Journal of Social Studies and Education
Vol. 4 No. 2 (2026), p. 193-205
Available online at https://serambi.org/index.php/jsse

The Impact of Pedagogical Deep Learning on Critical Thinking in Middle School Science

Lulu Royani Hamidah*, Meti Maspupah, Hadiansah
Universitas Islam Negeri Sunan Gunung Djati Bandung, Indonesia

DOI: https://doi.org/10.61987/jsse.v4i2.3029

Article History:
Received: 20 January 2026
Revised: 12 April 2026
Accepted: 9 May 2026

Keywords:
Pedagogical Deep Learning, Critical Thinking, Ecosystem Science, Science Education

****Correspondence Address:***
royanhl655@gmail.com

Abstract:
This study examined the effect of pedagogical deep learning on seventh-grade students' critical thinking skills in ecosystem science learning. Critical thinking is an essential competency for addressing complex scientific and environmental issues because students need to evaluate evidence, construct explanations, and make informed decisions. However, classroom practices often provide limited opportunities for students to engage in analysis, evaluation, reflection, and evidence-based reasoning. This study employed a quantitative quasi-experimental nonequivalent control-group design involving 56 seventh-grade students from SMP 8 Muhammadiyah Bandung. Data were collected using an essay-based critical thinking test and analyzed through descriptive statistics, independent-samples t-tests, normalized gain analysis, and ANCOVA. The results showed that both groups improved their critical thinking performance after instruction. However, the deep learning group demonstrated a higher mean improvement (19.14 points) than the comparison group (16.52 points). The difference in gain scores was statistically significant with a moderate effect size. ANCOVA analysis confirmed that the deep learning group maintained a significant posttest advantage after controlling for pretest performance. These findings indicate that pedagogical deep learning effectively supports higher-order cognitive development through meaningful, inquiry-oriented, and reflective learning processes. The study contributes empirical evidence for implementing Indonesia's pedagogical deep learning framework in science education and provides practical implications for designing reasoning-oriented classroom instruction. Future studies should investigate its effectiveness across broader contexts using larger and more diverse samples.

INTRODUCTION

Critical thinking in science education is a public capability, not merely a classroom outcome, because citizens increasingly encounter claims about health, technology, climate, biodiversity, and resource use that demand evidence-based judgment. The point is socially consequential: people who cannot distinguish evidence from assertion are more vulnerable to misinformation and poorly justified decisions. Schools matter because scientific literacy develops through repeated practice in questioning claims, examining sources, comparing explanations, and revising conclusions. Osborne and Pimentel (2023) argue that science education must prepare learners to evaluate scientific misinformation, while García-Carmona (2025) links critical and scientific thinking through evidence evaluation, inference, and metacognitive reflection. Therefore, deliberately strengthening critical thinking in lower-secondary science matters beyond examination performance. It prepares students to participate more responsibly in personal, environmental, and civic decisions where available evidence may often be incomplete, contested, or changing.

Ecosystem learning makes this need visible because ecological problems emerge from interacting biological, physical, and human factors rather than from single causes. Students must connect food webs, population change, habitat conditions, pollution, consumption, and the consequences of intervention before judging a proposed solution. This reasoning resembles the decisions citizens face when discussing environmental degradation, biodiversity loss, climate adaptation, or local resource management. Bossér (2024) shows that socio-scientific science teaching can promote functional scientific literacy by engaging learners with socially relevant evidence and values, while Sugrah et al. (2023) report gains in critical thinking when junior-high students examine socio-scientific issues. Suastrawan et al. (2021) likewise found that contextual science materials can strengthen seventh-grade critical thinking. Consequently, ecosystem instruction should not stop at naming components or recalling definitions; it should cultivate the capacity to analyze relationships, weigh evidence, anticipate consequences, and defend revisable conclusions.

