Inovasi Kurikulum *- p-ISSN 1829-6750 & e-ISSN 2798-1363*
Volume 22 No 4 (2025) 2611-2626

Inovasi Kurikulum
https://ejournal-hipkin.or.id/index.php/jik

Development of a socioscientific issues-based instrument for measuring critical thinking

Lusy Fatika Sari[1], Sri Latifah[2], Yani Suryani[3]
[1,2,3]Universitas Islam Negeri Raden Intan Lampung, Lampung, Indonesia
lusyfatika22@gmail.com[1], srilatifah@radenintan.ac.id[2], yanisuryani@radenintan.ac.id[3]

ABSTRACT

The development of 21st-century education requires students to possess critical thinking skills to cope with the complexity of modern life problems. However, various studies have shown that students' critical thinking skills in science learning remain relatively low because current learning processes do not yet emphasize analysis and reflection. Therefore, a valid and reliable assessment instrument is needed to measure and foster critical thinking skills effectively. This study aims to develop a Socioscientific Issues (SSI)-based assessment instrument to measure students' critical thinking skills regarding heat and temperature. The development employed the ADDIE model and involved 70 eleventh-grade students. The instrument covers three domains: cognitive, affective, and psychomotor, which were validated by experts and tested using the Rasch model. The study's results showed that the instrument was very suitable for the material, media, and evaluation. Both item and person reliability were in the good category, while teachers' and students' responses showed very positive results. The developed SSI-based instrument is considered valid, reliable, and feasible for comprehensively assessing students' critical thinking skills.

ARTICLE INFO

Article History:
Received: 22 Jul 2025
Revised: 2 Nov 2025
Accepted: 16 Nov 2025
Publish online: 25 Nov 2025
Keywords:
critical thinking; heat and temperature; Rasch model; Socioscientific Issues; SSI

Open access
Inovasi Kurikulum is a peer-reviewed open-access journal.

ABSTRAK

Perkembangan pendidikan abad ke-21 menuntut murid memiliki kemampuan berpikir kritis untuk menghadapi kompleksitas permasalahan kehidupan modern. Namun, berbagai penelitian menunjukkan bahwa kemampuan berpikir kritis murid dalam pembelajaran sains masih tergolong rendah akibat proses belajar yang belum menekankan analisis dan refleksi. Oleh karena itu, diperlukan instrumen penilaian yang valid dan reliabel untuk mengukur serta menumbuhkan kemampuan berpikir kritis secara efektif. Penelitian ini bertujuan mengembangkan instrumen penilaian berbasis Socioscientific Issues (SSI) guna mengukur kemampuan berpikir kritis murid pada materi suhu dan kalor. Pengembangan dilakukan menggunakan model ADDIE dengan melibatkan 70 murid kelas XI. Instrumen mencakup tiga aspek penilaian, yaitu kognitif, afektif, dan psikomotorik, yang divalidasi oleh ahli serta diuji menggunakan model Rasch. Hasil penelitian menunjukkan bahwa instrumen dinyatakan sangat layak untuk materi, media, dan juga evaluasi. Nilai reliabilitas item dan person berada pada kategori baik, sedangkan respons guru dan murid menunjukkan hasil sangat baik. Instrumen berbasis SSI yang dikembangkan dinyatakan valid, reliabel, dan layak digunakan untuk mengukur kemampuan berpikir kritis murid secara komprehensif.

Kata Kunci: *berpikir kritis; Rasch model; Socioscientific Issues; SSI; suhu dan kalor*

How to cite (APA 7)
Sari, L. F., Latifah, S., & Suryani, Y. (2025). Development of a socioscientific issues-based instrument for measuring critical thinking. *Inovasi Kurikulum, 22*(4), 2611-2626.

Peer review
This article has been peer-reviewed through the journal's standard double-blind peer review, where both the reviewers and authors are anonymised during review.

https://doi.org/10.64014/jik.v22i4.000

INTRODUCTION

21st-century educational developments demand that students possess higher-order thinking skills to navigate the complexities of modern life (García-Carmona, 2023). One essential skill students must master is critical thinking, which serves as the foundation for understanding concepts and making informed scientific decisions (Ochildinovna, 2024). Previous research reveals that students' critical thinking abilities in science learning remain relatively low because the learning process has not sufficiently emphasized analysis and reflection (Prabayanti *et al*., 2025). This condition underscores the urgent need for instructional strategies and assessment instruments capable of effectively fostering students' critical thinking skills (Arifin *et al*., 2025). Developing critical thinking instruments is vital to ensure these abilities can be measured validly and reliably within the context of science education (Afikah *et al*., 2024).

However, results from the 2022 Programme for International Student Assessment (PISA) indicate that critical thinking skills among Indonesian students are still subpar, with an average science literacy score of 383—significantly lower than the OECD average of 485. Only approximately 18% of students reached Level 2, which signifies a basic ability to interpret and recognize simple situations scientifically (source*:* https://www.oecd.org/en/publications/pisa-2022-results-volume-i_53f23881-en.html). These data suggest that most Indonesian students struggle with analyzing, evaluating, and solving problems—all of which are core components of critical thinking.

This issue is evident in physics education, particularly in the topic of temperature and heat, where students frequently harbor misconceptions, such as the notions of "hot heat" and "cold heat" (Sundari & Sarkity, 2021). Furthermore, physics instruction often focuses on mathematical problem-solving without practicing contextual analysis (Sartika *et al*., 2021). Consequently, students tend to memorize formulas without understanding their underlying meaning, leading to the perception of physics as a difficult and uninteresting subject. In reality, physics education requires the integration of knowledge, scientific attitudes, and practical skills, as mandated by National Education Standards, which state that graduate competencies must encompass cognitive, affective, and psychomotor domains. In practice, however, school-based assessments remain centered on lower-level cognitive domains (C1-C2), often neglecting the affective aspect (Kafii & Dwikoranto, 2023).

Meanwhile, the psychomotor aspect of laboratory skills has not received sufficient attention in learning activities. Therefore, evaluation plays a critical role in assessing the achievement of learning objectives in physics. Effective assessment should not only evaluate conceptual mastery but also students' attitudes and skills in applying physics concepts in the field (Zahroh *et al*., 2025). Comprehensive evaluation assists educators in obtaining a holistic view of students' abilities across various learning domains (Imania & Bariah, 2019). Thus, the development of assessment instruments that cover the three primary domains—cognitive, affective, and psychomotor—is essential to ensure that evaluation results reflect students' learning holistically. Several studies support the importance of the three-domain assessment in physics education. Learning evaluation should accommodate these three domains because assessments focusing solely on cognitive aspects fail to provide a comprehensive overview of student capabilities (Rahmawati *et al*., 2024).

