STEM Education, 5(4): 587–616
DOI: 10.3934/steme.2025028
Received: 20 December 2024
Revised: 07 May 2025
Accepted: 20 May 2025
Published: 10 June 2025

https://www.aimspress.com/journal/steme

Research article

A CIPP-based evaluation of critical thinking in science: Implementation of the Merdeka curriculum in Mataram public junior high schools

Nabilah Nabilah* and Slamet Suyanto

Department of Educational Research and Evaluation, Yogyakarta State University, Karang Malang, Caturtunggal, Yogyakarta Special Region, Indonesia; nabilah0011pasca.2022@student.uny.ac.id, slamet_suyanto@uny.ac.id

* **Correspondence:** Email: nabilah0011pasca.2022@student.uny.ac.id / nabilah.hasan2023@gmail.com; Tel: +62-853-384-98858.

Academic Editor: Ergun Gide

Abstract: This study evaluates the efficacy of the Merdeka Curriculum in promoting critical thinking within junior high school science teaching through the CIPP (Context, Input, Process, Product) evaluation framework. A mixed-methods approach utilizing a convergent parallel design was implemented, including 438 students and 26 science teachers from seven public junior high schools in Mataram. The data collection comprised standardized critical thinking assessments, semi-structured interviews, classroom observations, and document analysis.

This study finds that teachers demonstrate moderate proficiency in understanding critical thinking objectives (55.77%), yet they encounter challenges in translating these objectives into effective classroom practices. The classroom approach efficiently combines scientific procedures (69.23%), although it lacks structured discussions and inquiry-based learning, constraining students' capacity to cultivate reasoning abilities. Assessments of students indicate inadequate proficiency in reflective thinking and argument evaluation (55.79%), underscoring challenges in assessing assumptions and reasoning frameworks.

This research addresses a critical gap in understanding how curriculum implementation influences students' critical thinking development. However, the study has limitations. Future studies should consider using experimental or quasi-experimental designs, such as randomized controlled trials, to evaluate the effectiveness of specific curriculum strategies, like structured discussions or technology-integrated scaffolding. The results underscore the necessity for specialized teacher training, enhanced resource distribution, and organized assessment frameworks to maximize curriculum execution. These insights facilitate curriculum enhancement and guide policy decisions to reinforce science education in Indonesia.

Keywords: CIPP model, evaluation, critical thinking, science, junior high school, students, teachers

1. Introduction

Critical thinking is an essential competency in science education, enabling students to analyze, assess, and integrate knowledge to resolve issues efficiently [1]. Some related work includesidentifying key factors affecting college students' adoption of the e-learning system in mandatory blended learning environments [2]; from engagement to performance, the role of effort regulation in higher education online learning [3]; and ESERNet, learning spectrogram structure relationship for effective speech emotion recognition with swin transformer in classroom discourse analysis [4]. Another recent study indicates that promoting critical thinking in science education improves students' problem-solving skills, creativity, and scientific reasoning [5]. Nonetheless, despite the focus on critical thinking, numerous educational systems encounter challenges in delivering effective techniques to cultivate these skills in students [6]. Nonetheless, despite the focus on critical thinking, numerous educational systems encounter challenges in delivering effective techniques to cultivate these skills in students [6].

In acknowledgement of the significance of critical thinking, the Indonesian Ministry of Education implemented the Merdeka Curriculum in 2021 prioritized competency-based learning and student-centered pedagogies [7]. One of its fundamental elements, Profil Pelajar Pancasila, designates Bernalar Kritis (critical thinking) as a pivotal competency, intending to prepare students to examine information, evaluate logical frameworks, and make informed decisions. The Merdeka Curriculum diverges from prior Indonesian curricula that prioritized regimented subject delivery and standardized tests, instead adopting a competency-based education model that offers increased flexibility for schools and teachers [8]. Although prior research has investigated the overall application of the Merdeka Curriculum [9,10], there is a deficiency of studies that specifically assess its role in promoting critical thinking within junior high school scientific instruction. This study utilizes the CIPP assessment technique to evaluate the effectiveness of the curriculum in fostering students' critical thinking skills.

Moreover, the professional development of teachers and the modification of curricula continue to pose significant challenges in the integration of critical thinking abilities within science education. Research demonstrates that teachers have challenges in synchronizing instructional strategies with curriculum goals, owing to inadequate professional development and resource constraints [11,12]. Teacher training and curriculum adaptation can hinder the development of critical thinking skills, as suggested by several related studies—such as those examining the association between discussion strategies and students' critical thinking in asynchronous online discussion [13], exploring the relationship between teacher talk supports and student engagement from the perspective of students [14], and mmatrans: understanding students continuance intention: a comparison of live video learning, pre-recorded video learning, and hybrid video learning in the covid-19 pandemic [15].

Recent research indicates that numerous pre-service teachers possess a constrained conceptual grasp of critical thinking, frequently linking it exclusively to cognitive elements while overlooking its dispositional facets [5]. This signifies the necessity for extensive teacher training programs that encompass both cognitive and emotional aspects of critical thinking.

The Merdeka Curriculum promotes individualized learning and project-based instruction to enhance critical thinking [16]. Nevertheless, research suggests that numerous teachers encounter difficulties in properly applying these tactics in scientific classrooms [17]. This prompts apprehensions over the genuine development of analytical and reasoning skills among students as prescribed by the curriculum. Moreover, although critical thinking is globally acknowledged as a vital competency for lifelong learning and economic prosperity, studies indicate that many educational institutions still lack a coherent and enduring framework for its instruction [18]. This underscores the necessity of formulating systematic guidelines and assessment instruments to guarantee the effective incorporation of critical thinking abilities in science education.

The 2022 PISA report highlights this concern, revealing that merely 34% of Indonesian students attain at least Level 2 in scientific literacy, markedly below the OECD average of 76%, with nearly no Indonesian students reaching the highest performance levels (OECD, 2023). This highlights a significant deficiency in students' capacity to examine, evaluate, and reason scientifically, prompting questions regarding the efficacy of the Merdeka Curriculum in cultivating critical thinking abilities within science education.

In addition to its significance in science education, critical thinking is an essential life skill required for effective decision-making and innovation [19]. According to Piaget's theory, during the junior high school stage (Phase D), students undergo substantial cognitive growth, rendering it an essential period for the development of higher-order thinking skills [20]. The successful application of critical thinking in the classroom is significantly dependent on instructors' instructional methodologies, curricular coherence, and the accessibility of educational materials [21,22]. Research indicates that numerous teachers continue to have challenges in incorporating critical thinking into their scientific curricula, frequently depending on rote memory instead of inquiry-based learning [23]. Recent findings suggest that metacognitive and self-regulated learning strategies are essential for enhancing critical thinking in science education; yet, their implementation in classroom environments is still restricted [24].

Students frequently experience challenges in participating in critical thinking activities due to inadequate exposure to inquiry-based learning and insufficient support from teachers [6]. The prevailing passive learning culture in numerous schools hinders students from cultivating their analytical and reasoning skills. Consequently, a thorough assessment of the implementation of critical thinking in science education within the Merdeka Curriculum is necessary. This study seeks to assess the integration of critical thinking in junior high school science instruction within the Merdeka Curriculum, utilizing the CIPP assessment paradigm [25]. This research aims to investigate the following inquiries:

1. Context: To what extent do teachers comprehend the objectives and execution of critical thinking within the Merdeka Curriculum?
2. Input: What instructional strategies and resources foster critical thinking in junior high school science education?
3. Process: In what manner is critical thinking included in scientific classroom practices?

4. Product: What is the extent of students' proficiency in critical thinking within science education under the Merdeka Curriculum?

This research utilizes a mixed-methods approach with a parallel convergent design [26] to deliver a thorough evaluation of critical thinking integration in science education. The CIPP model (Context, Input, Process, Product) functions as a systematic framework for assessing the effectiveness of the curriculum in promoting critical thinking and pinpointing opportunities for enhancement. This study's findings will inform education policy debates and offer ideas for improving critical thinking instruction in science education.

2. Theoretical framework

The Concept of Merdeka Curriculum

The Merdeka Curriculum highlights flexible learning that prioritizes the demands of students and seeks to cultivate individuals with robust character and competence. This curriculum was implemented as part of an educational reform to establish a structured learning framework that enables schools to flexibly adopt competency-based education [27]. The fundamental components of the Merdeka Curriculum encompass competency-based education, an emphasis on important information, and adaptable learning pathways, contrasting with earlier Indonesian curricula that were predominantly content-driven and uniform. The phases of the Merdeka Curriculum denote stages of learning tailored to align with the competency levels of students as opposed to the conventional classroom system. The Junior High School level (Phase D) of the Merdeka Curriculum encompasses schooling for students aged 13 to 15 years, corresponding to grades 7 through 9 [28]. At this juncture, students undergo a cognitive shift from concrete to abstract thinking, as delineated in Piaget's developmental theory. Their capacity to analyze information and make evidence-based decisions begins to evolve, so enhancing critical thinking skills is crucial in scientific education [20]. The Merdeka Curriculum corresponds with the Pancasila Students Profile, which delineates six essential competencies: faith and morality, independence, collaboration, global variety, critical thinking, and creativity [29]. Critical thinking is essential in science education as it allows students to analyze information, assess logical arguments, and make educated judgments. Nevertheless, research demonstrates that numerous teachers continue to have difficulties in integrating teaching methodologies with this curriculum owing to insufficient professional development [24].

