Journal Corner of Education, Linguistics, and Literature
Vol. 5 No. 1 (August, 2025), pp. 94-112
https://doi.org/10.54012/jcell.v5i1.534

Development of Diagnostic Instruments for Critical Thinking Skills of Students in Physics Learning on Socio-Scientific Issues (SSI)

Dewi Tri Wulandari[1✉], Rahma Diani[2], Hendri Noperi[3]
[1 2 3]*Physics Education, Raden Intan State Islamic University, Lampung, Indonesia*
✉ email: dewitriwulandari41@gmail.com

Received:
August 2, 2025
Revised:
August 8, 2025
Accepted:
August 11, 2025
Published:
August 12, 2025

ABSTRACT

This study aimed to develop and test the feasibility of a diagnostic instrument for assessing students' critical thinking skills on socio-scientific issues (SSI) in physics education with the topic of global warming. This objective is motivated by the lack of valid, reliable, and contextually appropriate assessment instruments based on Socio-Scientific Issues (SSI) in physics education at the secondary school level, amid the low critical thinking competencies of Indonesian students as indicated by the 2018 PISA survey, and the need for an SSI approach as an effective strategy. The research method employs a research and development (R&D) design using the ADDIE model. The instrument was developed in the form of 20 essay questions based on SSI contexts and its validity was reviewed by experts and tested on 75 students from two high schools in rural and urban areas. Data were analyzed using Rasch modeling with the assistance of WINSTEPS software. Expert validation results showed that the instrument was categorized as "Highly Suitable." Rasch model analysis confirmed the instrument's unidimensionality (raw variance explained by observed measures: 30.9%). The analysis results showed that all items were deemed fit (Outfit MNSQ: 0.78–1.28; ZSTD:–1.51 to 1.81; and PTMEA Corr: 0.26–0.68). The instrument's reliability is high, with a Cronbach's alpha of 0.81, person reliability of 0.80, and item reliability of 0.91, using a structurally valid rating scale. However, significant DIF based on gender and region was found, indicating the need for further revision to ensure measurement fairness. This instrument is deemed valid and reliable for use as a diagnostic tool for critical thinking skills in SSI-based physics learning.
Keywords: *critical thinking; diagnostic instruments; socio-scientific issues; physics learning*

INTRODUCTION

Education is the main foundation in shaping superior human resources who are adaptive to the challenges of the times. Law Number 20 of 2003 on the National Education System states that education is a conscious and planned effort to create an active learning environment so that students can develop their full potential, both in the spiritual, intellectual, and life skills domains (David & Budianto, 2021). In the context of the 21st century, marked by technological advancements and globalization, education is not only required to transmit knowledge but also to develop higher-order thinking skills, particularly critical thinking skills (Agustihana & Suparno, 2019; Scott et al., 2021; Thornhill-Miller et al., 2023).

Critical thinking skills are a key element in the 4C competency framework (critical thinking, creativity, collaboration, communication) that students must possess to address the complex challenges of the Fourth Industrial Revolution (González-Pérez & Ramírez-Montoya, 2022; Prayogi, 2020; Silber- Varod et al., 2019). These skills include the ability to analysis arguments, evaluate information, and make decisions based on evidence and logic (Albano & Umberto Dello Iacono, 2019; Paul & Elder, 2019). Unfortunately, the results of the 2018 PISA survey by the Organisation for Economic Co-Operation and Development (OECD) show that Indonesia still ranks low in critical thinking competencies among students (Solihin et al., 2024).

One pedagogical approach proven to enhance critical thinking skills is the Socio-Scientific Issue (SSI) approach (Jamil et al., 2024; López-Fernández et al., 2022; Putri et al., 2024). This approach places students in real-world contexts that integrate scientific and social issues, enabling them to analysis, evaluate, and make decisions about contemporary issues in a reflective and scientifically informed manner (Fihani et al., 2021; Khoiri, Affandi, 2023). In physics education, the integration of SSI not only bridges abstract concepts with social reality but also builds stronger and more meaningful scientific literacy (Dewi et al., 2023; Febriani et al., 2023; Jusup et al., 2022).

The SSI approach in developing critical thinking skills cannot be accurately measured without valid, reliable, and contextual assessment instruments (Fita et al., 2021; Fitriyani et al., 2025). Diagnostic instruments play a crucial role in assessing students' critical thinking skills in physics education (Putra et al., 2023; Rapi et al., 2022). These instruments are designed to measure critical thinking skills, not only understanding of physics concepts but also the ability to analysis, evaluate, and apply knowledge critically, thereby promoting deeper learning and problem-solving skills (Bhakti et al., 2023; Pirinen et al., 2023). Based on a preliminary study conducted in two high schools in Lampung, it was found that the critical thinking assessment applied was not sufficiently varied and had not systematically integrated the SSI context. Survey results showed that 59% of students experienced difficulty in applying physics concepts to social issues. This fact underscores the urgent need for a diagnostic instrument capable of measuring students' critical thinking skills in the context of SSI-based physics learning.

Previous studies have confirmed that most teachers experience difficulties in developing valid and reliable assessment instruments to measure critical thinking skills (Ekawati et al., 2024; Payan-Carreira et al., 2022; Sari et al., 2024). Additionally, previous research has shown that the use of SSI-based instruments in chemistry education can enhance students' analytical and evaluative skills (Rahayu & Rosawati, 2023). Based on this previous research, it can be seen that research specifically developing diagnostic instruments for critical thinking skills in SSI-based physics learning is still unavailable. This gap indicates an urgent need for conceptual, valid, and reliable assessment instruments to support physics learning that focuses not only on cognitive outcomes but also on strengthening students' scientific reasoning and critical thinking skills. Therefore, this study was designed with the aim of developing and testing the feasibility of a diagnostic instrument in physics learning on Socio-Scientific Issues (SSI). SSI is designed to measure students' critical thinking skills in physics learning.

METHOD

Sample

The sample in this study consisted of 75 students (20 male and 55 female) in grade 11 from two areas between the village and the city. The village area was from SMA N 1 Adiluwih with 41 students, and the city area was from SMA Perintis 2 Bandar Lampung with 34 students. The instrument was tested in the odd semester of 2025/2026 in both schools. All students who were part of the sample in this study had received lessons on global warming.

Development Model

This study is a development study using a research and development (R&D) approach. The development process adopts the ADDIE model (Analysis, Design, Development, Implementation, and Evaluation) developed by Robert Maribe Branch, which consists of five development steps (Branch & Varank, 2009). This model was chosen because it provides a systematic, easy-to-understand framework suitable for designing products, including assessment instruments (Umami et al., 2021).

Figure 1. Research Stages

Instrument

The questions are based on socio-scientific issues related to global warming. The instrument consists of 20 essay questions covering 12 socio-scientific issues (SSI). The questions were developed using Bloom's revised taxonomy in terms of cognitive level (Farida et al., 2019).

