JURNAL EDUSCIENCE (JES)
p-ISSN : 2303 - 355X | e-ISSN : 2685 - 2217
PUBLISHED BY : LPPM of UNIVERSITAS LABUHANBATU

Critical Thinking Based on the Integration of *Batik* Local Wisdom in Science Learning: A Test Development

Nurul Iqdami Zuniari[1*], Antuni Wiyarsi[2*]

[1]Department of Science Education, State University of Yogyakarta, Indonesia
[2]Department of Chemistry Education, State University of Yogyakarta, Indonesia
Email: nuruliqdami.2023@student.uny.ac.id, antuni_w@uny.ac.id

ARTICLE INFO

Keywords:
Critical Thinking Test
Local Wisdom
Batik
Science Learning

ABSTRACT

Purpose-This study aims to develop an evaluation tool to measure students' critical thinking skills, specifically integrating the process of making *batik* local wisdom into science learning.

Methodology-This research uses the Research and Development (R&D) method with the 4D model, which includes four main stages: Define, Design, Develop, and Disseminate. A random sampling technique is used, which involves

160 students.

Findings-The instrument consists of 20 questions validated with Aiken's V values ranging from 0.93 to 0.97, indicating excellent content validity. All question items are declared fit based on the Rasch model analysis. The instrument had an item reliability of 0.91 and a respondent reliability of 0.73. Analysis of the distribution of threshold values (b) shows that most questions are in the medium category.

Contribution-This research integrates local wisdom, namely *batik*, into assessing critical thinking skills, an approach rarely encountered in previous research. In conclusion, the developed instrument effectively measures students' critical thinking skills. It can be used to measure them, especially in the context of learning about *batik* as part of local wisdom.

Received 11 December 2024; Received in revised form 18 December 2024; Accepted 12 May 2025
Jurnal Eduscience (JES) Volume 12 No. 3 (2025)
Available online xx June 2025

INTRODUCTION

Indonesian people are familiar with *batik*, one of the heritages owned by the country of Indonesia (Lestari & Hartadiyati, 2021). *Batik* is a form of local wisdom that has become one of Indonesia's unique identities at the international level(Puspasari et al., 2019). The recognition of *batik* as an intangible cultural heritage by the United Nations Educational, Scientific and Cultural Organization (UNESCO) in 2003 further strengthened *batik* as a symbol of Indonesian identity that needs to be preserved (Hakim, 2018). One way to maintain it is by integrating *batik* into learning materials, such as science lessons (Damayanti et al., 2017). Making *batik* can be integrated into science learning, namely temperature, heat, and expansion material because relevant scientific concepts provide students with contextual and relevant understanding (Utami, 2018). In addition, local wisdom, such as *batik*, can be used to implement the independent curriculum, which is the education application in Indonesia (Iwan, 2024).

Science learning is vital in developing students' knowledge while training objective thinking, fostering process skills, and deepening their understanding of the environment and scientific concepts (Sukma, 2020). Along with the times, science learning continues to transform to adapt to contemporary educational needs (Jannah & Atmojo, 2022). In the 21st-century education era, integrating digital technology is a key component of the learning process (Hikmat, 2022). The focus of education is now directed at developing four essential skills: creative thinking, critical thinking and problem-solving, communication, and collaboration, commonly referred to as the 4Cs (Partono et al., 2021).

Critical thinking skills are one of the main competencies students must have in the modern era (Waruwu et al., 2024). This skill helps students analyze and solve problems logically and becomes important in facing increasingly complex global challenges (Sulistiani, 2017). Critical thinking skills are very relevant in education, especially in learning Natural Sciences, which teaches students to understand natural phenomena scientifically and applicatively (Suratman, 2024). Critical thinking skills in science learning help them connect and understand scientific content that is often microscopic and abstract, requiring careful analysis, evaluation, and interpretation of students' thoughts (Yustiqvar et al., 2019).

In 21st-century learning, critical thinking skills are an important aspect that must be evaluated. These skills enable students to analyze, evaluate, and make decisions based on logical evidence (Zuniari et al., 2022).). Therefore, the evaluation should measure basic cognitive abilities and test students' ability to apply, analyze, and evaluate information in depth. A good evaluation will provide a comprehensive picture of students' abilities and help teachers design more effective learning strategies (Laila et al., 2024). Students' critical thinking skills can be trained and measured through three stages of learning: planning, implementation, and evaluation. At the planning stage, teachers prepare test instruments as an assessment technique to evaluate students' abilities. An instrument feasibility test is required for the assessment results to be accurate (Alyamuari, 2024). With the proper evaluation approach, learning can become more meaningful and relevant for students while preparing them to face real-world challenges.

Practical evaluation includes measuring learning outcomes and providing feedback for improving the learning process (Idrus, 2019). Through evaluation, educators can understand the extent to which students have mastered the material, identify weaknesses, and evaluate the effectiveness of the learning methods (Dakran, 2023). Evaluation is the final stage in the cognitive taxonomy, which involves the ability to assess information based on specific criteria. Thus, evaluation not only measures learning outcomes but also provides valuable feedback for teachers and students to improve the quality of learning (Naryatmojo, 2018).

Students' critical thinking skills are still low, especially in understanding materials related to measurements, such as temperature, heat, and expansion, which are considered difficult by 70% of students, compared to elemental, compound, and mixed (20%) and wave (10%) materials (Zuniari, 2025). This is due to the challenges in designing specific evaluation tools, such as evaluations to measure critical thinking skills with multiple-choice tests, which can accurately identify students' ability levels (Gierl et al., 2017). This limitation is an obstacle for grade VII students, who are at the stage of cognitive development where critical thinking skills should begin to be trained and developed. Thus, developing a comprehensive evaluation tool based on 21st-century learning needs is needed to support the success of the educational process.

The solution to overcome this problem is to develop an evaluation to measure critical thinking skills based on the integration of *batik* local wisdom in science learning to identify the level of students' critical thinking skills. This evaluation measures how students can analyze, evaluate, and solve problems involving relevant concepts in temperature, heat, and expansion materials, which are integrated with the *batik*-making process. This approach not only helps students understand scientific concepts in an applicable manner but also introduces and preserves local culture to make learning more contextual and meaningful.

Previous studies have highlighted the importance of critical thinking skills in science education, but most still focus on conventional approaches without integrating local culture. Another study on the development of culture-based learning media, Nurjanah (2024), showed that local wisdom can increase student engagement and understanding but has not been directed explicitly at evaluating critical thinking skills. In addition, research related to critical thinking evaluation tools by Zuniari (2022) shows success in measuring student abilities but less emphasis on aspects of integration with culture-based learning. This is in line with Suarjana's research (2020), namely developing evaluations with multiple choice instrument types to assess critical thinking skills in the context of science learning, but has not used cultural elements to increase student understanding. Siswanto (2022) developed a critical thinking skills e-instrument based on local wisdom but did not focus on *batik* as a cultural theme in developing critical thinking skills and not on temperature, heat, and expansion materials. In line with what Agustin (2018) has done, namely developing ethnoscience-laden test instruments to measure students' critical thinking skills, but the research does not focus on *batik* as a cultural theme in developing critical thinking skills, and not on temperature, heat, and expansion materials. The integration of batik in science learning not only improves the understanding of concepts but also students' critical thinking skills. For example, students can analyze the challenges of maintaining the temperature of the candle or determine the heat energy needed to make batik. It connects theory with practice, hones critical thinking skills, and encourages students to identify problems, collect data, and design efficient solutions, such as energy-saving methods in the coloring process (Zuniari, 2025).

