past experience
Metacognitive knowledge about the task	• High cognitive demands of certain physics concepts (perceived as difficult material) • Low cognitive demands of certain physics concepts (perceived as easy material)
Metacognitive knowledge about strategies	• Use of cognitive/metacognitive strategies • Use of competence-enhancing strategies • Use of avoidance strategies to cope with difficulties in processing physics tasks related to heat

Note. The table was adapted from Efklides and Vlachopoulos's (2012) framework.

Design

In this phase, the questionnaire was conceptually designed based on six interrelated factors derived from the three main indicators of metacognitive knowledge as outlined in Table 1. The development process focused on constructing items that capture the multidimensional aspects of metacognitive knowledge while remaining contextually relevant for high school students studying physics.

Each item was carefully developed to reflect specific aspects of self-knowledge, task knowledge, and strategy knowledge, in line with the theoretical framework proposed by Efklides and Vlachopoulos (2012). These indicators were translated into practical, measurable statements that students could relate to in their experience of learning heat-related physics concepts.

Twenty-eight items were formulated; each aligned with Bloom's taxonomy (Krathwohl, 2002) at the metacognitive level. The aim was to assess how students recognize, monitor, and regulate their thinking processes during physics learning, particularly in thermal expansion, temperature change, and heat absorption. The items were written to reflect real classroom scenarios, ensuring relevance and authenticity in students' responses.

The design process emphasized the connection between theory and application, ensuring that each item served as a valid reflection of metacognitive principles and a tool for capturing meaningful student data (Herdman et al., 1998). Figure 4 presents an example of the questionnaire developed in this phase.

Metacognitive Knowledge of Heat Concepts Questionnaire

Instructions: Please indicate how true each statement is for you by selecting the number that best represents your experience.

Section A: How true are the following statements for you?

No	Statement	1	2	3	4	5
1	I can easily recall the concept of heat even when I first encounter it.					
2	I can manage my time to understand the heat concept, even while working on other tasks.					
3	I find it difficult to choose the right steps to understand the heat concept because I can't find my own reference sources.					

Section B: How difficult do you find the following problems?

No	Statement	1	2	3	4	5
1	I can easily identify important information related to the heat concept, even when the procedure is complex.					
2	I can easily recognize factors that influence heat changes.					
3	I find it difficult to identify important information related to the heat concept, even in simple procedures.					

Section C: How often do you do what is described in each of the following statements?

No	Statement	1	2	3	4	5
1	I think of different ways to figure out the concept of heat.					
2	I consider rereading multiple times when I encounter a difficult heat concept.					
3	I give up when I don't understand what is required to comprehend the heat concept.					

Scale:
1 = Never
2 = Sometimes
3 = Often
4 = Very Often
5 = Always

Figure 4. The example of metacognitive knowledge questionnaire.

Note: This questionnaire was designed based on the three indicators of metacognitive knowledge—self-knowledge, task knowledge, and strategy knowledge. It follows the theoretical model of Efklides and Vlachopoulos (2012) and is tailored to address the specific cognitive demands involved in learning heat-related physics concepts.

Develop

After the initial instrument draft was completed, the development phase focused on establishing its content validity. A panel of six experts in relevant fields was invited to evaluate the instrument based on three primary dimensions: the relevance of content, the clarity and structure of item construction, and the appropriateness of language use. These evaluations were quantified using Aiken's V, a statistical measure to assess expert agreement. The result of 28 questionnaire items can be seen in Table 2.

All items achieved Aiken's V values equal to or greater than 0.78, which is considered an acceptable threshold indicating strong validity. This result affirmed the appropriateness of each item's content and formulation (Aiken, 1980; Retnawati, 2016). In addition to quantitative assessments, the experts provided detailed qualitative feedback. Suggestions included refining the wording of several items (item 3, item 4, and item 16) to ensure closer alignment with the intended metacognitive indicators, eliminating overlapping statements to avoid redundancy, improving the clarity of task instructions, and consistently adhering to standard Indonesian linguistic norms. These insights guided the revision process, which involved editorial and conceptual adjustments. The aim was to enhance the instrument's overall quality, clarity, and utility for its intended audience of secondary-level physics students.

