JPPIPA 11(6) (2025)

Jurnal Penelitian Pendidikan IPA

http://jppipa.unram.ac.id/index.php/jppipa/index

Problem-Based Questions to Assess Students' Critical Thinking on the Topic of Substances and Their Changes

Dwi Septiana Sari[1*], Indri Nurwahidah[1], Yeni Widiyawati[1]

[1]Science Education Study Program, Faculty of Science and Technology, Universitas Ivet, Semarang, Indonesia.

Received: December 18, 2024
Revised: May 12, 2025
Accepted: June 25, 2025
Published: June 30, 2025

Corresponding Author:
Dwi Septiana Sari
saridwiseptiana@gmail.com

DOI: 10.29303/jppipa.v11i6.10084

Abstract: 21st-century education demands critical thinking skills, which still need to be improved among Indonesian students, especially at the junior high school level. Substance matter and its changes have the potential to train critical thinking through analysis and real problem-solving but are often taught without in-depth understanding. Therefore, the development of problem-based questions is needed to measure and improve students' critical thinking skills. This research aims to develop a problem-based question instrument to measure students' critical thinking skills on substance topics and their changes. This type of research is research and development that involves modifying the Borg & Gall model. The instrument developed is in the form of 30 multiple-choice questions to determine the achievement of students' critical thinking skills. The questions developed are adjusted to critical thinking skills strategies, namely ((1) focus on questions; (2) analyze arguments; (3) ask and answer challenge clarification questions; (4) consider the credibility of a source; (5) observing and considering the results of observations; (6) making deductions and considering the results of deductions; (7) making induction and considering the results of induction; (8) making and evaluating decisions; (9) identifying assumptions; and (10) determining an action. The validation of evaluation experts, subject matter experts, and junior high school science teachers showed that problem-based questions were valid and feasible to measure students' critical thinking skills toward substance material and its transformation.

Keywords: Problem-based questions; Critical thinking skills; R&D; Assessment.

Introduction

21st-century education emphasizes the mastery of life skills (Sarıgoz, 2022), encompassing both cognitive and interpersonal competencies, to prepare students to meet global challenges and workplace competition (Haug & Mork, 2021; Mayarni & Nopiyanti, 2021). Cognitive skills include the ability to think critically and creatively (Siburian et al., 2019), while interpersonal skills involve collaboration and effective communication (Haug & Mork, 2021). In science education, critical thinking skills are considered essential for students to understand scientific concepts, evaluate phenomena, and analyze causal relationships in natural events (Manassero-Mas & Vázquez-Alonso, 2022; Santos, 2017). Critical thinking is central to developing scientifically literate individuals who can solve problems systematically, draw logical conclusions from evidence, and connect scientific knowledge to real-life contexts (Paul & Elder, 2019; Paulsen & Dankert, 2022; Supena et al., 2021; Sari et al., 2023; Suwistika et al., 2024). It also enables students to evaluate information logically, interpret data rationally, and formulate well-reasoned judgments (Albar & Southcott, 2021; Umrzokova & Pardaeva, 2020). Therefore, developing critical thinking skills in science learning is necessary to prepare students for real-world challenges (Albar & Southcott, 2021).

Despite its importance, studies have shown that Indonesian students particularly at the junior high school level still possess relatively low levels of critical thinking (Dewi & Prasetyo, 2016; Nurwahidah et al., 2020; Suprapto, 2016). This condition highlights a significant gap between the vision of 21st-century learning and its actual implementation in the classroom.

How to Cite:
Sari, D. S., Nurwahidah, I., & Widiyawati, Y. (2025). Problem-Based Questions to Assess Students' Critical Thinking on the Topic of Substances and Their Changes. *Jurnal Penelitian Pendidikan IPA*, *11*(6), 1028–1040. https://doi.org/10.29303/jppipa.v11i6.10084

While critical thinking is a key competency needed for global competitiveness, students often struggle to analyze information, evaluate arguments, and derive logical conclusions (OECD, 2019). For example, the 2022 PISA survey reported that the average science score of Indonesian students was 383, approximately 102 points below the global average, indicating weaknesses in analysis, evaluation, and reasoning skills. This gap suggests that without targeted intervention, students will remain unprepared for the demands of the modern world.

This evidence underscores the urgent need for effective instructional strategies and assessment methods that explicitly cultivate higher-order thinking. One contributing factor to the underdevelopment of critical thinking is the lack of assessment instruments specifically designed to foster such skills. Most science test items in Indonesia still emphasize factual recall rather than engaging students in analytical or creative thought (Nurwahidah et al., 2020; Shanta & Wells, 2022). Consequently, students find it difficult to logically and systematically interpret scientific phenomena, which undermines the essence of science education. Addressing this requires assessment tools that are aligned with the nature of science and challenge students to engage in deeper reasoning processes.

The topic of "substances and their changes" within the science curriculum encompasses core concepts such as physical and chemical properties, types of changes in matter, and the law of conservation of mass. These concepts offer rich opportunities for students to explore cause-and-effect relationships and distinguish between types of changes by observing simple phenomena (Bowen, 2022). For instance, the melting of ice illustrates a physical change, while the burning of paper into ash exemplifies a chemical transformation. When approached through inquiry and reasoning, such phenomena help students build strong conceptual frameworks and practice critical thinking skills. However, in practice, many classrooms still deliver this topic through rote methods, resulting in only surface-level understanding. This instructional gap further underscores the need for instructional and evaluative strategies that promote real-world problem-solving and critical thinking practice.

Problem-Based Learning (PBL) is one pedagogical model deemed effective in enhancing critical thinking. Through PBL, students engage with authentic, contextualized problems that prompt analysis, evaluation, and scientific solution development (Kardoyo et al., 2020). This model aligns well with the Independent Curriculum, which emphasizes higher-order thinking and real-world applicability. Moreover, PBL supports collaborative and reflective learning, both of which are integral to nurturing students' interpersonal abilities (Loyens et al., 2023). In assessments, problem-based questions derived from PBL contexts can stimulate higher-order cognitive processes, such as analysis, evaluation, and synthesis thus functioning not only as evaluative tools but also as meaningful learning experiences (Tawfik et al., 2021). Thus, the integration of PBL-based questions into assessment can serve a dual purpose: measuring critical thinking and simultaneously fostering it during the learning process.

Therefore, this study aims to develop a set of Problem-based questions on the topic of "substances and their changes" as an assessment instrument to measure and enhance the critical thinking skills of junior high school students. These instruments are expected to promote deeper conceptual understanding, improve problem-solving abilities, and address the urgent need for relevant and effective assessment tools aligned with the vision of 21st-century education and the principles of the Independent Curriculum.