This article aims to present a novelty from previous research, namely the development of an evaluation tool to measure students' critical thinking skills that explicitly integrates the process of making *batik* local wisdom into science learning. This tool serves to identify the level of students' critical thinking skills and support culture-based learning that is relevant to everyday life. By filling the gap in previous research, this evaluation can be an innovative model for measuring and improving students' critical thinking skills. Previous studies, such as those conducted by Zuniari (2022) and Suarjana (2020), have developed many evaluations to measure critical thinking skills, but none have specifically integrated local culture in this context. Based on these findings, this research offers novelty by focusing on integrating *batik* culture into science learning, which is expected to enrich students' learning experience while introducing and preserving local culture.

METHODOLOGY

Research Design

This research uses the Research and Development (R&D) method based on Sugiyono's approach (2019). This method was chosen because it is designed to develop new products that can solve real problems in education. According to Borg & Gall in Sa'diyah (2020), the R&D method is suitable for producing and testing the validity of learning instruments before they are widely implemented.

This test was developed based on critical thinking skills, which include five leading indicators: identification, explanation, inference, analysis, and evaluation. These indicators measure students' ability to think critically about temperature, heat, and expansion. The test is also integrated with *batik* as a cultural theme to provide a more relevant context in science learning. For example, students are asked to analyze the batik-making process regarding temperature, heat, and expansion. In this way, students learn scientific theory and relate it to their local cultural practice of *batik* making. The developed test consists of 20 multiple-choice items. Multiple-choice questions allow for deep understanding while encouraging critical thinking, analysis, and problem-solving by requiring students to assess options based on logical reasoning. In addition, this format explores various dual intelligences and prepares students for standardized exams that generally use multiple choice (Zahra, 2024).

The instrument validation process involves two experts—science material experts and instrument evaluation experts—who assess each question item based on three main aspects. First is the material aspect, which includes the suitability of the questions with critical thinking indicators, clarity of context and answer choices, and only one correct or most appropriate answer key. Second, the construction aspect includes the clarity of the question formulation, the accuracy of the sentence structure, the equality of the length of the answer choices, the effectiveness of the distractor, and the independence between the questions. Third, language, cultural, and ethical aspects were addressed to ensure adherence to Indonesian language standards, use of communicative language, and freedom from SARA content, political bias, violence, or local terms that could introduce bias.

After revision based on expert feedback, the instrument was empirically tested on 160 eighth-grade students who had studied temperature, heat, and expansion to evaluate its empirical validity, difficulty level, discriminating power, and overall reliability.

Test items were developed by integrating local cultural contexts, specifically traditional batik-making processes. Each item was designed based on critical thinking indicators (identification, inference, explanation, analysis, and evaluation) and contextualized within specific batik activities such as "nagging," "netbook," and "good." For example, items requiring students to analyze temperature conversions were framed around the boiling process during "good," while explanation-based items related to phase changes of wax during heating. This design aimed to assess students' critical thinking while engaging them in culturally relevant scenarios. This is in line with the opinions of Annisha (2024) and Nurjanah (2024), who stated that cultural elements in learning increase the involvement and relevance of material for students. The development model used is 4D (Define, Design, Develop, Disseminate) by Thiagarajan et al. (W. Ramansyah, 2012), which ensures the instrument's validity, reliability, and relevance to the learning needs based on local wisdom. The 4D model consists of four main stages: Define, Design, Develop, and Disseminate. These systematically ensure that the evaluation tools developed have validity, reliability, and relevance to local wisdom-based learning needs.

Define

The analysis stage was conducted to identify needs, analyze possible solutions, determine the test's purpose, and select the question format. It was conducted through observations, interviews, and literature studies to identify learning needs. The population in this study is junior high school students in Yogyakarta who are in culture-based schools (integrating local cultural values into their curriculum and learning activities), with a sample of 160 students who have studied temperature, heat, and expansion materials.

This study used a purposive sampling method to determine the subject. It obtained as many as 160 grade VIII students who had studied temperature, heat, and expansion materials and had a background in batik culture. Data will be collected in October 2024 through a Google Form link containing 20 multiple-choice questions based on critical thinking skills. Each question is developed based on critical thinking indicators such as identification, inference, explanation, analysis, and evaluation and is integrated with the batik process as a local cultural context.

Data analysis is carried out in several stages. The validity of the content was analyzed using Aiken's V, with rating categories ranging from very low to very high. Furthermore, the data was analyzed using the QUEST program with the Rasch Model approach to test the instrument's empirical validity, reliability, difficulty level, and the questions' differentiating power. Empirical validity was analyzed through MNSQ's INFIT value with a range of 0.77–1.33 and OUTFIT $t \leq 2.00$. The instrument's reliability is seen from the reliability coefficient of the item and the respondent. The difficulty level of the questions is determined based on the logit value, with the categories of very easy ($b < -2$) to complicated ($b > 2$). The differentiating power of questions is analyzed through the correlation between item and total scores to determine the ability of questions to distinguish between high and low-ability students.

Design

This stage includes the process of designing critical thinking skills questions that will be used; the design process includes making a matrix of questions according to the grid that has been designed and by the synthesis of critical thinking skills indicators, taking into account the local context in the formation of question instruments that have been carried out according to student needs.

Each indicator in this synthesis (identification, explanation, inference, analysis, evaluation) is translated into questions relevant to the real situation in the *batik*-making process. This aims to help students understand the material contextually and applicatively. The local context is integrated by incorporating cultural elements, such as the stages in *batik*-making, including the selection of materials for tools, such as canting, that are related to heat-conducting properties. The questions are designed to assess students' ability to analyze how heat transmission in different types of canting materials can affect the smoothness of the dyeing process on *batik* cloth. With this approach, students hone critical thinking skills and understand the application of science concepts, such as temperature, heat, and expansion, in everyday life related to local culture.

Develop

The next stage is preparing and validating critical thinking skills questions involving educational evaluation experts, physics education experts, and practitioners. This validation aims to ensure that the questions are by the indicators of critical thinking, learning objectives, and local context (Damayanti et al., 2017). Educational evaluation experts assess the validity and reliability of the instruments (An Nabil et al., 2022); physics education experts evaluate the suitability of the questions with science materials such as temperature, heat, and expansion, while practitioners provide input related to the relationship of the questions to learning practices and local cultural elements such as the *batik* process.

Validation was carried out using Aiken's V index (Aiken, 1980, 1985) to assess the suitability of each question item with the critical thinking skill indicator. The validators give ratings based on a scale of 1 to 5. The results of the assessment are calculated to determine the level of suitability of the questions. Questions with low index values are revised, replaced, and retested through several validation cycles. This approach ensures that the instrument is genuinely effective in measuring students' critical thinking skills in the context of science learning based on local wisdom (An Nabil et al., 2022). The V index has a value range between 0 and 1. Based on the V index calculation, each item can be grouped according to its index value, as listed in Table 1 (Syafitri et al., 2023).

