Prisma Sains: Jurnal Pengkajian Ilmu dan Pembelajaran Matematika dan IPA IKIP Mataram
https://e-journal.undikma.ac.id/index.php/prismasains/index
e-mail: prismasains.pkpsm@gmail.com

July 2026. Vol. 14, No. 3
p-ISSN: 2338-4530
e-ISSN: 2540-7899
pp. 1612-1626

Problem-Based Learning Effect on Elementary Students Critical Thinking with Science Literacy as Covariate

[1]*Mei Indra Jayanti, [1]Muh. Nasir, [2]Umar

[1]Biology Education Department, Faculty of Mathematics and Science Education, Universitas Nggusuwaru, Kota Bima, Indonesia
[2]Department of Pedagogy Post-Graduate, Universitas Muhammadiyah Bima, Bima, Indonesia

*Corresponding Author e-mail: meiindra15@gmail.com

Received: May 2026; Revised: June 2026; Published: July 2026

Abstract

Critical thinking skills among elementary school students in Indonesia remain at a suboptimal level, as evidenced by the 2022 PISA science literacy score of 383, substantially below the OECD average of 485. Despite a growing body of research on Problem-Based Learning (PBL) in Indonesian science education, studies that statistically control for students' initial cognitive profiles when estimating PBL's effect on critical thinking remain scarce at the elementary level, limiting the precision and credibility of existing effect estimates. This study aimed to examine the effect of PBL on the critical thinking skills of fifth-grade students at SDN 33 Lampe, Bima City, Indonesia, with science literacy as a covariate. A quasi-experimental post-test only control group design was employed, with two classes selected through cluster random sampling and randomly assigned to conditions: Class V.A (n = 31) as the PBL group and Class V.B (n = 29) as the control group receiving conventional instruction. Science literacy data were obtained from a cognitive diagnostic assessment administered before the intervention as an independent covariate, while critical thinking was measured using a validated essay instrument covering six dimensions of Facione (2015). Data were analyzed using Analysis of Covariance (ANCOVA) with estimated marginal means reported for each group. Results indicated a highly significant effect of instructional model on critical thinking ($F(1, 57) = 54.868$, $p < .001$, partial $\eta^2 = .490$), with the PBL group obtaining an adjusted mean of 17.71 points higher than the control group after controlling for science literacy. Science literacy also demonstrated a significant covariate effect (partial $\eta^2 = .178$). Within the scope and constraints of this single-school study, PBL was associated with significantly higher critical thinking scores compared to conventional instruction.

Keywords: Problem-based learning; Critical thinking; Science literacy; Covariate; Diagnostic assessment

How to Cite: Jayanti, M. I., Nasir, M., & Umar, U. (2026). Problem-Based Learning Effect on Elementary Students Critical Thinking with Science Literacy as Covariate. *Prisma Sains: Jurnal Pengkajian Ilmu Dan Pembelajaran Matematika Dan IPA IKIP Mataram*, *14*(3), 1612–1626. https://doi.org/10.33394/j-ps.v14i3.20580

https://doi.org/10.33394/j-ps.v14i3.20580

INTRODUCTION

Science education in primary schools plays a foundational role in fostering students' scientific reasoning, curiosity, and analytical abilities from an early age. Among the various competencies targeted in modern science curricula, critical thinking skills have been widely recognized as one of the most essential competencies for the 21st century (Ahmad & Siew, 2021; Yesnazar et al., 2024). Critical thinking facilitates students' ability not only to comprehend scientific concepts but also to critically evaluate evidence, construct coherent arguments, and make well-informed decisions in their daily lives (Ahmad & Siew, 2021; Dwyer, 2023; Sarwanto et al., 2021). Problem-Based Learning (PBL) has emerged as one of the most promising instructional models for addressing these challenges. Rooted in the principles of constructivist theory, PBL engages students in solving complex and ill-structured problems that reflect real-world phenomena, thereby stimulating active reasoning and collaborative inquiry (Aidoo, 2023; Doolittle et al., 2023; Santos et al., 2021).

However, the facts show that the critical thinking skills of elementary school students in Indonesia are still not at an optimal level, as reflected in Indonesia's performance on the Programme for International Student Assessment (De Bortoli et al., 2023; OECD, 2023). The 2022 PISA data show that Indonesian students' science literacy scores remain well below the OECD average, with an average score of 383 compared to the OECD average of 485 (OECD, 2023). These results indicate a persistent gap in students' ability to apply scientific knowledge in real-life contexts, which is closely linked to critical thinking skills (Hu & Bi, 2024; García-Carmona, 2023). This condition is exacerbated by the continued dominance of conventional instructional approaches in Indonesian classrooms, which tend to emphasize rote memorization over inquiry-based and problem-solving activities, thereby limiting students' opportunities to develop higher-order thinking skills (Anggraeny & Khongput, 2022; Kosasih et al., 2021). These conditions collectively underscore the urgency of adopting evidence-based instructional models such as PBL that can actively stimulate higher-order thinking and bridge the gap between current classroom practice and the competency demands of the 21st century.

Despite the growing body of literature on PBL in Indonesian science education, few studies at the elementary level have examined the effect of PBL on students' critical thinking while simultaneously controlling for students' prior science literacy using cognitive diagnostic assessment data. Most existing studies rely on simple pre-test post-test comparisons or independent samples t-tests, which do not account for heterogeneity in students' initial cognitive profiles. This results in effect estimates that may overstate or understate the true instructional contribution of PBL (Wijnen et al., 2017; Leppink & Pérez-Fuster, 2016). The present study addresses this gap by using science literacy data from a cognitive diagnostic assessment mandated under Indonesia's Merdeka Curriculum as a covariate in ANCOVA, thereby producing a statistically more precise and credible estimate of PBL's effect on critical thinking. The novelty of this study lies not in the analytical technique itself, but in the use of diagnostic assessment data as a scientifically justified covariate for evaluating the effect of PBL on critical thinking skills among fifth-grade elementary students, an approach that remains underutilized in the Indonesian elementary science education research context.