Table 2. The Aiken V result of expert validation

No Item	Assessment Items			Conclusion
	Content Aspect	Construction Aspects	Language Aspects	
1	0,87	0,97	0,96	Valid
2	0,87	0,98	0,96	Valid
3	0,79	0,98	0,97	Valid
4	0,79	0,98	0,96	Valid
5	0,83	0,98	0,95	Valid
6	0,83	0,96	0,95	Valid
7	0,86	0,98	0,96	Valid
8	0,88	0,96	0,95	Valid
9	0,86	0,96	0,96	Valid
10	0,91	0,98	0,96	Valid
11	0,86	0,98	0,96	Valid
12	0,92	0,96	0,95	Valid
13	0,83	0,96	0,95	Valid
14	0,83	0,98	0,97	Valid
15	0,79	0,98	0,98	Valid
16	0,79	0,98	0,97	Valid
17	0,79	0,98	0,96	Valid
18	0,83	0,96	0,95	Valid
19	0,87	0,94	0,97	Valid
20	0,87	0,96	0,97	Valid
21	0,92	0,96	0,97	Valid
22	0,92	0,98	0,96	Valid
23	0,86	0,98	0,96	Valid
24	0,86	0,98	0,93	Valid
25	0,83	0,98	0,97	Valid
26	0,85	0,97	0,96	Valid
27	0,86	0,98	0,97	Valid
28	0,86	0,96	0,97	Valid

Implement

Following these revisions, the finalized instrument was implemented in a field study to evaluate its performance in a real educational setting. The trial was conducted with 163 high school students from Bandung (see Figure 5), all of whom had completed instruction on heat in their physics curriculum. The instrument was administered over one week, providing ample time for students to engage with the questionnaire in a focused manner.

Figure 5. The trial location (Bandung).

Evaluate

After acquiring students' responses from the trial testing, construct validity can be assessed. To examine the instrument's construct validity, Rasch model analysis was conducted using WINSTEPS software. The analysis revealed that the raw variance explained by the measures was 37.5%, which surpasses the minimum threshold of 20 percent commonly accepted in Rasch-based validity studies (Chan et al., 2014; Sumintono, 2018). This finding indicates that the instrument effectively captures a coherent underlying construct and demonstrates good internal consistency (Boone & Noltemeyer, 2017).

Further investigation into the functioning of individual items was carried out using item fit statistics. These include the mean-square outfit (MNSQ), standardized z-scores (ZSTD), and point-measure correlations (PT Measure Corr), table 3 detailed all the values from 28 items.

Table 3. Construct Validity Test Using Rasch Model

No Item	Outfit Value		PT Measure Corr	Value Criteria	Interpretation
	MNSQ	ZSTD			
1	0,60	-4,58	0,50	Two Criteria Fulfilled	In accordance
2	0,67	-3,67	0,55	Two Criteria Fulfilled	In accordance
3	0,67	-3,64	0,53	Two Criteria Fulfilled	In accordance
4	1,33	2.73	0,46	Two Criteria Fulfilled	In accordance
5	1,36	3,27	0,45	Two Criteria Fulfilled	In accordance
6	0,80	-2,10	0,61	Two Criteria Fulfilled	In accordance
7	1,04	0,41	0,51	Two Criteria Fulfilled	In accordance
8	0,82	-1,77	0,36	Two Criteria Fulfilled	In accordance
9	0,79	-2,14	0,54	Two Criteria Fulfilled	In accordance
10	0,59	-4,68	0,72	Two Criteria Fulfilled	In accordance
11	0,71	-3,14	0,61	Two Criteria Fulfilled	In accordance
12	0,72	-3,02	0,70	Two Criteria Fulfilled	In accordance
13	0,83	-1,63	0,41	Three Criteria Fulfilled	In accordance
14	0,67	-3,50	0,56	Two Criteria Fulfilled	In accordance
15	0,70	-3,11	0,54	Two Criteria Fulfilled	In accordance
16	1,22	2,04	0,51	Two Criteria Fulfilled	In accordance
17	0,91	-0,87	0,28	Two Criteria Fulfilled	In accordance
18	1,78	6,15	-0,05	Three Criteria Not Met	It is not in accordance with
19	1,33	2,98	0,47	Two Criteria Fulfilled	In accordance
20	1,09	0,92	0,49	Two Criteria Fulfilled	In accordance
21	1,13	1,19	0,44	Three Criteria Fulfilled	In accordance
22	1,27	2,20	0,24	One Criteria Fulfilled	In accordance
23	0,98	-0,15	0,62	Three Criteria Fulfilled	In accordance
24	0,84	-1,51	0,63	Three Criteria Fulfilled	In accordance
25	1,10	1,00	0,04	Two Criteria Fulfilled	In accordance
26	1,25	2,33	0,08	Two Criteria Fulfilled	In accordance
27	1,63	5,16	0,23	Three Criteria Not Met	It is not in accordance with
28	1,31	2,75	0,35	One Criteria Fulfilled	In accordance