Defining Critical Thinking Mastery and Teacher Competence

Mastery of critical thinking denotes students' capacity to systematically analyze, assess, and utilize knowledge in problem-solving contexts [6]. This skill encompasses logical thinking, evidence-based arguments, and the capacity to derive conclusions from numerous sources of information [30]. In science education, mastery of critical thinking is associated with inquiry-based learning and problem-solving skills that improve students' comprehension and application of scientific topics [5]. Critical thinking in science education has gained increasing attention as teachers seek to prepare students for complex real-world problems. It is not merely a cognitive skill but also includes dispositional components, such as curiosity and open-mindedness, and metacognitive strategies like self-monitoring and reflection [19,31]

Teacher proficiency in promoting critical thinking encompasses the creation of inquiry-based lessons, the facilitation of conversations, and the evaluation of students' analytical abilities [24]. It also encompasses the capacity to establish learning environments that promote inquiry, discovery, and reflective thought [23]. Moreover, effective critical thinking instruction benefits from dialogic teaching and structured teacher-student interaction strategies. Recent studies suggest that teacher questioning techniques and classroom discourse significantly influence the development of students' analytical reasoning [2,13]. Research indicates that numerous teachers lack the requisite training and pedagogical tools to cultivate these competencies successfully, hence impacting the ability of students to participate in substantive scientific inquiry [5]. Studies from Indonesian contexts have also shown the potential of culturally grounded and language-responsive instructional models. For instance, the ethnoscience-integrated e-modules can enhance scientific reasoning [22], while other research highlights how questioning behavior in EFL classrooms influences students' engagement and cognitive responses [21]. These insights underscore the need for localized pedagogical strategies aligned with the Merdeka Curriculum's emphasis on critical thinking.

Marzano & Kendall's Taxonomy and Problem-Based Learning (PBL)

The incorporation of Marzano and Kendall's New Taxonomy into science education offers a systematic framework for comprehending cognitive processes. In contrast to Bloom's Taxonomy, which prioritizes hierarchical cognitive domains, Marzano and Kendall advocate for a systems-based framework that classifies thinking skills into three domains: cognitive system, metacognitive system, and self-system [31]. This taxonomy fosters critical thinking by facilitating students' engagement in the retrieval, understanding, analysis, and application of knowledge, crucial components for scientific inquiry, and problem-solving [5]. Recent studies demonstrate that the implementation of Marzano & Kendall's framework in science education improves students' capacity for critical information processing and the application of knowledge to intricate issues [30]. This model corresponds with the goals of the Merdeka Curriculum, which advocates for flexible learning and inquiry-based methodologies in science education.

Evaluation Framework: The CIPP Model

This study utilizes the Context, Input, Process, and Product (CIPP) evaluation methodology [25] to systematically evaluate the incorporation of critical thinking in science teaching within the Merdeka Curriculum. This framework is selected for its thorough methodology in assessing educational programs, facilitating a systematic evaluation of the curriculum's efficacy.

1. Context Evaluation: Evaluates teachers' comprehension of critical thinking goals within the Merdeka Curriculum.
2. Input Evaluation: Assesses the pedagogical strategies and resources available to enhance critical thinking in science education.
3. Process Evaluation: Assesses the implementation of critical thinking in educational activities.
4. Product Evaluation: Assesses students' proficiency in critical thinking abilities within the context of science education.

Research indicates that the CIPP model offers a comprehensive framework for curriculum assessment, guaranteeing that every facet of implementation, from planning to students' outcomes, is systematically appraised [32,33]. This study enhances the current literature by utilizing this approach to evaluate the efficacy of the Merdeka Curriculum in promoting critical thinking within science education and pinpointing areas for enhancement.

3. Method

3.1. Research design

This research utilized a convergent parallel mixed-methods design, combining quantitative and qualitative approaches to assess the integration of critical thinking abilities in science education within the Merdeka Curriculum. The Context, Input, Process, and Product (CIPP) evaluation methodology [25] was employed to systematically evaluate the efficacy of curriculum implementation, as seen in Figure 1. This methodology was chosen for its complete framework for assessing educational programs, guaranteeing that all facets of implementation, from planning to outcomes, are scrutinized. This evaluation primarily concentrates on the implementation process, while also indirectly examining factors that influence critical thinking development, including teacher-student interactions, classroom environment, lesson planning strategies, and students' engagement, through qualitative data collection methods such as interviews, documentation, and observational techniques. These elements offer supplementary context to comprehend the efficacy of the Merdeka Curriculum in promoting critical thinking.

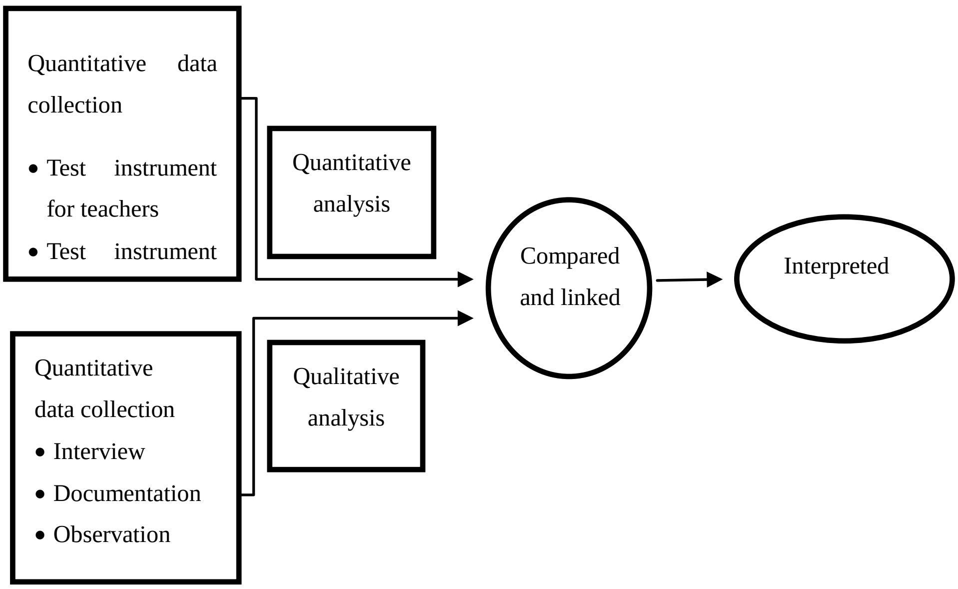

Figure 1. Research design with convergent mixed-parallel design.

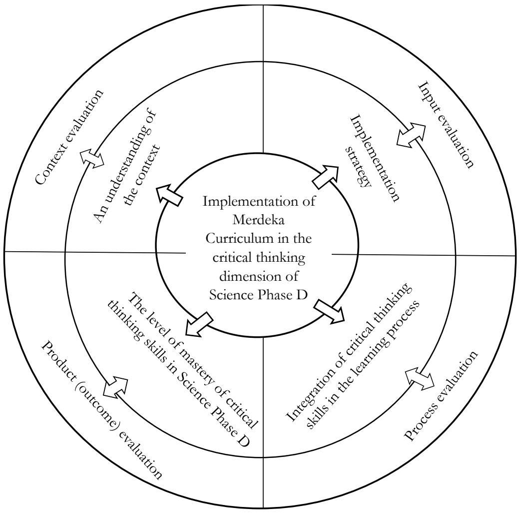

Figure 2. The CIPP evaluation model.

3.2. Participants and sampling criteria

The research was carried out in seven public junior high schools in Mataram City, Indonesia, which have been implementing the Merdeka Curriculum since 2022. The selection of these schools was predicated on their capacity to fulfill all inclusion criteria to guarantee quality observations and regional representation. A total of 26 science teachers and 438 students participated across these institutions. Teachers were chosen based on a minimum of three years of teaching experience in junior high school science and active participation in professional development programs pertinent to the curriculum. The selection of these individuals aimed to guarantee that the collected data represented a variety of experiences and instructional methodologies in the application of critical thinking abilities. Table 1 delineates the criteria for sample inclusion.

Table 1. Inclusion criteria for sampling selection.

Criteria	Remarks
The implementation duration of the Merdeka Curriculum	1. Schools that have consistently implemented the Merdeka Curriculum for a minimum duration of one full year, beginning from grade VII
Categories of implementation of the Merdeka Curriculum	1. Schools that have adopted the Merdeka Curriculum in the independent learning or independent transformation phase
The students' criteria	1. Students who have finished grade VIII and progressed to grade IX in the 2024 academic year, having completed two full years of the Merdeka Curriculum. This criterion facilitates a more representative assessment of students' critical thinking skills.
Teachers' criteria	1. Science teachers actively engaged in promoting the implementation of the Merdeka Curriculum. 2. Science teachers engaged in the implementation of the Merdeka Curriculum for grades VII and VIII in schools that fulfill the inclusion criteria.
Regional representation	1. The chosen schools are distributed across various regions in Mataram City, aiming to encompass diverse learning contexts, thereby enabling the evaluation results to reflect a more comprehensive condition.

3.3. Validity and reliability testing

3.3.1. Content validity

Expert judgment was utilized to establish content validity. A panel of five experts, comprising curriculum and assessment specialists as well as science education professionals, evaluated the instruments to confirm that the items were aligned with the constructs being measured and suitable for the study context. The panel assessed the clarity, relevance, and representativeness of items before the final implementation of the instruments.

3.3.2. Construct validity

Construct validity was evaluated through exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). EFA was conducted on the teacher instrument to identify the underlying factor structures. CFA was not performed due to the limited sample size (n = 26), which may result in model instability [34]. The Kaiser-Meyer-Olkin (KMO) test presented in Table 2 yielded a value of 0.3, suggesting that the instrument falls short of the optimal threshold for factor analysis. The Bartlett's Test of Sphericity yielded a significant result ($p < 0.05$), indicating that factor analysis may be applicable. Due to the multicollinearity issue (determinant = 0.0000013), factor extraction was conducted with caution. Four key factors were retained based on eigenvalues greater than 1 and the analysis of the Scree Plot, as illustrated in Figure 3.