Building on the theoretical and empirical foundations established above, this study proposes that students receiving PBL instruction will demonstrate significantly higher critical thinking skills compared to those receiving conventional instruction after statistically controlling for individual differences in science literacy a hypothesis grounded in the constructivist premise that learning through authentic, ill-structured problems stimulates deeper cognitive engagement and the internalization of higher-order reasoning processes central to critical thinking (Hmelo-Silver, 2004; Savery, 2015; Yew & Goh, 2016). Furthermore, it is hypothesized that science literacy will function as a significant positive predictor of critical thinking, consistent with empirical evidence demonstrating that students' prior scientific knowledge and reasoning capacity substantially mediate their ability to engage in evidence-based argumentation and analytical thinking within problem-based instructional contexts (Fitriani et al., 2019; Norris & Phillips, 2003; Osborne, 2014). The adjusted mean difference between groups is expected to reflect a practically meaningful and statistically significant advantage for PBL, in line with prior intervention studies reporting moderate to large effect sizes for PBL on higher-order thinking outcomes in elementary science education (Abrami et al., 2015; Pratiwi et al., 2019; Sumarni & Kadarwati, 2020). The scope of this study is limited to fifth-grade students in a public elementary school in Bima City, Indonesia, with findings intended to inform instructional design decisions within comparable Indonesian primary school contexts operating under the Merdeka Curriculum framework.

Therefore, this study aimed to examine the effect of Problem-Based Learning on fifth-grade elementary students’ critical thinking skills after controlling for their initial science literacy. Specifically, this study addressed the following research questions:

(1) Does Problem-Based Learning significantly affect students' critical thinking skills after science literacy is controlled as a covariate?

(2) Does science literacy significantly contribute to students' critical thinking skills?

The novelty of this study lies in the use of cognitive diagnostic assessment data from the Merdeka Curriculum as a scientifically justified covariate in estimating the effect of PBL on critical thinking among elementary school students. This approach provides a more precise estimate of instructional effectiveness than unadjusted comparisons commonly used in previous studies.

METHOD

Research Design

This study employed a quasi-experimental design with a post-test-only control group. This design was chosen based on practical considerations in educational research conducted in naturalistic classroom settings, where random assignment at the individual level is often not feasible (Krishnan, 2019; Muse & Baldwin, 2021). Two fifth-grade classes within the same school were selected, with one assigned as the experimental group receiving instruction through Problem-Based Learning (PBL), and the other designated as the control group receiving conventional instruction. The absence of a pre-test on the dependent variable (critical thinking) was intentional. Instead, students' science literacy scores, obtained from a cognitive diagnostic assessment administered before the intervention, were treated as a covariate in the ANCOVA. This approach effectively controls for initial differences in students' abilities while minimizing the risk of testing effects on the critical thinking variable (Manassero-Mas & Vázquez-Alonso, 2022). This approach is consistent with the recommendation (Wan, 2020) which states that covariates measured before the intervention and independent of the treatment serve as valid pre-test substitutes and are statistically more efficient in ANCOVA designs. The research design can be seen in the Figure 1.

V.A	X	O_1
V.B	-	O_2

Figure 1. Post-test-only control group research design

Notes:
X : Implementation of PBL
O_1 : Post-test for the experimental class
O_2 : Post-test for the control class

The PBL intervention was implemented in the experimental group over six weeks periode (12 learning sessions, each lasting 70 minutes) on the topic of harmonious ecosystems, as outlined in the fifth-grade elementary school science curriculum. The control group received conventional direct instruction covering the same material, consisting of teacher-centered explanations, textbook exercises, and question-and-answer sessions. All research procedures were conducted with written permission from the school principal. Because participants were elementary school students, informed consent was obtained from parents or guardians before data collection, and student assent was also sought. Participant confidentiality was maintained throughout the study, and all data were used exclusively for research purposes in accordance with institutional ethical guidelines. With only one class per condition, it is not possible to fully disentangle PBL effects from class-level effects, and this constitutes a major limitation that constrains the generalizability and causal strength of the findings.

Population and Sample

This study was conducted at SDN 33 Lampe, Bima City, Indonesia, during the 2024/2025 school year. The target population consisted of all fifth-grade students enrolled in three parallel classes at SDN 33 Lampe (N = 88). Cluster random sampling was used to select two of the three available fifth-grade classes as the study sample, treating each entire class as a sampling unit (cluster). The selection was conducted using a lottery procedure, in which class labels were written on identical pieces of paper, placed in a container, and drawn by a neutral party, resulting in Class V.A and Class V.B as the two selected clusters. A second draw was then conducted to randomly assign the two selected classes to the experimental conditions, so that Class V.A was assigned to the experimental group receiving PBL instruction and Class V.B was assigned to the control group receiving conventional instruction.

Research Instruments

Science literacy scores used as covariates in this study were obtained from a cognitive diagnostic assessment administered to all students in both sample classes at the beginning of the semester, before the implementation of the treatment intervention. The diagnostic instrument comprehensively measured four dimensions of cognitive competence, namely: (1) numeracy, (2) reading literacy, (3) science literacy, and (4) social literacy, in accordance with the cognitive diagnostic assessment framework recommended by the Merdeka Curriculum (Kemendikbudristek, 2022). For this study, only the science literacy subscale scores were extracted for use as covariates in the ANCOVA model. The science literacy subscale consists of 20 items designed to measure students' ability to explain scientific phenomena, interpret scientific data, and evaluate scientific evidence in accordance with the PISA science literacy framework (OECD, 2023).

Critical thinking was measured using a post-test instrument of 10 open-ended essay items assessing six dimensions of critical thinking as operationalized by Facione (2015): interpretation (items 1–2), analysis (items 3–4), evaluation (items 5–6), inference (items 7–8), explanation (item 9), and self-regulation (item 10). Each item was scored using a four-point analytic rubric (0–3), yielding a total possible score of 30 points, converted to a 0–100 scale. The rubric specified behavioral performance criteria for each score level, and rater training was conducted before scoring.