Table 1. Number of Items and Topics of Social-Scientific Issues

Cognitive Level	SSI Topic	Question Distribution
C4	Drought	1
C5	Household Electricity Consumption	2
C5	Biofuel Production and Use	3,4
C4, C5, C6	Motor Vehicle Emissions and Renewable Energy	5, 6
C4, C5, C6	Forest Fires	7, 8
C4, C5, C6	Drought and Agricultural Irrigation	9, 10
C4, C5, C6	Deforestation	11, 12, 13
C4, C5, C6	Flooding and Rising Sea Levels	14, 15, 16
C6	Urban Heat Island	17
C4, C5, C6	Clean Water Distribution during Droughts	18
C4, C5	The Role of Plants in Mitigating Global Warming	19
C4, C5	Changes in Rainfall Patterns and Their Impact on Health & the Economy	20

The number of questions using the cognitive dimension of Analysis (C4) is 7 questions, Evaluation (C5) is 7 questions, and Creation (C6) is 6 questions. One of the indicators that can be used to investigate critical thinking skills is from Facion, with indicators of interpretation, analysis, evaluation, inference, explanation, and self-regulation (Fithriyah & Sa'dijah, 2016).

Table 2. Critical Thinking Indicators and Item Indicators

Indikator Berpikir Kritis	Question Indicators
Interpretation	Analyzing the possible scientific causes of the 2025 dry season decline
	Analyzing the impact of increased CO_2 from deforestation on global warming
Analysis	Explaining how motor vehicle emissions cause global warming
	Analyzing the relationship between the greenhouse effect and increased forest fires
	Analyzing the relationship between rising global temperatures and drought in agricultural land
	Analyzing the relationship between global warming, melting polar ice caps, and tidal flooding
	Analyzing the results of experiments on the effects of plants on room temperature
Evaluation	Evaluating the most effective energy-saving measures to reduce global warming
	Evaluating the effectiveness of biofuels in reducing carbon emissions
	Evaluating government assistance options for addressing drought
	Evaluating whether logging bans are effective in reducing global warming
	Evaluating the most critical and long-term causes of tidal flooding
	Evaluating the effectiveness of clean water distribution by the Regional Disaster Management Agency (BPBD) during droughts
Inference	Evaluating the impact of food waste habits on global warming
Explanation	Designing a campaign slogan to reduce deforestation and global warming
Self-Regulation	Designing solutions for the effectiveness of biofuels in reducing carbon emissions
	Designing a school program to reduce vehicle emissions
	Designing proposals for local technology or policies to prevent forest fires
	Designing innovations to minimize the risk of tidal flooding
	Designing scientific solutions to reduce heat caused by the urban heat island effect

Assessment

Each item in this instrument is assessed using a 4-point scale ranging from 0 to 3. A score of 0 indicates the student's inability to answer the question, and a score of 3 indicates the student's ability to answer the question perfectly. Therefore, assessment guidelines for each item have been developed to facilitate teachers in evaluating answers to each item.

Data Analysis

The test results were analyzed using the WINSTEPS application, which supports Rasch modeling for ordinal data and provides detailed information about the quality of each item. The analysis was conducted through several analyses using Rasch modeling (Sumintono & Widhiarso, 2015).

1. Verification of the unidimensionality assumption, to ensure that the instrument measures only one main construct, namely critical thinking, in accordance with the unidimensionality principle of the Rasch model (Linacre, 2009).
2. Item fit test with the Rasch model (Model Fit). This test was conducted by examining the Outfit MNSQ, Outfit ZSTD, and Pt Measure Corr values (Boone, 2014). An item was categorized as a misfit if it met the following three criteria.
   a. Outfit Mean-Square Residual (MNSQ): $0.5 < y < 1.5$
   b. Outfit Standardized Mean-Square Residual (ZSTD): $-2.0 < z < +2.0$
   c. Point Measure Correlation (Pt Measure Corr): $0.4 < x < 0.8$
3. Rating Scale, this is done to test the effectiveness of the response categories used in the critical thinking instrument. This analysis evaluates whether the rating scale functions as intended by analyzing category thresholds, response distributions, and calibration steps to ensure meaningful progress in assessment.
4. Person-Item Map Analysis (Wright Map), to analyze the level of correspondence between the difficulty level of the items and the respondents' abilities to ensure that the instrument can measure effectively.
5. Item Separation Test, to assess how well the instrument separates questions into categories/levels of difficulty (e.g., easy, medium, difficult).
6. Person Separation Test, to assess how well the instrument can group students based on their ability levels (e.g., low, medium, high).
7. Reliability Test, to determine the consistency of the instrument in measuring students' critical thinking skills through the estimation of item and respondent reliability indices.
8. Difficulty Level Test, to categorize items into low, medium, or high difficulty levels to validate the instrument's suitability for the test group.
9. DIF (Differential Item Functioning) Test, to detect bias toward certain groups (e.g., based on gender) in each item. This analysis is important to ensure the fairness of the instrument.

Based on the analysis results, these items will be used to revise or refine the instrument so that it meets the validity and reliability criteria based on Rasch modeling standards. Overall, this analysis validates the diagnostic instrument as a powerful tool for measuring students' critical thinking skills in physics learning based on Socio-Scientific Issues.

FINDINGS AND DISCUSSION

This study began with a preliminary study through an analysis of the needs of students and educators. The results of a questionnaire distributed to students revealed that many students had difficulty applying physics concepts to solve problems that required critical thinking skills based on Socio-Scientific Issues (SSI). These difficulties arise due to the limited use of assessment tools designed to develop critical thinking skills, particularly in the context of Socio-Scientific Issues. To date, students have primarily been given routine questions that fail to sufficiently challenge their critical thinking abilities in depth.

Further interviews with physics educators highlighted the gap in current practices for measuring critical thinking skills based on Socio-Scientific Issues (SSI). Teachers emphasized the need for valid and reliable assessment tools to evaluate students' ability to critically respond to scientifically significant social issues, such as climate change and renewable energy. These instruments are expected to serve as a basis for designing more contextual learning strategies and providing meaningful feedback.

The number of validated items in this study was 20 essay questions administered to students to assess their critical thinking skills on Socio-Scientific Issues. This study was conducted in physics learning with the topic of global warming. The validity of the questions was validated by subject matter experts and evaluation experts. The results of the validation conducted by three subject matter experts and three evaluation experts are presented in Table 3

Table 3. Results of Content Expert and Evaluation Expert Validation

Validation		Percentage	Average	Category
Subject Matter Expert	Expert 1	91%	93%	Highly Qualified
	Expert 2	97%		
	Expert 3	90%		
Evaluation Expert	Expert 1	83%	91%	Highly Qualified
	Expert 2	100%		
	Expert 3	91%		

The results of the above validation show that the critical thinking instrument on Socio-Scientific Issues received a very satisfactory rating from the experts. These results were obtained after revising the product to improve clarity and the suitability of the developed questions as suggested by the experts. After revision, the instrument was then tested on students.