The present study connects two theoretical constructs: pedagogical deep learning as the instructional condition and critical thinking as the targeted cognitive outcome. In Indonesia, deep learning is framed as mindful, meaningful, and joyful learning that moves students through understanding, application, and reflection; Rahmawati et al. (2025) describe it as learner-centered, contextual, and transformative. This orientation is consistent with constructivist active learning, which treats learners as knowledge builders rather than recipients. Doolittle et al. (2023) define active learning as student-centered knowledge construction directed toward higher-order thinking. Critical thinking, meanwhile, involves interpretation, analysis, evaluation, inference, explanation, and self-regulation, with evidence and metacognition central to disciplined judgment. García-Carmona (2025) emphasizes this symbiosis between scientific and critical thinking. The resulting theoretical expectation is therefore that contextual inquiry, collaborative reasoning, evidence evaluation, and reflection should make students' reasoning more explicit, examinable, and progressively more self-correcting.

The general educational problem is that active-looking lessons do not necessarily produce active thinking. In many classrooms, students may work in groups, complete worksheets, or use digital media while still following predetermined procedures and reproducing teacher explanations. The source thesis identified this pattern: routine tasks were familiar, whereas opportunities to analyze evidence, evaluate alternatives, justify conclusions, and revise reasoning were limited. This matters because higher-order outcomes depend on cognitive demands, not surface participation. Lonergan et al. (2022) found that problem-based learning in secondary classrooms can support knowledge development but also revealed differences in learner readiness and challenges in problem-solving outcomes. Abu Khurma and El Zein (2024) similarly stress teacher competence, student readiness, classroom management, and learning environments. Thus, improving critical thinking requires structured epistemic work: students must know what claim is under examination, what evidence matters, and why conclusions follow.

Recent literature supports problem-centered and inquiry-oriented instruction, but warns against treating instructional labels as sufficient explanations of success. Loyens et al. (2023) found positive evidence for higher-order and critical thinking in problem- and project-based environments, yet identified inconsistent definitions, measures, omitted design features, and weak control conditions. Yu and Zin (2023) similarly concluded that problem-based learning supports critical thinking more consistently when critical-thinking processes are deliberately embedded. In science education, Urdanivia Alarcon et al. (2023) synthesized evidence that inquiry supports scientific knowledge construction through investigation and reasoning. Arifin et al. (2025) reported an overall effect of inquiry-based learning on critical thinking, alongside high heterogeneity across educational levels, strategies, technology use, duration, and assessment. Liu and Pásztor (2022) likewise found beneficial but heterogeneous effects of problem-based learning. Prior research therefore supports purposeful cognitive design while exposing substantial uncertainty about implementation, context, and measurement.

Indonesian and school-based evidence strengthens the rationale while revealing a specific gap. Mutiara et al. (2024) synthesized Sinta-indexed studies. They concluded that problem-based learning commonly improves critical thinking, whereas Rehman et al. (2023) experimentally reported stronger twenty-first-century skills under project-based instruction. Rehman et al. (2024) linked project-based learning with critical thinking, problem solving, collaboration, and engagement. In Indonesian science contexts, Fauziah et al. (2024) demonstrate the close relationship between critical thinking and scientific argumentation, and Aisy et al. (2024) found advantages for problem-based learning contextualized through socio-scientific issues. However, Wijnia et al. (2024) found only a small-to-moderate average motivational advantage for problem-, project-, and case-based approaches, with substantial heterogeneity. Existing evidence therefore concerns established pedagogies, other subjects or levels, reviews, or aggregated outcomes, leaving limited evidence about the newly formalized deep-learning framework in seventh-grade ecosystem science and individual critical-thinking change.

The state of the art lies in testing Indonesia's newly formalized pedagogical deep learning framework as an integrated instructional orientation rather than merely relabeling problem-based, project-based, or inquiry learning. Its novelty is contextual, conceptual, and analytical. Contextually, it examines seventh-grade ecosystem science, where causal interdependence and human-environment decisions provide authentic demands for evidence-based reasoning. Conceptually, it operationalizes mindful, meaningful, and joyful learning through tasks that require students to understand ecological relationships, apply concepts to contextual problems, articulate evidence, and reflect on conclusions. Analytically, it improves the transparency of the source thesis by treating the supplied individual-level Excel dataset as the source of record and recomputing descriptive and inferential statistics from raw scores. Summary tables can conceal transcription or formula errors. The study therefore contributes needed empirical evidence where Indonesian deep-learning scholarship remains largely conceptual, implementation-focused, or outside lower-secondary ecosystem science.