Table 1. Aiken's Value Category V

Value Range	Category
0,8 – 1	Very High
0,6 – 0,799	High
0,4 – 0,599	Moderat
0,2 – 0,399	Low
< 0.2	Very Low

Disseminate

The next stage is applying critical thinking skills integrated with local wisdom through making *batik*. In this study, the purposive sampling method was used to select students who have studied temperature, heat, and expansion materials and have a background in *batik* culture. The Rasch model objectively measures students' critical thinking skills by comparing individual abilities with the problem's difficulty level. This analysis allows the evaluation of the interaction between participants and questions and produces logit values that describe students' abilities and problem characteristics (Muntazhimah, 2020). This model ensures the quality of the questions by evaluating parameters such as validity, reliability, difficulty, and suitability of the questions using the Quest program. Questions that meet the criteria of INFIT MNSQ and OUTFIT t ≤ 2.00 are considered valid, while higher ones need revision (Pratama, 2020). Using the Rasch Model, this study can produce valid and reliable data to measure students' critical thinking skills in the context of learning based on local *batik* wisdom (Cahayu, 2024).

To test the validity using the QUEST program, as revealed by Setyawarno (2017), can be compared through the criteria below:

Table 2. INFIT MNSQ Value Criteria against the Rasch model

INFIT MNSQ Value	Information
>1.33	Does not match the model
0.77-1.33	Fit model
<0.77	Does not match the model

The reliability value of the Rasch model using the QUEST program includes estimates of the reliability of the items and respondents. QUEST is used to evaluate the quality of questions through a data analysis process, using a binary scale for students' answers (1 for true, 0 for false) (Mustafidah & Harjono, 2019). The parameters analyzed included validity, reliability, and differentiating power. Validity is measured through item-total correlation, while reliability includes the reliability of items and respondents to assess measurement consistency (Yulisharyasti et al., 2023). Differentiating power measures the ability of questions to distinguish students with high and low abilities. In contrast, the difficulty level is measured with a logit scale to ensure the questions are not too easy or difficult (Pratama, 2020). Items are suitable if they meet the criteria of INFIT MNSQ between 0.77-1.30 and OUTFIT t ≤ 2.00. Inappropriate items, such as OUTFIT t > 2.00, will be revised or removed (Bond, Yan, & Heene, 2020). Difficulty analysis ensures that the distribution of questions covers different levels of student ability, while reliability assesses the consistency of the overall test (Muntazhimah, 2020). This process ensures that questions can measure students' critical thinking skills with the relevance of the learning context based on local wisdom.

Table 3. Rasch Model Instrument Reliability Correlation Coefficient Criteria

Interval Score	Criterion
< 0.67	Weak
0,67 – 0,80	Simply
0,80 - 0,90	Good
0,91 – 0,94	Very good
> 0.94	Special

The difficulty level of items in the QUEST program can be assessed based on the criteria thresholds listed in Table 4.

Table 4. Difficulty Level Criteria

Threshold Value	Information
$b > 2$	Very difficult
$1 < b \leq 2$	Difficult
$-1 \leq b \leq 1$	Medium
$-2 \leq b < -1$	Easy
$b < -2$	Very easy

FINDINGS

The instrument validation process is carried out to ensure that the questions developed have good content validity. This is done by involving expert validators who assess the suitability and quality of the questions. The validity test of the instrument includes material, construction, and language. Several revisions related to the wording are appropriate to the question. The analysis of the validity of the questions assessed by the validators showed that Aiken's V value was in the range of 0.93 to 0.97, which is very high, indicating a valid result. Based on the validation results, several questions were revised, especially in the language section, to make it easier for students to understand, and adjusted to the corrections and suggestions given by experts and practitioners. In addition, there are grammatical revisions to certain terms in the context of the *batik*-making process, such as "*nagging*" and "lord," which are italicized to provide emphasis and make it easier for students to understand. The results of the INFIT MNSQ value analysis in the QUEST program can be seen in Table 5.

The results of the analysis showed that the INFIT MNSQ value was in the range of 0.80 to 1.24, which showed that the 20 questions were by the model because the INFIT MNSQ value was in the range of 0.77 to 1.33 and Outfit t, where the question item is considered valid if Outfit t ≤ 2. The analysis results in Table 5 show that all items of the fit question correspond to the Rasch model and are valid for use. Therefore, it can be concluded that 20 questions were tested, all of which meet the validity criteria and are suitable for use.

Table 5. Results of MNSQ INFIT Value Analysis

Item	INFIT MNSQ	Description	OUTFIT t	Description
1.	1.00	Item Fit	0.5	Valid
2.	0.80	Item Fit	-2.0	Valid
3.	1.05	Item Fit	0.2	Valid
4.	1.00	Item Fit	-0.1	Valid
5.	1.21	Item Fit	1.7	Valid
6.	1.05	Item Fit	0.2	Valid
7.	0.98	Item Fit	-0.5	Valid
8.	1.02	Item Fit	-0.3	Valid
9.	0.84	Item Fit	-1.7	Valid
10.	0.94	Item Fit	-0.8	Valid
11.	1.05	Item Fit	0.3	Valid
12.	1.04	Item Fit	0.4	Valid
13.	0.85	Item Fit	-1.6	Valid
14.	0.91	Item Fit	-0.2	Valid
15.	1.16	Item Fit	1.6	Valid
16.	0.93	Item Fit	-0.9	Valid
17.	1.22	Item Fit	1.7	Valid
18.	1.04	Item Fit	0.1	Valid
19.	0.96	Item Fit	-0.3	Valid
20.	1.03	Item Fit	0.2	Valid

The reliability value of the Rasch model using the QUEST program is seen in the item estimate and the reliability of the case estimate (Pratama, 2020). The results of the reliability analysis obtained in developing critical thinking skills instruments can be seen from the output of the Quest program in Table 6.

Table 6. Reliability of Critical Thinking Skills Test Instruments

No.	Reliability	Value	Category
1.	Reliability of item estimate	0,91	Very good
2.	Reliability of case estimate	0,73	Simply

An item estimation reliability value of 0.91 indicates an excellent category, contributing to many items that fit the model. The higher the reliability, the more items fit the model. However, the estimated reliability value of the test takers was only 0.73, which was quite enough. This reliability value is influenced by the number of test takers, and previous research by Radja (2023) showed that the reliability of test takers increased along with the increase in the number of participants. To overcome the problem of suboptimal reliability, this study involved 160 participants. This is in line with the opinion of Alwi (2015), who stated that to ensure the stability of the respondent size in the trial, the number of respondents should be 10% of the number of items in the measuring instrument. This step aims to increase the stability of the estimation of participants' abilities based on the recommendations of previous research, which showed that the reliability of test participants increased with the increase in the number of participants. In the development of instruments, the number of test subjects is determined based on the number of items, which is 5 to 10 times the number of question items (Sumin, 2023).