Before conducting further analysis using the Rasch model, it is important to ensure that the instrument meets the assumptions of unidimensionality and local independence. The dimensionality test using the Winstep program is shown in the item output: dimensionality. The results of the analysis can be seen in Figure 2.

```
TABLE 23.0 INPUTRASCHDEWI.xlsx                      ZOU781WS.TXT  Jul 31 2025 09:16
INPUT: 75 PERSON  20 ITEM  REPORTED: 75 PERSON  20 ITEM  4 CATS MINISTEP 5.10.1.0
---------------------------------------------------------------------------------

     Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = ITEM information
units
                                           Eigenvalue   Observed   Expected
Total raw variance in observations     =     28.9616 100.0%         100.0%
  Raw variance explained by measures   =      8.9616  30.9%          30.8%
    Raw variance explained by persons  =      5.6704  19.6%          19.5%
    Raw Variance explained by items    =      3.2913  11.4%          11.3%
  Raw unexplained variance (total)     =     20.0000  69.1% 100.0%   69.2%
    Unexplned variance in 1st contrast =      2.8395   9.8%  14.2%
    Unexplned variance in 2nd contrast =      2.0214   7.0%  10.1%
    Unexplned variance in 3rd contrast =      1.9454   6.7%   9.7%
    Unexplned variance in 4th contrast =      1.7340   6.0%   8.7%
    Unexplned variance in 5th contrast =      1.6113   5.6%   8.1%
```

Figure 2. Unidimensionality Analysis of the Instrument

The analysis results show that the total variance explained by the instrument is 30.9%, exceeding the threshold of 20% and fulfilling the assumption of unidimensionality. Meanwhile, the unexplained variance in the first contrast is 9.8%, which is still below the 15% threshold, indicating no significant interference from additional dimensions (Linacre, 2009; Sumintono & Widhiarso, 2015). These findings confirm that the Socio-Scientific Issue-based critical thinking instrument meets the assumption of unidimensionality, validating its use in physics education.

Item fit analysis shows that most items are consistent with Rasch modeling. According to Boone, the mean-square residual (MNSQ), z-standard (ZSTD), and point measure correlation (PTMEA CORR) are criteria used to assess item fit (Boone, 2014). The results of the item fit test for the developed items can be seen in Table 4 below.

Table 4. Results of the Item Fit Test

Item	Measure	MNSQ Outfit	ZSTD Outfit	PTMEA CORR	Description
1IN	0.21	1.09	0.60	0.26	fit
2EV	0.45	1.23	1.43	0.33	fit
3EV	-1.22	0.99	-0.04	0.50	fit
4SR	-1.09	0.97	-0.15	0.65	fit
5AN	-0.41	0.78	-1.51	0.45	fit
6SR	-0.35	1.22	1.41	0.68	fit
7AN	-0.53	0.87	-0.85	0.47	fit
8SR	0.45	1.09	0.62	0.39	fit
9AN	-0.47	0.80	-1.37	0.27	fit
10EV	0.45	0.80	-1.34	0.55	fit
11IN	0.82	0.83	-1.10	0.39	fit
12EV	0.69	1.14	0.90	0.37	fit
13EX	0.27	0.91	-0.58	0.57	fit
14AN	-0.17	1.02	0.20	0.33	fit
15EV	0.60	0.92	-0.49	0.45	fit
16SR	-0.76	1.28	1.81	0.53	fit
17SR	-0.11	1.27	1.72	0.65	fit
18EV	0.82	0.98	-0.07	0.47	fit
19AN	0.30	1.05	0.38	0.63	fit

Item	Measure	MNSQ Outfit	ZSTD Outfit	PTMEA CORR	Description
20IF	0.06	0.78	-1.50	0.30	fit
Mean	0,00				
P.SD	0.59				

Overall, the 20 items developed were deemed fit for the Rasch model because they met at least one of the three eligibility criteria, namely Outfit MNSQ, Outfit ZSTD, and PTMEA CORR. The Outfit MNSQ values for all items fell within the ideal range of 0.5–1.5, indicating response adequacy (Boone, 2014; Linacre, 2009). The Outfit MNSQ values for all items fell within the ideal range of 0.5–1.5, indicating response adequacy (Boone, 2014; Linacre, 2009). The Outfit ZSTD criterion was used to detect outliers or extreme discrepancies. The ideal value is around 0, with a tolerance of ±2 considered acceptable (Sumintono & Widhiarso, 2015). Meanwhile, although the PTMEA CORR values for some items are below the ideal value of 0.4 (e.g., items 1IN, 2EV, 9AN, and 20IF), the items are still considered valid because they meet the MNSQ and ZSTD criteria. However, these items are recommended for revision to enhance construct strength and discriminative ability in the future (Linacre, 2012).

The evaluation of the scoring scale in the scoring rubric was conducted by examining the Andrich Thresholds to determine whether the score categories effectively differentiated the ability levels of the students. Ideally, the threshold distance should range from 1.4 to 5.0 logits. The threshold analysis can be seen in Table 5.

Table 5. Analysis of Andrich Thresholds for the Rating Scale

Rating Transition	Lower Threshold	Upper Threshold	Andrich Threshold Distance	Recommendation
Rating 0 to 1	0	-2.20	2.20	Maintain
Rating 1 to 2	-2.20	0.16	2,36	Maintain
Rating 2 to 3	0.16	2.04	2.20	Maintain

The analysis results show that all categories of the assessment scale have a threshold distance above 1.4 logits, meeting the criteria for structural validity without any indication of the need to merge or separate categories. The probability curve shows a clear, sequential distribution of categories, each with a dominant peak, indicating that each category is used and understood well by respondents. There are no flat categories or categories that have lost their peaks, reinforcing that the scale structure has been calibrated appropriately. Thus, this rating scale is valid and reliable in effectively measuring variations in student ability.

The Wright Map provides a comprehensive representation of the relationship between student ability and item difficulty. The Wright Map illustrates the relationship between student ability and item difficulty on a single logit scale, where high-ability individuals and difficult items are positioned at the top, while easy items and low-ability students are at the bottom (Boone, 2014).

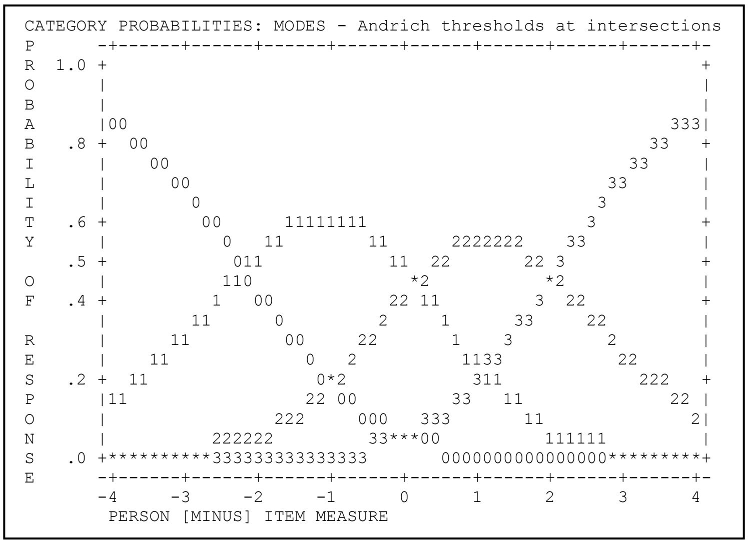

Figure 3. Andrich Threshold Graph at Category Intersections

The Wright Map in Figure 4 shows that the distribution of respondents is at logit 0–2, indicating relatively good ability, while the items are spread across varying levels of difficulty. This balanced distribution indicates that the instrument is able to measure student ability effectively and proportionally.