Knowledge, attitude, and skill assessments are complementary components of the evaluation process; therefore, instruments spanning all three domains are required to ensure holistic and accurate assessment. However, instruments based on Socio-Scientific Issues (SSI) that measure learning skills across these three domains remain very limited, particularly for temperature and heat. Consequently, the development of SSI-based instrumentation is crucial for more deeply measuring students' critical thinking skills while simultaneously assessing their scientific attitudes and practical skills. Based on a preliminary study conducted through a needs analysis questionnaire administered to teachers and students at SMA Muhammadiyah 2 Bandar Lampung, several physics teachers indicated a need for new approaches to

https://doi.org/10.64014/jik.v22i4.206

facilitate students' understanding of temperature and heat. Teachers believe that learning media using an SSI approach is more effective in engaging students' interest than conventional methods. Furthermore, the tests currently in use measure only conceptual mastery and do not yet probe students' critical thinking skills in depth.

Simultaneously, survey results from 70 students revealed that their critical thinking skills require improvement. Only 41.4% of students are accustomed to questioning information before accepting it as truth, and only 25.7% frequently evaluate various perspectives before concluding. Conversely, 60% of students expressed interest in discussing the application of physics concepts to social problem-solving, and 50% desired test items that challenge their critical thinking. Thus, there is a need to develop SSI-based instrumentation to measure critical thinking skills in temperature and heat to make physics learning more contextual, meaningful, and capable of honing students' higher-order critical thinking. While previous studies have developed critical thinking instruments, research that integrates SSI with all three assessment domains simultaneously is still lacking (Mappalesye *et al*., 2021). The novelty of this research lies in the development of SSI-based critical thinking instruments for temperature and heat that encompass cognitive and psychomotor aspects concurrently.

This approach is expected to provide a complete descriptive profile of students' critical thinking skills and offer a practical contribution to teachers in implementing 21st-century assessments (Fitroty *et al*., 2023). This study aims to develop SSI-based test instruments to measure students' critical thinking skills across the cognitive, affective, and psychomotor domains related to temperature and heat. Additionally, this research aims to test the validity and reliability of the developed instruments and evaluate teacher and student responses to their application as physics evaluation tools. The results of this research are expected to provide both theoretical and practical benefits. Theoretically, this study enriches the literature on applying the SSI approach to developing comprehensive assessment instruments for physics education. Practically, the results serve as a reference for teachers and researchers in designing assessments that measure holistic critical thinking and as evaluation material for schools to improve the quality of instruction.

LITERATURE REVIEW

Assessment Instruments

Assessment instruments are crucial tools in the educational evaluation process, serving to measure the extent to which learning objectives have been successfully achieved. Based on their format, assessment instruments are categorized into two primary types: test and non-test instruments (Pitaloka *et al*., 2023). Test instruments are typically employed to measure a student's cognitive domain through written questions, such as multiple-choice, short answers, or essays. Conversely, non-test instruments are utilized to assess the affective and psychomotor domains through observation, interviews, reflective journals, attitude scales, self-assessments, or performance rubrics (Ahmad *et al*., 2025). A high-quality test instrument must possess high reliability and validity to ensure that the measurement results accurately reflect the student's actual abilities (Darman *et al*., 2024).

In modern developmental research, Rasch model analysis is frequently utilized to evaluate item characteristics, item fit, and instrument reliability. This is because the Rasch model provides in-depth information regarding item difficulty, respondent ability, and the relationship between the two within a single logit scale (Widodo *et al*., 2024). Non-test instruments play a vital role in measuring students' attitudes, interests, and skills that written tests cannot capture. Through non-test assessments, teachers can evaluate affective and psychomotor aspects contextually and authentically (Pitaloka *et al*., 2023). This

https://doi.org/10.64014/jik.v22i4.000

type of assessment supports the implementation of the *Kurikulum Merdeka* (Independent Curriculum), which emphasizes a balance between knowledge, attitudes, and skills (Zarkasyi *et al*., 2023).

Critical Thinking Skills

Critical thinking is a mental process involving the ability to analyze information, evaluate arguments, and make logical decisions based on available evidence (Septiany *et al*., 2024). This capability encompasses verifying the accuracy of information and formulating conclusions that can be rationally justified. According to widely used theoretical frameworks, critical thinking skills are categorized into five primary indicators. First, elementary clarification involves providing fundamental explanations, including the ability to focus on specific questions and to formulate clear, concise statements. Second, basic support refers to the capacity to establish a sound basis for decision-making by evaluating the credibility of sources and identifying relevant supporting reasons. Third, inference is the ability to derive logical conclusions based on the synthesis of available evidence and information.

Fourth, advanced clarification entails providing further depth through precise definitions and the rigorous verification of conceptual consistency. Fifth, strategies and tactics involve the ability to select effective problem-solving strategies and to systematically evaluate alternative decisions (Lestari *et al*., 2021). Furthermore, critical thinking is demonstrated through activities such as assessing the strength of arguments, identifying hidden assumptions, and testing the logical consistency of statements. This process requires individuals to perform an in-depth analysis of information before reaching a decision, ensuring that the resulting conclusion is grounded in valid and accountable evidence (Septiany *et al*., 2024). These skills are closely linked to the ability to evaluate alternative solutions and examine the relevance of the data utilized. Consequently, students with well-developed critical thinking skills are capable of drawing more accurate, systematic, and well-supported conclusions.

Socioscientific Issues (SSI)

The Socioscientific Issues (SSI) approach is a science education strategy that integrates contemporary social issues such as climate change, social justice, and technological ethics into learning activities. This approach ensures that students do not merely study scientific concepts but also learn to think critically, evaluate evidence, and make decisions grounded in social values (Macalalag *et al*., 2024). Within this framework, socio-technical issues are selected based on their real-world relevance and moral controversy, allowing students to engage in argumentative discussions and ethical reflections within a scientific context (Selamat, 2021). Furthermore, as a pedagogical approach, SSI strengthens students' scientific literacy through active engagement in solving global and local problems rooted in social reality, while fostering awareness of scientific and social responsibility (Arjaya & Surata, 2024). Moreover, SSI-based learning encourages students to develop critical thinking skills by analyzing social issues intertwined with science, such as global warming, environmental pollution, or climate change.