The QUEST program's output can be used to review the difficulty level of items obtained from developing critical thinking skills instruments. The difficulty level of an item can be found by checking the results of the item estimate (Threshold) analysis, as seen in Figure 1.

Based on the figure, the analysis of the questions' difficulty level shows that most are in the medium difficulty category, with question number 5 included in the difficult category (threshold value > 1.0). Most of the other questions are in the range of -1.0 to 1.0, indicating that they have a moderate level of difficulty.

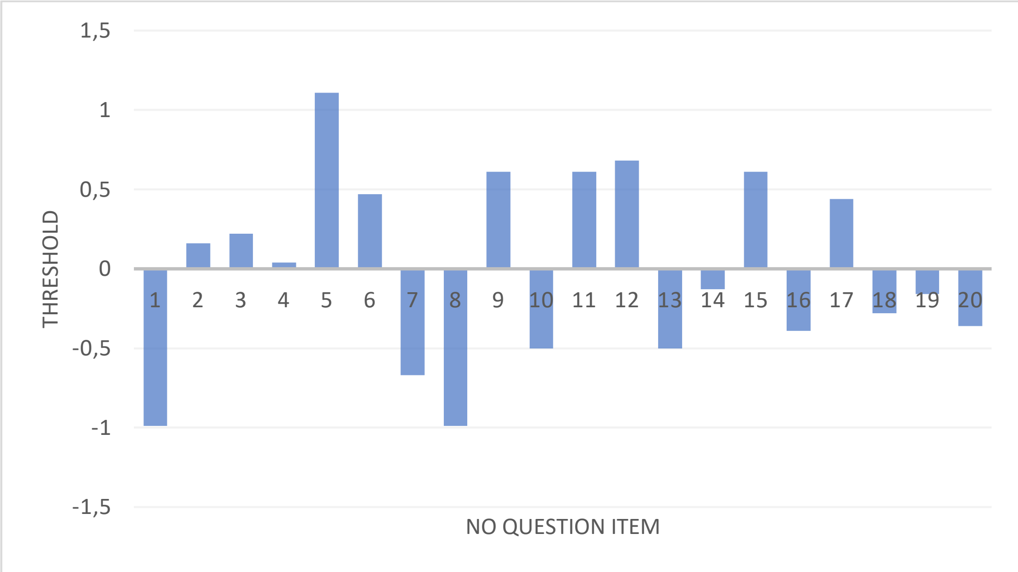


Figure 1. Item Difficulty Spread

DISCUSSION

This research aims to develop an instrument of critical thinking skills integrated with local wisdom in making *batik*. The development of evaluation to measure students' critical thinking skills based on integrating *batik* local wisdom provides strategic benefits in education, especially in finding out the extent of students' ability to analyze, evaluate, and solve problems. With this evaluation, data showed that students' critical thinking skills in temperature, heat, and expansion were in the medium to low category. These findings confirm the need for improvement and innovation in learning media to support improving students' critical thinking skills.

The instrument is declared valid based on the construct's and face's validity. The validity of the construct was obtained through expert and practitioner responses with Aiken's V values ranging from 0.93 to 0.97, indicating good aspects of the material, design, and language. The instrument meets the criteria of items that correspond to indicators, clear answer choices, and communicative and grammatical language.

The IRT model analysis showed that the INFIT MNSQ value ranged from 0.80 to 1.24, which indicated that the 20 questions were by the model. The INFIT MNSQ value range is within the expected limit of 0.77 to 1.33. The reliability analysis showed that the reliability value of the item was 0.91, and the reliability of the students was 0.73. Thus, the instrument was declared feasible to measure students' critical thinking skills.

The graph analysis of the distribution of each question item's threshold value (b) shows that most of the questions are in the medium category, namely in the range of value b between –1.0 and 1.0. This shows that, in general, the questions prepared have a proportionate level of difficulty and are based on the abilities of most students. This category is ideal for measuring students' critical thinking abilities, as they are not too easy but not too challenging. One question item is included in the difficult category, namely question number 5, with a threshold value of more than 1.0. This question item shows that the assessment also includes the ability of students at a higher cognitive level to distinguish students with above-average abilities. Overall, the distribution of the difficulty level of the questions dominated by the medium category and interspersed with one difficult question reflects the quality of the instrument. This is in line with the principle of learning evaluation, which prioritizes the proportion of difficulty of questions in order to be able to accommodate variations in students' abilities fairly and evenly (Sawaluddin & Muhammad, 2020).

The following is an explanation of 20 questions with critical thinking skills indicators designed to help students develop their ability to analyze, evaluate, and make conclusions based on existing information. These questions will measure various aspects of critical thinking skills.

Identification Indicators

The instrument with identification indicators has four question items designed to measure students' ability to recognize and determine the correct answer based on the information provided. These questions closely relate to the learning context, integrating scientific theories with practical applications in the *batik*-making process. Here are the four questions:

Question number 1 (Nyungging process and temperature concept)

> "*The process of making batik involves the process of nagging, which is making patterns on the fabric to be batiked. In the process of nagging, a light is added under the pattern fabric, which is used to help see the pattern's details. If Irma is in the process of nagging and feels that the air is hotter, the possible cause is...*"

There are four different answer options in this question. Presented with the phenomenon of *batik*-making, students could identify the influence of heat energy on the user's comfort when weaving *batik*. Students are expected to be able to identify the leading cause of the increase in temperature felt during the *nagging* process by connecting the phenomenon that occurs, namely the use of bright lights, with the concept of temperature. This question aims to enable students to identify the reason for the occurrence of an event, which is by one of Ennis' indicators, namely in the basic clarification indicator, students can identify the reason for the occurrence of an event (Soleh, Susyla, & Syopiana, 2019). In the question above, students are asked to identify what happens in the *nagging* process and choose the appropriate answer, namely how the electrical energy used by bright lights is converted into heat energy, thus affecting the temperature of the room

Question number 3 (Movement of particles)

> "*In everyday life, temperature can be measured using a thermometer. In general, thermometers are made of glass pipes filled with a liquid substance. Joji will measure a substance that has a temperature of -45°C. The thermometer filler material that might reach that temperature is...*"

There are four different answer options for this question. Presented with a phenomenon related to how thermometers work, students can identify the appropriate thermometer filling material to measure very low temperatures. Students are expected to be able to identify the appropriate thermometer filling material by paying attention to the concept of temperature on particle movement and the working principle of the thermometer. This question aims to enable students to identify the appropriate liquid substance by paying attention to the particles' characteristics and the thermometer's working principle. This is Ennis's opinion, which Nufus (2020) states: a person is said to have carried out critical thinking activities if he can identify and find answers according to the context.