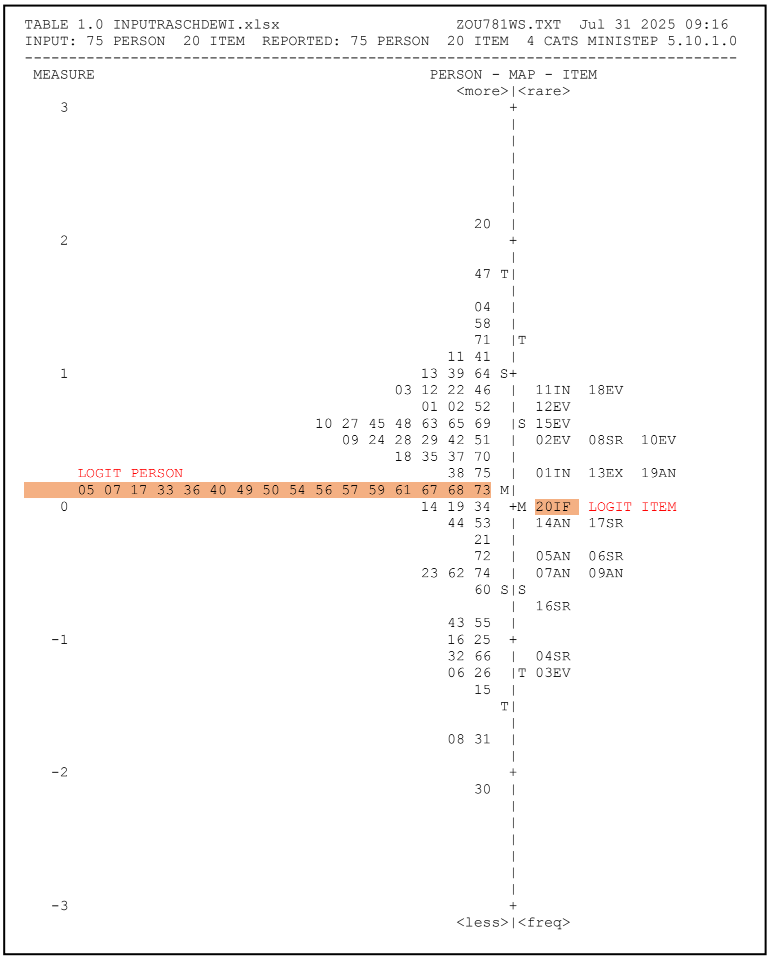

Figure 4. Wright Map Analysis

The most difficult items are found at logit +0.9, namely 18EV and 11IN, while the easiest items are found at logit -1, namely 3EV. This means that 18EV and 11IN are designed for students with high abilities, while 3EV is suitable for students with lower abilities. Adjusting the difficulty level of such items is important to ensure that all students can be effectively reached by the measurement instrument. This mapping helps in evaluating the suitability of the items with the students' responses. Items that are too easy or too difficult for most students can be considered suboptimal and need revision (Sumintono & Widhiarso, 2015).

Tabel 6. Person &Item Separation dan Statistik Reliabilitas

Parameter	Measure	SD	Separation	Reliability	Alpha Cronbach
Person (N=75)	0.16	2.12	2.00	0.80	0.81
Item (N=20)	0.00	0.61	3.15	0.91	

The results of the Person & Item Separation analysis show that a person separation value of 2.00 indicates the instrument's ability to distinguish students into three different ability groups, with a participant reliability of 0.80. This reflects the consistency and stability of individual ability measurements. In addition, the Cronbach's alpha value of 0.81 confirms that the items have good internal consistency in measuring the desired critical thinking construct. On the other hand, the item separation of 3.15 and item reliability of 0.91 indicate that the items have a wide range of difficulty levels and high consistency in their measurement (Boone, 2014).

Item difficulty levels were classified based on logit distribution using the mean and standard deviation (SD) as references. Items with a logit > mean + 1 SD were categorized as very difficult, while items with a logit < mean – 1 SD were classified as very easy (Sumintono & Widhiarso, 2015). Based on the output "Summary of 20 Measured Items" in Table 4, the mean logit value is 0.00 and the SD is 0.53 logit. Using this reference, the items are classified into four difficulty categories as shown in Table 7.

Table 7. Item Classification Based on Difficulty Level

Category	Range	Number of Questions
Very difficult	>1 SD	11IN, 18EV, 12EV, 15EV
Difficult	Mean s/d 1 SD	2EV, 8SR, 10EV, 19AN, 13EX, 1IN, 20IF
Easy	-1SD s/d mean	7AN, 9AN, 5AN, 6SR, 14AN, 17SR
Very easy	<-1SD	3EV, 4SR, 16SR

Based on the analysis of question difficulty levels, most items are distributed in the very difficult and easy categories, with 7 and 6 items in each category, respectively. Four items, namely 11IN, 18EV, 12EV, and 15EV, are in the very difficult category with a logit >0.53 logit. Meanwhile, three items, namely 3EV, 4SR, and 16SR, are in the very easy category with a logit <-0.53 logit. This distribution shows that the instrument has a sufficiently high level of difficulty to measure students' critical thinking skills. This is consistent with the nature of critical thinking ability tests, which target students' higher-order critical thinking skills.

An item is said to have bias/dif in a category if its probability value is less than 0.05 (Sumintono & Widhiarso, 2015). DIF analysis was conducted based on gender (male and female) and region (rural and urban).