Through these contexts, students not only comprehend scientific principles but also learn to rationally weigh various perspectives in scientific decision-making (Fihani *et al*., 2021). In the practice of selecting issues for SSI, the literature emphasizes several consistent criteria. Issues must be authentic and currently discussed in society or the media, relevant to students' lives, so that the decisions made are meaningful, and inherently controversial to allow for diverse viewpoints and open analytical space (Viehmann *et al*., 2024). Additionally, SSI issues should provide opportunities for ethically safe open discussions and maintain a clear connection with scientific and technological concepts, enabling instruction to bridge scientific facts with social contexts. National studies reinforce the importance of these criteria to ensure that selected issues are locally relevant and capable of stimulating students' critical thinking skills within the context of science education in Indonesia (Rosmayuni *et al*., 2024).

https://doi.org/10.64014/jik.v22i4.206

Temperature and Heat

Temperature and heat are fundamental concepts in physics related to thermal energy and its transfer. Temperature serves as a measure of the degree of hotness or coldness of an object, determined by the average kinetic energy of its constituent particles. In contrast, heat is the energy that is transferred due to a temperature gradient between two systems. Mastery of these two concepts is essential, as they form the basis for various physical phenomena, including phase changes, thermal expansion, and the mechanisms of heat transfer, namely conduction, convection, and radiation. However, students frequently harbor misconceptions, such as conflating temperature with heat or erroneously believing that an object "contains" heat (Hara *et al*., 2023).

These difficulties are often exacerbated by instructional practices that remain centered on formulas and mathematical computations rather than conceptual understanding and real-world implementation (Sartika *et al*., 2021). A context-based or **Socioscientific Issues (SSI)** approach is considered effective in addressing these issues, as it bridges the concepts of temperature and heat with real-world challenges such as global warming, energy consumption, and climate change. Integrating the study of temperature and heat with social and environmental contexts not only facilitates deep conceptual understanding for students but also fosters critical thinking skills and scientific awareness regarding global dilemmas.

METHODS

This study employs a Research and Development (R&D) approach to generate a specific product and evaluate its effectiveness. The developmental framework follows the ADDIE model, which consists of five stages: Analysis, Design, Development, Implementation, and Evaluation. The research was conducted at SMA Muhammadiyah 2 Bandar Lampung during the even semester of the 2024/2025 academic year, involving 70 eleventh-grade students as subjects. The research stages are detailed as follows.

1. **Analysis Phase**: This stage was conducted to identify the necessity for instrument development by collecting data through questionnaires distributed to teachers and eleventh-grade students. The data obtained were used to evaluate instructional needs and to determine the urgency of developing the instrument. The results of this needs analysis served as the foundation for designing the initial product and refining the instrument to align with the practical requirements of physics education.
2. **Design Phase**: Activities in this phase included developing the instrument blueprints for the cognitive, affective, and psychomotor domains, selecting the essay test format, and drafting test items based on Ennis's critical thinking indicators. Additionally, SSI-based student worksheets (LKPD) were developed to support psychomotor assessment. The initial design was subsequently consulted with research supervisors to obtain constructive feedback before proceeding to the development stage.
3. **Development Phase**: This stage involved the validation of the instrument and the SSI-based worksheets by subject matter experts, media specialists, and evaluation experts to assess content, readability, and media presentation. Feedback from these experts was utilized to revise the product. This phase resulted in a refined version of the instrument, incorporating the validators' recommendations.
4. **Implementation Phase**: The instrument was administered to eleventh-grade students to assess its practicality and item quality. Further revisions were made in response to feedback from both teachers and students.
5. **Evaluation Phase**: This phase involved analyzing the trial results, focusing on validity and reliability

https://doi.org/10.64014/jik.v22i4.000

using the Rasch model. The analysis determined the feasibility of the instrument as a comprehensive tool for measuring students' critical thinking skills, resulting in a final product that is valid, reliable, and ready for use in SSI-based physics instruction.

After the validators completed and validated the questionnaires, the researcher analyzed the assessment data to determine the instrument's feasibility level. Subsequently, the trial data were further analyzed using the Rasch model via the Winsteps 5.0.3.4 software. Rasch analysis was conducted to evaluate the instrument's validity and reliability. Validity testing included overall validity through summary statistics, item validity through item fit order analysis, and construct validity. Instrument reliability was examined through Cronbach's Alpha values and the reliability indices for both items and respondents. Furthermore, item difficulty levels were analyzed using logits and visualized through a wright map.

RESULTS AND DISCUSSION

Analysis Phase

In this stage, the researcher conducted a needs analysis as the foundation for developing a critical thinking assessment instrument integrated with Socioscientific Issues (SSI) regarding the topic of temperature and heat. The analysis was performed by distributing questionnaires to physics teachers and 70 eleventh-grade students at SMA Muhammadiyah 2 Bandar Lampung during the 2024/2025 academic year. The results indicated that teachers require instruments capable of assessing critical thinking skills comprehensively, as existing assessments remain limited to conceptual mastery. Furthermore, 60% of the students expressed interest in discussing the application of physics concepts within the social domain, while 50% desired test items that challenge their critical thinking abilities. Based on these findings, an SSI-based instrument focusing on the cognitive, affective, and psychomotor domains was developed to enhance the meaningfulness, contextual relevance, and effectiveness of physics learning and to hone students' higher-order thinking skills.

Design Phase

The design phase was categorized into three components: material design, assessment instrument design, and Student Worksheet (LKPD) design. The details of each component are as follows.