Most of the 28 items in the questionnaire met at least two of these statistical criteria, which suggests that the instrument is well-calibrated and valid for assessing metacognitive knowledge. Although most items showed appropriate fit, two items, item 18 and item 27, did not meet any criteria and were flagged as misfitting. Item 18, which addresses avoidance strategies, suggests that some students give up when faced with difficulties in understanding heat concepts. Item 27, which deals with metacognitive strategy use, indicates that students may struggle to consider multiple approaches related to heat-related conceptual problem-solving. These misfit items highlight areas for further refinement but do not detract from the overall structural validity of the instrument. In addition to validity, reliability analysis was conducted to evaluate the consistency of the instrument. The result of Rasch analysis can be seen in Table 4.

Table 4. Instrument Reliability Test Using Rasch Model

Types of Reliability	Values	Interpretation
Person Reliability	0,82	Very special
Item Reliability	0,98	Very good

The person reliability value was 0.82, indicating a high level of consistency in student responses. The item reliability value was 0.98, which reflects an extreme degree of measurement precision across the instrument items. These results affirm that the instrument possesses structural validity and yields consistent and reliable measurements of students' metacognitive knowledge (Carver, 1974; Taherdoost, 2016; Ziegler & Brunner, 2016).

The findings of this evaluation are in line with those of previous studies. For example, Cotterall and Murray (2009) highlight that metacognitive knowledge can reshape students' learning beliefs and improve their ability to plan, monitor, and evaluate their learning processes. Pintrich (2002) further emphasizes the essential role of metacognitive knowledge in education, especially in developing learners' self-regulation skills. Additionally, Efklides and Vlachopoulos (2012) argue that with proper measurement and training, metacognitive knowledge can significantly enhance students' ability to manage their learning effectively.

Discussion

The primary finding of this study is the successful development and validation of a psychometrically sound instrument for assessing students' metacognitive knowledge in the specific context of heat-related physics concepts. The evidence confirms that the questionnaire is a valid and reliable tool, addressing a clear gap in the existing educational literature for domain-specific metacognitive assessments.

The instrument's strong psychometric properties are supported by multiple analyses. The content validity (Aiken's V ≥ 0.79) mirrors the thresholds reported by Retnawati (2016) and Azwar (2012), indicating high expert agreement and relevance of the instrument items. These results suggest that the developed questionnaire successfully captures the multidimensional aspects of metacognitive knowledge while remaining contextually relevant to secondary physics education. The construct validity findings from the Rasch analysis (37.5% explained variance) surpass the recommended 20% benchmark (Widhiarso & Sumintono, 2016), confirming the instrument's coherence in measuring metacognitive knowledge. Similar psychometric strength has been reported by Yildiz and Yildirim (2020) in the validation of metacognitive awareness tools for science learning contexts. The identification of misfit items, particularly those related to avoidance strategies (item 18) and multi-approach problem solving (item 27), reflects patterns noted by Pintrich (2002), who found that some metacognitive strategies are inconsistently employed by students with lower confidence or weaker mastery of content. These misfits suggest that further refinement may be required to ensure consistent interpretation across varying student profiles. High reliability scores (person reliability = 0.82; item reliability = 0.98) are consistent with findings from Boone and Noltemeyer (2017), who emphasize that such values indicate excellent measurement stability. This consistency supports the instrument's suitability for repeated use in both research and classroom assessment. Overall, the findings demonstrate that the developed instrument addresses an existing gap in the literature by providing a psychometrically sound tool for assessing metacognitive knowledge in heat-related physics concepts. The results also reinforce the theoretical claims of Cotterall and Murray (2009) and Efklides and Vlachopoulos (2012), who argue that well-designed metacognitive assessments can enhance students' ability to plan, monitor, and regulate their learning processes in complex subject areas.