Content validity was established through expert review by three validators: one specialist in elementary science content, one specialist in science education, and one specialist in educational measurement. The Content Validity Index (CVI) was calculated using the proportion of items rated as essential by at least two of three validators, yielding a mean scale-level CVI (S-CVI) of 0.87, exceeding the recommended threshold of 0.80 (Lynn, 1986; Polit & Beck, 2006). It is important to note that the S-CVI of .87 does not directly exceed the per-item critical CVR of .99 for $N = 3$ validators (Lawshe, 1975); rather, the validity decision is based on the S-CVI threshold recommended for small validator panels (Lynn, 1986). Construct validity was supported by the theoretical alignment of each item with the six-factor structure of Facione's (2015) critical thinking framework, as verified through expert review during the content validation process. Inter-rater reliability was assessed through independent double-scoring by two trained raters, yielding Cohen's $\kappa = .84$, indicating strong agreement (Landis & Koch, 1977). Internal consistency yielded Cronbach's $\alpha = 0.82$ (George & Mallery, 2003).

Data Analysis

Data were analyzed by using Analysis of Covariance (ANCOVA) with critical thinking post-test scores as the dependent variable, instructional group (PBL vs. conventional) as the independent variable, and science literacy scores as the covariate. The following ANCOVA assumptions were systematically verified before analysis: (1) normality of residuals, assessed using Kolmogorov-Smirnov tests; (2) homogeneity of variances, examined using Levene's tests; (3) homogeneity of regression slopes, tested by examining the Group × Science Literacy interaction term; (4) linearity of the covariate-dependent variable relationship; (5) independence of the covariate from treatment, established procedurally and verified by baseline equivalence testing; and (6) absence of extreme outliers, verified through inspection of standardized residuals (Stevens, 2016; Tabachnick & Fidell, 2019). An independent samples t-test was conducted on science literacy scores to examine baseline group equivalence on the covariate before the intervention. Effect sizes were reported using partial eta squared (η^2p), interpreted per Cohen (1988): small (0.01), medium (0.06), large (0.14). Following the significant ANCOVA result, estimated marginal means (adjusted means) with 95% confidence intervals were reported for each group to quantify the adjusted between-group difference after covariate control. All analyses were conducted using Jamovi version 2.3.21.0 with $\alpha = 0.05$.

RESULTS AND DISCUSSION

The results of the analysis of the content validity and reliability of the science literacy instrument indicate content validity, as confirmed by expert evaluations from two experts in elementary school science education and an expert in learning assessment, as shown in the Figure 2.

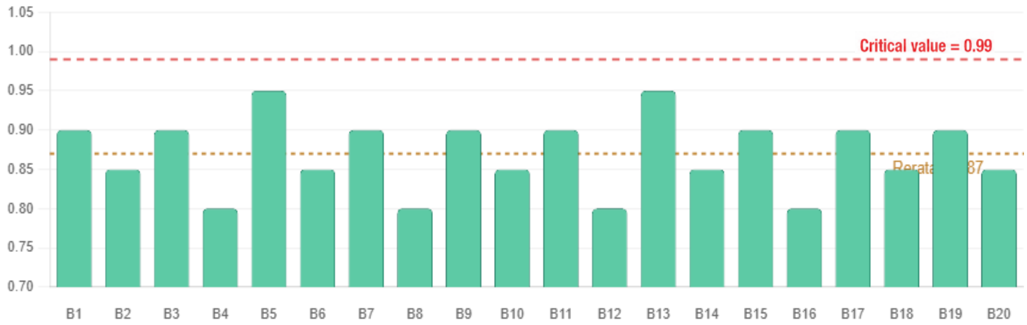

Figure 2. Distribution of CVR by science literacy item

The overall mean Content Validity Ratio (CVR) of the scientific literacy instrument was 0.87, exceeding the critical Content Validity Index (CVI) threshold of 0.80. This finding confirms that, collectively, the instrument items were judged to be relevant, representative, and essential by both expert validators for measuring students' scientific literacy. The distribution of CVR values across items, ranging from 0.80 to 0.95 as illustrated in Figure 2, indicates a high level of inter-validator consistency and the absence of items with substantially disputed relevance. These results confirm that the developed scientific literacy instrument possesses adequate content validity to function as a covariate measurement instrument in this study.

Meanwhile, the analysis of the instrument's reliability using Cronbach's alpha revealed satisfactory internal consistency (Cronbach's $\alpha = 0.78$). A summary of the test results is shown in Figure 3.

Figure 3. Reliability analysis of science literacy test items

Figure 3 illustrates the reliability and stability analysis of the scientific literacy instrument through the visualization of Cronbach's alpha coefficient values obtained when each of the 20 items is individually removed from the instrument (α if item deleted). The horizontal axis is constrained to the range of 0.75–0.79 to allow for a more precise observation of variability among values, while the vertical red line at $\alpha = 0.780$ represents the overall alpha coefficient of the instrument, serving as a reference point for comparison.

The testing of the critical thinking skills assessment instrument, as the dependent variable, incorporates both the item discrimination index and the item difficulty level. The results of the analysis are presented across three distinct color-coded zones corresponding to three categories: very good, good, and adequate.