Question number 5 (Temperature of Batik Stages)

> "*A batik maker needs a specific temperature for the batik process. Here are the temperatures required for the different stages:*
> *- Damping: High temperatures, ranging from 75–80°C, to block fabrics.*
> *- Klowong: Medium temperature, ranging from 60–75°C, to make thin lines or motifs.*
> *- Paraffin: Low temperatures, ranging from 50–60°C, to make batik for specific motifs.*
> *However, the thermometer used showed a temperature of 185°F on the Fahrenheit scale and 358°K on the Kelvin scale. Batik faced difficulties because the temperature did not match the desired category. If the temperature listed is 185°F on the Fahrenheit scale and 358°K on the Kelvin scale, the likelihood of this is...*"

In this question, there are four different answer options. Students are expected to be able to identify the appropriate temperature scale by converting the degree scale to the thermometer. This question aims to enable students to clarify questions by converting the appropriate temperature scale and identify how the temperature scale conversion can be used in appropriate treatment in the *batik*-making process, such as to block fabric, *batik* thin motif lines, or make specific motifs. In Ennis' opinion, students are said to have critical thinking skills when they can clarify questions or identify the reasons for an event (Soleh et al., 2019).

Question number 13 (Candle Form Change)

> *"In mating batik, the heat from the electric stove transfers to the pan and melts the wax. When the wax applied to the batik fabric is left unheated, the wax will harden again. The cause of the wax hardening again after being applied to the batik cloth and left without heating is..."*

There are four different answer options for this question. Presented with an analysis of heat (heat) in the *batik*-making process, students can identify changes in the form of substances that occur.

Students are expected to be able to identify by providing conclusions about the change in form in the *batik*-making process. In Ennis' opinion, students are said to have critical thinking skills when they can identify the conclusion of the occurrence of a phenomenon (Soleh et al., 2019).

Inference Indicator

In the instrument with the inference indicator, two questions pass the OUTFIT t value of ≤ 2.00 to be used in drawing conclusions or making decisions based on the available information. These questions closely relate to the learning context, integrating scientific theories with practical applications in the *batik*-making process. Here are the two questions:

Question number 8 (Concept of Calories)

> *"If an object is heated, it will receive additional heat so that its temperature increases or increases. On the other hand, if the substance is cooled, it releases heat, which causes the temperature to drop. The relationship between the statement and the reason is .... "*

There are four different answer options in this question. Students are expected to be able to draw conclusions about whether the statement is in accordance with the concept of heat that they have learned. This question aims to enable students to question the proof of the concept of heat in objects by concluding that a statement is true or false. This is to Facione's statement Anjani (2023), which is that the inference indicator can foster the skill of questioning evidence and predicting alternatives with the final result of making a decision or conclusion.

Question number 2 (Types of Thermometers)

> *"Attention to the thermometer characteristics statement below!*
> *1. Has a freezing point of -144°C and a boiling point of 78°C*
> *2. Has a freezing point of -39°C and a boiling point of 357°C*
> *3. Easy to see because it is shiny*
> *4. It takes a long time to read the temperature scale*
> *Febi wants to measure the temperature of an object using a mercury thermometer. Here is a description of the characteristics of the water thermometer. The correct mercury is shown in the number.*

In this question, there are four different answer options. Students are expected to be able to identify the characteristics that correspond to the mercury thermometer from the four statements presented. According to Facione's statement, Anjani (2023), by the inference indicator, this skill includes the ability to predict possible alternatives and draw conclusions based on available evidence. Students are expected to be able to identify the characteristics of mercury thermometers based on the available statements. According to the inference indicator, according to Facione, this skill includes the ability to predict alternatives and infer based on evidence (Anjani, 2023).

Question number 17 (Heat transfer)

> *"In the process of boiling batik cloth using a stick is an example of heat transfer by ...."*

In this case, students are expected to be able to associate the activity of boiling batik cloth using a stick with the concept of proper heat transfer. By the inference indicator, according to Facione (in Anjani, 2023), this skill includes the ability to predict possible alternatives and draw conclusions based on available evidence so that students can determine that the process is an example of heat transfer by conduction.

Question number 18 (Movement of particles and working principle of thermometer)

> "*In making batik, the drying process is essential for the final result of the batik motif. It still contains high moisture after the fabric is dyed and removed from the dye solution. To ensure that the dye adheres well and the batik motif is clear, the fabric must be dried effectively, usually by hanging it in the sun or an area with good air circulation. Sunlight helps accelerate water evaporation from the fabric, making the dye harden and permanent. The drying process shows the process of heat transfer in a way....*"

In this question, there are four different answer options. Presented with phenomena related to the fabric drying process, students can conclude the heat transfer events. Students are asked to choose the right type of heat transfer, which involves understanding the physical processes that occur during fabric drying. Students must predict alternative types of heat transfer that correspond to the phenomenon of drying fabrics, i.e., whether it is convection, conduction, or radiation. Students are also expected to make decisions or conclusions based on information about drying fabrics in the sun or areas with good airflow. An indicator of inference is the ability of students to obtain and identify concepts or elements in a conclusion (Zuniari et al., 2022). This is to Facione's statement by Anjani (2023) that the inference indicator can foster the skill of questioning evidence and predicting alternatives, with the final result of making a decision or conclusion.

Explanatory Indicators

In the instrument with identification indicators, three question items pass the OUTFIT t ≤ 2.00 value so that it can be used in explaining a science concept related to the *batik*-making process. These questions closely relate to the learning context, integrating scientific theories with practical applications in the *batik*-making process. Here are the four questions:

Question number 4 (How Thermometers Work)

> "*Rina uses a digital air thermometer that can show the temperature of the surrounding environment. The device shows an air temperature of 30°C. The way the device works to measure the air temperature is ....*"

There are four different answer options in this question. Students are expected to be able to explain how a device works, related to how a thermometer works. This question asks students to explain how a digital air temperature gauge works, which requires them to justify or explain the procedure used by the device to measure air temperature. This is according to Facione's statement Anjani (2023), which states that explanatory indicators can cultivate the skill of declaring results, justifying procedures, presenting arguments, or making self-corrections.

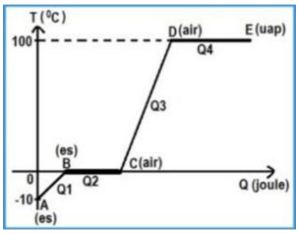
Question number 12 (Movement of particles and the working principle of the thermometer)

> "*Sasi put ice cubes in a container filled with water. After 30 minutes, the water and ice temperatures become the same, with more water and less ice left. The correct statement is....*"

In this question, there are four different answer options. Presented with an analysis of the temperature balance in a mixture of ice and water, students can explain that the temperature of the water will be the same as the temperature of the remaining ice after reaching temperature equilibrium. This question asks students to identify or conclude the experiment's results, i.e., changes in water and ice temperatures after 30 minutes, and then declare the results by choosing the correct answer. This is according to Facione's statement Anjani (2023), which states that explanatory indicators can cultivate the skill of declaring results, justifying procedures, presenting arguments, or making self-corrections.

Question number 15 (The relationship of heat with temperature changes)

"*Look at the following graph of the relationship between heat and temperature change!*



The graph shows the process of melting ice, which occurs at the temperature of the substance during the melting phase, and the phase of temperature change as follows....."

There are four different answer options for this question. Presented with a graph showing the relationship between heat and temperature change during ice melting, students are expected to explain which segments of the graph represent the melting phase and which represent temperature change.

This question requires students to interpret the meaning of each line segment (A–B, B–C, C–D), then declare the correct process at each stage by selecting the most appropriate answer. This aligns with Facione's statement in Anjani (2023) states that explanation indicators foster the ability to state results, justify procedures, present arguments, or make self-corrections.