This validated instrument has significant practical implications for science educators. It can be used as a formative assessment tool to diagnose specific areas where students lack metacognitive awareness. For example, teachers can identify whether students struggle more with assessing task difficulty, knowing which learning strategies to apply, or accurately judging their own understanding. These insights can inform the design of targeted interventions, such as classroom activities that prompt students to reflect on their thinking or workshops that introduce new learning strategies for complex physics problems. Ultimately, the questionnaire can facilitate a more explicit classroom focus on the process of learning, empowering students to become more effective and independent learners.

Implications

The validated metacognitive knowledge questionnaire developed in this study offers significant contributions for both physics education practice and research by providing teachers with a reliable diagnostic tool to identify students' self-awareness, task understanding, and strategy use in the context of heat-related physics concepts, enabling more targeted instructional interventions and formative assessment. Its domain-specific nature fills an important gap in existing metacognitive measurement tools, allowing researchers to more accurately investigate how students regulate their thinking when confronting misconceptions in thermal physics. Furthermore, the instrument can be used in teacher education and curriculum development to strengthen the integration of metacognitive elements into physics instruction, promote reflective learning, and support the design of lessons that explicitly develop students' ability to plan, monitor, and evaluate their cognitive processes.

Limitations and Suggestions

This study has several limitations that should be considered when interpreting the findings. The validation process involved 163 students from a limited geographical area in Bandung, which restricts the generalizability of the instrument to broader populations with different academic backgrounds and learning environments. The Rasch analysis also identified two items that did not fit the measurement model, indicating that some statements related to avoidance strategies and multi-approach problem solving may be interpreted inconsistently by students. Additionally, the instrument relies on self-report data, which may be influenced by social desirability tendencies or students' inaccurate self-assessment of their own metacognitive processes.

Based on these limitations, future studies are encouraged to test the questionnaire with more diverse samples across different schools and regions to strengthen its external validity. Researchers may refine or replace misfitting items by conducting cognitive interviews or pilot studies that help clarify student interpretations of complex metacognitive behaviors. Triangulation using complementary methods such as learning journals, classroom observations, or performance-based tasks could provide a more holistic understanding of students' metacognition. Further development of the instrument for additional physics topics and the creation of digital or adaptive versions may also enhance its usability for both classroom practice and educational research

CONCLUSION

This research has successfully developed and validated a metacognitive knowledge questionnaire focused on the concept of heat in physics, demonstrating it to be a reliable and valid instrument for assessing students' metacognitive understanding. The instrument is structured around three key indicators of metacognitive knowledge: self-awareness (perceived fluency and difficulty), task characteristics (perceived low and high demand), and strategy use (competence-enhancing, cognitive, and avoidance strategies). Through a rigorous validation process involving 163 students, the instrument showed strong content and construct validity, as well as high reliability across its 28 items. The overall robustness and usability of the instrument are confirmed by the fact that 26 of the 28 statements met the required psychometric criteria. These findings support the questionnaire's effectiveness in capturing students' reflective thinking and strategic awareness when learning physics, particularly within the challenging context of heat concepts. The primary contribution of this study is a scientifically tested tool that can be used to measure and understand students' metacognitive processes. In practical terms, educators can use this instrument as a formative assessment to diagnose specific areas where students may lack metacognitive awareness, such as in estimating task difficulty or selecting effective strategies. The results can guide the design of targeted interventions, including classroom activities that prompt reflection or workshops on new learning techniques. By facilitating a more explicit focus on the process of learning, the tool can help teachers foster greater academic self-awareness and independence in their students. However, to enhance the generalizability of these findings, further trials involving more diverse student populations across different schools are recommended. Additionally, expanding the development of similar instruments for other topics in science would help build a more comprehensive understanding of students' metacognitive development, ultimately supporting more effective and targeted instructional practices.