Table 2. KMO and Bartlett's test results for the teacher test instrument.

Kaiser- Meyer- Olkin (KMO) Measure of Sampling Adequacy	*Overall MSA*	0.3
Bartlett's Test of Sphericity	*chisq*	235.8664
	df	190
	p. value	0.01177472
Positive determinant	$1.323512e^{-6}$	

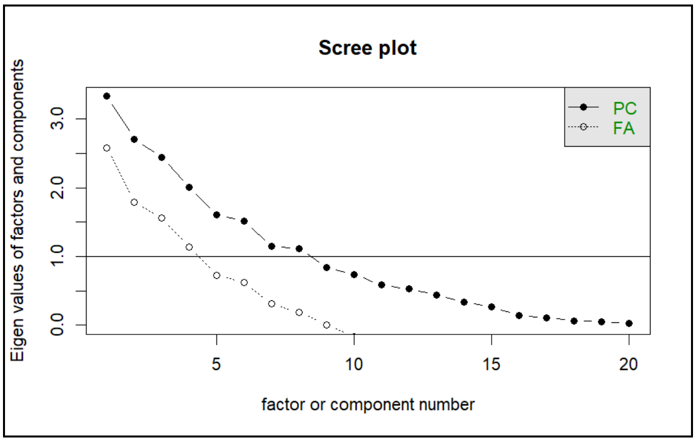

Figure 3. Scree plot for the teacher test instrument.

CFA was performed after EFA on the students' instrument to validate the factor structure. Table 3 indicates a KMO value of 0.81, signifying eligibility for factor analysis, and Bartlett's test yielded significant results. The CFA findings indicated a satisfactory model fit, with an RMSEA of 0.032, CFI of 0.909, and SRMR of 0.044, validating that the factor structure conformed to theoretical assumptions, as illustrated in Figure 4.

Table 3. KMO and Bartlett's test results for the students' test instrument.

Kaiser- Meyer- Olkin (KMO) Measure of Sampling Adequacy	*Overall MSA*	0.81
Bartlett's Test of Sphericity	*chisq*	1001.158
	df	190
	p. value	2.543626e-110
Positive determinant	0.09719999	

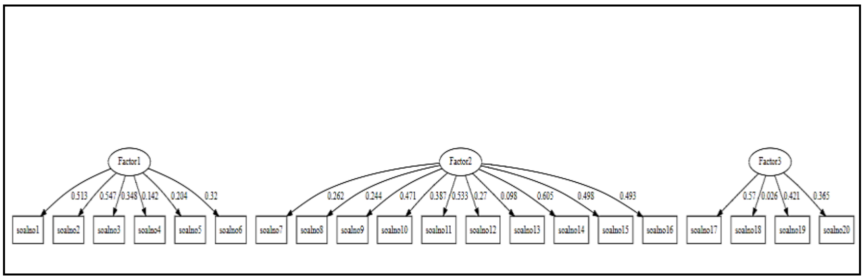

Figure 4. CFA path diagram for students test instrument.

3.3.3. Reliability assessment

The dependability of the instrument was evaluated with Cronbach's Alpha. The teacher instrument demonstrated a reliability coefficient of 0.49, reflecting moderate internal consistency, while the students' instrument exhibited a reliability coefficient of 0.73, showing good internal consistency. Despite the teacher instrument demonstrating moderate reliability, the EFA results yielded further insights into factor structures, hence reinforcing the instrument's validity.

3.4. Quantitative data collection

This study utilizes a critical thinking assessment for both teachers and students. The instrument employed comprises multiple-choice questions aimed at assessing the diverse aspects of critical thinking as delineated in the Critical Thinking Profile of the Merdeka Curriculum. Table 4 presents the teacher's assessment framework emphasizing context, input, and process evaluation, whereas Table 5 delineates the students' assessment framework that evaluates the output (result) of critical thinking skills.

Table 4. The teacher's test instrument framework.

Elements of Critical Thinking	Sub-elements	Indicators	No Item
Elements of acquiring and processing information and ideas	Posing inquiries	Recognizing the comprehension of the critical thinking element outcomes for Phase D, based on the example questions aligned with the Phase D critical thinking dimension outcomes (context).	1
		Identify the appropriate learning technique to stimulate students' curiosity regarding the connection between a topic and its practical application in everyday life (input)	2,4
		Enhance the critical thinking process to assist students in deriving conclusions from the supplied evidence (process)	3,5
	Identify, clarify, and process information and ideas	Identify essential components of critical thinking pertinent to Phase D scientific education (context)	6
		Determine the appropriate learning technique to promote critical thinking in understanding the application of scientific concepts in daily life, utilizing information presented through images or news articles (input).	7,9
		Enhance critical thinking abilities by integrating Phase D scientific concepts with practical applications to evaluate errors arising from knowledge regarding a phenomenon associated with Phase D science concepts (process)	8,10
Elements of analyzing and evaluating logic and procedures	Examine and assess logic and methodologies	Identify case studies that exemplify scientific concepts from Phase D science, aligning with the components of critical thinking to enhance the analytical process concerning Phase D science topics in everyday contexts.	11
		Identify suitable learning strategies to promote critical thinking in decision-making based on knowledge of the supplied Phase D scientific concepts (input)	12,13
		Enhance critical thinking abilities by integrating Phase D science ideas with practical applications in daily life to evaluate suitable strategies for developing preventive measures based on the knowledge provided through Phase D science concepts (process).	14,16
Elements of reflective thinking and thought processes	Contemplate and assess their cognition.	Identify case studies grounded in scientific principles from Phase D science to enhance students' reflective process (context).	15
		Identify the suitable learning approach to promote critical thinking, facilitating the articulation of the assumptions that underpin their thinking or arguments concerning a topic or problem in Phase D science ideas (input).	17,19
		Enhance critical thinking abilities by integrating Phase D science principles with practical applications to evaluate disparities in expert viewpoints on a scientific concept, thereby acquiring a more comprehensive understanding (process).	18,20

Table 5. The students' test instrument framework.

Elements of Critical Thinking	Sub-elements	The outcome of the critical thinking dimension in Phase D	Indicators	No Item
Elements of information acquisition and concept processing	Identifying, clarifying, and processing information and ideas	Identifying, elucidating, and managing information and concepts	Identify a problem arising from the application of science in everyday life, utilizing information offered through images or news articles.	1,2
			Analyze facts regarding a scientific application in everyday life, utilizing images or news articles as sources.	3,4
	Posing inquiries	Inquiring for clarification and interpretation of information, as well as determining the reasons and implications of said information.	Selecting the appropriate inquiry type depending on the information provided on a phenomenon.	5,6
Elements of analyzing and evaluating logic and procedures	Examine and assess logic and methodologies	Utilize several arguments to arrive at a conclusion or decision	Determine the appropriate issue according to a specified criterion.	7,8
			Condense information derived from the provided data.	9,10
			Render decisions informed by the data about the introduced Phase D scientific concept.	11,12
			Examine inaccuracies that arise from the information provided on a phenomenon associated with the concept of Phase D science.	13,14
			Ascertain the requisite actions to formulate preventive strategies based on the provided facts.	15,16
Elements of reflective thinking and thought processes	Contemplate and assess their cognition.	Articulates the underlying assumptions, acknowledges the inclination and ramifications of cognitive bias, and endeavors to contemplate diverse viewpoints	Explain the presuppositions that inform their reasoning or arguments concerning a topic or issue inside Phase D science concepts.	17,18
			Evaluate the many perspectives of specialists on a scientific issue to acquire a more thorough understanding.	19,20

3.5. Qualitative data collection

Semi-structured interviews were performed with three teachers and three to five students from each school who consented to participate. The interviews aimed to investigate teachers' instructional strategies, problems, and perceptions concerning the incorporation of critical thinking skills, whereas students' interviews concentrated on their learning experiences. Each interview lasted 30 to 45 minutes, was audio-recorded with consent, and transcribed verbatim for analytical purposes. The interview protocol comprised both structured and open-ended questions, formulated as a guideline (Appendix A.1).

Classroom observations were undertaken alongside interviews to evaluate teacher-student interactions and the execution of critical thinking tasks. Observations served not as the major data source, but rather to corroborate and validate the findings from interviews and document analyses. This method corresponds with the suggestion from Creswell [26] to employ triangulation to augment study validity. Documents, including lesson plans, instructional materials, and assessment rubrics, were examined to verify alignment with curriculum objectives and assess the integration of critical thinking skills in instructional design. Figure 5 illustrates the data collection method undertaken by students.

Figure 5. Data collection method undertaken by students.

3.6. Ethical considerations

This study complied with ethical research standards and secured legal authority via official research permits from the university to the Mataram City Education Office, subsequently obtaining consent letters from the Mataram City Department of Education and Culture for the participating schools. To facilitate clear communication about the research objectives, direct meetings were held with each school, enabling instructors to comprehend the research procedure and the confidentiality of the gathered data. Confidentiality and anonymity were rigorously upheld, safeguarding participant identities. Furthermore, participants were apprised that their participation in this study was optional and that they might withdraw at any moment without incurring any repercussions. Figure 6 illustrates the hearing process involving the school.

Figure 6. The hearing process involving the school.

3.7. Data analysis techniques

The research included both quantitative and qualitative data analysis to guarantee a thorough assessment of the curriculum's efficacy. Quantitative data were examined through descriptive statistics to interpret the test results of teachers and students. The statistical measures employed comprised means, standard deviations, and frequency distributions to encapsulate the data. The instruments' validity and reliability were established by EFA and CFA, as detailed in the Instrumentation and Data Collection section. The conclusive data analysis concentrated on elucidating patterns in instructors' and students' responses.