Method

This research is a Research and Development (R&D) using the Borg and Gall model. The study was conducted at SMP Negeri 1 Kaliwungu Kudus from July to September 2024, with the research subjects being grade VIII students. This research aims to develop problem-based questions on substance matter and its changes as an evaluation instrument to measure the critical thinking ability of junior high school students. The research procedure follows the Borg and Gall model which due to time and resource considerations, is implemented only up to the seventh of ten stages: (1) preliminary study, (2) planning, (3) initial product development, (4) product validation, (5) limited trials, (6) revision of the main product, and (7) limited product dissemination (Borg & Gall, 1983).

Preliminary study

The first stage is the preliminary study, consisting of field and literature studies. Field studies were conducted to identify the needs of teachers and students related to assessment instruments that can measure critical thinking skills in differentiated science learning. Based on the results of field studies, it is known that teachers need evaluation instruments that not only measure conceptual understanding but also train critical thinking skills through a problem-based approach. Meanwhile, substances and their changes were chosen because they are relevant to everyday life, such as the phenomenon of physical and chemical change. Literature studies were carried out by reviewing theories, research, and references related to critical thinking skills, problem-based learning (PBL), and substance material and its changes. The results of this study were the basis for compiling a question grid.

Planning

The second stage is planning, where the learning outcomes of science in grade VIII are mapped based on the Independent Curriculum and adjusted to the critical thinking skills strategy from Nitko & Brookhart (2011). This strategy includes (1) focusing on the question, (2) analyzing the argument, (3) asking and answering challenge clarification questions, (4) considering the credibility of a source, (5) observing and considering the results of observations, (6) making deductions and considering the results of the deduction, (7) making inductions and considering the results of the induction, (8) making and assessing decisions, (9) identifying assumptions, and (10) determining an action. This step is continued with preparing question indicators and developing a grid that refers to the strategy. The details of the question numbers based on critical thinking ability indicators are presented in Table 1. After the planning stage, it proceeds to the initial product development stage.

Initial product development

In the initial product development stage, researchers compiled 30 problem-based multiple-choice questions, including substance matter and its changes. The questions are designed to be contextual, relevant to students' lives, and able to measure critical thinking skills. In addition, product validation instruments were developed to assess the quality of questions regarding material/substance, construction, and language.

Product validation

The next stage is product validation, involving material experts, evaluation experts, and junior high school science teachers. The material expert assesses the suitability of the questions with the basic competencies in the Independent Curriculum. Meanwhile, the evaluation expert assesses whether the questions are able to measure critical thinking skills appropriately. Science teachers provide input related to the practicality of using questions in classroom learning. Feedback and suggestions from validators are then used to revise the product to ensure question quality and alignment with evaluation objectives.

Limited trials

The fifth stage in this study is a limited trial involving 27 students of class VIII A at SMP N 1 Kaliwungu Kudus. This stage aims to empirically validate the developed problem-based questions related to substances and the changes that are developed. Furthermore, the analysis of question items was carried out using Winstep software with a Rasch measurement modeling approach to obtain more in-depth and accurate data related to the quality of the questions. Students are also asked to provide responses, criticisms, and suggestions related to the quality of problem-based questions that have been developed. The results of this trial are used to identify the weaknesses of the questions and as a basis for future product revisions.

Revision of the main product

The sixth stage is revising the main product, where researchers analyze data from limited trial results and input from validators and students. This revision aims to improve problem-based question instruments to meet the question quality criteria regarding material/substance, construction, and language. This product revision produces a problem-based question instrument ready to be used to evaluate students' critical thinking skills.

Limited product dissemination

The seventh stage is limited dissemination. The final product is disseminated to science teachers in several selected schools at this stage. Teachers are given guidance on using problem-based question instruments in the learning evaluation process. This stage aims to introduce the products that have been developed to potential users and get additional feedback for further improvements.

Data collection and analysis

The research data consists of qualitative and quantitative data. Qualitative data include responses, criticisms, and suggestions from material experts, evaluation experts, science teachers, and students related to the quality of the questions developed. Students' responses were categorized into two categories, namely "Good" and "No." Quantitative data were obtained from the validator's assessment of the questions' content, presentation, and language aspects. These assessments are tabulated, averaged, and converted into interval data on a scale of four to facilitate analysis. The reference for score changes based on assessment indicators is presented in Table 2.

Data analysis was carried out using qualitative and quantitative techniques. Qualitative data were analyzed through tabulation, simplification, and formulation based on input from various parties to improve the question instruments. Meanwhile, quantitative data was analyzed using descriptive statistics by calculating the average validation score for each assessment indicator. The results of this analysis are used to evaluate the quality of the product and determine the necessary revisions before the product is disseminated. This combination of qualitative and quantitative analysis ensures that the problem-based question instruments developed are valid, practical, and feasible for assessing students' critical thinking skills in a contextualized manner.

To enrich data analysis and improve the comprehensiveness of the research findings, several additional analyses are recommended. First, a Differential Item Functioning (DIF) analysis can be conducted to examine potential bias in test items across various subgroups, such as gender or different levels of academic achievement. This helps ensure that the developed instrument functions equitably for all students. Second, interpreting the Wright Map can provide a visual representation of the alignment between item difficulty and student ability, which is essential in evaluating whether the test items are appropriately distributed across the ability spectrum. Third, a Test Information Function (TIF) analysis can be employed to determine the precision and effectiveness of the instrument in measuring student abilities at different levels. In addition, a thematic analysis of qualitative feedback from students and teachers can be used to identify common issues encountered during the test, such as linguistic ambiguity, confusing distractors, or misalignment with the intended content. Lastly, to ensure consistency among validators, the assessment criteria were standardized and clearly defined to minimize subjective differences in judgment.