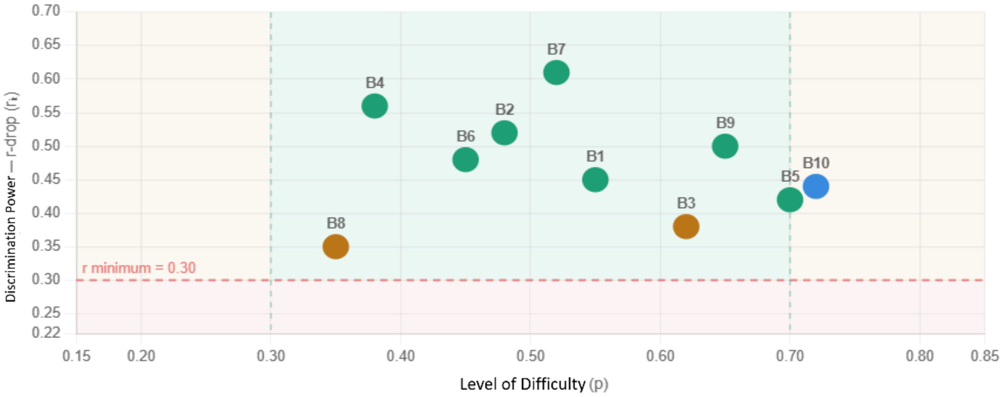

Figure 4. Psychometric Map of the CTS Instrument Test

Item analysis was conducted to examine the discrimination and difficulty properties of each item. As shown in Figure 4, discrimination indices ranged from r-it = 0.32 to r-it = 0.61, with all 10 items exceeding the minimum acceptable threshold of 0.30, indicating that each item adequately differentiates between students with higher and lower levels of critical thinking. Most items are concentrated within the green zone, which represents an optimal combination of strong discrimination ($r_{(it)} \geq 0.40$) and moderate difficulty ($0.30 \leq p \leq 0.70$), while a smaller proportion falls within the amber zone, indicating acceptable discrimination that still meets the minimum feasibility criteria. This distribution pattern is consistent with the reported range of discrimination indices ($r_{(it)}$ = 0.32–0.61) and collectively supports the conclusion that the instrument demonstrates adequate psychometric quality for use in research aimed at measuring elementary school students' critical thinking skills.

Content validity was established through expert review by three validators: one specialist in elementary science content, one specialist in science education, and one specialist in educational measurement. The Content Validity Index (CVI) was calculated using the proportion of items rated as essential by at least two of three validators, yielding a mean scale-level CVI (S-CVI) of 0.87, exceeding the recommended threshold of .80 (Lynn, 1986; Polit & Beck, 2006). It is important to note that the S-CVI of .87 does not directly exceed the per-item critical CVR of 0.99 for N = 3 validators (Lawshe, 1975); rather, the validity decision is based on the S-CVI threshold recommended for small validator panels (Lynn, 1986). Construct validity was supported by the theoretical alignment of each item with the six-factor structure of Facione's (2015) critical thinking framework, as verified through expert review during the content validation process. Inter-rater reliability was assessed through independent double-scoring by two trained raters, yielding Cohen's $\kappa = .84$, indicating strong agreement (Landis & Koch, 1977). Internal consistency yielded Cronbach's $\alpha = 0.82$ (George & Mallery, 2003).

Figure 5. Distribution of the CVR by item on the CTS test

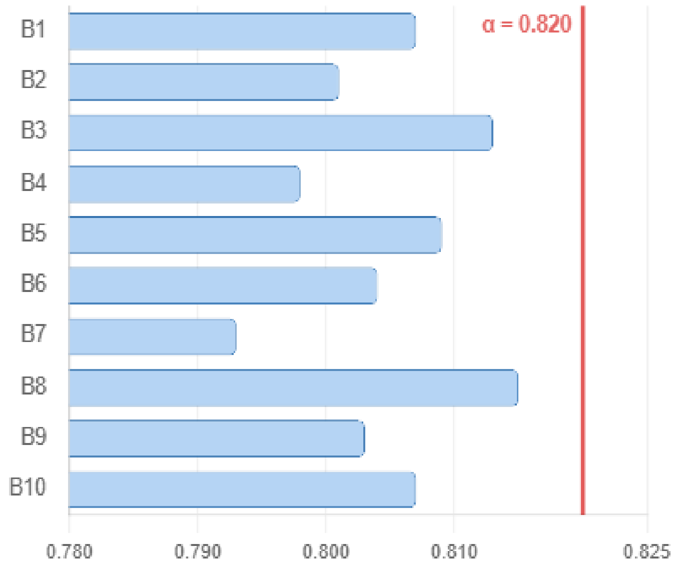

Figure 6. Item reliability analysis for the CTS

Scoring consistency was ensured through rater training and the use of a four-point analytic rubric with clearly specified behavioral criteria for each score level. Internal consistency of the critical thinking instrument was then assessed using Cronbach's alpha. As shown in Figure 6, the overall alpha for the 10-item scale was 0.82, indicating good internal consistency (George & Mallery, 2003). The red dashed line in Figure 6 represents this overall alpha value as a reference point. Inspection of the alpha-if-item-deleted values confirmed that no individual item, when removed, produced a meaningful increase in the overall alpha coefficient, suggesting that all 10 items contribute constructively to the reliability of the scale.

Before inferential analysis, descriptive statistics for both the critical thinking post-test scores and the science literacy covariate scores were computed separately for each group and are presented in Table 1.

Table 1. Descriptive statistics of critical thinking and science literacy scores by group

	Model	Critical thinking	Scientific literacy
N	PBL	31	31
	CONV	29	29
Mean	PBL	85.161	69.129
	CONV	65.172	60.655
Median	PBL	80	73
	CONV	70	67
Standard deviation (SD)	PBL	9.957	15.756
	CONV	9.495	14.707
Skewness	PBL	0.061	-1.111
	CONV	-0.725	-0.500
Std. error (SE) skewness	PBL	0.421	0.421
	CONV	0.434	0.434
Kurtosis	PBL	-0.973	0.515
	CONV	0.565	-0.904
Std. error (SE) kurtosis	PBL	0.821	0.821
	CONV	0.845	0.845

As shown in Table 1, the PBL group (n = 31) demonstrated a higher mean critical thinking score (M = 85.16, SD = 9.96, Median = 80) compared to the control group (n = 29; Mean = 65.17, SD = 9.50, Median = 70). The distribution of critical thinking scores in the PBL group was approximately symmetric (skewness = 0.061, SE = 0.421), while the control group showed a slight negative skew (skewness = -0.725, SE = 0.434). Both distributions were considered sufficiently normal for ANCOVA purposes, consistent with the formal normality test results reported subsequently. Regarding the science literacy covariate, the PBL group obtained higher mean scores (Mean = 69.13, SD = 15.76, Median = 73) compared to the control group (Mean = 60.66, SD = 14.71, Median = 67). The science literacy distribution in the PBL group exhibited a moderate negative skew (skewness = -1.111, SE = 0.421), indicating a slight concentration of scores at the higher end of the scale. Although this skewness value approaches the conventional threshold of ±1.0, it is considered acceptable given that ANCOVA is robust to moderate violations of univariate normality when residual normality is satisfied (Field, 2018; Stevens, 2016).