Before measuring the evaluation according to the achievement criteria, the outcomes of the multiple-choice instrument responses from teachers and students were initially assessed. Instruments for teachers and students are designed based on the elements and sub-elements of critical thinking within the Pancasila students' profile, which are then developed into indicators and items aligned with the evaluated characteristics. The formula employed to ascertain the value of each instrument is as follows:

$$\text{Final score} = \frac{\textit{number of item answered correctly}}{\textit{number of test items}} \times 100$$

The parameters of achievement for assessing the outcomes of the Merdeka Curriculum implementation in the critical thinking dimension of science at the junior high school level (Phase D) in public junior high schools in Mataram City are determined by the average percentage of the total score across each evaluative aspect of this implementation. The percentage outcomes of the final score are subsequently referenced in the evaluation measurement criteria table, modified to align with the percentage and category sections of the Learning and Assessment Guidelines for Early Childhood Education, Primary Education, and Secondary Education, Revised Edition of 2024, as delineated in Table 6.

Table 6. The parameters of achievement.

Interval	Category of Aspect Achievement	
	Context, Input, Process	***Product (Outcome)***
81%-100%	Very well achieved. Teachers can be given further challenges to implement innovations in learning strategies or mentor other teachers.	Very well achieved. Students have exceeded learning targets.
61%-80%	Achieved. The teacher has achieved the evaluation objectives, yet there may still be room for improvement.	Achieved. Students have achieved the learning objectives, though there may be some areas for further improvement.
41%-60%	Almost well achieved. The teacher needs to improve some aspects that are still lacking in implementation.	Almost well achieved. Students need to improve their comprehension of some key elements.
21%-40%	Not yet achieved. Teachers should identify certain aspects that are poorly understood or have not been successfully implemented.	Not yet achieved. Students need to revisit most of the learning material, focusing on understanding the concepts that are not well understood.
0%-20%	Not achieved. Teachers need to conduct in-depth reflections to identify challenges faced in understanding and implementing the Merdeka Curriculum in the Critical Thinking dimension.	Not achieved. Students need to identify challenges or difficulties faced in understanding the Critical Thinking dimension.

Source : (Ministry of Education, Culture, Research, and Technology 2024, pp. 32–33)

4. Findings and discussions

4.1. Context evaluation

The percentage distribution of scores for each evaluated indicator is presented in the bar chart. Figure 7 is a summary of the key findings based on the evaluation. The bar chart illustrates the percentage distribution of scores for each assessed indicator.

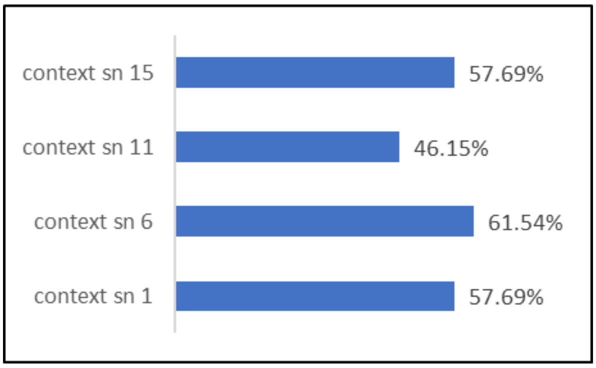

Figure 7. The percentage distribution of scores for each evaluated indicator in the context evaluation.

Teachers demonstrate moderate proficiency in understanding critical thinking objectives (55.77%) classified as "almost well achieved", yet they struggle to implement them effectively in classroom instruction. Nevertheless, a more profound investigation revealed differing degrees of understanding. Teachers demonstrated a satisfactory comprehension of the overarching aims of critical thinking (60.25%, achieved), but encountered difficulties in incorporating reflective learning (46.15%, nearly achieved). Challenges were identified in the capacity to devise effective questioning tactics and scaffold debates, which are crucial for enhancing students' analytical reasoning [20].

Vygotsky's Zone of Proximal Development (ZPD) posits that effective scaffolding necessitates teachers to facilitate students in tasks that progressively shift responsibility to them [20]. Nevertheless, research reveals that numerous teachers are deficient in the ability to facilitate talks proficiently. A teacher observed the following:

> “Students encounter difficulties with intricate problem-solving tasks due to their lack of experience in reasoning beyond rote memorization” (Teacher 3).

Another remarked:

> “The critical thinking aspect is not consistently utilized, contingent upon the material.” Early-grade students concentrate mostly on fundamental concepts” (Teacher 5).

This corresponds with research highlighting the necessity for teachers to get organized professional development in scaffolding methodologies [35]. Pre-service teachers frequently acknowledge the significance of critical thinking, although they conceive it with differing levels of profundity [5]. This research underscores deficiencies in reflective learning and questioning methodologies, along with the current study's conclusions that teachers encounter challenges in proficiently incorporating reflective questioning frameworks. This highlights the imperative for specialized teacher training programs in scaffolding and inquiry-based learning.

Subsequent qualitative analysis, as seen in Figure 8, identified obstacles in the implementation of critical thinking instruction:

- The interview results corroborate the quantitative findings, indicating that the primary barrier in integrating the critical thinking dimension is students’ preparedness and instructor adaptation tactics.
- Critiques of Execution: Certain teachers believe that the incorporation of critical thinking within the Pancasila Students Profile is impractical for implementation at the junior secondary school level. Teachers reported that most students continue to face difficulties in meeting the anticipated critical thinking benchmarks, with only approximately 40% of students in a class capable of independently analyzing and synthesizing information. The analytical results indicate that the category "A Critique of the Implementation" comprises the highest proportion, accounting for 25.97% of the total utterances in the interviews.
- Adaptation in Scientific Education. Teachers indicated that not all aspects of the Pancasila Students Profile are utilized in every science lecture, contingent upon the subject matter being addressed. In grade 7, the emphasis on critical thinking is sometimes minimal; nevertheless, it becomes relevant in more intricate subjects such as the classification of living organisms or light waves. Teachers emphasized that, in numerical resources, the primary emphasis remains on comprehending fundamental concepts rather than cultivating critical thinking abilities. The table

analysis results indicate that the "Adaptation of the Merdeka Curriculum" category exhibits a significant distribution (14.60% - 21.90%).

- Critical Thinking as a Historical Concept in Science. Certain teachers acknowledge that the notion of critical thinking has historically been integral to science education, although under the Merdeka Curriculum, this notion is delineated as a distinct component within the Pancasila Students Profile.

	A Critique of the Implementatio... 8 9	Adaptation of the Merdeka Curriculum 6 7	Merdeka Curriculum, new terminology in old concepts 6 7	Totals
1: Int... 9	3,33 25,97%	1,67 14,60%		5 14,29%
2: Int... 9	1,67 12,99%	1,67 14,60%	1,67 15,50%	5 14,29%
3: Int... 10	2,5 19,48%	1,25 10,95%	1,25 11,63%	5 14,29%
4: Int... 10	2 15,58%	1 8,76%	2 18,60%	5 14,29%
5: Int... 13	1,67 12,99%	1,67 14,60%	1,67 15,50%	5 14,29%
6: Int... 10	1,67 12,99%	1,67 14,60%	1,67 15,50%	5 14,29%
7: Int... 6		2,5 21,90%	2,5 23,26%	5 14,29%
Totals	12,83 100,00%	11,42 100,00%	10,75 100,00%	35 100,00%

Figure 8. Outcomes from co-document versus group codes table derived from Atlas.ti.

Furthermore, an analysis of lesson plans indicated that when teachers incorporate problem-solving exercises, these predominantly emphasize procedural tasks rather than open-ended, reflective discourse. This underscores the necessity for enhanced integration of critical thinking instruction. The results underscore the imperative for professional development programs that provide teachers with tools to facilitate discussions and promote inquiry-based learning effectively.

4.2. Input evaluation

The bar chart illustrates the percentage distribution of scores for each evaluated indicator. Figure 9 below summarizes the key findings derived from the evaluation aspects.

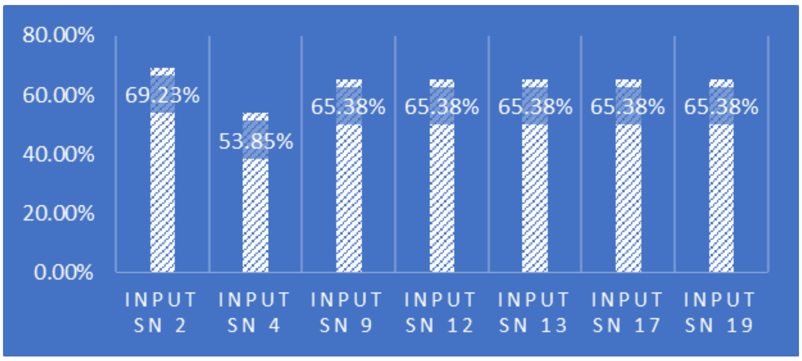

Figure 9. Distribution of scores for each assessed indicator in the input evaluation.

The evaluation results indicated that teachers utilized structured learning strategies to enhance critical thinking, attaining an average score of 64.29%. Nonetheless, their ability to stimulate students' curiosity about real-world applications of science presented a challenge, reflected by a score of 53.85%, which indicates near achievement. Teachers exhibited competence in developing instructional strategies, attaining a score of 66.41%. Nonetheless, their efficacy in choosing real-world examples to improve engagement in learning was limited, reflected by a score of 49.28%, suggesting only moderate success. The results support previous studies emphasizing the importance of context-based learning for enhancing critical thinking skills [10,30,36]. The challenges of real-world contextualization indicate a deficiency in teachers' capacity to effectively scaffold learning, a critical component emphasized in Vygotsky's Scaffolding Theory [20,31].