Table 8. DIF Based on Gender and Region for the Developed Test Items

Item	Name	DIF Gender (*Probability*)	DIF Region (*Probability*)
1	01IN	0.2497	0.6283
2	02EV	0.0544	0.0837
3	03EV	0.8287	0.2870
4	04SR	0.5590	1.0000
5	05AN	0.3443	1.0000
6	06SR	0.0272	0.0080
7	07AN	0.1740	0.6746
8	08SR	0.0589	0.7192
9	09AN	0.2505	0.6864
10	10EV	0.4464	0.4781
11	11IN	0.7264	0.5426
12	12EV	0.2208	0.0103
13	13EX	0.8464	0.5544
14	14AN	0.1703	0.8989
15	15EV	0.0933	0.7752
16	16SR	0.8155	0.5492
17	17SR	0.6634	0.8593
18	18EV	0.4563	1.0000
19	19AN	0.6259	0.7980
20	20IF	0.5154	0.4548

Based on the output of the Winstep application in Table 7, item 6SR has a probability value of 0.0272, which is less than 0.05, meaning that 6SR has a bias based on gender. In Figure 6, item 6SR has a value of 0.0080 and item 12EV has a value of 0.0103. This means that items 6SR and 12EV exhibit bias based on region. DIF may arise if an item is more relevant or easier to understand for one particular group compared to others. These differences underscore the need for further evaluation to ensure that the items remain fair in measurement

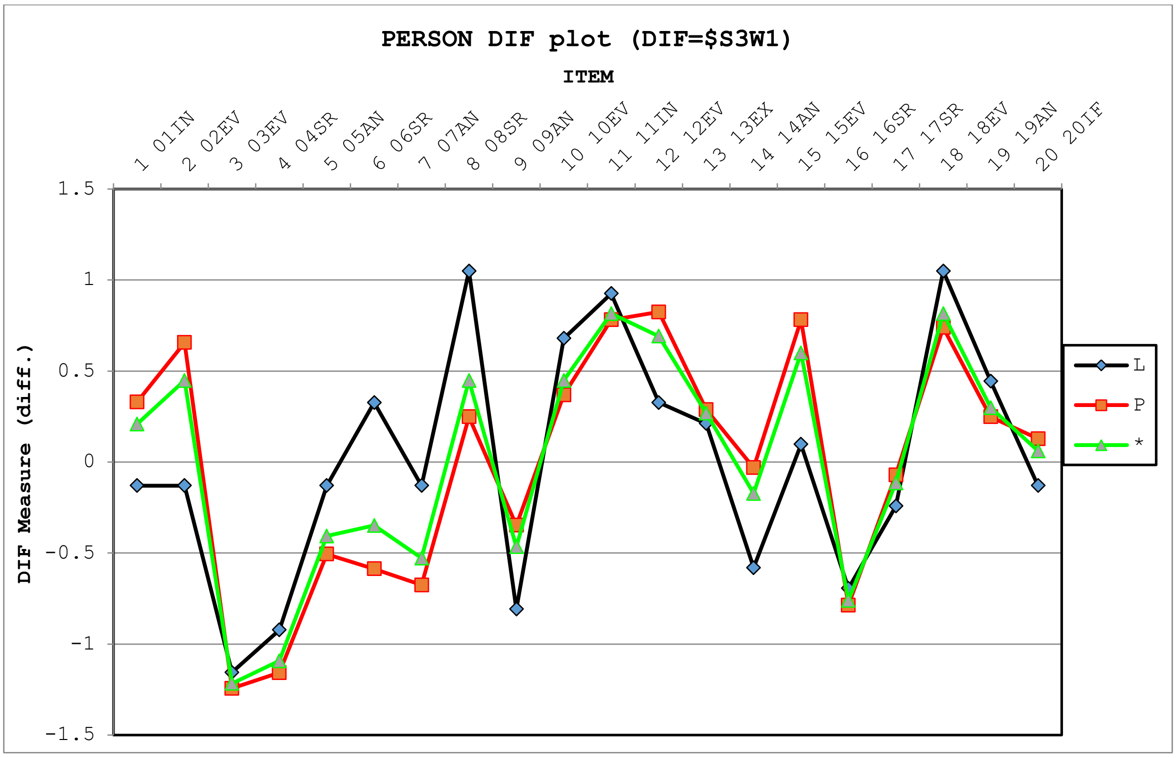

Figure 5. DIF Measure Plot Graph for Men and Women in completing test items

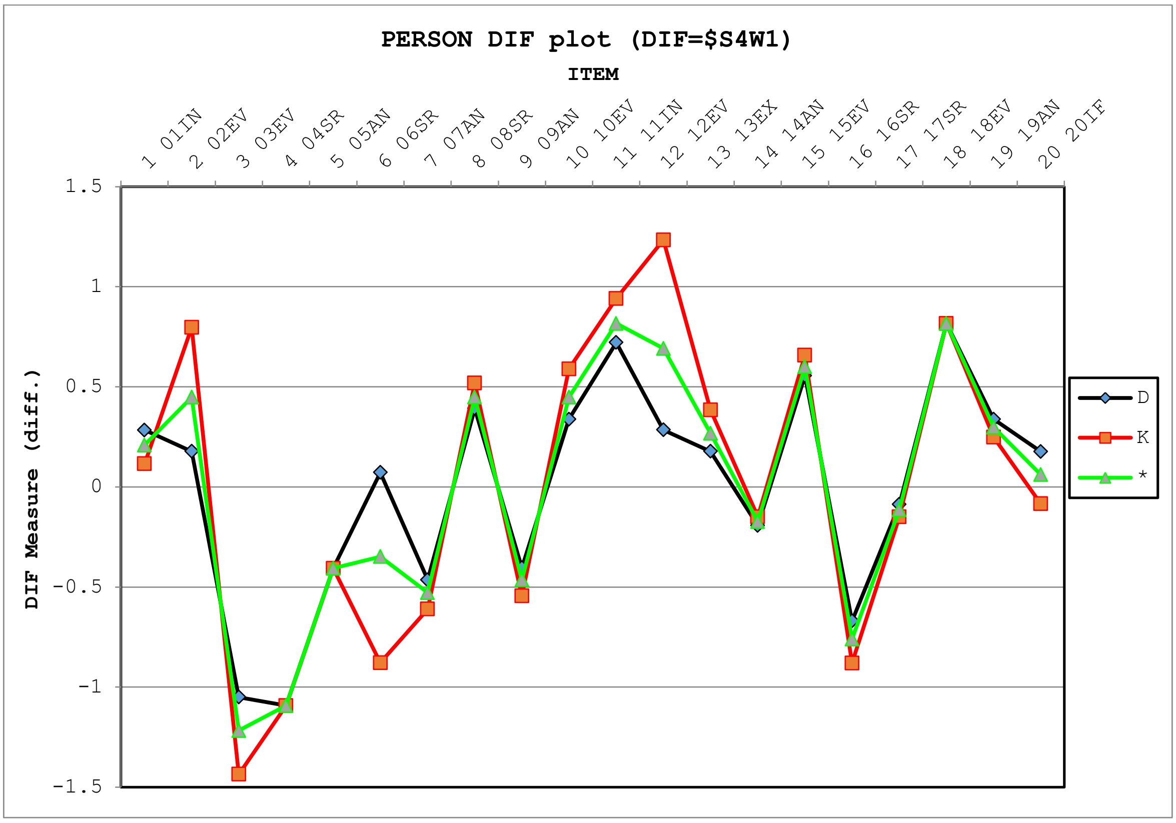

Figure 6. DIF Measure Plot Graph for Regions in completing test items

Overall, the significant DIF in Figure 7, in question 6SR, shows that the question does not function fairly between male and female groups. Where the question is more advantageous for women than for men. Meanwhile, Figure 8 shows that questions 6SR and 12EV show that the questions function unfairly between rural and urban areas. Question 6SR appears to favor urban areas over rural areas, while question 12EV favors rural areas over urban areas. Therefore, revisions are needed to ensure that the items measure equally without favoring a particular group, in order to maintain the fairness and validity of the instrument (Adams et al., 2018).