AUTHOR CONTRIBUTIONS STATEMENT

HN contributed to defining the main idea of the research and led the article writing process. SDA was responsible for developing the research instrument and conducting the empirical testing. AS provided valuable insights and critical feedback for refining the manuscript. MZ performed the data analysis and enhanced the article’s visual structure, including figures and flowcharts.

REFERENCES

Agnezi, L. A. (2023, September). Development of The Online Assessment Instrument for Fluid, Temperature, and Heat to measure the problem solving skills of high school students. In *Journal of Physics: Conference Series* (Vol. 2582, No. 1, p. 012053). IOP Publishing. https://doi.org/10.1088/1742-6596/2582/1/012053

Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and psychological measurement*, *40*(4), 955-959. https://doi.org/10.1177/001316448004000419

Beck, B., Peña-Vivas, V., Fleming, S., & Haggard, P. (2019). Metacognition across sensory modalities: Vision, warmth, and nociceptive pain. *Cognition*, *186*, 32-41. https://doi.org/10.1016/j.cognition.2019.01.018

Boone, W. J., & Noltemeyer, A. (2017). Rasch analysis: A primer for school psychology researchers and practitioners. *Cogent Education*, *4*(1), 1416898. https://doi.org/10.1080/2331186X.2017.1416898

Carver, R. P. (1974). Two dimensions of tests: Psychometric and edumetric. *American Psychologist*, *29*(7), 512. https://doi.org/10.1037/h0036782

Chan, S. W., Ismail, Z., & Sumintono, B. (2014). A Rasch model analysis on secondary students’ statistical reasoning ability in descriptive statistics. *Procedia-Social and Behavioral Sciences*, *129*, 133-139. https://doi.org/10.1016/j.sbspro.2014.03.658

Cotterall, S., & Murray, G. (2009). Enhancing metacognitive knowledge: Structure, affordances and self. *System*, *37*(1), 34-45. https://doi.org/10.1016/j.system.2008.08.003

Duncan, T., Pintrich, P., Smith, D., & Mckeachie, W. (2015). Motivated strategies for learning questionnaire (MSLQ) manual. *National Center for Research to Improve Postsecondary Teaching and Learning.*

Efklides, A., & Vlachopoulos, S. P. (2012). Measurement of metacognitive knowledge of self, task, and strategies in mathematics. *European Journal of Psychological Assessment*. https://doi.org/10.1027/1015-5759/a000145

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist, 34*(10), 906. https://doi.org/10.1037/0003-066X.34.10.906

Herdman, M., Fox-Rushby, J., & Badia, X. (1998). A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. *Quality of life Research*, *7*(4), 323-335. https://doi.org/10.1023/A:1008846618880

Hikmah, F. N., Sukarelawan, M. I., Nurjannah, T., & Djumati, J. (2021). Elaboration of high school student's metacognition awareness on heat and temperature material: Wright map in Rasch model. *Indonesian Journal of Science and Mathematics Education, 4*(2), 172–182. https://doi.org/10.24042/ijsme.v4i2.9488

Jalil, S., Ali, M. S., & Haris, A. (2018, June). Development and validation of science process skills instrument in physics. In *Journal of Physics: Conference Series* (Vol. 1028, No. 1, p. 012203). IOP Publishing. https://doi.org/10.1088/1742-6596/1028/1/012203

Joshi, A., Kale, S., Chandel, S., & Pal, D. K. (2015). Likert scale: Explored and explained. *British journal of applied science & technology*, *7*(4), 396. https://doi.org/10.9734/BJAST/2015/14975

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into practice*, *41*(4), 212-218. https://doi.org/10.1207/s15430421tip4104_2

Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of applied measurement*, *3*(1), 85-106.