To formally examine baseline group equivalence on the covariate, an independent samples t-test was conducted on science literacy scores before the intervention. Results are presented in Table 2.

Table 2. Independent samples t-test results for science literacy covariate baseline equivalence

Variable	t	df	p	Cohen's d	MD	95% CI
Science literacy	2.150	58	0.036*	0.555	8.474	[0.583, 16.365]

Results indicated a statistically significant difference between groups on science literacy before the intervention (t(58) = 2.150, p = .036, Cohen's d = 0.555, 95% CI [0.026, 1.076]), with the PBL group demonstrating higher mean science literacy (M = 69.13) than the control group (M = 60.66). The mean difference was 8.474 points (95% CI [0.583, 16.365]). Rather than representing a threat to validity, this finding directly reinforces the methodological necessity of including science literacy as a covariate. Because the two groups differed significantly on the covariate before treatment, any unadjusted comparison of post-test critical thinking scores would yield a biased estimate of PBL's instructional effect. ANCOVA corrects for this pre-existing difference by statistically equating both groups on science literacy before estimating the between-group difference in critical thinking, thereby producing adjusted mean estimates that more accurately reflect the true effect of the instructional treatment (Maxwell & Delaney, 2004; Stevens, 2016; Tabachnick & Fidell, 2019). Where the covariate differs significantly between groups is precisely the situation in which ANCOVA provides its greatest methodological advantage over unadjusted group comparisons (Maxwell & Delaney, 2004).

Before hypothesis testing was conducted, all statistical assumptions required by the ANCOVA test were systematically verified. First, the normality of the residual distribution was tested using three complementary tests: the Shapiro-Wilk test, the Kolmogorov-Smirnov test, and the Anderson-Darling test. The results of the normality tests for each method are shown in the Table 3.

Table 3. Results of the normality test

	Statistic	p
Shapiro-Wilk	0.964	0.071
Kolmogorov-Smirnov	0.100	0.591
Anderson-Darling	0.473	0.235

Three normality tests yielded p-values above the significance level of α = 0.05, indicating that the assumption of normality is met. A visual inspection to clarify the data distribution pattern was conducted by examining the following Q-Q plot (see Figure 7).

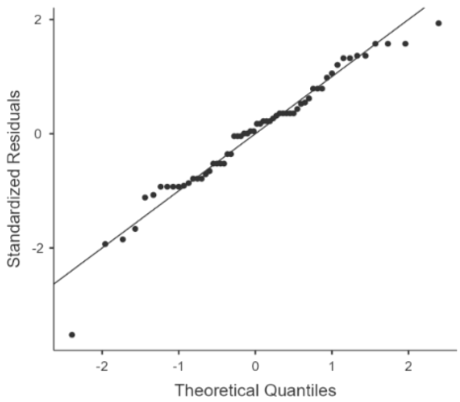

Figure 7. Q-Q Plot

The standardized residual Q-Q plot confirms that the data points are consistently scattered along the theoretical diagonal line, with a symmetrical distribution around the zero point providing strong visual confirmation that the residual distribution is approximately normal. The second test homogeneity of variance across groups was assessed using two complementary tests: the Levene test and Bartlett's test.

Table 4. Homogeneity test results

	Statistic	df	df2	p
Levene's	0.231	1	58	0.632
Bartlett's	0.064	1		0.800

The Levene's test yielded F = 0.231, p = 0.632, while the Bartlett's test produced a value of 0.064 with p = 0.800. Both tests resulted in p-values substantially exceeding the 0.05 significance level, indicating that the data are homogeneous; in other words, there is no significant difference in variance between the PBL group and the control group. The fulfillment of all prerequisite tests, namely normality and homogeneity of variance, supports the use of ANCOVA as the primary analytical technique and ensures that the resulting statistical inferences can be interpreted with confidence.

The Effect of Problem-Based Learning on Critical Thinking Skills

The results of the hypothesis test examining the effect of the PBL model on critical thinking skills, with science literacy as a covariate, using ANCOVA, are presented in Table 5.

Table 5. Results of the ANCOVA test

	Sum of Squares	df	Mean Square	F	p	η^2p
Overall model	5329.177	2	2664.589	43.885	< .001	
Model	4352.822	1	4352.822	54.868	< .001	0.490
Scientific literacy	976.355	1	976.355	12.307	< .001	0.178
Residuals	4521.976	57	79.333			

Based on the ANCOVA results presented in the table above, the overall model was highly significant, with F = 54.868, $p < 0.001$, and partial $\eta^2 = 0.490$. The partial η^2 value of 0.490 indicates a very large effect size, suggesting that the Problem-Based Learning (PBL) model accounted for 49.0% of the variance in fifth-grade elementary students' critical thinking skills in science, after statistically controlling for differences in students' scientific literacy.

The covariate of scientific literacy also demonstrated a significant effect on critical thinking, with F = 12.307, $p < 0.001$, and partial $\eta^2 = 0.178$. This partial η^2 value indicates a large effect size, meaning that scientific literacy explains 17.8% of the variance in critical thinking ability after controlling for other variables. These findings confirm that scientific literacy and critical thinking are two fundamental cognitive constructs in science learning.

Although the ANCOVA results confirm a significant effect of the instructional model on critical thinking skills, ANCOVA only provides information regarding the presence or absence of differences between the compared groups (experimental and control) and the magnitude of these effects, without explicitly quantifying the direction of differences between specific group pairs. Therefore, a post hoc analysis using Tukey's Honestly Significant Difference (HSD) test was employed as a follow-up comparison procedure, adjusted for the influence of the covariate.