The research results indicate that the Socio-Scientific Issue-based critical thinking instrument is valid and reliable for measuring students' critical thinking skills in physics education. The instrument demonstrates good reliability with a Cronbach's Alpha value of 0.81, reflecting high consistency, in line with recommendations in the literature (Barbera et al., 2020). The research results indicate that the Socio-Scientific Issue-based critical thinking instrument is valid and reliable for measuring students' critical thinking skills in physics education. The instrument demonstrates good reliability with a Cronbach's Alpha value of 0.81, reflecting high consistency, in line with recommendations in the literature (Avinç & Doğan, 2024; Sumintono & Widhiarso, 2015).

Item suitability analysis based on the Rasch model also shows that all 20 items fall into the fit category. The three main indicators for item fit testing, namely Outfit MNSQ, Outfit ZSTD, and PT Measure COR, are within the recommended tolerance range (Boone, 2014; Linacre, 2012). This indicates that all items are consistent with the Rasch model and measure the same construct consistently, thereby supporting the construct validity of the instrument.

Andrich threshold analysis shows that the transition between categories is above the Andrich threshold distance (1.4 logits). This indicates that the categories are effective and can be used in measurement (Adams et al., 2018; Cantó-Cerdán et al., 2021; Kalkbrenner, 2021). Additionally, the Differential Item Functioning (DIF) analysis revealed potential bias in item 6SR based on gender, with a probability value of 0.0272 (< 0.05). This item tends to favor women over men. Furthermore, the DIF analysis also identified potential bias in items 6SR and 12EV based on region. Specifically, item 6SR favors urban areas over rural areas, while item 12EV favors rural areas over urban areas. The presence of significant DIF in certain items can threaten fairness in measurement, so revisions are needed to ensure that the instrument remains fair and free from bias between groups (Sari et al., 2024).

The distribution of item difficulty levels also provides important information about critical thinking skills. Based on logit analysis, most items fall into the difficult (4 items) and very difficult (7 items) categories, while the rest are spread across the easy (6 items) and very easy (3 items) categories. This distribution indicates that the instrument tends to be challenging and is intended to measure critical thinking skills. Although this is consistent with the instrument's objectives, it is necessary to ensure that the distribution of difficulty levels remains proportional so that the instrument can measure the entire range of students' abilities fairly and comprehensively (Boone, 2014; Sahin Kursad & Yalcin, 2024).

This finding is reinforced by the Wright map, which illustrates the relationship between student ability and item difficulty level. Most students are in the 0 to 2 logit ability range, while items are evenly distributed above and below that range. This indicates that the instrument has sufficient discriminating power to differentiate students based on their ability level.

Overall, the person separation value of 2.00 and person reliability of 0.80 indicate that the instrument is capable of distinguishing at least three different ability groups. Meanwhile, the item separation value of 3.15 and item reliability of 0.91 indicate a good distribution of difficulty levels and consistency in measurement between items. This reliability strength confirms that the instrument was developed with the appropriate structure and item composition to measure students' critical thinking constructs in the context of Socio-Scientific Issues (Kleemola et al., 2022; Noperi et al., 2024; Quinn et al., 2020).

Discussion

These findings align with Ekawati's research which emphasizes the importance of providing systematically designed assessment instruments to measure critical thinking (Ekawati et al., 2024). Similarly, Sari's findings indicate that most teachers face difficulties in developing valid and reliable instruments that effectively measure critical thinking (Sari et al., 2024). On the other hand, this study also expands on the findings Rahayu's which focused on the use of SSI-based instruments in chemistry learning (Rahayu & Rosawati, 2023). This study contributes further in the context of physics learning and offers a new perspective by developing a specialized diagnostic instrument that uses current Socio-Scientific Issues as a context to trigger critical thinking. These results are also in line with international studies by Sadler which emphasize that SSI is an effective learning context for fostering science literacy and critical thinking in the 21st century (Sadler, 2004).

Theoretically, this instrument is based on critical thinking skills indicators, namely interpretation, analysis, evaluation, inference, explanation, and self-regulation (Fithriyah & Sa'dijah, 2016). These indicators are applied concretely in the context of SSI-based physics learning, such as questions linking deforestation to increased CO_2 levels, or asking students to design scientific solutions to address the urban heat island effect. This approach demonstrates a strong connection between theory and practice. In other words, this instrument not only aligns with the conceptual framework of critical thinking but can also be practically applied by educators in the field (Fita et al., 2021; Fitriyani et al., 2025). In the classroom, teachers can utilize this instrument as an evaluative tool capable of mapping students' critical thinking achievements based on current social issues, while also sparking in-depth discussions and reflections among students (Mardianto, 2019; Viehmann et al., 2024). These results are further supported by research Fitriani and Rahayu which found that context-based approaches like SSI can enhance scientific argumentation skills and higher-order thinking (Fitriani & Festiyed, 2023; Rahayu & Rosawati, 2023).

In terms of scientific contribution, this study has clear novelty. To the best of the authors' knowledge, there has been no research in Indonesia that specifically develops and validates diagnostic instruments for assessing students' critical thinking skills in physics education based on Socio-Scientific Issues. Therefore, this study fills a gap in the literature by developing and testing an essay-based instrument using Facione's indicators and Bloom's Taxonomy at the higher levels (C4–C6) through Rasch model analysis. This approach to instrument development, which combines higher-order cognitive dimensions and critical thinking indicators, is rarely applied simultaneously in physics learning assessments (Fithriyah & Sa'dijah, 2016; Oktoviani et al., 2019). This positions the instrument not only as a measurement tool but also as an educational vehicle for character development and the cultivation of scientific literacy among students (Dewi et al., 2023; Febriani et al., 2023; Jusup et al., 2022). Additionally, the design of context-based essay questions in SSI provides a higher level of analytical depth compared to conventional multiple-choice tests, which often only measure knowledge aspects without comprehensively assessing critical thinking processes (Fita et al., 2021; Putra et al., 2023).

However, this study has a number of limitations that need to be considered. First, the physics topics used are still limited to the theme of global warming, so that generalization of this instrument to other topics in the physics curriculum cannot be done automatically without adaptation. Second, the instrument trial only involved two high schools in the Lampung region, so that expansion of the sample in different geographical and social contexts is needed to test the consistency of the results. Third, since this study was conducted only once (cross-sectional), it is not yet possible to determine whether the use of this instrument truly has a long-term impact on improving students' critical thinking skills. Fourth, since the questions are in essay form, there is a possibility that students may answer based on what they believe to be correct, rather than their actual understanding. This could affect the accuracy of the data obtained.