Table 1. Indicators and distribution of problem-based questions on the topic of substances and their changes

Category	Strategy	Description	Number Question
Giving a simple explanation	Focus on questions	a) Identify or formulate questions	22
		b) Formulate or select the right criteria to evaluate the material presented	27
		c) Stay focused on the problem and its context	1
	Analyze arguments	a) Identify conclusions	10
		b) Identify both explicit and implicit reasons behind the argument	3
		c) Identify similarities and differences between two or more arguments	4
		d) Identify and distinguish relevant from irrelevant information in an argument	7
		e) Explain the logical structure of an argument	23
		f) Summarize the arguments.	6, 18
	Asking and answering challenge clarification questions	a) Provide a simple explanation	5
		b) Provide relevant examples	24
Assess basic information support	Consider the credibility of the source	Assessing the credibility of a source is taking into account a) The expertise of the person providing evidence b) The clarity and appeal of the information presented c) The relevance and consistency of information from multiple sources d) The reputation of the source of information e) Potential biases due to the source's affiliations or interests f) The adaptability of the information across different contexts	25
	Observing and considering the results of the observations	a) Involves some degree of inference or assumption	19
		b) Considers the time interval between observation and reporting	12
		c) Report the results of the observations	20
		d) recording the results of observations,	21
		e) Use appropriate and accurate evidence	29
		f) using good access,	15
		g) using technology	2
		h) accountability for the results of observations	26
Drawing conclusions	Making deductions and evaluating the results of deductions	Applying logical reasoning when analyzing reports and drawing conclusions	11

Category	Strategy	Description	Number Question
	Formulating inductions and assessing the outcomes	a) Formulate a hypothesis	30
		b) Design experiments	9
		c) Draw conclusions based on factual evidence	13
		d) Draw conclusions based on experimental results	14
	Making and evaluating decisions based on evidence and values	Determine whether a conclusion is based on values, and when those values are applicable in decision-making	16
Doing more explanation	Identifying underlying assumptions	Identify assumptions that influence reasoning or belief and construct arguments accordingly	8
Applying strategies and tactics in solving problems	Defining and implementing appropriate actions	Define problems, generate and assess alternative solutions, approach the problem holistically, take action, and evaluate the outcomes	17

Table 2. Results of conversion of problem-based question quality assessment scores

Aspect	Score interval	Category
Material/ Substance	$13 \leq \overline{M} \leq 16$	Excellent
	$10 \leq \overline{M} < 13$	Good
	$7 \leq \overline{M} < 10$	Enough
	$4 \leq \overline{M} < 7$	Less
Construction	$16,35 \leq \overline{M} \leq 20$	Excellent
	$12,5 \leq \overline{M} < 16,25$	Good
	$8,75 \leq \overline{M} < 12,5$	Enough
	$5 \leq \overline{M} < 8,75$	Less
Language	$16,35 \leq \overline{M} \leq 20$	Excellent
	$12,5 \leq \overline{M} < 16,25$	Good
	$8,75 \leq \overline{M} < 12,5$	Enough
	$5 \leq \overline{M} < 8,75$	Less

Result and Discussion

This study aims to develop problem-based questions on substances and their transformation as an instrument to measure the critical thinking ability of junior high school students. The development procedure follows the Borg & Gall (1983) model, which is modified into seven stages, namely: (1) preliminary study, (2) planning, (3) initial product development, (4) product validation, (5) limited trial, (6) product revision based on limited trial results, and (7) limited dissemination.

Based on the preliminary study results, information was obtained that the Independent Curriculum emphasizes the development of students' critical thinking skills. Nonetheless, its implementation still requires a more rigorous and structured evaluation mechanism to effectively cultivate these competencies. One of the topics that can potentially develop students' critical thinking skills is the topic of substances and their changes. This topic includes the definition of substances, types of substances (elements, compounds, mixtures), properties of substances (physical and chemistry), and changes in substances (changes in physical and chemistry).

The observation results show that most of the questions contained in the textbook tend to focus on the level of basic abilities, such as remembering (C1), understanding (C2), and applying (C3). These questions are essential to test students' basic knowledge but are not sufficient to train critical thinking skills, including analysis, evaluation, and problem-solving. Therefore, the development of problem-based questions is needed to complement the existing evaluation instruments.

In addition, field studies reveal that students often have difficulty understanding the relevance of the topic of substances and their changes to daily life. Problem-based questions are a relevant alternative to encourage students to relate the concepts they learn to actual phenomena, such as evaporation, combustion, or mixing simple materials. Thus, learning will become more meaningful, in alignment with the fundamental principles promoted by the Independent Curriculum.

Literature studies were conducted to collect information underlying the development of assessment instruments. The study was conducted on competencies in the junior high school science curriculum, substance materials, their changes, critical thinking skills, and problem-based learning approaches (PBL). Based on the learning outcomes in the Independent Curriculum for phase D (grade VII junior high school), the material and its changes have great potential to train students' critical thinking skills. The learning outcomes emphasized an understanding of the process of identifying properties and characteristics of substances, physical and chemical changes, and simple separation of mixtures. These learning experiences are designed to enable students to engage in identifying the properties and characteristics of substances, differentiate between physical and chemical changes, and perform basic separation of mixtures. Meanwhile, the material discussed on the topic of substances and their changes is the definition of substances, types of substances (elements, compounds, mixtures), properties of substances (physical and chemistry), and changes in substances (physical and chemical changes).

This material is closely related to everyday phenomena that students often experience, such as melting ice, boiling water, or burning paper. Through this topic, students can be trained to identify, analyze, and evaluate information based on real-life situations. Literature reviews also show that critical thinking skills include indicators such as focusing on questions, analyzing arguments, considering the credibility of sources, making deductions and inductions, and assessing decisions (Nitko & Brookhart, 2011). These indicators serve as the foundation for constructing problem-based assessment instruments that target higher-order thinking skills.

The planning stage begins with the preparation of the initial design of the instrument through needs analysis, goal formulation, and framework development. This design is prepared based on preliminary and literature studies' results. This study uses the critical thinking skills strategy from Nitko & Brookhart (2011) as a reference. This strategy consists of (1) Focus on questions, (2) Analyze arguments, (3) Ask questions and answer clarifying challenges, (4) Consider the credibility of sources, (5) Observe and consider observation results, (6) Make deductions and inductions, (7) Assess decisions, (8) Identify assumptions, (9) Determine an action.

The next step in planning is to prepare a grid of questions that refer to indicators of critical thinking and learning outcomes in the Independent Curriculum. This grid contains topic components, indicators of critical thinking skills, question types, and forms of problem-based stimulus. The preparation of this grid aims to ensure that the instruments developed are systematic, measurable, and aligned with curriculum objectives and cognitive skill development.

Based on the prepared question grid, the researcher developed 30 problem-based multiple-choice questions. The trial was carried out on 27 SMP Negeri 1 Kaliwungu grade VIII students. The problem-based questions tested consisted of 30 items with dichotomous scores (1 for correct answers, 0 for incorrect answers). Each question is accompanied by a stimulus in the form of a description of contextual phenomena relevant to student's daily lives, such as changes in the form of water, metal rust, or the process of separating mixtures. These questions are designed to train predetermined indicators of critical thinking skills. In addition, validation instruments are also prepared to assess the quality of questions from the aspects of material/substance, construction, and language. The development of this question involves using contextual stimuli in the form of texts, tables, graphs, or science phenomena that require students to analyze information, evaluate data, and draw logical conclusions. In the material aspect, the questions are designed based on the learning objectives determined based on the learning outcomes of science subjects in phase D and refer to critical thinking strategies. The question sentences are written firmly and unambiguously in terms of construction, with explicit instructions so students can understand what is being asked. The stimulus used is varied and relevant to the context of daily life to encourage students to think deeply. The language aspect is also a significant concern, where question sentences and answer choices are written using standard, communicative, and in accordance with the level of student development. An example of the question display is presented in Table 3.