The post hoc test results comparing the PBL group and the control (conventional) group, using estimated marginal means with Tukey adjustment, revealed a highly significant difference in mean scores (Mean Difference = 17.710, SE = 2.391, $t(57) = 7.407$, $p(\text{Tukey}) < .001$). This finding indicates that, after controlling for initial differences in science literacy as a covariate, students who received PBL instruction achieved, on average, 17.71 points higher in critical thinking scores than those who received conventional instruction. The magnitude of this difference provides strong evidence that PBL is more effective than conventional learning in fostering students' critical thinking skills within the context of elementary science education.

These findings are consistent with a substantial body of research confirming the effectiveness of Problem-Based Learning (PBL) in fostering critical thinking skills. Hmelo-Silver (2004), in a comprehensive review, concluded that PBL consistently enhances higher-order thinking abilities, as its problem-centered instructional structure compels students to engage in deep cognitive processes, including problem analysis, evaluation of alternative solutions, and the synthesis of new knowledge. Furthermore, Yew and Goh (2016) emphasized that the key mechanism underlying the effectiveness of PBL lies in the cognitive elaboration that occurs during collaborative inquiry, wherein students not only construct conceptual understanding but also develop metacognitive capacities, which are essential components of critical thinking.

This effect size (partial $\eta^2 = 0.490$) can be explained by several contextual factors: (1) the use of authentic problems contextualized around local environmental issues enhances students' relevance and intrinsic motivation (Savery, 2015); (2) the six-week intervention period provided sufficient time for the internalization of critical thinking processes (Sinaga, 2024); and (3) statistical control for science literacy via ANCOVA yielded effect estimates that were more free from initial confounding factors. In the context of science education in Indonesian elementary schools, these results demonstrate the effectiveness of PBL in enhancing learning motivation and higher-order thinking skills (Arrohima, 2025; Handayani et al., 2023). The uniqueness of this study lies in the combination of three methodological innovations: the use of ANCOVA with covariates from cognitive diagnostic assessment data in accordance with the Merdeka Curriculum guidelines, as well as the reporting of Tukey's post hoc test, which provides an explicit quantification of differences between groups (Mean Difference = 17.710). This combination yields far stronger empirical evidence compared to previous studies that used only a simple t-test without controlling for covariates.

Estimated marginal means (adjusted means) for each group after controlling for science literacy are presented in Table 6.

Table 6. Estimated marginal means of critical thinking

			95% Confidence Interval (CI)	
Model	**Mean**	**SE**	**Lower**	**Upper**
PBL	84.060	1.630	80.796	87.324
CONV	66.350	1.688	62.970	69.729

As shown in Table 6, after statistically controlling for individual differences in science literacy, the PBL group obtained an adjusted mean critical thinking score of 84.060 (SE = 1.630, 95% CI [80.796, 87.324]), compared to an adjusted mean of 66.350 (SE = 1.688, 95% CI [62.970, 69.729]) for the control group. The adjusted mean difference between groups was 17.710 points (SE = 2.391, $t(57) = 7.407$, $p < .001$), indicating that students who received PBL instruction scored substantially higher on critical thinking after accounting for their initial science literacy levels. These findings indicate that, within the constraints of this single-school study, PBL was associated with significantly higher critical thinking scores compared to conventional instruction after controlling for science literacy.

The Role of Science Literacy as a Covariate

The study revealed an interesting methodological finding: the significant contribution of science literacy as a covariate to critical thinking skills (partial $\eta^2 = 0.178$). This finding empirically confirms the theoretical proposition that science literacy serves as a cognitive foundation that facilitates the development of critical thinking in science education (OECD, 2023; Rahayuni, 2016). Students with higher science literacy on the initial diagnostic assessment demonstrated better critical thinking skills on the post-test, regardless of the instructional model they received. This is consistent with the PISA theoretical framework, which positions science literacy as an integrative capacity encompassing epistemic, conceptual, and procedural knowledge dimensions that directly support the critical thinking process (OECD, 2023). In this context, the Programme for International Student Assessment (PISA) framework conceptualizes scientific literacy as an integrative capacity encompassing three core competency dimensions: explaining phenomena scientifically, evaluating and designing scientific enquiry, and interpreting data and scientific evidence. These three dimensions are inherently associated with higher-order cognitive processes and substantially overlap with key components of critical thinking, particularly within the domains of analysis, evaluation, and inference. Students who demonstrate higher levels of scientific literacy are, therefore, structurally equipped with more advanced cognitive capacities, enabling them to engage more effectively in critical thinking processes within scientific contexts (Khery et al., 2022; Yanto & Sari, 2025).

The finding that science literacy accounts for 17.8% of the variance in critical thinking is consistent with and reinforces a number of previous empirical studies. Thus, the study of prospective teacher students found a significant positive correlation between science literacy and critical thinking skills, and concluded that science literacy serves as a cognitive foundation that facilitates the development of critical thinking in the context of science education (Basri, 2025; Wahdiyana, 2025). This is in line with that, Sadler et al., (2017) found that engagement with socioscientific issues requiring scientific literacy consistently enhances students' critical thinking skills, as the process of evaluating controversial scientific claims forces students to integrate scientific knowledge with argumentative reasoning. In the context of elementary school, Fragkiadaki & Ravanis, (2021) found that children with a stronger conceptual understanding of science demonstrated greater ability to construct causal explanations and evaluate claims, which are core dimensions of critical thinking skills.