1. **Material Design:** The developed materials focus on the concepts of temperature and heat in accordance with the *Kurikulum Merdeka* (Independent Curriculum), emphasizing the correlation between physics concepts and social or environmental issues through the SSI approach. The materials were structured according to the needs analysis, which highlighted the need for contextual evaluation instruments oriented toward higher-order thinking. The material outline includes: 1) The definitions of temperature and heat; 2) Heat transfer via conduction, convection, and radiation; 3) The application of temperature and heat concepts in daily life; 4) The connection between these phenomena and global issues, such as global warming and the greenhouse effect.
2. **Assessment Instrument Design:** The assessment instruments were developed to comprehensively measure students' critical thinking skills across three domains.
    a. Cognitive Domain**:** An SSI-based essay test utilizing contextual stimuli related to global warming and climate change.
    b. Affective Domain**:** A Likert-scale questionnaire designed to evaluate students' scientific attitudes toward science-related social issues.
    c. Psychomotor Domain**:** Observation sheets and self-assessment rubrics linked to a simple experimental activity on the greenhouse effect using incandescent lamps.

https://doi.org/10.64014/jik.v22i4.206

Each instrument was designed based on indicators derived from Ennis's Critical Thinking Theory and Simpson's Taxonomy for the psychomotor domain. These instruments undergo validation by subject matter, evaluation, and media experts before being tested using the Rasch Model.

3. **Student Worksheet (LKPD) Design:** To complement the psychomotor assessment, an SSI-based LKPD was developed to support a simple greenhouse effect laboratory activity. This worksheet assists students in conducting observations, recording experimental results, analyzing data, and deriving scientific conclusions relevant to environmental issues. The LKPD serves as a contextual learning tool that fosters critical, collaborative, and reflective thinking skills during the experimental process. The design features a systematic and visually engaging layout to ensure ease of use. Generally, the LKPD structure consists of an introductory section (identity, instructions, and learning objectives), a concept map linking temperature and heat topics to climate change issues, and a practical activity: a simple greenhouse effect experiment, equipped with observation sheets and prompts for scientific attitude reflection.

Figure 1. Design layout of the Socioscientific Issues-based student worksheet (LKPD): (a) LKPD cover page, (b) Learning objective flow, and (c) Excerpt of the greenhouse effect laboratory activity.
Source: 2025 Research

The design of the Student Worksheet (LKPD) is presented in **Figure 1**, illustrating (a) the LKPD cover page, (b) the concept map for temperature and heat, and (c) an excerpt from the greenhouse effect laboratory activity as an application of temperature and heat concepts within a social context.

Development Phase

In this stage, the previously designed instruments, consisting of essay questions, Likert scales, and self-assessment rubrics, along with supporting materials such as the LKPD and psychomotor assessment sheets, were validated by experts to evaluate their content, construct, and linguistic feasibility. The validation process involved subject matter and instrument evaluation experts, as well as media specialists, to ensure the alignment of the instruments with the learning objectives and the Socioscientific Issues (SSI) context. The validation results from the subject matter, evaluation, and media experts are presented below.

https://doi.org/10.64014/jik.v22i4.000

Table 1. Results of Subject Matter, Evaluation, and Media Expert Validation

Validation	Aspect	Percentage	Average	Category
Subject Matter Expert	Clarity of Content	84,00%	86,41%	Very Worthy
	Visual Conformity	86,67%		
	Language	88,57%		
Evaluation Expert	Construct	80,00%	80,00%	Worthy
	Language	80,00%		
	Evaluation Presentation	80,00%		
Media Expert	Visual Display	84,00%	86,55%	Very Worthy
	Psychomotor Questions	89,09%		

Source: 2025 Research

Based on the validation results presented in **Table 1**, the instrument underwent several adjustments regarding its content and achieved an average score of 86.41%, placing it in the "Highly Feasible" category. Furthermore, the validation results from the evaluation expert showed an average score of 80.00%, also categorized as "Highly Feasible", while the media expert provided an average rating of 86.55%, similarly falling within the "Highly Feasible" category.

Implementation Phase

The implementation phase involved administering the instrument to eleventh-grade students to assess its practicality and item quality, followed by further revisions based on feedback from both teachers and students. This stage aimed to evaluate the extent to which the test instruments, LKPD, and other supporting materials could be effectively utilized in SSI-based physics instruction for the topic of temperature and heat. Teacher and student responses were obtained using the instruments at SMA Muhammadiyah 2 Bandar Lampung. Teachers noted that the instruments and LKPD were highly relevant to the learning outcomes and capable of assessing students' critical thinking skills, scientific attitudes, and psychomotor domains. Additionally, teachers provided positive feedback, stating that the SSI-based learning activities were engaging and contextual, helping them understand the relationship between temperature and heat concepts and global issues.

https://doi.org/10.64014/jik.v22i4.206

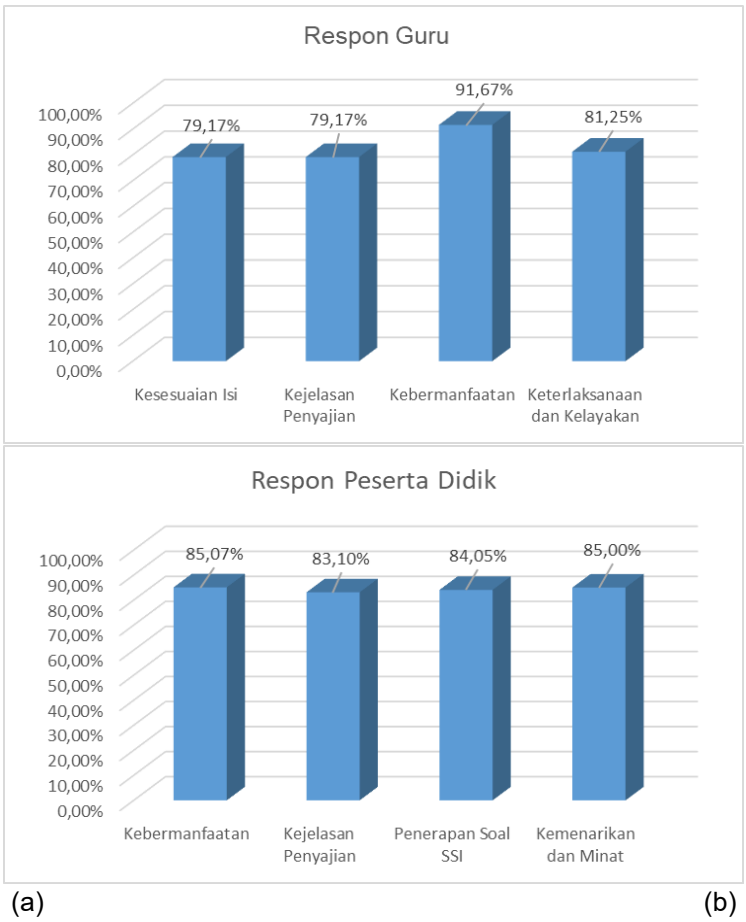

Figure 3. Teacher (a) and Student (b) Response Results toward Socioscientific Issues-based Instruments and Student Worksheets (LKPD)

Source: 2025 Research

The responses from teachers and students regarding the SSI-based instruments and LKPD are presented in **Figure 3**. The results indicate a highly positive assessment. Teachers evaluated the developed test instruments and LKPD as relevant to the learning outcomes and indicators for critical thinking skills, scientific attitudes, and psychomotor skills. The layout, language, and instructions were deemed clear and easy to understand, thereby assisting teachers in assessing student abilities comprehensively, with an average response score of 82.82%, categorized as "Excellent". Similarly, students provided positive feedback, perceiving the SSI-based LKPD as engaging, contextual, and capable of linking physics concepts with social issues such as global warming and the greenhouse effect. This was reflected in an average response score of 84.31%, which also falls into the "Excellent" category.