Table 3. Example of the problem based question

Question number	Indicator	Question
4	Seeing the Equation and the difference between two or more arguments;	4. Pay attention to the following statements! (1) The weight of water in one bottle is lighter than the weight of water in one large bucket. (2) The weight of water in one bottle is the same as the weight of water in one large bucket. (3) The density of water in one bottle is less than the density of water in one large bucket. (4) The density of water in one bottle is the same as the density of water in one large bucket. Numbers indicate the correct statements... a. (1) and (3) b. (1) and (4) c. (2) and (3) d. (2) and (4)

Question number	Indicator	Question
9	Design experiments	9. A student wants to make his ice cream at home. The ingredients he has are fresh milk, sugar, and salt. He also has several containers and a cooler. According to Changes in the State of Matter, the following are the most appropriate steps for making ice cream. (a) Mix all the ingredients, put them in a container, and then place them in the freezer. (b) Chill the milk first, then mix it with the sugar and salt. Place the mixture in an airtight container, then in a container containing ice cubes and salt. Stir regularly. (c) Heat the milk first, then mix it with the sugar and salt. Chill the mixture before placing it in the freezer. (d) Combine the milk, sugar, and salt, then beat until frothy. Place the mixture in an airtight container and place it in the freezer.

The product validation stage includes one material expert lecturer (ME), one evaluation expert lecturer (EE), and two junior high school science teachers (T1 & T2). The material expert assesses the suitability of the questions to the learning outcomes and objectives. In contrast, the evaluation expert analyzes the extent to which the questions can measure critical thinking skills appropriately. This includes examining alignment with critical thinking indicators such as analysis, evaluation, and inference. Science teachers provide input on the practicality of using instruments in the classroom. Their feedback covers the ease of implementation, clarity of instructions, and time efficiency. The quality of the problem-based question products developed in this study was evaluated in three aspects, namely the material/substance aspect, construction, and language. These aspects ensure that the content is conceptually accurate, structurally coherent, and linguistically accessible to the students. A comparison of the average scores of each aspect of the assessment obtained from validators and junior high school science teachers is presented in Table 4. This comparison is essential to identify any discrepancies in perception between expert validators and field practitioners regarding the quality and applicability of the developed instruments.

Table 4. Average scores of each aspect of the assessment

Aspects	ME	EE	T1	T2
Material/ substance	13	12	12	12
Construction	15	14	14	14
Language	15	15	16	15

The limited trial phase was carried out on 27 grade VIII A SMP N 1 Kaliwungu Kudus students to obtain empirical validation data on problem-based questions on substance topics and their changes developed. Data was analyzed using *Winstep software* with a Rasch measurement modeling approach. The statistical summary table of the output of the analysis of question items is presented in Figure 1.

```
     SUMMARY OF 27 MEASURED Person
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT         OUTFIT    |
|          SCORE      COUNT     MEASURE   S.E.      MNSQ   ZSTD   MNSQ   ZSTD |
|------------------------------------------------------------------------------|
| MEAN      15.4       25.0         .43     .52      1.00   -.08   1.02    .04 |
|  SEM        .6         .0         .15     .01       .05    .21    .11    .19 |
| P.SD       2.9         .0         .78     .03       .27   1.08    .57    .95 |
| S.SD       3.0         .0         .79     .03       .27   1.10    .58    .97 |
| MAX.      21.0       25.0        2.04     .61      1.73   2.22   2.86   2.71 |
| MIN.       9.0       25.0       -1.31     .50       .62  -2.04    .50  -1.12 |
|------------------------------------------------------------------------------|
| REAL RMSE    .55 TRUE SD     .55  SEPARATION  1.00  Person RELIABILITY  .50 |
|MODEL RMSE    .52 TRUE SD     .58  SEPARATION  1.11  Person RELIABILITY  .55 |
| S.E. OF Person MEAN = .15                                                    |
-------------------------------------------------------------------------------
Person RAW SCORE-TO-MEASURE CORRELATION = 1.00
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .53  SEM = 1.99
STANDARDIZED (50 ITEM) RELIABILITY = .71

     SUMMARY OF 22 MEASURED Item
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT         OUTFIT    |
|          SCORE      COUNT     MEASURE   S.E.      MNSQ   ZSTD   MNSQ   ZSTD |
|------------------------------------------------------------------------------|
| MEAN      15.2       27.0         .00     .51       .98    .04   1.02    .18 |
|  SEM       1.4         .0         .32     .04       .03    .12    .10    .18 |
| P.SD       6.5         .0        1.46     .17       .13    .56    .47    .82 |
| S.SD       6.6         .0        1.50     .18       .13    .58    .48    .84 |
| MAX.      26.0       27.0        2.38    1.03      1.17   1.07   2.75   2.42 |
| MIN.       4.0       27.0       -3.11     .41       .68   -.93    .34  -1.11 |
|------------------------------------------------------------------------------|
| REAL RMSE    .55 TRUE SD    1.35  SEPARATION  2.46  Item   RELIABILITY  .86 |
|MODEL RMSE    .54 TRUE SD    1.36  SEPARATION  2.50  Item   RELIABILITY  .86 |
| S.E. OF Item MEAN = .32                                                      |
-------------------------------------------------------------------------------
  MAXIMUM EXTREME SCORE:      3 Item 12.0%
```

Figure 1. Summary Statistic

Based on the Summary Statistic, a *Person measure score* of 0.43 logit was obtained, indicating that the average ability of students was relatively lower than the difficulty level of the questions. In Rasch modeling, logit scores close to 0 or negative indicate students have difficulty answering questions. This means that the whole item tends to be difficult for students who are trial participants. This can indicate that the difficulty level of the questions needs to be balanced with the student's ability. The value of Person *reliability* = 0.50 (weak) shows that the consistency of students' ability to answer questions is relatively weak. Person reliability measures the extent to which students consistently respond to questions. This score indicates that there is a reasonably high variation in students' abilities, or there may be some students who answer questions inconsistently. This score also shows that the question instrument needs to map students' abilities well fully.