Several limitations must be acknowledged. This study was conducted at a single school with one class per condition and a limited total sample of 60 students, which substantially limits generalizability. The absence of a critical thinking pre-test means that pre-existing differences in critical thinking between classes cannot be directly ruled out. Teacher and classroom effects cannot be fully excluded, as each condition was taught by a different teacher. The critical thinking instrument was developed specifically for the harmonious ecosystems topic and its transferability to other science topics or grade levels requires empirical verification. Additionally, construct validity of the critical thinking instrument in this study was established through expert-based content validation and internal consistency evidence; examination of the factor structure through EFA or CFA on a larger sample is recommended in future research to further strengthen construct validity evidence. Results should therefore be interpreted as indicative rather than definitive, and caution is warranted against strong causal claims.

CONCLUSION

This study provides empirical evidence that the Problem-Based Learning (PBL) model exerts a statistically significant and superior effect on the critical thinking skills of fifth-grade elementary school students in science learning compared to conventional instruction, after controlling for initial differences in scientific literacy between groups as a covariate. This finding is supported by a very large effect size (partial $\eta^2 = 0.490$) and an adjusted mean difference of 17.71 points, further confirmed by Tukey's post hoc test ($p < 0.001$). These results indicate that PBL is not merely an alternative classroom activity but rather a substantively and measurably more effective instructional model for enhancing elementary students' critical thinking capacities in the context of science education. A further finding reveals that scientific literacy functions as a strong cognitive predictor of critical thinking (partial $\eta^2 = 0.178$, $p < 0.001$). This suggests that students' prior abilities to understand, interpret, and evaluate scientific phenomena inherently facilitate the development of critical thinking skills in science learning. Methodologically, this study supports the utilization of cognitive diagnostic assessment data, as mandated by the Kurikulum Merdeka, as a valid and meaningful supporting variable in instructional processes.

RECOMMENDATION

Based on the findings of this study, it is recommended that future research explore the moderating role of science literacy in the relationship between Problem-Based Learning (PBL) and critical thinking using moderation regression analysis. Additionally, replicating this research design with a larger sample is necessary to strengthen the generalizability of the findings. Further research should also focus on identifying the specific dimensions of critical thinking that are most responsive to PBL interventions, thereby providing clearer operational guidance for teachers in designing instructional practices. Several challenges need to be anticipated in future studies, particularly the limited transferability of critical thinking instruments developed within the context of specific science topics, such as the theme of "harmonious ecosystems." Therefore, cross-content and cross-grade-level research remains essential to expand the generalizability of these findings.

ACKNOWLEDGMENTS

The author would like to thank the principal, the homeroom teacher, and the fifth-grade students of SDN 33 Lampe for their participation and assistance throughout the research process.

FUNDING INFORMATION

This research received no external funding

AUTHOR CONTRIBUTIONS STATEMENT

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Mei Indra Jayanti	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	
Muh. Nasir		✓		✓		✓				✓		✓	✓	
Umar	✓			✓		✓	✓			✓	✓		✓	

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

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

Ahmad, J., & Siew, N. (2021). Curiosity towards STEM education: A questionnaire for primary school students. *Journal of Baltic Science Education, 20*, 289–304. https://doi.org/10.33225/jbse/21.20.289

Aidoo, B. (2023). Teacher educators' experience adopting problem-based learning in science education. *Education Sciences*, *13*(11), 1113. https://doi.org/10.3390/educsci13111113

Anggraeny, E., & Khongput, S. (2022). Teachers' perceptions and practices of critical thinking instruction in Indonesian senior high schools: A case study. *TEFLIN Journal, 33(1), 1–26*. https://doi.org/10.15639/teflinjournal.v33i1/1-26

Arrohima, H. (2025). Analisis implementasi model problem-based learning (PBL) dalam meningkatkan motivasi belajar peserta didik di sekolah dasar. *Journal of Pedagogical Perspectives in Education (J-PPE), 1*(2), 127–132.

Basri, A. (2025). Analisis pengaruh literasi sains terhadap kemampuan berpikir kritis siswa pada pembelajaran IPAS di sekolah dasar. *Al-Ubudiyah: Jurnal Pendidikan dan Studi Islam, 6*(2), 228–236.

Chotimah, C. (2025). Implementasi asesmen diagnostik dalam pembelajaran IPAS Kurikulum Merdeka pada aspek penilaian berpikir kritis siswa. *Jurnal Didaktika Pendidikan Dasar*. https://doi.org/10.26811/didaktika.v9i1.1678

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

De Bortoli, L., Underwood, C., & Thomson, S. (2023). *PISA in brief 2022: Student performance and equity in education*. https://doi.org/10.37517/978-1-74286-727-4

Doolittle, P., Wojdak, K., & Walters, A. (2023). Defining active learning: A restricted systemic review. *Teaching and Learning Inquiry*. https://doi.org/10.20343/teachlearninqu.11.25

Dwyer, C. (2023). An evaluative review of barriers to critical thinking in educational and real-world settings. *Journal of Intelligence, 11*. https://doi.org/10.3390/jintelligence11060105

Facione, P. A. (2015). *Critical thinking: What it is and why it counts*. Insight Assessment.

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.

Fitriani, H., Asy'ari, M., Zubaidah, S., & Mahanal, S. (2019). Exploring the prospective teachers' critical thinking and critical analysis on science learning. *Jurnal Pendidikan IPA Indonesia, 8*(3), 433–442. https://doi.org/10.15294/jpii.v8i3.19434

Fragkiadaki, G., & Ravanis, K. (2021). The unity between intellect, affect, and action in a child's learning and development in science. *Learning, Culture and Social Interaction, 29*, 100495. https://doi.org/10.1016/j.lcsi.2021.100495

García-Carmona, A. (2023). Scientific thinking and critical thinking in science education. *Science & Education*, 1–19. https://doi.org/10.1007/s11191-023-00460-5

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference* (4th ed.). Allyn & Bacon.

Handayani, N., Indraswari, R., Shaluhiyah, Z., & Kusumawati, A. (2023). Pengasahan kemampuan kognitif sehat pada siswa sekolah dasar melalui permainan keterampilan berpikir kritis. *Kacanegara Jurnal Pengabdian Pada Masyarakat, 6717*, 11–16.

Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review, 16*(3), 235–266. https://doi.org/10.1023/B:EDPR.0000034022.16470.f3

Hu, X., & Bi, H. (2024). Exploring and validating the componential model of students' scientific critical thinking in science education. *Thinking Skills and Creativity*. https://doi.org/10.1016/j.tsc.2024.101695

Kemendikbudristek. (2022). *Panduan pembelajaran dan asesmen pendidikan anak usia dini, pendidikan dasar, dan menengah*. Badan Standar, Kurikulum, dan Asesmen Pendidikan.

Khery, Y., Sarjan, M., Ahzan, S., & Efendi, I. (2022). Konseptualisasi literasi sains mengacu pada kerangka sains PISA sejak tahun 2000. *Educatoria: Jurnal Ilmiah Ilmu Pendidikan, 2*(4), 200–231.

Kosasih, A., Supriyadi, T., Firmansyah, M. I., & Rahminawati, N. (2021). Higher-order thinking skills in primary school: Teachers' perceptions of Islamic education. *Journal of*

Ethnic and Cultural Studies. https://doi.org/10.29333/ejecs/994

Krishnan, P. (2019). A review of the non-equivalent control group post-test-only design. *Nurse Researcher, 26*(2), 37–40. https://doi.org/10.7748/nr.2018.e1582

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*(4), 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Leppink, J., & Pérez-Fuster, P. (2016). Why we need structured peer feedback in problem-based learning courses. *Medical Teacher, 38*(7), 702–706. https://doi.org/10.3109/0142159X.2015.1105941

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research, 35*(6), 382–385.

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*, 1–27. https://doi.org/10.1007/s43545-022-00546

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). Lawrence Erlbaum Associates.

Muse, A., & Baldwin, J. (2021). Quasi-experimental research design. In *The Encyclopedia of Research Methods in Criminology and Criminal Justice*. https://doi.org/10.1002/9781119111931.ch61

Norris, S. P., & Phillips, L. M. (2003). How literacy in its fundamental sense is central to scientific literacy. *Science Education, 87*(2), 224–240. https://doi.org/10.1002/sce.10066

OECD. (2023). *PISA 2022 results factsheets: Indonesia*. OECD.

Osborne, J. (2014). Teaching scientific practices: Meeting the challenge of change. *Journal of Science Teacher Education, 25*(2), 177–196. https://doi.org/10.1007/s10972-014-9384-1

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health, 29*(5), 489–497. https://doi.org/10.1002/nur.20147

Pratiwi, S. N., Cari, C., Aminah, N. S., & Affandy, A. N. (2019). Problem-based learning with argument mapping to improve critical thinking skills. *Journal of Physics: Conference Series, 1155*(1), 012063. https://doi.org/10.1088/1742-6596/1155/1/012063

Rahayuni, G. (2016). Hubungan keterampilan berpikir kritis dan literasi sains pada pembelajaran IPA terpadu dengan model PBM dan STM. *Jurnal Penelitian dan Pembelajaran IPA, 2*(2), 131–146.

Sadler, T. D., Foulk, J. A., & Friedrichsen, P. J. (2017). Evolution of a model for socio-scientific issue teaching and learning. *International Journal of Education in Mathematics, Science and Technology*. https://doi.org/10.18404/ijemst.55999

Santos, S. Dos, Reis, P., Reis, J., & Tavares, F. (2021). Two decades of PBL in teaching computing: A systematic mapping study. *IEEE Transactions on Education, 64*, 233–244. https://doi.org/10.1109/te.2020.3033416

Sarwanto, S., Fajari, L., & Chumdari, C. (2021). Critical thinking skills and their impacts on elementary school students. *Malaysian Journal of Learning and Instruction*. https://doi.org/10.32890/mjli2021.18.2.6

Savery, J. R. (2015). Comparative pedagogical models of problem-based learning. *Interdisciplinary Journal of Problem-Based Learning, 9*(2), 81–97.

Sinaga, R. (2024). Analisis penggunaan metode diskusi kelompok dalam meningkatkan keterampilan berpikir kritis di SD Swasta Sint. Yoseph Kabanjahe. *Jurnal Pendidikan: Media, Strategi, dan Metode, 1*(2), 77–83.

Stevens, J. P. (2016). *Applied multivariate statistics for the social sciences* (6th ed.). Routledge.

Sumarni, W., & Kadarwati, S. (2020). Ethno-STEM project-based learning: Its impact on critical and creative thinking skills. *Jurnal Pendidikan IPA Indonesia, 9*(1), 11–21. https://doi.org/10.15294/jpii.v9i1.21754

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.

Wahdiyana, V. Y. M. (2025). Efektivitas literasi sains pada pembelajaran biologi. *MetaBio: Jurnal Pendidikan, 7*(1), 50–58.

Wan, F. (2020). Analyzing pre-post designs using the analysis of covariance models with and without the interaction term in a heterogeneous study population. *Statistical Methods in Medical Research, 29*, 189–204. https://doi.org/10.1177/0962280219827971

Wijnen, M., Loyens, S. M. M., Smeets, G., Kroeze, M., & van der Molen, H. (2017). Comparing problem-based learning students to students in a lecture-based curriculum: Learning strategies and the relation with self-study time. *European Journal of Psychology of Education, 32*(3), 431–447. https://doi.org/10.1007/s10212-016-0296-7

Yanto, N., & Sari, N. I. (2025). Exploring scientific literacy in science classrooms: A literature study. *VENN: Journal of Sustainable Innovation on Education, Mathematics and Natural Sciences*, 164–173.

Yesnazar, A., Zhorabekova, A., Kalzhanova, A., Zhuzimkul, B., & Almukhanbet, S. (2024). Methodological system for the formation of meta-subject skills of primary school students in the context of STEM education. *Frontiers in Education*. https://doi.org/10.3389/feduc.2024.1340361

Yew, E. H. J., & Goh, K. (2016). Problem-based learning: An overview of its process and impact on learning. *Health Professions Education, 2*(2), 75–79. https://doi.org/10.1016/j.hpe.2016.01.004