The qualitative analysis corroborates these findings. Thematic coding revealed significant challenges in teacher preparedness, particularly in adapting to the Merdeka Curriculum and employing effective questioning techniques. The category of Teacher Readiness indicates that, despite training initiatives, numerous teachers continue to face difficulties in independently developing inquiry-based lessons that promote critical thinking among students. This aligns with Dewey's reflective thinking model, which posits that effective critical thinking instruction necessitates structured opportunities for students to reflect on real-world issues [20,35]. Themes such as a critique of critical thinking implementation and initiatives to improve teacher capacity highlight disparities in teachers' interpretations and applications of critical thinking principles in classrooms, indicating differing levels of metacognitive engagement as classified in Marzano's New Taxonomy [31].

Teacher support networks, such as the Merdeka Mengajar Platform (PMM) and MGMP mentorship, significantly contributed to addressing implementation challenges. Teachers indicated that involvement in these training initiatives offered essential insights for developing instruction focused on improving critical thinking. Variations in training quality indicate a necessity for more structured and consistent professional development programs [37]. A teacher remarked:

> "Centralized training has been particularly beneficial, especially through PMM. Furthermore, valuable insights are gained from best practices in MGMP" (Teacher 7).

Another stated:

> "The initial transformation was very difficult because everything had to be prepared independently, and learning resources were still very limited" (Teacher 8).

The findings support the constructivist learning theory from Piaget (1950) and Vygotsky (1978), highlighting the necessity for ongoing professional development for teachers to effectively promote student-centered learning [20,31]. The variation in training outcomes underscores the significance of scaffolding, indicating that professional learning communities must offer structured and flexible guidance to assist teachers in effectively internalizing critical thinking instruction. Figure 10 presents the results of co-document analysis compared to group codes derived from Atlas.ti.

	A Critique of the Implementation of Critical Thinking... 8 9	Adaptation of the Merdeka Curriculum 6 7	Efforts to improve teacher capacity 13 14	The challenge of teacher readiness 6 6	Totals
1: Interview... 9	2,33 20,38%	1,17 11,42%	3,5 16,99%		7 14,29%
2: Interview... 9	1,4 12,23%	1,4 13,70%	2,8 13,59%	1,4 20,79%	7 14,29%
3: Interview... 10	2,8 24,45%	1,4 13,70%	2,8 13,59%		7 14,29%
4: Interview... 10	2 17,47%	1 9,79%	1 4,85%	3 44,55%	7 14,29%
5: Interview... 13	1,75 15,28%	1,75 17,13%	3,5 16,99%		7 14,29%
6: Interview... 10	1,17 10,19%	1,17 11,42%	2,33 11,33%	2,33 34,65%	7 14,29%
7: Interview... 6		2,33 22,84%	4,67 22,65%		7 14,29%
Totals	11,45	10,22	20,6	6,73	49

Figure 10. The results of co-document analysis compared to group codes derived from Atlas.ti.

The qualitative findings underscore the significance of connecting training with classroom implementation. The Adaptation category of the Merdeka Curriculum indicates that teachers recognize the curriculum's flexibility; however, they need more explicit guidelines and contextualized examples for effective integration. This is consistent with recent findings from the OECD, which emphasize the necessity for structured guidance in competency-based curricula [36]. Findings indicate that motivation is significant, as teachers participating in professional learning communities exhibit greater confidence in implementing innovative strategies. This aligns with Vygotsky's Zone of Proximal Development (ZPD), indicating that teachers need sufficient support systems to move from guided instruction to independent curriculum adaptation [20,31].

4.3. Process evaluation

The bar chart in the data analysis section presents the percentage distribution of scores for each evaluated indicator. Figure 11 below summarizes the key findings derived from the evaluation aspects.

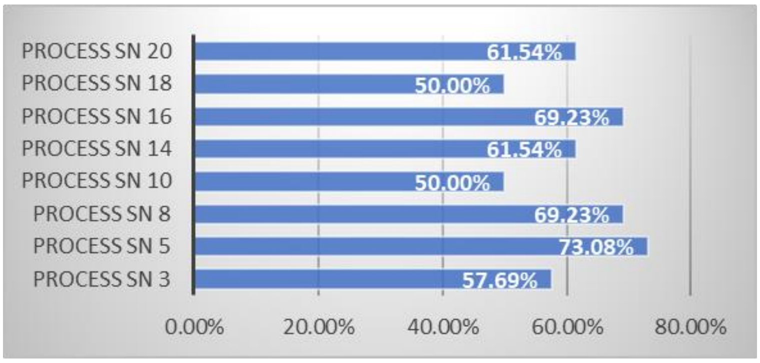

Figure 11. Distribution of scores for each assessed indicator in the process evaluation.

The findings from the process evaluation suggest that the incorporation of critical thinking skills into junior high school science lessons (Phase D) was moderately successful, achieving an average score of 61.54%. Teachers exhibited the highest effectiveness in facilitating practicum-based learning (69.23%, achieved), while their effectiveness in employing questioning techniques (54.18%, almost well achieved) and promoting deeper discussions among students (51.32%, almost well achieved) was comparatively lower. This indicates that although hands-on activities improve analytical reasoning, additional interactive approaches are necessary to enhance conceptual understanding. This is consistent with Marzano & Kendall's taxonomy, highlighting the significance of knowledge application and metacognitive involvement in the learning process [31].

	Efforts to conduct differentiated learning 5 5	Learning and differentiation challenges 10 10	The scientific approach is most dominantly used 7 8	Use of learning resources in learning... 2 2	Totals
1: Interview... 9		4 23,95%	2 13,61%		6 14,29%
2: Interview... 9		4,5 26,95%	1,5 10,20%		6 14,29%
3: Interview... 10	2,4 28,57%	1,2 7,19%	1,2 8,16%	1,2 54,55%	6 14,29%
4: Interview... 10		3 17,96%	3 20,41%		6 14,29%
5: Interview... 13	1 11,90%	2 11,98%	2 13,61%	1 45,45%	6 14,29%
6: Interview... 10	2 23,81%	2 11,98%	2 13,61%		6 14,29%
7: Interview... 6	3 35,71%		3 20,41%		6 14,29%
Totals	8,4 100,00%	16,7 100,00%	14,7 100,00%	2,2 100,00%	42 100,00%

Figure 12. Results of the co-document compared to the group codes table.

The values indicate the proportion of responses reflecting the presence of these factors among various interviewees. The findings indicate that although efforts to implement differentiated learning were present, their success varied across different contexts. The prevalence of the scientific approach indicates a systematic, albeit somewhat inflexible, methodology, underscoring the need for more adaptable strategies to promote critical thinking. Interaction between teachers and students, as well as the classroom environment, are critical components of the educational process. Data from interviews indicated that large class sizes and low student engagement hindered effective implementation. A teacher stated:

> "Students struggle to conclude. I frequently inquire, 'What did you learn today?' yet their responses remain insufficient" (Teacher 6).

Another observed that the scientific approach emphasizes process skills. Teacher 2 asserts that the implementation of learning media and questioning strategies markedly enhances active engagement in students. Vygotsky's scaffolding theory suggests that teachers should first model questioning strategies, and then gradually transfer responsibility to students [20]. The zone of proximal development (ZPD) emphasizes the significance of guided participation in fostering critical thinking [20].

The findings demonstrate that differentiated learning, highlighted in the 2024 Curriculum Guide by the Ministry of Education, Culture, Research, and Technology, is crucial within the classroom setting. Differentiated instruction enables teachers to modify their teaching strategies according to the varied needs of students, thereby ensuring the achievement of learning objectives [27]. Teachers indicated difficulties in implementing it due to constraints related to class size. An educator noted that the primary challenge lies in differentiated instruction. While there has been progress in differentiating content and assessment difficulty levels, the differentiation of processes remains difficult, particularly in large class settings. Diagnostic tests, encompassing both subject-specific and non-academic assessments, are conducted to evaluate students' abilities.

4.3.1. Students' engagement and inquiry-based learning

Despite the implementation of differentiated learning strategies, teachers observed that student engagement was still variable. The absence of active participation was linked to restricted students-initiated inquiry, as numerous classroom discussions remained predominantly teacher-directed. A teacher noted that the varied conditions of students significantly impact differentiation strategies. This presents a challenge; for instance, when differentiating content or products, adjustments must be made according to class size and students' conditions. The findings are consistent with OECD findings, which emphasizes that effective education systems foster inquiry-based and students-centered learning to enhance reasoning skills [35]. Analysis indicates that teachers encounter significant challenges in specific competencies related to critical thinking instruction:

1. Creating open-ended questions that encourage higher-order thinking
2. Promoting dialogues that stimulate students to question assumptions
3. Offering support that progressively enhances students' autonomy in reasoning

Recent studies emphasize the importance of active learning methods, including peer assessment and collaborative inquiry, in improving students'' critical thinking skills [6]. This study demonstrates that instructional strategies significantly influence the development of students' analytical skills, particularly within Southeast Asian contexts. The efficacy of these methods is corroborated by OECD findings, which highlight the significance of inquiry-based instruction in enhancing students' reasoning skills across various educational systems [36].


Figure 13. Implementation of learning in a selected school.