Evaluation Phase

The evaluation phase was conducted to assess the quality of the instruments following the field trials. This process involved processing the data from test results and student responses. The assessment data obtained from the SSI-based critical thinking test instruments were analyzed using the Rasch model. The analysis results indicated that 25 essay items, 15 scientific attitude items, and 10 psychomotor items were valid. In terms of reliability, the essay items achieved a reliability index in the "Excellent" category, the scientific attitude items also reached the "Excellent" category, and the psychomotor items were classified in the "Good" category.

https://doi.org/10.64014/jik.v22i4.000

Data Description and Analysis of Research Implementation

Unidimensionality

Unidimensionality analysis was conducted to ensure that each instrument measures only a single intended ability construct. Based on the Raw Variance Explained by Measures results for the three instruments—29.4% for the critical thinking essay test, 30.7% for the scientific attitude scale, and 32.3% for the psychomotor skill instrument—all values exceeded the minimum threshold of 20%. Meanwhile, the Unexplained Variance values for the three instruments were 13.2%, 12.6%, and 10.8%, respectively, all of which remained below the 15% limit. These results indicate that all three instruments fulfill the criteria for unidimensionality, confirming that each item consistently measures a single underlying construct.

Item Difficulty

Item difficulty refers to the challenge level of a specific test item, indicating how easy or difficult it is for respondents to provide a correct or high-scoring answer.

Table 2. Instrument Item Difficulty

Instrument Realm	Range Division Reference	Range Logit Item	Category	Question Item Code
Critical Thinking Essay Test	> +1 SD	> +0,44	Very difficult	E25, E18, E20, E19, E22, E24
	0,00 s/d +1 SD	0,00 s/d 0,44	Difficult	E17, 321, E23
	-1 SD s/d 0,00	-0,44 s/d 0,00	Easy	E15, E10, E1, E13, E14, E16, E9, E5, E11, E12, E6
	< -1 SD	< -0,44	Very easy	E7, E4, E2, E3, E8
Scientific Attitude Scale	> +1 SD	> +0,36	Very difficult	I14, I12
	0,00 s/d +1 SD	0,00 s/d 0,36	Difficult	I15, I13
	-1 SD s/d 0,00	-0,36 s/d 0,00	Easy	I9, I11, I6, I3, I10
	< -1 SD	< -0,36	Very easy	I8, I4, I7, I5, I2, I1
Practical Skills Instrument	> +1 SD	> +0,40	Very difficult	S7, S9, S10
	0,00 s/d +1 SD	0,00 s/d 0,40	Difficult	S1
	-1 SD s/d 0,00	-0,40 s/d 0,00	Easy	S6, S3, S8, S4
	< -1 SD	< -0,40	Very easy	S5, S2

Source: Primary Data, 2025

Item difficulty for each instrument is summarized in **Table 2**. Based on the Rasch Model analysis, the three developed instruments—the critical thinking essay test (cognitive), the scientific attitude scale (affective), and the practical skills instrument (psychomotor)—demonstrate an even distribution of items across four difficulty categories: easy, difficult, very difficult, and very easy. This distribution indicates that each instrument has good discriminative power and can measure students across a wide range of ability levels. In the critical thinking essay test (cognitive domain), 6 items were categorized as very difficult, 3 items as difficult, 10 items as easy, and 5 items as very easy. This proportion indicates a balanced distribution of

https://doi.org/10.64014/jik.v22i4.206

items: approximately 36% are classified as difficult or very difficult, whereas 64% are classified as easy or very easy.

This condition indicates that while the majority of students could answer the questions successfully, several items effectively differentiated higher-order critical thinking abilities. Thus, the instrument demonstrates strong discriminative ability in distinguishing variations in student competence. For the scientific attitude scale (affective domain), the item composition consists of 2 very difficult items, 2 difficult items, 5 easy items, and 5 very easy items. This distribution creates a balance between items requiring deep reflection on scientific attitudes and statements that are more easily accepted by students. Consequently, this scale not only measures positive inclinations toward social and environmental issues but also assesses the consistency and depth of students' scientific attitudes toward SSI phenomena.

Regarding the practical skills instrument (psychomotor domain), the analysis indicates 3 items in the very difficult category, 1 item in the difficult category, 4 items in the easy category, and 2 items in the very easy category. This proportion shows that approximately 40% of the items are difficult or very difficult, while 60% are easy or very easy. Such a distribution indicates that the instrument can assess students' practical skills across diverse competency levels, from basic to advanced. Overall, the proportion of difficult and easy items across the three instruments ranges from 35%–40% for the difficult categories and 60%–65% for the easy categories, indicating a balanced difficulty level. Therefore, the developed instruments meet the ideal criteria of Rasch measurement theory, providing an even distribution of difficulty levels that comprehensively measure students' abilities across the cognitive, affective, and psychomotor domains.

Item Fit

Item fit indicates the extent to which a test item functions optimally in measuring the construct it is intended to measure. If an item is identified as a "misfit," it indicates that students may have encountered misunderstandings or misconceptions regarding that specific item. The item fit results for the instruments are summarized in **Table 3**.