The value *of the reliability item* obtained of 0.86 is included in the good category. The reliability item shows the extent to which the question consistently measures student ability. This value indicates that the quality of the questions is quite good, with an adequate and reliable distribution of the difficulty level of the questions for further measurement. Meanwhile, *the Alpha Cronbach* value obtained at 0.53 shows that the overall reliability of the instrument is still low. Cronbach's Alpha measures the internal consistency of all questions. This value indicates that there is a weakness in the connection between questions. The factors contributing to this low score may be that some questions need to match the student's ability or be too complicated, which causes students to answer randomly or inconsistently.

The results of the item measure presented in Figure 2 show the question's difficulty level and the question item's feasibility. In contrast, the category of the question's difficulty level is presented in Table 5.

Item STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASURE-AL CORR.	PTMEASURE-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
4	4	27	2.38	.56	1.16	.54	2.75	2.42	-.22	.25	85.2	85.2	S4
9	5	27	2.09	.52	.77	-.72	.68	-.66	.54	.27	81.5	81.5	S9
18	5	27	2.09	.52	.97	.00	1.00	.14	.28	.27	81.5	81.5	S18
20	9	27	1.22	.43	.97	-.15	1.07	.34	.33	.32	70.4	69.8	S20
2	10	27	1.04	.42	1.16	1.07	1.15	.72	.14	.33	63.0	67.3	S2
22	10	27	1.04	.42	1.08	.58	1.17	.80	.21	.33	63.0	67.3	S22
24	11	27	.87	.42	.92	-.59	.87	-.64	.44	.33	66.7	65.4	S24
25	12	27	.70	.41	1.01	.13	1.03	.22	.32	.34	63.0	64.4	S25
1	13	27	.53	.41	1.02	.22	1.07	.47	.30	.34	59.3	63.6	S1
3	14	27	.36	.41	.95	-.41	.89	-.66	.42	.34	63.0	63.2	S3
19	15	27	.19	.41	.96	-.25	1.01	.13	.37	.34	66.7	64.1	S19
8	16	27	.02	.42	1.13	.91	1.09	.52	.20	.34	51.9	65.7	S8
13	17	27	-.16	.42	.96	-.23	1.18	.90	.33	.34	77.8	68.3	S13
15	17	27	-.16	.42	1.04	.31	1.10	.54	.27	.34	70.4	68.3	S15
21	17	27	-.16	.42	.85	-.93	.78	-1.06	.53	.34	70.4	68.3	S21
23	19	27	-.53	.45	.85	-.69	.74	-.94	.53	.33	74.1	73.4	S23
6	21	27	-.97	.49	1.15	.63	1.40	1.07	.08	.31	74.1	78.5	S6
17	21	27	-.97	.49	1.17	.71	1.30	.87	.09	.31	66.7	78.5	S17
16	23	27	-1.51	.56	.68	-.85	.46	-1.11	.67	.27	85.2	85.1	S16
7	24	27	-1.87	.63	.98	.11	1.07	.33	.25	.25	88.9	88.8	S7
5	26	27	-3.11	1.03	.90	.19	.34	-.28	.36	.15	96.3	96.3	S5
14	26	27	-3.11	1.03	.90	.19	.34	-.28	.36	.15	96.3	96.3	S14
10	27	27	-4.35	1.83	MINIMUM MEASURE				.00	.00	100.0	100.0	S10
11	27	27	-4.35	1.83	MINIMUM MEASURE				.00	.00	100.0	100.0	S11
12	27	27	-4.35	1.83	MINIMUM MEASURE				.00	.00	100.0	100.0	S12
MEAN	16.6	27.0	-.52	.67	.98	.04	1.02	.18			73.4	74.6	
P.SD	7.2	.0	1.97	.46	.13	.56	.47	.82			11.6	10.4	

Figure 2. Item Measure

Table 5. Problem-based question difficulty level developed

Item Number	Nilai Measure	Category
5, 10, 14, 11, 12	> - 2.00 logit	It's very easy
6, 7, 13, 15, 16, 17, 21, 23	- 2.00 logit to 0	Easy
1, 2, 3, 8, 19, 20, 22, 24, 25	- 2.00 logit to 0	Difficult
4, 9, 18	> +2.00 logit	Very difficult

Based on *the* item measure table, item number 4 has a value of +2.38 logit, indicating the most challenging question item, and item numbers 10, 11, and 12, with a measured value of -4.35 logit, is the most straightforward question. Of the 30 questions developed, five were in the very easy category, 8 in the easy category, 9 in the difficult category, and 3 in the very difficult category. Meanwhile, to determine the feasibility of an instrument item, it can be seen from the data of Infit and Outfit Mean Square (MNSQ) values, Outfit Z-Standard (ZSTD) values, and Point Measure Correlation (Pt Mean Corr) values. The criteria for determining the suitability (fit or not) of an item are (a) The accepted MNSQ value is 0.5 < the MNSQ < 1.5; (b) The accepted ZSTD value is -2.0 < ZSTD < +2.0; and (c) The accepted Pt Mean Corr value is 0.4 < Pt Measure Corr < 0.85 (Sumintono & Widhiarso, 2013). Most questions have INFIT MNSQ and OUTFIT MNSQ values in the ideal range. However, several questions have extreme scores, such as question number 4 with an OUTFIT MNSQ score of 2.75 and ZSTD 2.42 (misfit). The possible cause is that the question needs to be clarified). In addition, for number 21 with INFIT, MNSQ is 1.78, and ZSTD is 1.06 (not fit). This may be because the question items do not match the student's ability. Questions with an MNSQ close to 1.00 show a good fit, for example, questions 18 and 24.

For Pt Mean Corr data, all questions have positive scores, ranging between 0.08 to 0.67. The highest score is question number 16 has a Pt Mean Corr value of 0.67, showing that this question is very good at measuring students' abilities. Meanwhile, question number 4 has a Pt Mean Corr value of -0.22, which means it has a low correlation value, which indicates that this question may need to be revised. The results of this analysis are mostly in accordance with the Rasch measurement model and are suitable for evaluating students' critical thinking skills. However, two question items (number 4 and 21) need to be corrected or revised so that the instrument has better validity and reliability. The difficulty level of the questions varies, ranging from easy moderate, to complex, looks evenly, and is drawn through *the Wright* map or *Person-Item Map* on Figure 3. This variation in difficulty level is essential to ensure that the instrument can measure students' abilities comprehensively and fairly.