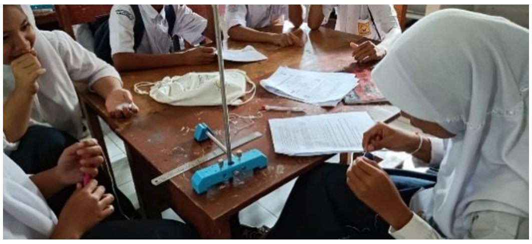

Figure 14. Implementation of learning in a selected school.

A critical thinking observation tool for secondary education offers a systematic method for analyzing classroom interactions, highlighting the necessity for standardized frameworks in assessing instructional effectiveness [18]. Document analysis indicated that although classroom activities included students' discussions, these were predominantly teacher-led, rather than fostering student-driven inquiry, underscoring the necessity for a more student-centered dialogue.

According to Dewey's theory (1933), the lack of sufficient exposure to real-world applications restricts students' capacity to link conceptual understanding with practical implementation [20]. Additionally, teachers who exhibit a lack of confidence in implementing inquiry-based instruction may revert to conventional, teacher-centered approaches, thereby diminishing opportunities for students engagement in critical thinking [36]. The study of critical thinking skills among Malaysian and Vietnamese undergraduates [6] highlights the impact of cultural and educational contexts on cognitive development. The findings highlight the necessity of tailoring teacher training to specific contextual learning environments. The OECD findings indicate that nations with high-performing students in science education typically implement inquiry-based and students-centered learning methodologies. This underscores the necessity for Indonesia to enhance its pedagogical strategies through international best practices [38].

A document study revealed that although certain instructional materials incorporated case-based learning, most did not provide explicit guidance for promoting students' questioning of assumptions or analysis of multiple perspectives. This finding aligns with constructivist learning principles, emphasizing that knowledge is actively constructed through engagement with complex, real-world problems [20]. The implementation of constructivist teaching strategies, including problem-based learning and peer collaboration, may improve students' engagement in critical discourse and the development of higher-order thinking skills [36].

4.4. Product (outcome) evaluation

The average mastery of critical thinking skills among students was 55.79%, classified as "almost well achieved". Significant differences were observed among various skill domains. The national competency standard (Kriteria Ketuntasan Minimum, KKM) for science subjects is established at 75. An average score of 55.79 indicates that students have not achieved the necessary level of mastery. This underscores the necessity for focused instructional interventions. This result not only highlights a cognitive gap, but also underscores the significance of dispositional and metacognitive dimensions in critical thinking, which are essential for sustained reflective habits [19,31]. Addressing this limitation requires not just instructional redesign, but also the integration of innovative educational technologies. For instance, recent advancements in AI-powered platforms offer analytics for assessing argumentation patterns, while gamified tools foster self-monitoring and metacognitive awareness in students. Structured digital environments can guide students in problem identification and evaluation [39]. Similarly, a report from OECD underscores the role of digital literacy and adaptive feedback systems in supporting reflection and decision-making within student-centered curricula [35] such as Merdeka Belajar. These tools represent promising avenues to bolster reflective skills among students in more personalized and scalable ways. Recent studies have also demonstrated how AI-driven learning analytics and adaptive technologies can enhance students' critical thinking and reflective processes [2,4,13,14,40].

4.4.1. Acquisition and processing of information and ideas

The fundamental element of critical thinking, involving the collection and analysis of information and ideas, indicated that students demonstrated moderate competence in identifying and managing information (average score: 58.26%, categorized as "needs improvement"). Yet, students encountered significant challenges in assessing the relevance of scientific applications in practical contexts, achieving only 29%. The qualitative analysis revealed a persistent theme of "Challenges in Acquiring and Processing Information and Ideas" across the interviews and qualitative data. A participant remarked:

> "Question 1 was the clearest for me, as the phrase 'if the mouse population decreases' illustrated its consequences within the food chain." Students 1. Another individual noted, "I recalled that question 2 contained a clue in the tropical forest data, which aided my comprehension of the question" (Individual 2).

The findings align with the study of Rubenstein et al. (2020), who state that students utilize inductive tactics, including re-reading and identifying essential terms, to address challenges [39], which aligns with the problem identification component of Marzano's New Taxonomy [31]. Students demonstrated limited ability in crafting relevant scientific inquiries, attaining scores of 45.66% and 50.23% on different question items. The primary concern identified was the assessment of information about practical science applications, suggesting that students encountered difficulties in applying classroom knowledge to real-world contexts. Dewey's reflective thinking model (1933) emphasizes the importance of directly teaching reflection, enabling students to link theoretical concepts with real-world applications [20]. A contextual learning approach combined with problem-based instruction offers structured opportunities for students to engage in deeper inquiry.

	8: Interview_... 8	9: Interview_... 10	10: Interview... 9	11: Interview... 7	12: Interview... 6	13: Interview... 10	14: Interview... 9	Totals
Difficultie in Acquiring and Processing Information and Ideas ◇ 2 5	1,5 16,67%	1 11,11%	1,29 14,29%				2 22,22%	5,79 9,18%
Difficulties in Analyzing and Evaluating Reasoning and Procedures ◇ 2 9	3 33,33%	2 22,22%	1,29 14,29%			2,25 25,00%	2 22,22%	10,54 16,72%
Difficulties Reflecting on Thought Processes and Reasoning ◇ 1 3				1,8 20,00%	1,8 20,00%	1,13 12,50%		4,73 7,50%
Independent learning activity ◇ 2 7	1,5 16,67%	1 11,11%	1,29 14,29%	1,8 20,00%	1,8 20,00%	1,13 12,50%	1 11,11%	9,51 15,10%
Oppotunity to asking questions ◇ 2 5		2 22,22%	1,29 14,29%			1,13 12,50%	1 11,11%	5,41 8,59%
Perception about science subjects ◇ 2 6	1,5 16,67%	1 11,11%	1,29 14,29%	1,8 20,00%	1,8 20,00%	1,13 12,50%		8,51 13,51%
student's favourite teaching methods ◇ 3 14	1,5 16,67%	2 22,22%	2,57 28,57%	3,6 40,00%	3,6 40,00%	2,25 25,00%	3 33,33%	18,52 29,40%
Totals	9 100,00%	9 100,00%	9 100,00%	9 100,00%	9 100,00%	9 100,00%	9 100,00%	63 100,00%

Figure 15. Results from the co-document and group codes table derived from the Atlas.ti.

Rogers' (1969) motivational theory of experiential learning, alongside contextualized and problem-based learning approaches, elucidates how meaningful learning can enhance students' engagement [20]. Thus, highlighting learning relevant to students' lives, which is self-initiated and permits self-evaluation, will significantly enhance their intrinsic motivation to engage in critical thinking and deeper information processing.

4.4.2. Analysis and evaluation of reasoning and procedures

The second component of critical thinking, which involves analyzing and evaluating reasoning and procedures, demonstrated an average mastery level of 56.48%, indicating it was nearly well achieved. Students encountered difficulties in identifying core problems, with rates of 41.55% and 44.29%, as well as in making informed decisions based on the provided information, at 47.03%. These challenges indicate that students require explicit instruction in logical reasoning and systematic problem analysis. Marzano's New Taxonomy states that analytical reasoning is a multidimensional cognitive process necessitating scaffolding through directed inquiry and demonstration [31]. The qualitative analysis shown in Figure 17 demonstrates that the theme 'Difficulties in Analyzing and Evaluating Reasoning and Procedures' represents a notable challenge. Students faced difficulties in:

1. Identifying the primary issue of the problem
2. Making decisions based on the available information

In contrast, students exhibited an adequate capacity to conclude the provided data, attaining scores of 57.76% and 68.95%. This indicates that although students can summarize knowledge, they may not consistently assess its accuracy critically. A student remarked:

> "I frequently encounter diseases like acid reflux and diabetes, yet I have not completely understood their symptoms" (Students 3),

This finding emphasizes the distinction between rote learning and profound comprehension. This corroborates results indicating that scientific reasoning requires explicit training instead of being presumed to evolve naturally [6]. Moreover, students exhibited a notable skill in decision-making (81.28%), reflecting their ability to choose suitable replies when confronted with structured options. Their ability to analyze reasoning errors showed significant variation across various question items, with percentages of 34.02% and 60.50%, respectively. This discrepancy suggests that students can recognize faulty reasoning in familiar contexts but struggle to apply this skill to unfamiliar problems. The importance of explicit learning strategies in scientific reasoning, allowing students to move beyond mere memorization of concepts to engage in comprehensive analysis [30,41].

4.4.3. Analysis of cognitive processes and logical reasoning

The third component, reflection and metacognition, attained an overall score of 50.40%, signifying that students can assess their cognitive processes to some extent, but require additional support for systematic reflection. An indication evaluating the ability to clarify underlying assumptions received a moderate score of 68.26%, indicating that students can express implicit reasoning when the topics are familiar. In contrast, a second comparable indicator, requiring systematic thinking based on an experiment, produced a significantly lower score of 21.23%. This indicates that students struggle to identify implicit assumptions in scientific investigation and controlled variables.

Interviews with students validated this discovery. A participant exhibited confidence in responding to a question on an electrical circuit, stating:

> "I recognized that the essential factor in answering this was the nail's ability to conduct electricity" (Students 4).

Another student misinterpreted the same concept, asserting:

> "I thought the lamp would be brighter because of the nail" (Students 5).

The responses demonstrate the necessity for organized hypothesis-testing efforts, as outlined in Marzano's Knowledge Utilization framework. The results align with OECD research, demonstrating that explicit education in metacognitive methods enhances introspection and self-regulation [42]. Integrating structured peer discussions, reflective journaling, and debate-based activities into science instruction may enhance students' capacity to critically assess their reasoning processes. Rogers' (1969) theory posits that meaningful learning with personal relevance enhances students' ability to engage in reflection and self-evaluation [20].