The results of this research have important implications, both theoretically, practically, and methodologically. Theoretically, this study confirms that the integration of the SSI approach with essay-based assessment can bridge the gap between cognitive assessment and the strengthening of scientific character. Practically, physics teachers in high schools can use this instrument as a tool for evaluation and the development of reflective learning that encourages students to think deeply about scientific issues and their impact on real life.

CONCLUSION

This study successfully developed and tested the feasibility of a diagnostic instrument to measure students' critical thinking skills in physics learning based on Socio-Scientific Issues (SSI). Through the ADDIE stages, the developed instrument showed very high validity and was declared "highly feasible." The feasibility test using the Rasch model confirmed the construct validity of the instrument through indicators of unidimensionality, item fit, the feasibility of the rating scale, and high reliability (α =

0.81). Based on these results, this instrument is feasible for use as a diagnostic tool for measuring students' critical thinking skills in the context of physics learning based on SSI.

REFERENCES

Adams, D., Sumintono, B., Mohamed, A., & Noor, N. S. M. (2018). E-learning readiness among students of diverse backgrounds in a leading Malaysian higher education institution. *Malaysian Journal of Learning and Instruction*, *15*(2), 227–256. https://doi.org/10.32890/mjli2018.15.2.9

Agustihana, S., & Suparno, S. (2019). Development of HOTS Oriented Cognitive Problems in Thermodynamics for Senior High Schools. *Jurnal Penelitian Fisika Dan Aplikasinya (JPFA)*, *9*(1), 44–54. https://doi.org/10.26740/jpfa.v9n1.p44-54

Albano, G., & Umberto Dello Iacono. (2019). Sebuah alat perancah untuk mendorong argumentasi dan bukti dalam matematika : beberapa studi kasus. *Educational Technology in Higher*, *16*(1), 1-12. https://doi.org/10.1186/s41239-019-0134-5

Avinç, E., & Doğan, F. (2024). Digital literacy scale: Validity and reliability study with the rasch model. *Education and Information Technologies*, *29*(17), 22895–22941. https://doi.org/10.1007/s10639-024-12662-7

Barbera, J., Naibert, N., Komperda, R., & Pentecost, T. C. (2020). Clarity on Cronbach's alpha use. *Journal of Chemical Education*, *98*(2), 257–258. https://doi.org/10.1021/acs.jchemed.0c00183

Bhakti, Y. B., Arthur, R., & Supriyati, Y. (2023). Development of an assessment instrument for critical thinking skills in Physics: a systematic literature review. *Journal of Physics: Conference Series*, *2596*(1), 1-8. https://doi.org/10.1088/1742-6596/2596/1/012067

Boone, C. (2014). *Property and political order in Africa: Land rights and the structure of politics*. Cambridge University Press.

Branch, R. M., & Varank, İ. (2009). *Instructional design: The ADDIE approach* (Vol. 722). Springer.

Cantó-Cerdán, M., Cacho-Martínez, P., Lara-Lacárcel, F., & García-Muñoz, Á. (2021). Rasch analysis for development and reduction of Symptom Questionnaire for Visual Dysfunctions (SQVD). *Scientific Reports*, *11*(1), 1-10. https://doi.org/10.1038/s41598-021-94166-9

David, D., & Budianto, A. (2021). Legal policy in the sky of higher education as a strategy for empowering Indonesia society. *Proceedings of the 1st International Conference on Law, Social Science, Economics, and Education, ICLSSEE 2021, March 6th 2021, Jakarta, Indonesia*.

Dewi, A. I. K., Suyono, S., & Erman, E. (2023). Effectiveness of Socioscientific Issues (SSI) Based Learning to Improve Argumentation Skills. *Jurnal Penelitian Pendidikan IPA*, *9*(1), 279–283. https://doi.org/10.29303/jppipa.v9i1.2866

Ekawati, E. Y., Sabrilla Almas Azzahra, & Hanun Fithriyah. (2024). Construction of Critical Reasoning Skills Assessment Instruments as Diagnostic Assessments in Physics Learning with Polytomus Scoring. *Journal of Education Research and Evaluation*, *8*(2), 338–349. https://doi.org/10.23887/jere.v8i2.69259

Farida, R., Alba, A., Kurniawan, R., & Zainuddin, Z. (2019). Pengembangan Model Pembelajaran Flipped Classroom Dengan Taksonomi Bloom Pada Mata Kuliah Sistem Politik Indonesia. *Kwangsan: Jurnal Teknologi Pendidikan*, *7*(2), 104-122. https://doi.org/10.31800/jtp.kw.v7n2.p104--122

Febriani, Jumadi, J., & Dwandaru, W. S. B. (2023). Socio-scientific issues in physics learning to improve students' critical thinking skills. *Revista Mexicana de Física E*, *20*(1), 1–6. https://doi.org/10.31349/RevMexFisE.20.010202
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*, 186–192.
Fita, M. N., Jatmiko, B., & Sudibyo, E. (2021). The Effectiveness of Problem Based Learning (PBL) Based Socioscientific Issue (SSI) to Improve Critical Thinking Skills. *Studies in Learning and Teaching*, *2*(3), 1–9. https://doi.org/10.46627/silet.v2i3.71
Fithriyah, I., & Sa'dijah, C. (2016). *Analisis kemampuan berpikir kritis siswa kelas IX-D SMPN 17 Malang*.
Fitriani, E., & Festiyed, F. (2023). The effect of STEM integrated physics E-Modules to improve critical thinking ability of class XI studentd of SMAN 2 Lubuk Sikaping. *Physics Learning and Education*, *1*(2), 124–129. https://doi.org/10.24036/ple.v1i2.57
Fitriyani, F., Atmojo, I. R. W., & Yamtinah, S. (2025). Socio Scientific Issue Approach to Enhance Critical Thinking Skills: A Prisma Systematic Literature Review. *Multidisciplinar (Montevideo)*, *3*(3), 1-10. https://doi.org/10.62486/agmu2025212
González-Pérez, L. I., & Ramírez-Montoya, M. S. (2022). Components of Education 4.0 in 21st century skills frameworks: systematic review. *Sustainability*, *14*(3), 1-31. https://doi.org/10.3390/su14031493
Jamil, M., Hafeez, F. A., & Muhammad, N. (2024). Critical thinking development for 21st century: Analysis of Physics curriculum. *Journal of Social & Organizational Matters*, *3*(1), 1–10.
Jusup, M., Holme, P., Kanazawa, K., Takayasu, M., Romić, I., Wang, Z., Geček, S., Lipić, T., Podobnik, B., & Wang, L. (2022). Social physics. *Physics Reports*, *948*, 1–148. https://doi.org/10.1016/j.physrep.2021.10.005
Kalkbrenner, M. T. (2021). A Practical Guide to Instrument Development and Score Validation in the Social Sciences: The MEASURE Approach. *Practical Assessment, Research & Evaluation*, *26*, 1.
Khoiri, Affandi, A. (2023). *Model ILESSI-DCF "Inquiry Lesson based Ethno Socioscientific Issues-Dieng Culture Festival."* Perkumpulan Rumah Cemerlang Indonesia.
Kleemola, K., Hyytinen, H., & Toom, A. (2022). Exploring internal structure of a performance-based critical thinking assessment for new students in higher education. *Assessment & Evaluation in Higher Education*, *47*(4), 556–569. https://doi.org/10.1080/02602938.2021.1946482
Linacre, J. M. (2009). Unidimensional models in a multidimensional world. *Rasch Measurement Transactions*, *23*(2), 1209.
Linacre, J. M. (2012). Winsteps® Rasch measurement computer program user's guide. *Beaverton, Oregon: Winsteps. Com*.
López-Fernández, M. del M., Gonzalez-Garcia, F., & Franco-Mariscal, A. J. (2022). How can socio-scientific issues help develop critical thinking in chemistry education? A reflection on the problem of plastics. *Journal of Chemical Education*, *99*(10), 3435–3442. https://doi.org/10.1021/acs.jchemed.2c00223
Mardianto. (2019). Peran Guru Di Era Digital Dalam Mengembangkan Self Regulated Learning Siswa Generasi Z Untuk Pencapaian Hasil Pembelajaran Optimal. *Prosiding Seminar Nasional & Call Paper* , *1*, 150–158.