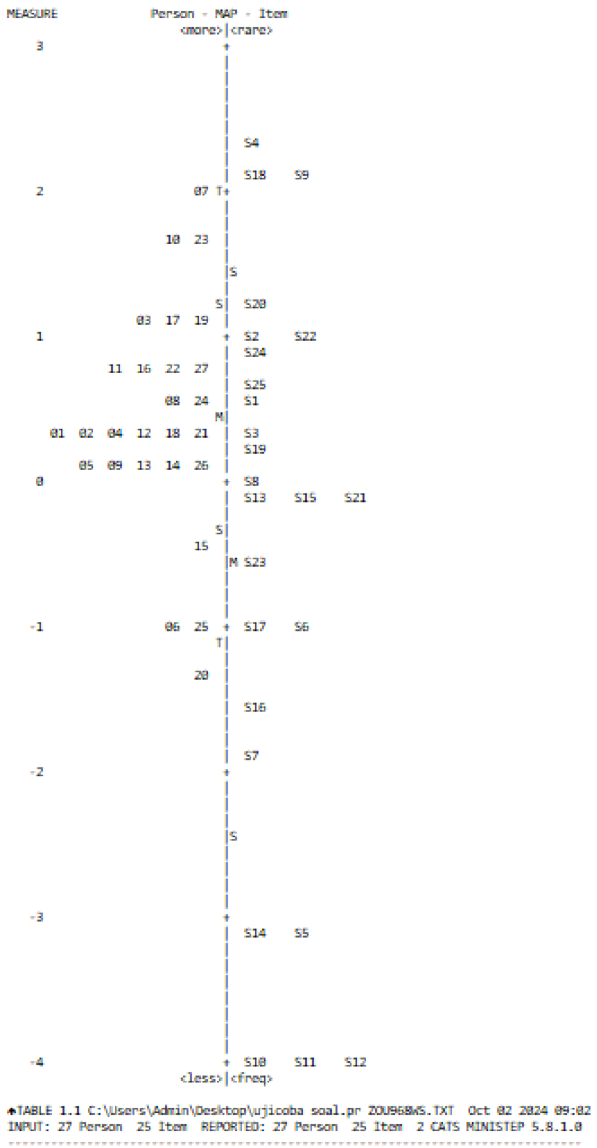

Figure 3. Person-Map-Item output

Based on the output of Person-Map-Item**,** the distribution of student's ability levels and difficulty levels of the questions is displayed on the same logit scale. The students with the highest ability are in logit **+3**, while those with the lowest ability are in **logit -3**, with the majority of students concentrated in the **0 to +1** logit range. This shows that most students have moderate ability, although some students have very high and very low ability.

On the other hand, the difficulty level of the questions varies from the most difficult to the easiest. Question number 4 has a value of +3 logits, making it the most difficult question, and can only be answered by students with the highest ability. In contrast, questions 10, 11, and 12 have a value of -4 logits, which shows that these questions are very easy and can be done by almost any student. Most questions are spread in the logical range of -1 to +1, which is entirely on the ability of most students.

The imbalance was seen in questions that were too difficult **(**number 4) and questions that were too easy **(**numbers 10, 11 and 12**)**. These questions need to be reviewed because they could be more effective in distinguishing students' abilities. Questions that are too difficult may be caused by excessive ambiguity or complexity, while questions that are too easy do not provide significant information in the evaluation. However, most questions around logits 0 to +1 are good enough to distinguish students with intermediate abilities. Overall, the instrument has a fairly good distribution of difficulty, although some problems need to be improved to create a more optimal balance.

The revision stage is carried out based on the results of limited trials and input from validators. Examples of product revisions are presented in Table 6. The revised final product is disseminated in limited dissemination to science teachers in several schools. Teachers are given guidance regarding using problem-based question instruments as an evaluation tool in science learning. This dissemination aims to introduce the product and get additional feedback for further refinement.

Table 6. Example of product revision of problem-based question instruments

Question Number	Commentary	Revision Results
4	Question sentences should be corrected so that they are easier for students to understand	Correcting sentence redaction 4. Pay attention to the following statements! (1) The weight of water in one bottle is lighter than the weight of water in one large bucket. (2) The weight of water in one bottle is the same as the weight of water in one large bucket. (3) The density of water in one bottle is less than the density of water in one large bucket. (4) The density of water in one bottle is the same as the density of water in one large bucket. Numbers indicate the correct statements... a. (1) and (3) b. (1) and (4) c. (2) and (3) d. (2) and (4)

Problem-based questions can improve students' critical thinking skills, especially in the material "Substances and Their Changes," by relating science concepts to daily life. The questions developed often focus on events students usually experience or do, such as making ice cream, boiling water, or sublimating camphor. Presenting relevant and contextual problems encourages students to actively think, analyze, and solve issues related to physical and chemical phenomena (Abrami et al., 2015). The questions are usually accompanied by initial information—such as reading passages, images, or graphics—linked to the topic being studied, prompting students to process data and connect it to prior knowledge.

Problem-Based Learning (PBL) is a constructivist approach that engages students in authentic problem-solving and fosters deeper conceptual understanding (Yu & Zin, 2023). In junior high school science, PBL can make abstract topics such as "Substances and Their Changes" more meaningful by embedding them in real-world contexts, including environmental issues or everyday science phenomena (Smith et al., 2022). This contextualization increases student interest and motivation, an important factor given the decline in engagement when instruction relies solely on traditional methods (Reyes, 2025). Well-designed PBL activities enhance critical thinking as students analyze information, evaluate options, and apply concepts to solve non-routine problems (Yu & Zin, 2023). Within a PBL framework, the teacher serves as a facilitator or expert learner (Prasad & O'Malley, 2022), guiding students to build understanding through inquiry and reasoning, thereby fostering scientific literacy (Miterianifa et al., 2021).

This approach requires students to go beyond memorization, applying critical thinking skills to answer questions (Sari et al., 2019). They must observe, gather data, and draw conclusions grounded in relevant scientific concepts (Gürses et al., 2022). Over time, this process develops their understanding and sharpens critical thinking skills essential for scientific problem-solving (Jiménez Pérez et al., 2025). Asyari et al. (2016) found that PBL improves critical thinking because it involves students in problem-solving that requires analysis and evaluation. Similarly, Santos-Meneses et al. (2023) note that problem-based questions push students to think deeply and critically, connecting knowledge to real-life contexts rather than relying solely on memorization (Hanzlová & Kudrnáč, 2024).

Designing effective written problem-based questions requires adherence to several best practices. First, questions should be rooted in contexts that are familiar or meaningful to seventh graders, linking chemical concepts to their experiences or societal issues to boost relevance and engagement (Muhibbuddin et al., 2023). Second, they should be open-ended or ill-structured, allowing for multiple approaches or partial solutions and encouraging flexible thinking. Incorporating higher-order prompts, such as "why" and "how" instead of "what", guides students to explain their reasoning and consider alternatives. Questions should also align with curriculum goals (e.g., scientific inquiry) and progress from guided problems to independent inquiries as students' skills develop. Scaffolding through hints, diagrams, or preliminary questions can activate prior knowledge without revealing the answer. While specific guidelines vary, research consistently emphasizes including explicit critical thinking elements, such as evaluating claims or designing solutions, to make PBL more effective (Yu & Zin, 2023).