Thus, students demonstrated competence in gathering information and making decisions, but they still need to improve their reasoning evaluation, error analysis, and reflective thinking skills. The results highlight the importance of educational systems that explicitly develop critical thinking abilities. The integration of inquiry-based and problem-based learning methodologies enhances students' ability to apply scientific principles in practical contexts [18]. Additionally, scaffolding strategies, particularly those integrated into inquiry-based science learning, may help foster student engagement and reasoning [2,31]. A structured, evidence-based learning approach is necessary for enhancing students' critical thinking skills and equipping them for future academic and professional challenges. Being unable to address deficient critical thinking skills could affect students' problem-solving abilities and career readiness, as well as constrain their adaptability to the changing demands of the workforce and higher education.

5. Conclusions, implications, and recommendations

5.1. Conclusions

This study's results indicate that the Merdeka Curriculum implementation is shown to be effective in enhancing students' critical thinking skills, particularly when applied consistently and supported by appropriate scaffolding. The process and evaluation factors of critical thinking demonstrate positive outcomes in science learning. Nonetheless, challenges remain in the implementation of project-based learning and the integration of contextual approaches that require further enhancement. Resource limitations and the necessity to enhance teachers' competencies are factors influencing the effectiveness of curriculum implementation. Consequently, systematic enhancements are required to optimize students' outcomes.

5.2. Implications

The results of this study hold significant implications for educational policy, institutions, and instructors, particularly for the implementation of the deep learning methodology. This study's findings underscore the necessity for policymakers to enhance support for teachers in the execution of project-based learning and critical thinking, which are integral to the framework of deep learning. This research indicates that the effective execution of this strategy in schools is significantly contingent upon teachers' preparedness and the provision of sufficient educational materials. As schools migrate to deeper learning practices, it is imperative to maintain and adapt instructional methods that have demonstrated efficacy in fostering critical thinking. This research offers teachers insights into the significance of innovative learning strategies to enhance the effectiveness of fostering critical thinking among students in response to evolving educational methodologies. Through ongoing enhancement, the education system is anticipated to become more effective in attaining its objectives.

5.3. Recommendations

Based on the research findings, several recommendations can be proposed to enhance the efficacy of the Merdeka Curriculum implementation. Teacher training must emphasize inquiry-based learning methodologies and Socratic questioning to enhance students' critical thinking abilities. The incorporation of technology in education, including AI-driven learning aids and gamification, is anticipated to enhance student engagement. Yet, its execution must correspond with the school's infrastructure and digital preparedness, guaranteeing equitable access for all students. Furthermore, the integration of AI in education must maintain academic integrity and ethical standards, ensuring it functions as a tool to enhance learning rather than supplant essential cognitive processes. Third, educational authorities and institutions need to ensure the availability of learning resources that facilitate the application of the Merdeka Curriculum, including interactive modules and access to online learning platforms. Future studies should incorporate structured observations and peer assessments to evaluate how critical thinking skills are more accurately applied in classroom settings. Additionally, students' evaluation may include diagnostic evaluations for project-based critical thinking abilities and analysis of assignment portfolios. Moreover, educational institutions and authorities may conduct curriculum evaluations, and frequent assessments using measurable success criteria, such as improved instructor proficiency and student assessment results.

6. Limitations and future research

6.1. Limitations

This study employed an evaluative methodology utilizing the CIPP model, incorporating teacher interviews, analysis of lesson plan documents, and classroom observations for data triangulation. Nonetheless, certain methodological constraints must be acknowledged. The observation could not be comprehensively executed due to the time constraints imposed by the school, resulting in incomplete documentation of certain aspects of the learning dynamics. Second, this study predominantly depends on data derived from teacher and students reports, which may exhibit perceptual bias. To mitigate this potential bias, the research corroborates the findings using document analysis and comprehensive interviews. The scope of this study is confined to schools in Mataram City, hence the findings may not be entirely generalizable to other regions with distinct educational environments.

6.2. Future research

Future research should undertake a longitudinal assessment of the Merdeka Curriculum implementation to ascertain its enduring effects on students' critical thinking abilities. Cross-regional investigations are essential to examine how varying educational environments influence the effective execution of this curriculum. Moreover, subsequent studies could enhance the analysis through a more comprehensive mixed-method approach, incorporating prolonged classroom observations and experimental experiments to evaluate the efficacy of learning methodologies. Future research should adopt experimental or quasi-experimental designs to compare the effectiveness of various instructional approaches, such as inquiry-based learning versus traditional methods, or to test the impact of specific scaffolding strategies like structured discussions or technology-integrated tasks. Such designs would allow for a more rigorous evaluation of causality and better align with the evidence-based intervention focus of STEM education.

Author contributions

Nabilah Nabilah: Conceptualization, Formal analysis, Resources, Writing – original draft； Slamet

Suyanto: Conceptualization, Supervising, Reviewing. All authors have read and approved the final version of the manuscript for publication.

Use of Generative-AI tools declaration

The authors declare that we did not use Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgement

We sincerely thank the reviewers for their constructive and detailed feedback. Their insights helped us clarify our definitions, tighten our argumentation, and improve the structure of the manuscript.

Conflict of interest

The authors declare there is no conflict of interest in any part of this article.

Ethics declarations

The author declares that the research process and data collection were approved by Yogyakarta State University, Indonesia and approved by the Education Department of Mataram City.

References

1. Kong, M. and Pek, L., Impact of critical thinking skills on learning outcome: a review of researchers' findings. *Quantum J. Soc. Sci. Humanit.*, 2023, 4(6): 28–36. https://doi.org/10.55197/qjssh.v4i6.297
2. Zhang, Z., Cao, T., Shu, J. and Liu, H., Identifying key factors affecting college students' adoption of the e-learning system in mandatory blended learning environments. *Interact. Learn. Environ.*, 2022, 30(8): 1388–1401. https://doi.org/10.1080/10494820.2020.1723113
3. Han, S., Zhang, Z., Liu, H., Kong, W., Xue, Z., Cao, T., et al., From engagement to performance: the role of effort regulation in higher education online learning. *Interact. Learn. Environ.*, 2024, 32(10): 6607–6627. https://doi.org/10.1080/10494820.2023.2273486
4. Liu, T., Wang, M., Yang, B., Liu, H. and Yi, S., ESERNet: Learning spectrogram structure relationship for effective speech emotion recognition with swin transformer in classroom discourse analysis. *Neurocomputing*, 2025, 612: 128711. https://doi.org/10.1016/j.neucom.2024.128711
5. Altun, E. and Yildirim, N., What does critical thinking mean? Examination of pre-service teachers' cognitive structures and definitions for critical thinking. *Think. Ski. Creat.*, 2022, 49: 101367. https://doi.org/10.1016/j.tsc.2023.101367
6. Van Le, H. and Chong, S.L., The dynamics of critical thinking skills: a phenomenographic exploration from malaysian and vietnamese undergraduates. *Think. Ski. Creat.*, 2023, 51: 101445. https://doi.org/10.1016/j.tsc.2023.101445
7. Ministry of Education, Culture, Research and Technology, *Panduan pembelajaran dan asesmen pendidikan anak usia dini, pendidikan dasar, dan menengah*, 2022.
8. Curriculum Standards and Education Assessment Board, *Kajian Akademik Kurikulum untuk*

Pemulihan Pembelajaran, 2022.
9. Riyan Rizaldi, D. and Fatimah, Z., Merdeka curriculum: characteristics and potential in education recovery after the COVID-19 pandemic conditions of the creative commons attribution license. *Int. J. Curric. Instr.*, 2022, 15(1): 260–271.
10. Warsihna, J., Ramdani, Z., Amri, A., Kembara, M.D., Steviano, I., Anas, Z., et al., Tantangan dan strategi implementasi kurikulum merdeka pada jenjang sd: sebuah temuan multi-perspektif. *Kwangsan J. Teknol. Pendidik.*, 2023, 11(1): 296–311. https://doi.org/10.31800/jtp.kw.v11n1.p296--311
11. Pratikno, Y., Hermawan, E. and Arifin, A.L., Human resource 'Kurikulum Merdeka' from design to implementation in the school: what worked and what not in Indonesian education. *J. Iqra' Kaji. Ilmu Pendidik.*, 2022, 7(1): 326–343. https://doi.org/10.25217/ji.v7i1.1708
12. Rohmah, Z., Hamamah, H., Junining, E., Ilma, A. and Rochastuti, L.A., Schools' support in the implementation of the Emancipated Curriculum in secondary schools in Indonesia. *Cogent Educ.*, 2024, 11(1): 2300182. https://doi.org/10.1080/2331186X.2023.2300182
13. Zhang, Z., Zhang, E., Liu, H. and Han, S., Examining the association between discussion strategies and learners' critical thinking in asynchronous online discussion. *Think. Ski. Creat.*, 2024, 53: 101588. https://doi.org/10.1016/j.tsc.2024.101588
14. Yang, B., Xie, C., Liu, T., Xu, J. and Li, W., Exploring the relationship between teacher talk supports and student engagement from the perspective of students' perceived care. *Interact. Learn. Environ.*, 2024, 32(10): 6393–6412. https://doi.org/10.1080/10494820.2023.2263045
15. Wang, X., Liu, T., Wang, J. and Tian, J., Understanding learner continuance intention: a comparison of live video learning, pre-recorded video learning and hybrid video learning in COVID-19 Pandemic. *Int. J. Human–Computer Interact.*, 2022, 38(3): 263–281. https://doi.org/10.1080/10447318.2021.1938389
16. Ministry of Education, Culture, Research and Technology, *Panduan Pembelajaran dan Asesmen Pendidikan Anak Usia Dini, Pendidikan Dasar, dan Pendidikan Menengah Edisi Revisi Tahun 2024*, Revisi. 2024.
17. Ariesanti, D., Mudiono, A. and Arifin, S., Analisis implementasi kurikulum merdeka dan perencanaan pembelajaran di sekolah dasar. *SENTRI J. Ris. Ilm.*, 2023, 2(6): 1896–1907. https://doi.org/10.55681/sentri.v2i6.995
18. Shafer, D., A critical thinking thematic framework and observation tool for improved theory and developing secondary teachers' instructional practice: Proof of concept. *Think. Ski. Creat.*, 2025, 56: 101787. https://doi.org/10.1016/j.tsc.2025.101787
19. Lau, J.Y.F., *An Introduction to Critical Thinking and Creativity: Think More, Think Better*, John Wiley & Sons, Inc, 2011. https://doi.org/10.1002/9781118033449
20. Schunk, D.H., *Learning Theories An Educational Perspective*, 6th ed. Boston: Pearson, 2012. https://doi.org/10.1007/BF00751323
21. Zainudin, A., Vianty, M. and Inderawati, R., The practice and challenges of implementing critical thinking skills in EFL teachers questioning behavior. *English Rev. J. English Educ.*, 2019, 8(1): 51–58. https://doi.org/10.25134/erjee.v8i1.2112
22. Sari, M.P., Muttaqiin, A., Putri, R.E. and Oktavia, R., Integrating ethnoscience on critical-thinking oriented web-based e-module of secondary school science. *J. Penelit. Pendidik. IPA*, 2024, 10(1): 371–384. https://doi.org/10.29303/jppipa.v10i1.5928
23. Eze, I.F., Iwu, C.G. and Dubihlela, J., Lecturers' perspectives concerning the variables that hinder critical thinking development in the classroom. *Int. J. Res. Bus. Soc. Sci.*, 2022, 11(10): 341–349. https://doi.org/10.20525/ijrbs.v11i10.2232
24. Liao, W., Liu, W. and Yuan, R., How is teachers' critical thinking defined, approached, measured,