Table 3. Instrument Item Fit

Instrument Realm	Item	MNSQ Outfit	MNSQ Criteria	ZSTD	Correlation (Pt. Measure Corr)	PM Criteria	Fit/Misfit
Critical Thinking Essay Test	E25	1.82	Good	4.4	0.42	Good	Fit
	E1	1.34	Good	2.1	0.18	Good	Fit
	E15	1.19	Good	1.1	0.53	Very good	Fit
	E20	1.19	Good	1.1	0.53	Very good	Fit
	E22	1.13	Good	0.8	0.56	Very good	Fit
	E2	1.12	Good	0.6	0.52	Very good	Fit
	E3	1.04	Good	0.5	0.55	Very good	Fit
	E18	1.04	Good	0.5	0.55	Very good	Fit
	E10	1.04	Good	0.5	0.49	Very good	Fit
	E9	1.00	Good	0.3	0.46	Very good	Fit
	E16	0.97	Good	-0.3	0.49	Very good	Fit
	E23	0.95	Good	-0.3	0.48	Very good	Fit
	E21	0.94	Good	-0.6	0.45	Very good	Fit

https://doi.org/10.64014/jik.v22i4.000

Instrument Realm	Item	MNSQ Outfit	MNSQ Criteria	ZSTD	Correlation (Pt. Measure Corr)	PM Criteria	Fit/Misfit
	E19	0.87	Good	-0.7	0.46	Very good	Fit
	E6	0.84	Good	-1.0	0.55	Very good	Fit
	E17	0.80	Good	-1.3	0.52	Very good	Fit
	E8	0.77	Good	-1.3	0.51	Very good	Fit
	E14	0.66	Good	-2.0	0.56	Very good	Fit
	E13	0.59	Good	-3.2	0.54	Very good	Fit
	E12	0.59	Good	-3.2	0.54	Good	Fit
	E24	0.78	Good	-1.1	0.50	Very good	Fit
	E4	1.15	Good	0.6	0.42	Very good	Fit
	E5	1.14	Good	0.8	0.51	Very good	Fit
	E7	0.85	Good	-0.8	0.49	Very good	Fit
	E11	0.90	Good	-0.5	0.47	Very good	Fit
Scientific Attitude Scale	I6	1.17	Good	1.11	0.44	Good	Fit
	I5	1.14	Good	0.95	0.48	Very good	Fit
	I15	1.09	Good	0.83	0.55	Very good	Fit
	I9	1.08	Good	0.63	0.58	Very good	Fit
	I10	1.07	Good	0.41	0.47	Very good	Fit
	I2	1.05	Good	0.31	0.53	Very good	Fit
	I13	1.04	Good	0.29	0.69	Very good	Fit
	I14	0.99	Good	-0.27	0.69	Very good	Fit
	I3	0.91	Good	-0.92	0.62	Very good	Fit
	I12	0.88	Good	-1.23	0.73	Very good	Fit
	I8	0.85	Good	-1.42	0.54	Very good	Fit
	I11	0.81	Good	-1.78	0.74	Very good	Fit
	I7	0.80	Good	-1.89	0.57	Very good	Fit
	I1	0.70	Good	-2.31	0.44	Good	Fit
	I4	0.70	Good	-2.31	0.44	Good	Fit
Practical Skills Instrument	S1	1.23	Good	1.23	0.53	Good	Fit
	S2	1.19	Good	1.19	0.39	Good	Fit
	S4	1.13	Good	1.10	0.58	Very good	Fit
	S3	1.09	Good	0.96	0.62	Very good	Fit
	S9	0.99	Good	-0.21	0.59	Very good	Fit
	S5	0.91	Good	-0.49	0.64	Very good	Fit
	S10	0.87	Good	-0.91	0.59	Very good	Fit
	S7	0.81	Good	-1.46	0.64	Very good	Fit
	S6	0.75	Good	-1.86	0.64	Very good	Fit
	S8	0.73	Good	-1.96	0.53	Good	Fit

Source: Primary Data, 2025

https://doi.org/10.64014/jik.v22i4.206

Table 3 presents the results of the item fit analysis based on three primary parameters: Outfit Mean Square (MNSQ), Z-Standardized (ZSTD), and Point Measure Correlation (PT-MEA), with the following details.

1. **Outfit MNSQ Parameter:** This parameter is utilized to assess the extent to which an item aligns with the Rasch model. Based on the analysis, all items across the three instruments—the critical thinking essay, the scientific attitude scale, and the practical skills instrument—yielded Outfit MNSQ values within the ideal range of 0.5 to 1.5. This demonstrates that all items fulfill the criteria for a good fit with the Rasch model. In the critical thinking instrument, the highest value was observed in item E25 (1.82), while the lowest was in item E13 (0.59). Although item E25 is slightly above the ideal threshold, it remains acceptable as it does not exhibit extreme distortion. For the scientific attitude scale, the highest Outfit MNSQ was for item I6 (1.17) and the lowest for item I12 (0.88), while in the practical skills instrument, the highest value was for item S1 (1.23) and the lowest for item S8 (0.73). These values indicate that no items require revision, as all fall within the fit category.
2. **ZSTD (Z-Standardized) Parameter:** This describes the statistical deviation of empirical data from the model. The ideal ZSTD value ranges from −2 to +2, indicating that the item does not deviate significantly from the Rasch model. In the critical thinking instrument, the highest value was found in item E25 (4.4) and the lowest in item E13 (−3.2). Although some values slightly exceed the limits, these deviations are considered minor and do not affect the overall model fit. For the scientific attitude scale, ZSTD values ranged from −3.2 to 1.11, whereas for the practical skills instrument, ZSTD values ranged from −1.96 to 1.23. These results signify the absence of extreme outliers; thus, all items are considered highly compatible with the Rasch model.
3. **Point Measure Correlation (PT-MEA):** This measures the relationship between participant responses and the ability levels estimated by the Rasch model. The ideal range for PT-MEA correlation is 0.3 to 0.8, where higher values indicate better discriminating power. In the critical thinking essay test, the lowest correlation was found in item E1 (0.18) and the highest in item E22 (0.56). In the scientific attitude scale, correlations ranged from 0.44 to 0.74, whereas in the practical skills instrument, correlations ranged from 0.39 to 0.64. All items demonstrated positive correlations, signifying that participant responses are consistent with the measured ability constructs.

Based on these three parameters (Outfit MNSQ, ZSTD, and PT-MEA), all items in the three instruments are declared fit with the Rasch model. No items needed to be deleted or revised, as no significant discrepancies were found. These results confirm that the instruments possess a high level of quality, with items capable of measuring constructs consistently and reliably in accordance with the Rasch model assumptions. Therefore, all items across the three domains are deemed feasible for measuring students' cognitive abilities, scientific attitudes, and practical skills.