Assessment of student responses to written problem-based questions should evaluate not only the correctness of the final answer but also the quality of the reasoning process. Clear, criterion-based rubrics—such as those developed by Reynders et al. (2020) that define performance levels in analysis, use of evidence, and logical coherence, help teachers communicate expectations and provide consistent feedback. A rubric might allocate points for identifying relevant concepts, interpreting data accurately, explaining relationships between variables, and justifying conclusions with chemical reasoning. Although open-ended questions take longer to grade, they allow for partial credit and reveal students' thought processes. In larger classes, combining question types (e.g., higher-order multiple choice or short-answer alongside open problems) can still target critical thinking by including scenario-based items or requiring explanations. Formative applications, such as class discussions of model answers or peer reviews, can strengthen critical thinking before summative assessments.

In conclusion, problem-based questions on "Substances and Their Changes" not only deepen students' understanding of scientific concepts but also cultivate critical thinking skills. This approach fosters active, creative, and analytical engagement with scientific problems, enhancing both conceptual mastery and higher-order thinking.

Conclusion

This study succeeded in developing a problem-based question instrument on the topic of substances and their changes to measure the critical thinking skills of junior high school students by following seven stages of the modified Borg & Gall model. A total of 30 questions were developed and validated by material experts, evaluations, and science teachers, showing good quality in material, construction, and language. The trial using Rasch modeling on 27 students showed a varied distribution of problem difficulty. However, two questions (4 and 21) must be revised to match the student's abilities. Although the instrument's reliability still needs to be improved, most of the questions are feasible to evaluate students' critical thinking skills comprehensively, with a fairly good distribution of difficulty levels. The revised instrument is disseminated on a limited basis to science teachers as a more contextual evaluation tool and by the principles of the Independent Curriculum.

Acknowledgments

This research is funded by the Ministry of Research, Technology, and Higher Education in the Beginner Lecturer Research Grant in the 2024 fiscal year. We would like to thank the Ministry of Research, Technology, and Higher Education for funding the research; teachers and students of SMP N 1 Kaliwungu Kudus, as well as other parties who have contributed to the implementation of this research that we cannot mention one by one.

Author Contributions

D.S.S, I.N, and Y.M arranged the research design. Y.W provided input on the research design that had been prepared. D.S.S, and I.N developed a problem-based question and conducted product validation with experts. D.S.S, I.N, and Y.W interpret and analyze the results of product validation. D.S.S, I.N and Y.W revised problem-based question product that was developed. All authors contributed in arrange research articles.

Funding

This research is funded by the Ministry of Research, Technology, and Higher Education in the Beginner Lecturer Research Grant in the 2024 fiscal year.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this research. This study was funded by the Ministry of Research, Technology, and Higher Education through the Beginner Lecturer Research Grant for the 2024 fiscal year. The funder had no role in this manuscript's design, data collection, analysis, interpretation, or writing.

References

Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, C. A., & Persson, T. (2015). Strategies for teaching students to think critically: A meta-analysis. *Review of Educational Research*, *85*(2), 275–314. https://doi.org/10.3102/0034654314551063

Albar, S. B., & Southcott, J. E. (2021). Problem and project-based learning through an investigation lesson: Significant gains in creative thinking behaviour within the Australian foundation (preparatory) classroom. *Thinking Skills and Creativity*, *41*(100853). https://doi.org/10.1016/j.tsc.2021.100853

Asyari, M., Al Muhdhar, M. H. I., Susilo, H., & Ibrohim, I. (2016). Improving critical thinking skills through the integration of problem based learning and group investigation. *International Journal for Lesson and Learning Studies*, *5*(1), 36–44. https://doi.org/10.1108/IJLLS-10-2014-0042

Borg, W. R., & Gall, M. D. (1983). *Educational research an introduction* (4th ed.). Logman, Inc.

Bowen, R. S. (2022). Student perceptions of " critical thinking ": insights into clarifying an amorphous construct. *Chemistry Education Research and Practice*, *23*(3), 725–741. https://doi.org/10.1039/x0xx00000x

Dewi, N. D. L., & Prasetyo, Z. K. (2016). Pengembangan instrumen penilaian IPA untuk memetakan critical thinking dan practical skill peserta didik SMP. *Jurnal Inovasi Pendidikan IPA*, *2*(2), 213–222. https://doi.org/10.21831/jipi.v2i2.11963

Gürses, A., Şahin, E., & Güneş, K. (2022). Investigation of the Effectiveness of the Problem-Based Learning (PBL) model in teaching the concepts of "heat, temperature and pressure" and the effects of the activities on the development of scientific process skills. *Education Quarterly Reviews*, *5*(2), 67–73. https://doi.org/10.31014/aior.1993.05.02.469

Hanzlová, R., & Kudrnáč, A. (2024). Developing critical thinking test for adolescents: A validity and reliability study from the Czech Republic. *Thinking Skills and Creativity*, *53*(August). https://doi.org/10.1016/j.tsc.2024.101613

Haug, B. S., & Mork, S. M. (2021). Taking 21st century skills from vision to classroom: What teachers highlight as supportive professional development in the light of new demands from educational reforms. *Teaching and Teacher Education*, *100*, 103286. https://doi.org/10.1016/j.tate.2021.103286

Jiménez Pérez, E. del P., León Urrutia, M., García Guirao, P., & Cerezo Guzmán, M. V. (2025). Design and validation of CRISENSE, a novel critical competence assessment tool for Spanish adolescents and young adults. *Thinking Skills and Creativity*, *56*(March 2023). https://doi.org/10.1016/j.tsc.2024.101708

Kardoyo, Nurkhin, A., Muhsin, & Pramusinto, H. (2020). Problem-based learning strategy: Its impact on students' critical and creative thinking skills. *European Journal of Educational Research*, *9*(3), 1141–1150. https://doi.org/10.12973/EU-JER.9.3.1141

Loyens, S. M. M., van Meerten, J. E., Schaap, L., & Wijnia, L. (2023). Situating Higher - Order , Critical , and Critical - Analytic Thinking in Problem - and

Project - Based Learning Environments : A Systematic Review. In *Educational Psychology Review* (Vol. 35, Issue 2). Springer US. https://doi.org/10.1007/s10648-023-09757-x

Manassero-Mas, M. A., & Vázquez-Alonso, Á. (2022). An empirical analysis of the relationship between nature of science and critical thinking through science definitions and thinking skills. *SN Social Sciences*, *2*(12), 1–27. https://doi.org/10.1007/s43545-022-00546-x

Mayarni, M., & Nopiyanti, E. (2021). Critical and analytical thinking skill in ecology learning : A correlational study. *JPBI (Jurnal Pendidikan Biologi Indonesia)*, *7*(1), 63–70. https://doi.org/10.22219/jpbi.v7i1.13926

Miterianifa, Ashadi, Saputro, S., & Suciati. (2021). A conceptual framework for empowering students' critical thinking through Problem Based Learning in chemistry. *Journal of Physics: Conference Series*, *1842*(1), 1–9. https://doi.org/10.1088/1742-6596/1842/1/012046

Muhibbuddin, M., Artika, W., & Nurmaliah, C. (2023). Improving critical thinking skills through higher order thinking skills (HOTS)-based science. *International Journal of Instruction*, *16*(4), 283–296. https://doi.org/10.29333/iji.2023.16417a

Nitko, A. J., & Brookhart, S. M. (2011). *Educational assesment of students*. Pearson Education, Inc.