and evaluated in empirical studies? A methodological review. *Think. Ski. Creat.*, 2024, 57: 101795. https://doi.org/10.1016/j.tsc.2025.101795
25. Stufflebeam, D.L. and Coryn, C.L., *Evaluation theory, models, & applications*, 2nd ed. John Wiley & Sons, 2014.
26. Creswell, J.W., *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*, Fourth Edi. Pearson, 2012.
27. Ministry of Education, Culture, Research and Technology, *Panduan Pengembangan Kurikulum Satuan Pendidikan Edisi Revisi Tahun 2024*, Edisi Revi. 2024.
28. Curriculum Standards and Education Assessment Board, *Tahapan Implementasi Kurikulum Merdeka di Satuan Pendidikan*, 2022.
29. Ministry of Education, Culture, Research and Technology, Riset, *Dimensi, Elemen, dan Subelemen Profil Pelajar Pancasila pada Kurikulum Merdeka*, 2022.
30. Ma, S., Tiruneh, D.T. and Spector, J.M., Critical thinking conceptualization in K-12: A case study of middle school teachers. *Soc. Sci. Humanit. Open*, 2023, 8(1): 100517. https://doi.org/10.1016/j.ssaho.2023.100517
31. Marzano, R.J. and Kendall, J.S. eds., *The new taxonomy of educational objectives*, Corwin Press, 2006.
32. Irene, E.A., Evaluation of teacher education curricula and its relevance to licensure examination using context, input, process and product (CIPP) model. *Soc. Sci. Humanit. Open*, 2023, 8(1): 100607. https://doi.org/10.1016/j.ssaho.2023.100607
33. Zhao, Y., Li, W., Jiang, H., Siyiti, M., Zhao, M., You, S., et al., Development of a blended teaching quality evaluation scale (BTQES) for undergraduate nursing based on the context, input, process and product (CIPP) evaluation model: A cross-sectional survey. *Nurse Educ. Pract.*, 2024, 77: 103976. https://doi.org/10.1016/j.nepr.2024.103976
34. Watkins, M., *A Step-by-Step Guide to Exploratory Factor Analysis with R and RStudio*, 2020. https://doi.org/10.4324/9781003120001
35. OECD, *PISA 2022 Results (Volume II): learning during – and from – disruption*, 2023. https://doi.org/10.1787/a97db61c-en
36. OECD, *PISA 2022 Results (Volume I): the state of learning and equity in education*, Paris, 2023. https://doi.org/10.1787/53f23881-en
37. Kapitariyani Kimpo Ellen and Lala Bumela Sudimantara, Examining emancipated curriculum development in middle schools: a case study. *Panyonara J. English Educ.*, 2023, 5(2): 165–188. https://doi.org/10.19105/panyonara.v5i2.8779
38. OECD, *Curriculum flexibility and autonomy: promoting a thriving learning environment*, Paris: OECD Publishing, 2024. https://doi.org/10.1787/eccbbac2-en
39. Rubenstein, L.D.V., Callan, G.L., Speirs Neumeister, K. and Ridgley, L.M., Finding the problem: How students approach problem identification. *Think. Ski. Creat.*, 2020, 35: 100635. https://doi.org/10.1016/j.tsc.2020.100635
40. Lee, N.Y., Wang, Z. and Lim, B., The development of critical thinking: what university students have to say. *Teach. High. Educ.*, 2021, 1(14): 1–14. https://doi.org/10.1080/13562517.2021.1973412
41. Bhaw, N., Kriek, J. and Lemmer, M., Insights from coherence in students' scientific reasoning skills. *Heliyon*, 2023, 9(7). https://doi.org/10.1016/j.heliyon.2023.e17349
42. OECD, *Adapting curriculum to bridge equity gaps*, Paris: OECD Publishing, 2021. https://doi.org/10.1787/6b49e118-en

Author’s biography

M.Ed. Nabilah Nabilah is a master's graduate in educational research and evaluation with a concentration in policy research and evaluation from Universitas Negeri Yogyakarta, Indonesia. Her research interests include education, science education, and educational program evaluation.

Prof. Dr. Slamet Suyanto, M.Ed is a professor in the educational research and evaluation program at Universitas Negeri Yogyakarta, Indonesia. His research interests include science education, learning, curriculum, and biology education. He has extensive experience in various aspects of science education and educational research, and is actively engaged in conducting research and publishing scholarly work in the field of education.

Appendix

A.1. Interview guideline for teachers.

Evaluation Aspects	Evaluation Sub-Aspects	No	Questions for Teachers
Context	• Identification of the objectives of the critical thinking dimension contained in the Merdeka Curriculum • Understanding of the dimensions of critical thinking, elements, and sub-elements contained in the critical thinking dimension is adjusted to the context of science learning in Phase D to facilitate students' needs to have critical thinking skills in public junior high schools in Mataram City.	1	What is your comprehension of the Merdeka Curriculum?
		2	What is your interpretation of the critical thinking aspect within the Pancasila students’ profile?
		3	To what extent have you incorporated the critical thinking aspect into Phase D scientific education?
		4	What is the demographic profile of the students you instruct? Do they possess diverse critical thinking abilities?
		5	What methods can be employed to ascertain requirements for the implementation of the Merdeka Curriculum to enhance the development of critical thinking skills in students within the framework of Phase D scientific education?
Input	• Examining the methodology employed to integrate the components and subcomponents of students'' critical thinking dimensions in Phase D science education. • Examining the learning strategies employed to enhance critical thinking skills in Phase D science education. • The preparedness of human resources and educational materials to facilitate the execution of the critical thinking aspect of Phase D science. • The Merdeka Curriculum policy encompasses the design aspects related	6	How explicit is the direction you received concerning the Phase D science learning outcomes?
		7	What resources do you utilize for instruction, and how do they facilitate the enhancement of critical thinking abilities?
		8	Do you employ diverse pedagogical methods to assist students with deficient critical thinking abilities? In what manner?
		9	How do you tailor your instructional methods to accommodate the unique demands of students with varying abilities?
		10	In what ways do training and professional development enhance your ability to teach critical thinking skills?

	to critical thinking dimensions and learning outcomes in Phase D science.		
Process	Evaluating the effectiveness of the techniques devised and executed to enhance critical thinking skills in the Phase D science education process within public junior high schools in Mataram City.	11	Which pedagogical approaches do you employ to foster critical thinking among students? Could you provide examples?
		12	What methods do you employ to assess pupils' critical thinking abilities throughout the educational process? Are these assessments formal (tests, quizzes) or informal (observations, discussions)?
		13	What is the primary problem you encounter in instructing critical thinking skills? How do you surmount them?
		14	How do you evaluate the effectiveness of your methods in enhancing pupils' critical thinking abilities?
		15	Comprehensive interview concerning a specific facet of deficient critical thinking

A.2. Interview guideline for teachers.

Evaluation Aspects	Evaluation Sub-Aspects	No	Questions for Students
Product (Outcome)	Assess the attainment of skills according to the components and subcomponents within the critical thinking dimension.	1	In what manner did the instruction you received enhance your critical thinking abilities? Could you provide some instances from your experience?
		2	Did you feel compelled to inquire and critically evaluate material during class? Could you provide instances in which you were motivated or a lack thereof?
		3	How do you comprehend the subjects examined in class? What techniques or instruments did you employ to enhance your comprehension?
		4	Did the instructional approaches employed in the classroom facilitate your comprehension of the lesson? What are the reasons for or against?
		5	What actions do you take when you struggle to comprehend the lesson?
			Comprehensive interview concerning a facet of deficient critical thinking