Question number 19 (Application of the Black Principle)

"*During the nembok process (covering the fabric with wax), the wax heated in the pan becomes liquid. The melting of the candle occurred because ....*"

There are four different answer options in this question. Students are asked to explain why the candle heated in the pan becomes liquid, which requires them to present an argument or scientific explanation related to changing the shape of the candle due to heating by choosing the correct answer. This is according to Facione's statement Anjani (2023), which states that explanatory indicators can cultivate the skill of declaring results, justifying procedures, presenting arguments, or making self-corrections.

Analytical Indicators

The instrument with identification indicators has four question items designed to measure students' ability to recognize and determine the correct answer based on the information provided. These questions closely relate to the learning context, integrating scientific theories with practical applications in the *batik*-making process. Here are the four questions:

Question number 6 (Concept of Heat)

"*Three batik makers measured the temperature of the water heated in the good process using three different thermometers. The first batik recorded a temperature of 149°F, while the second and third batik started inserting the fabric when the water temperature reached 25°C and 209°K. The first batik finishes faster than the others. To ensure that all three batik makers can complete the job simultaneously, all that must be done is... .*"

There are four different answer options for this question. Presented with a phenomenon related to temperature conversion in the *batik* process, students can analyze the temperature difference and determine the right temperature to equalize the completion time. Students are asked to assess and choose the correct arguments regarding the temperature that must be achieved so that all three *batik* makers can complete the work at the same time. This problem requires students to consider the information provided (temperature differences in Fahrenheit, Celsius, and Kelvin) and assess the most appropriate arguments based on physics concepts related to temperature and its conversion. This is similar to Facione's statement by Anjani (2023), which states that analytical indicators can carefully examine an idea and assess the correctness of an argument.

Question number 7 (Movement of particles and the working principle of the thermometer)

> "*Nadia had difficulties opening the jar with a metal lid. To solve the problem, he decided to pour hot water over the lid of the jar. After this process, the lid of the jar becomes easier to open. This is caused by.....*"

In this question, there are four different answer options. Presented with problems on the concept of heat transfer, students can analyze the effect of metal expansion on the ease of opening the lid of the jar. Students were asked to assess the arguments about the expansion effect on metal lids and jars after exposure to hot water. They need to evaluate the most appropriate explanations based on the physics concept of material expansion, particularly the difference in thermal properties between metals and other materials such as glass or plastic. This is similar to Facione's statement by Anjani (2023), which states that analytical indicators can carefully examine an idea and assess the correctness of an argument.

Question number 14 (Factors causing expansion)

> "*In the process of making batik, there is a stage of drying batik cloth that has been boiled to dry. Batik fabrics that are sun-dried in hot places tend to dry quickly compared to those in the shade. This suggests that the evaporation process can be accelerated in a way that ... .*"

In this question, there are four different answer options. Presented with the identification of changes in the form of substances in the concept of heat, students can analyze ways that can accelerate the evaporation process. Students are asked to examine ideas or concepts related to accelerating the evaporation process, which involves understanding the effects of temperature, pressure, airflow, and surface area on the evaporation rate. This question tests their understanding of the basic concepts in physics related to the change of the form of matter. This is similar to Facione's statement by Anjani (2023), which states that analytical indicators can carefully examine an idea and assess the correctness of an argument.

Question number 16 (Heat Transfer)

> "*In making batik, the canting part called nyamplung is made of copper. Copper is used as the main material of nyamplung because it has properties that can withstand the heat of the candle, so that the candle does not freeze quickly and clog the canting. The properties of copper that make it suitable for use in nyamplung are ... .*"

There are four different answer options for this question. By examining phenomena related to heat transfer in copper materials, students can analyze the properties of copper materials that can affect their ability to conduct heat. Students were asked to examine ideas related to copper's properties that make it suitable for use in the *batik* process, especially related to copper's ability to conduct heat. This question tests students' understanding of thermal conductivity and how the material's properties affect its use in practice. This follows Facione's statement by Anjani (2023), which states that analytical indicators can carefully examine an idea and assess the correctness of an argument.

Evaluation Indicators

The instrument with identification indicators has four question items designed to measure students' ability to recognize and determine the correct answer based on the information provided. These questions closely relate to the learning context, integrating scientific theories with practical applications in the *batik*-making process. Here are the four questions:

Question number 9 Application of the Heat Formula

> "*Batik boils water at a temperature of 100°C, which is insufficient for the medical process. He then added 300 g of water with a temperature of 0°C. If the mixture's temperature is 40°C, then the mass of hot water added is .....*"

In this question, there are four different answer options. Presented with initial mass data and temperature changes, students can calculate the added hot water mass results. The problem is included in providing the expected solution by solving the given problem because students are asked to calculate the mass of hot water based on known temperature data and cold water mass, using the Black principle concept (heat equilibrium).

This is based on Facione's statement, Oktaviyanti & Fadly (2023), that the evaluation indicators can provide the expected solutions and solve the problems given.

Question number 10 (Heat of Object Type)

"*Pay attention to the following type of heat table!*

Heat Type (J/Kg°C)	*Thing*
235	*Silver*
380	*Copper*
450	*Iron*
920	*Aluminum*

Metal X with a mass of 4.2 kg is heated to a temperature of 140°C, then put into 9.2 kg of water at a temperature of 30°C. After a while of equilibrium, the final temperature of the mixture becomes 40°C. If it is known that the heat of the water type is 4,200 J/kg°C, then it can be concluded that the type of metal material X is ... ."

In this question, there are four different answer options. Presented with heat data of object types, students can determine the type of metal material by comparing the type of heat in the table with the results of calculating the temperature of the mixed equilibrium. Students were asked to analyze the data, use Black's basic concept (heat equilibrium), and calculate the calorific value of metal type X to match it with the data in the table. This process requires the application of physical formulas, mathematical calculations, and drawing logical conclusions based on the results of calculations. This is to Facione's statement by Oktaviyanti and Fadly (2023), which states that analytical indicators can carefully examine an idea and assess the correctness of an argument.

Question number 11 (Calculating the Number of Calories)

"*A batik craftsman wants to melt wax on batik cloth (pe-lorod-an process). It requires water with an initial temperature of 20°C and a mass of 0.5 kg, which needs to be heated to 100°C and then evaporated. The water type heat is 4,200 J/kg°C, and the water vapor heat is 2,300 kJ/kg. The amount of heat required to evaporate the water is equal to.....*"

In this question, there are four different answer options. Presented with data on the type of water and water vapor heat, students can calculate the number of calories using the heat formula. The problem is included in solving the given problem because students must apply heat, calculate the total energy needed to heat water to the boiling point, and add heat to convert it into steam. Solving this problem involves understanding physics concepts and mathematical abilities to complete the calculation steps correctly. This is to Facione's statement by Oktaviyanti and Fadly (2023), which states that analytical indicators can carefully examine an idea and assess the correctness of an argument.