Item dan Person Reliability

The reliability analysis results indicate that the three developed instruments possess a high level of consistency. In the critical thinking essay test (cognitive domain), the item reliability was 0.87, and the person reliability was 0.84, both falling into the "Good" category. This indicates that the test items and student responses are stable and consistent. Furthermore, the scientific attitude scale (affective domain) achieved an item reliability of 0.84 (Good) and a person reliability of 0.78 (Fair/Sufficient), signifying that the statement items are sufficiently reliable despite variations in student responses. As for the practical skills instrument (psychomotor domain), the item reliability was 0.85 (Good) and the person reliability was 0.68 (Fair/Sufficient), showing that the observation items are consistently assessed even though

https://doi.org/10.64014/jik.v22i4.000

participant abilities vary. Overall, all three instruments are deemed reliable and can consistently measure the cognitive, affective, and psychomotor domains.

Reliability Testing

The results of the reliability analysis across the three instrument domains demonstrate that all instruments achieved an optimal level of reliability. The critical thinking essay test attained the highest reliability value of 0.86, categorized as "Excellent," followed by the scientific attitude scale with a value of 0.81, also categorized as "Excellent." Meanwhile, the practical skills instrument showed a reliability of 0.71, placed in the "Good" category. These values indicate that the instruments possess sufficient internal consistency and are feasible for measuring student ability levels within their respective domains.

Discussion

The analysis using the Rasch model demonstrates that all items across the three instruments—the critical thinking essay test, the scientific attitude scale, and the practical skills instrument—exhibit strong fit with the model. This finding indicates that the developed instruments successfully fulfill the five indicators of critical thinking according to Ennis's framework, which is widely utilized in contemporary research: *elementary clarification, basic support, inference, advanced clarification, and strategies and tactics* (Lestari *et al*., 2021). Furthermore, the reliability values and the consistency of response patterns indicate measurement stability across domains. This is aligned with Boone's theory in *"Rasch Analysis in the Human Sciences"* regarding Item Response Theory (IRT), which emphasizes the alignment between item difficulty and respondent ability to produce objective measurements.

The findings of this study are consistent with previous research asserting the importance of valid and reliable assessment instruments for measuring critical thinking in science education. Systematically designed assessment instruments can reveal critical thinking abilities more accurately (Afikah *et al*., 2024; Yokhebed et al., 2025). Meanwhile, other findings suggest that the Rasch model is effective in validating item characteristics and respondent consistency across various educational assessment contexts (Avinç & Doğan, 2024; Darman *et al*., 2024; Pitaloka *et al*., 2023). These results reinforce the present study's findings that the developed instruments possess high psychometric quality. The SSI (Socioscientific Issues) context utilized in these instruments—covering global warming, climate change, and energy consumption—meets the criteria for selecting issues that are authentic, relevant, and capable of stimulating critical analysis (Viehmann *et al*., 2024).

These issues provide opportunities for students to correlate the concepts of temperature and heat with social problems in their environment (Noperi et al., 2024; Rosmayuni *et al*., 2024). This aligns with the objectives of the SSI approach, which aims to develop scientific reasoning by analyzing social issues grounded in scientific concepts. This study contributes to the development of physics learning assessments, particularly for the topic of temperature and heat. Instruments that integrate the three domains of assessment and are analyzed using the Rasch model have the potential to serve as comprehensive evaluation tools that meet 21st-century educational demands. The SSI approach renders assessments more authentic by bridging physics concepts with social issues familiar to students' lives. Consequently, educators can obtain a more holistic overview of students' critical thinking abilities when they encounter relevant scientific problems in daily life.

CONCLUSION

The results of this study demonstrate that the developed instruments fulfill the criteria for reliability and validity, with all items confirmed to fit the Rasch model. Furthermore, the findings indicate that these

https://doi.org/10.64014/jik.v22i4.206

instruments effectively measure students' critical thinking skills comprehensively by integrating the three assessment domains. Feedback from both educators and students further confirms that the instruments are practical, user-friendly, and relevant to the context of physics learning based on socioscientific issues (SSI). Consequently, the resulting instruments can be used by teachers as authentic assessment tools to more effectively measure and develop students' thinking abilities during instruction. Future research could be directed toward developing similar instruments for other topics within physics or other scientific disciplines. Additionally, further studies should test the effectiveness of these instruments through long-term instructional experiments and explore the integration of such instruments into digital platforms to enhance accessibility and implementation across diverse school contexts.

AUTHOR'S NOTE

The author declares that there is no conflict of interest regarding the publication of this article. The author affirms that the data and content of this article are free from plagiarism.

REFERENCES

Afikah, A., Rohaeti, E., & Jumadi, J. (2024). Measuring critical thinking skills in science learning: Students' perception versus actual performance. *TEM Journal*, *13*(4), 3266-3275.

Ahmad, N. Q., Arthur, R., Rahayu, W., & Rahmadhani, E. (2025). Pengembangan instrumen kemampuan pemecahan masalah Matematika siswa menggunakan model Rasch. *Juring (Journal for Research in Mathematics Learning)*, *8*(1), 61-74.

Arifin, Z., Sukarmin, Saputro, S., & Kamari, A. (2025). The effect of inquiry-based learning on students' critical thinking skills in science education: A systematic review and meta-analysis. *Eurasia Journal of Mathematics, Science and Technology Education*, *21*(3), 1-24.

Arjaya, I. B. A., & Surata, S. P. K. (2024). A systematic review: Trends socioscientific issues in climate change materials. *Jurnal Santiaji Pendidikan (JSP)*, *14*(1), 52-65.

Avinç, E., & Doğan, F. (2024). Digital literacy scale: Validity and reliability study with the rasch model. *Education and Information Technologies, 29*(17), 22895-22941.

Darman, D. R., Suhandi, A., Kaniawati, I., Samsudin, A., & Wibowo, F. C. (2024). Development and validation of Scientific Inquiry Literacy Instrument (SILI) using rasch measurement model. *Education Sciences*, *14*(3), 1-28.

Fihani, N., Hikmawati, V. Y., & Mu'minah, I. H. (2021). Pendekatan Socio-Scientific Issue (SSI) untuk meningkatkan keterampilan berpikir kritis siswa SMA pada konsep Virus. *Prosiding Seminar Nasional Pendidikan*, *3*(3), 186-192.