Nurwahidah, I., Widiyawati, Y., Sari, D. S., Masykuri, M., & Budiyanto, C. W. (2020). Development of science test to measure HOTS and digital literacy of junior high school students on the topic of city noise. *Edusains*, *12*(2), 203–213.

OECD. (2019). Programme for international student assessment (PISA) results from PISA 2018. In *Organisation for Economic Co-operation and Development*. http://www.oecd.org/pisa/ Data

Paul, R., & Elder, L. (2019). *The nature and functions of critical &creative thinking*. books.google.com. https://books.google.com/books?hl=en&lr=&id=KjWbDwAAQBAJ&oi=fnd&pg=PA2&dq=systemic+thinking+chemistry+learning&ots=ZJiLtKsyEV&sig=bXT7CYnBlJ1Mj6MGBRyYJxVOINk

Paulsen, V. H., & Dankert, S. (2022). Students ' reasoning when faced with test items of challenging aspects of critical thinking. *Thinking Skills and Creativity*, *43*(November 2021), 1–13. https://doi.org/10.1016/j.tsc.2021.100969

Prasad, S., & O'Malley, C. (2022). An introductory framework of Problem-Based Learning (PBL) and perspectives on enhancing facilitation approaches. *HAPS Educator*, *26*(3), 52–58. https://doi.org/10.21692/haps.2022.016

Reyes, R. L. (2025). Integrating real-world problems into chemistry curricula: Enhancing relevance and student engagement. *Forum for Education Studies*, *3*(2), 2177. https://doi.org/10.59400/fes2177

Reynders, G., Lantz, J., Ruder, S. M., Stanford, C. L., & ... (2020). Rubrics to assess critical thinking and information processing in undergraduate STEM courses. In *International Journal of …*. Springer. https://doi.org/10.1186/s40594-020-00208-5

Santos-Meneses, L. F., Pashchenko, T., & Mikhailova, A. (2023). Critical thinking in the context of adult learning through PBL and e-learning: A course framework. *Thinking Skills and Creativity*, *49*(June), 101358. https://doi.org/10.1016/j.tsc.2023.101358

Santos, L. F. (2017). The role of critical thinking in science education. *Journal of Education and Parctice*, *8*(20), 159–173.

Sari, D. S., Widiyawati, Y., & Nurwahidah, I. (2019). Pengembangan instrumen integrated science test untuk mengukur kemampuan berpikir kritis peserta didik SMP. *Prosiding Seminar Nasional Sains Dan Entrepreneurdhip VI*, 1–9.

Sari, D. S., Widiyawati, Y., Nurwahidah, I., & Setiawan, T. (2023). STEM critical thinking assessment for measuring students ' critical thinking skills in the automotive chemistry course. *Jurnal Penelitian Pendidikan IPA*, *9*(7), 5289–5295. https://doi.org/10.29303/jppipa.v9i7.4750

Sarıgoz, O. (2022). Analysis of masters degree attaining teachers' opinions on developing 21st-century life skills of students. *African Educational Research Journal*, *10*(4), 400–409. https://doi.org/10.30918/aerj.104.22.055

Shanta, S., & Wells, J. G. (2022). T/E design based learning: assessing student critical thinking and problem solving abilities. *International Journal of Technology and Design …*, *32*, 267–285. https://doi.org/10.1007/s10798-020-09608-8

Siburian, J., Corebima, A. D., Ibrohim, & Saptasari, M. (2019). The correlation between critical and creative thinking skills on cognitive learning results. *Eurasian Journal of Educational Research*, *2019*(81), 99–114. https://doi.org/10.14689/ejer.2019.81.6

Smith, K., Maynard, N., Berry, A., Stephenson, T., Spiteri, T., Corrigan, D., Mansfield, J., Ellerton, P., & Smith, T. (2022). Principles of Problem-Based Learning (PBL) in STEM Education: Using expert wisdom and research to frame educational practice. *Education Sciences*, *12*(10), 1–20. https://doi.org/10.3390/educsci12100728

Sumintono, B., & Widhiarso, W. (2013). *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial*. Trim Komunikata Publishing House.

Supena, I., Darmuki, A., & Hariyadi, A. (2021). The influence of 4C (constructive, critical, creativity, collaborative) learning model on students' learning outcomes. *International Journal of Instruction*, *14*(3), 873–892. https://doi.org/10.29333/iji.2021.14351a

Suprapto, N. (2016). What should educational reform in Indonesia look like?-learning from the PISA science scores of East-Asian countries and Singapore. *Asia-Pasific Forum on Science Learning and Teaching*, 1–20.

Suwistika, R., Ibrohim, I., & Susanto, H. (2024). Improving critical thinking and creative thinking skills through POPBL learning in high school student. *JPBI (Jurnal Pendidikan Biologi Indonesia)*, *10*(1), 115–122. https://ejournal.umm.ac.id/index.php/jpbi/article/view/30172

Tawfik, A. A., Gish-lieberman, J. J., Gatewood, J., & Arrington, T. L. (2021). How k -12 teachers adapt problem-based learning. *The Interdisciplinary Journal of Problem Based Learning*, *15*(1), 1–19. https://doi.org/10.14434/ijpbl.v15i1.29662

Umrzokova, G., & Pardaeva, S. (2020). Developing teacher` professional competence and critical thinking is a key factor of increasing the quality of education. *Mental Enlightenment Scientific Methodological Journal*, *2*, 66–75. mentaljournaljspu.uz/index.php/mesmj/article/view/31

Yu, L., & Zin, Z. M. (2023). The critical thinking-oriented adaptations of problem-based learning models: a systematic review. *Frontiers in Education*, *8*. https://doi.org/10.3389/feduc.2023.1139987