Question number 20 (Calculating the Length Coefficient)

"*A copper rod with an initial length of 50 cm has an initial temperature of 15°C. The coefficient of expansion of copper length is 1.8×10−5/°C. Calculate the length of the copper rod at 65°C.*"

In this question, there are four different answer options. Presented with data on the initial length, temperature difference, and length expansion coefficient, students can calculate the final result of copper length using the length expansion formula. The problem is included in solving the given problem because students are asked to calculate the length of the copper rod based on temperature changes using available data. Solving this problem requires the accurate analysis and application of the concept of length expansion. This is to Facione's statement by Oktaviyanti and Fadly (2023), which states that analytical indicators can carefully examine an idea and assess the correctness of an argument.

CONCLUSION

Based on the results of data analysis and discussion of the research that has been conducted, it can be concluded that the validity of the instrument shows good results, with an Aiken's V value between 0.93–0.97, including aspects of material, design, and language by indicators, clear answer choices, and communicative and grammatical language. Twenty instruments met the model criteria with INFIT MNSQ values in the range of 0.77 to 1.33, passed with an OUTFIT t ≤ 2.00. The reliability of the item estimate value of 0.91 indicates a very good category. However, the reliability of the case estimate of test participants is only 0.73, which is quite enough. The difficulty level of the instrument shows that question number 5 is classified as difficult, while other questions are in the medium category.

REFERENCES

Agustin, N., Sudarmin, S., Sumarti, S. S., & Addiani, A. K. (2018). Desain Instrumen Tes Bermuatan Etnosains untuk Mengukur Kemampuan Berpikir Kritis Siswa SMA. *Jurnal Inovasi Pendidikan Kimia*, *12*(2).

Alwi, I. (2015). Kriteria Empirik dalam Menentukan Ukuran Sampel pada Pengujian Hipotesis Statistika dan Analisis Butir. Formatif: Jurnal Ilmiah Pendidikan MIPA, 2(2).

Alyamuari, A., Liliawati, W., & Saepuzaman, D. (2024). Item Analysis Of The Critical Thinking Test Instrument On Temperature And Heat Using Rasch Modelling. Jurnal Eduscience, 11(2), 279-289.

An Nabil, N. R., Wulandari, I., Yamtinah, S., Ariani, S. R. D., & Ulfa, M. (2022). Analisis Indeks Aiken untuk Mengetahui Validitas Isi Instrumen Asesmen Kompetensi Minimum Berbasis Konteks Sains Kimia. *PAEDAGOGIA*, *25*(2), 184. https://doi.org/10.20961/paedagogia.v25i2.64566

Anjani, S., Muspita, T. D., & Salsabila, N. (2023). Analysis Of Students' Critical Thinking Ability Class IV Civics Subjects At State Elementary School 106154 Rantang City. *International Journal of Students Education*, *2*(1), 17-20.

Annisha, D. (2024). Integrasi Penggunaan Kearifan Lokal (Local Wisdom) dalam Proses Pembelajaran pada Konsep Kurikulum Merdeka Belajar. *Jurnal Basicedu*, *8*(3), 2108–2115. https://doi.org/10.31004/basicedu.v8i3.7706

Cahayu, A., Siburian, J., Hamidah, A., & Falani, I. A Rasch Model Analysis in Measuring Students' Critical Thinking Skills in Science Learning Model Problem-Based Learning (PBL) Based on Local Wisdom Article history Abstract Keyword Rasch Model Critical Thinking Science Learning Problem-Based Local Wisdom. *Jurnal Pendidikan Sains Universitas Muhammadiyah Semarang*, *12*(1), 2339–0786. https://doi.org/10.26714/jps.12.1.2024.1-8

Dakran, D., Zulhimma, Z., Harahap, W. A. A., & Royhanuddin, F. (2024). Evaluasi dalam Proses Pembelajaran di SMA Negeri 1 Panyabungan Selatan. *Cognoscere: Jurnal Komunikasi dan Media Pendidikan*, *2*(2), 32-44.

Damayanti, C., Rusilowati, A., & Linuwih, S. (2017). Pengembangan Model Pembelajaran IPA Terintegrasi Etnosains untuk Meningkatkan Hasil Belajar dan Kemampuan Berpikir Kreatif. *Journal of Innovative Science Education*, *6*(1), 116-128.

Gierl, M. J., Bulut, O., Guo, Q., & Zhang, X. (2017). Developing, Analyzing, and Using Distractors for Multiple-Choice Tests in Education: A Comprehensive Review. *Review of educational research*, *87*(6), 1082-1116. https://doi.org/10.3102/0034654317726529

Hakim, L. M. (2018). Batik Sebagai Warisan Budaya Bangsa dan Nation Brand Indonesia. *Nation State: Journal of International Studies (NSJIS)*, *1*(1), 60–89.

Hikmat, H. (2022). The Readiness of Education in Indonesia in Facing The Society Era 5.0. *Jurnal Basicedu*, *6*(2), 2953–2961. https://doi.org/10.31004/basicedu.v6i2.2526

Idrus, L. (2019). Evaluasi dalam Proses Pembelajaran. *Adaara: Jurnal Manajemen Pendidikan Islam*, *9*(2), 920-935.

Iwan, I., Sumitro, S. B., Ibrohim, I., & Rohman, F. (2024). The Effect Of PBL Integrated With Igya Ser Hanjob's Local Wisdom To Improve Students Solving Skills In Conservation Biology Courses. Jurnal Eduscience, 11(1), 20-32.

Jannah, D. R. N., & Atmojo, I. R. W. (2022). Media Digital dalam Memberdayakan Kemampuan Berpikir Kritis Abad 21 pada Pembelajaran IPA di Sekolah Dasar. *Jurnal Basicedu*, *6*(1), 1064–1074. https://doi.org/10.31004/basicedu.v6i1.2124

Laila Laila, Alawiyah Nabila, & Eka Widyanti. (2024). Konsep Dasar Evaluasi Pembelajaran. *Jurnal Manajemen Dan Pendidikan Agama Islam*, *2*(5), 252–262. https://doi.org/10.61132/jmpai.v2i5.536

Lestari, F. I., & WH, E. H. (2021, November). Efektifitas Penggunaan Pewarna Alami Batik pada Pembelajaran Materi Angiospermae Bermuatan Sustainability terhadap Hasil Belajar Kognitif Siswa di SMP. In *Seminar Nasional Sains & Entrepreneurship* (Vol. 1, No. 1).

Muntazhimah, M., Putri, S., & Khusna, H. (2020). Rasch Model untuk Memvalidasi Instrumen Resiliensi Matematis Mahasiswa Calon Guru Matematika. *JKPM (Jurnal Kajian Pendidikan Matematika)*, *6*(1), 65-74.

Mustafidah, H., & Harjono, H. (2019). Implementasi Program QUEST untuk Menganalisis Butir Soal bagi Guru-Guru SMP Muhammadiyah 2 Karanglewas. *JPPM (Jurnal Pengabdian Dan Pemberdayaan Masyarakat)*, *3*, 321. https://doi.org/10.30595/jppm.v3i2.3378

Naryatmojo, D. L. (2018, December). Penggunaan Taksonomi Bloom dalam Pembelajaran Keterampilan Menyimak Bermuatan Pendidikan Karakter Profetik Untuk Mengukur Keberhasilan Hasil Belajar Mahasiswa. In *Proceeding Seminar Nasional Pertemuan Ilmiah Bahasa Dan Sastra Indonesia*(Vol. 40, No. 601, pp. 601-620).