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.

Oktoviani, V., Widoyani, W. L., & Ferdianto, F. (2019). Analisis kemampuan pemahaman matematis siswa SMP pada materi sistem persamaan linear dua variabel. *Edumatica : Jurnal Pendidikan Matematika*, *9*(1), 39–46. https://doi.org/10.22437/edumatica.v9i1.6346

Paul, R., & Elder, L. (2019). *The miniature guide to critical thinking concepts and tools*. Rowman & Littlefield.

Payan-Carreira, R., Sacau-Fontenla, A., Rebelo, H., Sebastião, L., & Pnevmatikos, D. (2022). Development and validation of a critical thinking assessment-scale short form. *Education Sciences*, *12*(12), 1-19. https://doi.org/10.3390/educsci12120938

Pirinen, P., Lehtinen, A., & Holmes, N. G. (2023). Impact of traditional physics lab instruction on students' critical thinking skills in a Finnish context. *European Journal of Physics*, *44*(3), 35702.

Prayogi, R. D. (2020). Kecakapan Abad 21: Kompetensi Digital Pendidik Masa Depan. *Manajemen Pendidikan*, *14*(2), 144–151. https://doi.org/10.23917/jmp.v14i2.9486

Putra, P. D. A., Sulaeman, N. F., Supeno, & Wahyuni, S. (2023). Exploring students' critical thinking skills using the engineering design process in a physics classroom. *The Asia-Pacific Education Researcher*, *32*(1), 141–149. https://doi.org/10.1007/s40299-021-00640-3

Putri, J. K., Gunawan, C. W., & Wiyatmo, Y. (2024). Implementation of guided inquiry to increase students' interest in learning physics and critical thinking skills. *Jurnal Pendidikan Fisika Dan Teknologi*, *10*(2), 257–263. https://doi.org/10.29303/jpft.v10i2.7091

Quinn, S., Hogan, M., Dwyer, C., Finn, P., & Fogarty, E. (2020). Development and validation of the student-educator negotiated critical thinking dispositions scale (SENCTDS). *Thinking Skills and Creativity*, *38*(1), 1-17. https://doi.org/10.1016/j.tsc.2020.100710

Rahayu, S., & Rosawati, E. E. (2023). The development of higher-order thinking skills (HOTS) assessment instrument in chemistry using socioscientific issues context: A preliminary trial. *AIP Conference Proceedings*, *3*(1), 1-10.

Rapi, N. K., Suastra, I. W., Widiarini, P., & Widiana, I. W. (2022). The influence of flipped classroom-based project assessment on concept understanding and critical thinking skills in physics learning. *Jurnal Pendidikan IPA Indonesia*, *11*(3), 351–362. https://doi.org/10.15294/jpii.v11i3.38275

Sadler, T. D. (2004). Informal reasoning regarding socioscientific issues: A critical review of research. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *41*(5), 513–536. https://doi.org/10.1002/tea.20009

Sahin Kursad, M., & Yalcin, S. (2024). Effect of Differential Item Functioning on Computer Adaptive Testing Under Different Conditions. *Applied Psychological Measurement*, *48*(7–8), 303–322. https://doi.org/10.1177/01466216241284295

Sari, E., Suryawati, E., & Mahadi, I. (2024). *Diagnostic assessment to identify learning styles and critical thinking in independent learning Abstract :* *10*(3), 860–873. https://doi.org/10.22219/jpbi.v10i3.36557

Scott, I. A., Hubbard, R. E., Crock, C., Campbell, T., & Perera, M. (2021). Developing critical thinking skills for delivering optimal care. *Internal Medicine*

Journal, *51*(4), 488–493. https://doi.org/10.1111/imj.15272

Silber- Varod, V., Eshet- Alkalai, Y., & Geri, N. (2019). Tracing research trends of 21st- century learning skills. *British Journal of Educational Technology*, *50*(6), 3099–3118. https://doi.org/10.1111/bjet.12753

Solihin, R. R., Susanto, T. T. D., Fauziyah, E. P., Yanti, N. V. I., & Ramadhania, A. P. (2024). The Efforts of Indonesian Government In Increasing Teacher Quality Based On PISA Result In 2022: A Literature Review. *Perspektif Ilmu Pendidikan*, *38*(1), 57–65. https://doi.org/10.21009/PIP.381.6

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan rasch pada assessment pendidikan*. Trim komunikata.

Thornhill-Miller, B., Camarda, A., Mercier, M., Burkhardt, J.-M., Morisseau, T., Bourgeois-Bougrine, S., Vinchon, F., El Hayek, S., Augereau-Landais, M., & Mourey, F. (2023). Creativity, critical thinking, communication, and collaboration: Assessment, certification, and promotion of 21st century skills for the future of work and education. *Journal of Intelligence*, *11*(3), 1-32. https://doi.org/10.3390/jintelligence11030054

Umami, R., Rusdi, M., & Kamid, K. (2021). Pengembangan instrumen tes untuk mengukur higher order thinking skills (HOTS) berorientasi programme for international student asessment (PISA) pada peserta didik. *JP3M (Jurnal Penelitian Pendidikan Dan Pengajaran Matematika)*, *7*(1), 57–68. https://doi.org/10.37058/jp3m.v7i1.2069

Viehmann, C., Fernández Cárdenas, J. M., & Reynaga Peña, C. G. (2024). The Use of Socioscientific Issues in Science Lessons: A Scoping Review. *Sustainability (Switzerland)*, *16*(14), 1-29. https://doi.org/10.3390/su16145827