Fitroty, I. Y., Wahyuni, E. A., Ahied, M., Hartiningsih, T., & Rakhmawan, A. (2023). Pembelajaran problem based learning berpendekatan socio-scientific issue berbantuan buletin untuk meningkatkan berpikir kritis siswa. *Jurnal Natural Science Educational Research*, *6*(3), 1-11.

García-Carmona, A. (2023). Scientific thinking and critical thinking in science education. *Science and Education*, *34*(1), 227-245.

Hara, A. K., Astiti, K. A., & Lantik, V. (2023). Analisis penguasaan konsep Fisika pada materi suhu dan kalor pasca pembelajaran online di kelas XI SMA Negeri 12 Kota Kupang. *Jurnal Ilmu Pendidikan (JIP) STKIP Kusuma Negara*, *14*(2), 118-126.

Imania, K. A., & Bariah, S. K. (2019). Rancangan pengembangan instrumen penilaian pembelajaran berbasis daring. *Jurnal Petik*, *5*(1), 31-47.

Kafii, M. S., & Dwikoranto, R. S. (2023). Analisis validitas instrumen tes keterampilan berpikir kritis siswa

https://doi.org/10.64014/jik.v22i4.000

pada materi gelombang berjalan dan gelombang stasioner. *Inovasi Pendidikan Fisika. IPF*, *12*(3), 111-118.

Lestari, H., Sopandi, W., Sa'ud, U. S., Musthafa, B., Budimansyah, D., & Sukardi, R. R. (2021). The impact of online mentoring in implementing RADEC learning to the elementary school teachers' competence in training students' critical thinking skills: A case study during COVID-19 pandemic. *Jurnal Pendidikan IPA Indonesia*, *10*(3), 346-356.

Macalalag Jr, A. Z., Kaufmann, A., Van Meter, B., Ricketts, A., Liao, E., & Ialacci, G. (2024). Socioscientific issues: Promoting science teachers' pedagogy on social justice. *Disciplinary and Interdisciplinary Science Education Research*, *6*(1), 1-16.

Mappalesye, N., Sari, S. S., & Arafah, K. (2021). Pengembangan instrumen tes kemampuan berpikir kritis dalam pembelajaran Fisika. *Jurnal Sains dan Pendidikan Fisika*, *17*(1), 69-82.

Noperi, H., Wiliyanti, V., Yanti, F. A., Syamsuri, M. M. F., & Yusandika, A. D. (2024). Validasi instrumen literasi sains dengan model Rasch untuk socio-scientific issues. *Natural: Jurnal Ilmiah Pendidikan IPA, 11*(2), 63-75.

Ochildinovna, N. D. (2024). The importance of critical thinking in modern education. *World Bulletin of Social Sciences*, *34*(1), 127-131.

Pitaloka, D. A. E., Kusuma, I. Y., Pratiwi, H., & Pradipta, I. S. (2023). Development and validation of assessment instrument for the perception and attitude toward Tuberculosis among the general population in Indonesia: A Rasch analysis of psychometric properties. *Frontiers in Public Health*, *11*(1), 1-10.

Prabayanti, E., Usman, U., Khaeruddin, K., & Setiawan, T. (2024). Analysis of students' critical thinking abilities in physics learning: A case study at SMAN 5 Sidrap. *Jurnal Pendidikan Fisika*, *12*(3), 141-151.

Rahmawati, S., Junaenah, S., Alfiyyah, A. A., & Nugraha, E. (2024). Problematika menilai ranah kognitif, afektif, dan psikomotorik pada mata pelajaran IPAS kurikulum merdeka. *Pedadidaktika: Jurnal Ilmiah Pendidikan Guru Sekolah Dasar*, *11*(4), 613-622.

Rosmayuni, A. A. A. P., Suma, K., & Suja, I. W. (2024). E-Modul pembelajaran IPA berbasis isu-isu sosial sains untuk meningkatkan literasi sains dan literasi sosial peserta didik kelas VIII. *Indonesian Journal of Instruction*, *5*(2), 204-215.

Sartika, S., Mansyur, J., & Kendek, Y. (2021). Miskonsepsi siswa SMA Negeri 1 Palu tentang suhu dan kalor. *Jurnal Kreatif Online*, *9*(2), 158-165.

Selamat, I. N. (2021). Keterampilan abad ke-21 pada pembelajaran sains dengan konteks socio-scientific issues di Indonesia: Tinjauan literatur sistematis. *Jurnal Pendidikan dan Pembelajaran IPA Indonesia*, *11*(2), 14-21.

Septiany, L. D., Puspitawati, R. P., Susantini, E., Budiyanto, M., Purnomo, T., & Hariyono, E. (2024). Analysis of high school students' critical thinking skills profile according to Ennis indicators. *IJORER: International Journal of Recent Educational Research*, *5*(1), 157-167.

Sundari, P. D., & Sarkity, D. (2021). Keterampilan berpikir kritis siswa SMA pada materi suhu dan kalor dalam pembelajaran fisika. *Journal of Natural Science and Integration*, *4*(2), 149-161.

Viehmann, C., Fernández Cárdenas, J. M., & Reynaga Peña, C. G. (2024). The use of socioscientific issues in science lessons: A scoping review. *Sustainability*, *16*(14), 1-12.

Widodo, W., Degeng, I. N. S., Pali, M., & Hitipeuw, I. (2024). Development and validation of the student well-being scale for elementary school using Rasch model analysis. *Journal of Educational, Health and Community Psychology*, *13*(2), 1-12.

Yokhebed, Y., Karmadi, R. M. D., & Nastiti, L. R. (2025). Validity and reliability analysis of a socioscientific issues-based critical thinking self-assessment instrument using the Rasch model. *Journal of Biological Education Indonesia (Jurnal Pendidikan Biologi Indonesia), 11*(1), 73-82.

https://doi.org/10.64014/jik.v22i4.206

Zahroh, F. L., Muflih, M., & Hilmiyati, F. (2024). Pembuatan instrumen evaluasi pembelajaran: Analisis literatur review. *Jurnal Paris Langkis*, *5*(1), 319-328.

Zarkasyi, A. H., Himam, A. S., Lubis, R. H., Prameswari S. K., & Badriya, I. (2023). Kurikulum merdeka for Arabic language learning according to Piaget's cognitive development theory. *Inovasi Kurikulum, 20*(2), 305-316.

https://doi.org/10.64014/jik.v22i4.000