Molenda, M. (2003). In search of the elusive ADDIE model. *Performance improvement*, *42*(5), 34-37. https://doi.org/10.1002/pfi.4930420508

Novia, H., Kaniawati, I., Rusli, A., & Rusdiana, D. (2019). The development of metacognitive awareness related to the implementation of metacognitive-based learning. In *Journal of Physics: Conference Series* (Vol. 1170, No. 1, p. 012034). IOP Publishing. https://doi.org/10.1088/1742-6596/1170/1/012034

Peskin, J., & Astington, J. W. (2004). The effects of adding metacognitive language to story texts. *Cognitive development*, *19*(2), 253-273. https://doi.org/10.1016/j.cogdev.2004.01.003

Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory into practice*, *41*(4), 219-225. https://doi.org/10.1207/s15430421tip4104_3

Radhakrishna, R. B. (2007). Tips for developing and testing questionnaires/instruments. *The Journal of Extension*, *45*(1), 25.

Retnawati, H. (2016). Proving content validity of self-regulated learning scale (The comparison of Aiken index and expanded Gregory index). *REiD (Research and Evaluation in Education)*, *2*(2), 4. https://doi.org/10.21831/reid.v2i2.11029

Samsudin, A., Cahyani, P. B., Rusdiana, D., Efendi, R., Aminudin, A. H., & Costu, B. (2021). Development of a Multitier Open-Ended Work and Energy Instrument (MOWEI) Using Rasch Analysis to Identify Students' Misconceptions. *Cypriot Journal of Educational Sciences*, *16*(1), 16-32. https://doi.org/10.18844/cjes.v16i1.5504

Sari, K. M. (2019, June). Metacognitive Knowledge and Critical Thinking Biology 11th of Public Senior High School in Bogor. In *Journal of Physics: Conference Series* (Vol. 1241, No. 1, p. 012056). IOP Publishing. https://doi.org/10.1088/1742-6596/1241/1/012056

Sukarelawan, M., Sulisworo, D., Kuswanto, H., & Rofiqah, S. A. (2021). Heat and Temperature Metacognition Awareness Inventory: A Confirmatory Factor Analysis. *International Journal of Evaluation and Research in Education, 10*(2), 389–395. https://doi.org/10.11591/ijere.v10i2.20917

Sumintono, B. (2018, February). Rasch model measurements as tools in assesment for learning. In *1st International Conference on Education Innovation (ICEI 2017)* (pp. 38-42). Atlantis Press. https://doi.org/10.2991/icei-17.2018.11

Suryana, T. G. S., Setyadin, A. H., Samsudin, A., & Kaniawati, I. (2020, February). Assessing multidimensional energy literacy of high school students: an analysis of rasch model. In *Journal of Physics: Conference Series* (Vol. 1467, No. 1, p. 012034). IOP Publishing. https://doi.org/10.1088/1742-6596/1467/1/012034

Taherdoost, H. (2016). Validity and reliability of the research instrument; how to test the validation of a questionnaire/survey in a research. *International journal of academic research in management (IJARM)*, *5*. https://doi.org/10.2139/ssrn.3205040

Teng, M. F. (2025). *Metacognition in language teaching*. Cambridge University Press.
Van Eck, N., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *scientometrics*, *84*(2), 523-538. https://doi.org/10.1007/s11192-009-0146-3
Wenden, A. L. (1998). Metacognitive knowledge and language learning. *Applied linguistics*, *19*(4), 515-537. https://doi.org/10.1093/applin/19.4.515
Widhiarso, W., & Sumintono, B. (2016). Examining response aberrance as a cause of outliers in statistical analysis. *Personality and Individual Differences*, *98*, 11-15. https://doi.org/10.1016/j.paid.2016.03.099
Zahran, M., Samsudin, A., Suhandi, A., Aminudin, A. H., Wibowo, F. C., Prahani, B. K., & Kapıcı, H. Ö. (2025). Development of Augmented Reality for Special Education Needs (ARSEN) to Enhance Students' Conceptions of Solar System. *Multidisciplinary Science Journal*, *7*(11), 2025531-2025531. https://doi.org/10.31893/multiscience.2025531
Ziegler, M., & Brunner, M. (2016). Test standards and psychometric modeling. In *Psychosocial skills and school systems in the 21st century: Theory, research, and practice* (pp. 29-55). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-28606-8_2