Nufus, H., & Kusaeri, A. (2020). Analisis Tingkat Berpikir Kritis Siswa dalam Memecahkan Masalah Geometri. *JPMI (Jurnal Pendidikan Matematika Indonesia)*, *5*(2), 49-55.

Oktaviyanti, R., & Fadly, W. (2023). Analisis Kemampuan Berpikir Kritis Peserta Didik pada Materi Klasifikasi Makhluk Hidup dan Benda Tak Hidup. *Faktor: Jurnal Ilmiah Kependidikan*, *10*(1), 77-88.

Pratama, D. (2020). Analisis Kualitas Tes Buatan Guru Melalui Pendekatan Item Response Theory (IRT) Model Rasch. *Tarbawy : Jurnal Pendidikan Islam*, *7*(1), 61–70. https://doi.org/10.32923/tarbawy.v7i1.1187

Puspasari, A., Susilowati, I., Kurniawati, L., Utami, R. R., Gunawan, I., & Sayekti, I. C. (2019). Implementasi Etnosains dalam Pembelajaran IPA di SD Muhammadiyah Alam Surya Mentari Surakarta. *SEJ (Science Education Journal)*, *3*(1), 25–31. https://doi.org/10.21070/sej.v3i1.2426

Radja, A. P., Bano, V. O., & Ina, A. T. (2023). Analisis Kualitas Butir Soal Hasil Belajar Peserta Didik Berdasarkan Tingkat Kesukaran, Daya Beda dan Efektivitas Pengecoh di SMAN 1 Pandawai. *Jurnal Edusavana*, *1*(1), 30-41.

Sawaluddin, S., & Muhammad, S. (2020). Langkah-langkah dan teknik evaluasi hasil belajar Pendidikan Agama Islam. Jurnal PTK Dan Pendidikan, 6(1).

Sa'diyah, H., Alfiyah, H. Y., Ar, Z. T., & Nasaruddin, N. (2020). Model Research and Development dalam Pembelajaran Pendidikan Agama Islam. *EL-BANAT: Jurnal Pemikiran dan Pendidikan Islam*, *10*(1), 42-73.

Setyawarno, D. (2017). Panduan Penggunaan Program Quest untuk Analisis Butir Soal Hasil Belajar Bahasa Model Konvergen dan Divergen. Yogyakarta: Universitas Negeri Yogyakarta.

Siswanto, J. (2022). Pengembangan E-Instrumen Keterampilan Berpikir Kritis Berbasis Kearifan Lokal untuk Siswa Kelas VI SD. *Jurnal Kualita Pendidikan*, *3*(2), 2774–2156. Retrieved from https://bit.ly/PTS_MTK6_SDNSempu

Soleh, R., Susyla, D., & Syofiana, M. (2019). Developing Students Critical Thinking Ability Through Lesson Study. *Journal of Physics: Conference Series*, *1320*, 12005. https://doi.org/10.1088/1742-6596/1320/1/012005

Suarjana, I. M., Lasmawan, I. W., & Gunamantha, I. M. (2020). Pengembangan Instrumen Kemampuan Berpikir Kritis dan Sikap Peduli Lingkungan Tema 8 Peserta Didik Kelas IV Sd, *4*(2).

Sugiyono. (2019). *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. Bandung: Alfabeta.

Sukma, G. (2020). Application of Problem-Based Learning (PBL) Learning Model in Improving Student Learning Outcomes in Natural Sciences Subjects of Material Changes in Objects in Class II MI Al-Islah Sidoarjo. *Indonesian Journal of Science Learning (IJSL)*, *1*(1), 26–31.

Sulistiani, E., & Masrukan, M. (2017, February). Pentingnya Berpikir Kritis dalam Pembelajaran Matematika Untuk Menghadapi Tantangan MEA. In *PRISMA, Prosiding Seminar Nasional Matematika* (pp. 605-612).

Sumin, S., Retnawati, H., & Sayadi, W. (2023). Development and Validation of the Islamic Work Exemplary Scale in Indonesia. *Islamic Guidance and Counseling Journal*, *6*(2).

Syafitri, A., Anggraini, D. P., Parinduri, W. M., Rambe, T. R., & Rambe, N. (2023). Analisis Validitas Isi pada Instrumen Penilaian Akhir Semester Mata Pelajaran IPAS di SD. *School Education Journal Pgsd Fip Unimed*, *13*(4), 344. https://doi.org/10.24114/sejpgsd.v13i4.54309

Utami, P., Agustin, W., Wahyuni, S., Rayendra, & Bachtiar, W (2018). Pengembangan Modul Fisika Berbasis Potensi Lokal "Batik Lumbung dan Tahu Tamanan" untuk Siswa SMA di Kecamatan Tamanan Bondowoso (Materi Suhu Dan Kalor). *Jurnal Pembelajaran Fisika*, *7*(1), 62-69.

W. Ramansyah. (2012). *Strategi Pembelajaran*. Bangkalan: UTM Press.

Waruwu, L., Gulo, Y., Halawa, S., & Zalukhu, N. M. (2024). Analisis Mendalam terhadap Perubahan Keterampilan Berpikir Kritis Siswa melalui Kurikulum Merdeka. *Journal of Education Research*, *5*(3), 3783-3789.

Yulisharyasti, L., Nurdin, A., Aulia, N., Arfa, F. A. H., & Fadjryani. (2023). Analyzing The Quality Of Measurement Instruments Of Multiple Choice Questions On Class XI Economics Material In Public High School 3 Gorontalo Through Classical Test Theory And Rasch Models. *Parameter: Journal of Statistics*, *3*(1), 28–38. https://doi.org/10.22487/27765660.2023.v3.i1.16417

Yustiqvar, M., Gunawan, G., & Hadisaputra, S. (2019, December). Green Chemistry-Based Interactive Multimedia on Acid-Base Concept. In *Journal of Physics: Conference Series* (Vol. 1364, No. 1, p. 012006). IOP Publishing.

Zahra, A. A., & At-Taqiyyah, A. K. (2024). Mengukur kemampuan kognitif tingkat tinggi dengan pertanyaan pilihan ganda. *Jurnal Pelita Nusantara*, *1*(4), 462-470.

Zuniari, N. I., Wiyarsi, A., Wilujeng, I., & Rahmawati, L. (2025). Integration of Batik in Science Learning to Overcome Students' Low Critical Thinking Skills in Junior High School: Teacher's Perspective. *Formatif: Jurnal Ilmiah Pendidikan MIPA*, *15*(1).

Zuniari, N. I., Ridlo, Z. R., Wahyuni, S., Ulfa, E. M., & Dharmawan, M. K. S. (2022). The Effectiveness of Implementation Learning Media Based on Augmented Reality in Elementary School in Improving Critical Thinking Skills in Solar System Course. In *Journal of Physics: Conference Series* (Vol. 2392). Institute of Physics. https://doi.org/10.1088/1742-6596/2392/1/012010