The research problem is whether students participating in pedagogical deep learning demonstrate greater critical-thinking improvement than students receiving the comparison instruction in a seventh-grade ecosystem unit. The directional hypothesis is that the deep-learning class will show a larger pretest-to-posttest gain and a higher posttest level after accounting for baseline performance. This expectation follows from evidence that critical thinking develops when learners confront meaningful problems, investigate information, evaluate evidence, construct explanations, discuss alternatives, and regulate their reasoning (Felton et al., 2022; Loyens et al., 2023; Arifin et al., 2025). The study more directly connects those mechanisms to Indonesia's mindful, meaningful, and joyful framework while testing outcomes at the individual level. Because intact classes rather than individual students were assigned to conditions, any observed advantage should be interpreted as a plausible instructional association under quasi-experimental conditions, not definitive causal proof.

RESEARCH METHODS

This study employed a quantitative quasi-experimental nonequivalent control-group design. This design was selected because the study involved intact classroom groups in a natural school setting where individual random assignment was not feasible. Quasi-experimental approaches are appropriate for educational research because they allow researchers to examine instructional effects while maintaining ecological validity in authentic classroom contexts (Creswell & Creswell, 2023; Strunk & Mwavita, 2022). The research was conducted at SMP 8 Muhammadiyah Bandung during the second semester of the 2025/2026 academic year. The population consisted of 140 seventh-grade students distributed across five intact classes. Cluster sampling was applied to select two comparable classes. Class VII Ibnu Thufail received the pedagogical deep learning intervention, whereas Class VII Ibnu Bathutah received comparison instruction. Each class consisted of 28 students, resulting in a total analytical sample of 56 participants. Both groups completed an identical critical thinking pretest before instruction and an identical posttest after the instructional period. Because assignment occurred at the class level rather than the individual level, findings are interpreted as instructional associations within the study context rather than definitive causal effects.

The deep learning condition was implemented through contextual ecosystem problems and an inquiry-oriented learning sequence designed to promote understanding, application, and reflection. Deep learning approaches emphasize meaningful knowledge construction through active engagement, contextual problem solving, collaboration, and metacognitive reflection rather than passive information acquisition (Rahmawati et al., 2025; Doolittle et al., 2023). Lessons began with environmental stimuli, including images, videos, or authentic ecosystem cases, followed by guiding questions that encouraged students to identify problems, activate prior knowledge, and construct initial explanations. Students worked collaboratively using learning worksheets to examine information, identify relationships between biotic and abiotic components, analyze ecological interactions, evaluate environmental problems, and develop evidence-based explanations. Groups presented their reasoning, responded to peer questions, and revised conclusions during reflection activities. The comparison condition maintained conventional teacher-guided instruction through explanation, examples, note-taking, question-and-answer sessions, and structured practice activities. Both groups studied identical ecosystem concepts and achieved the same curriculum objectives. Therefore, the comparison condition represented a realistic instructional alternative rather than a no-treatment control. This approach follows recommendations that educational intervention studies should compare meaningful classroom practices rather than artificial control conditions (Loyens et al., 2023).

Students' critical thinking was measured using an essay-based assessment developed specifically for the ecosystem learning unit. Essay assessments were selected because they provide opportunities to evaluate higher-order reasoning processes, including explanation, justification, inference, and evaluation of evidence, which are difficult to capture through simple selected-response formats (García-Carmona, 2025). The source study initially developed a 30-item instrument and retained 20 items after item-level evaluation, including validity, discrimination index, and difficulty-level analysis. The retained items represented six critical thinking dimensions adapted from Ennis' framework: focus, reason, inference, situation, clarification, and overview. The 30-item pilot instrument produced Cronbach's alpha = .951, indicating strong internal consistency. However, because the thesis did not report reliability statistics separately for the final 20-item version, this study does not claim the pilot coefficient as evidence of reliability for the final instrument. Scores from both pretest and posttest were converted to a 0–100 scale. This procedure follows current recommendations that instrument validity evidence should be reported according to the specific version and purpose of measurement rather than transferred automatically from pilot forms (AERA et al., 2024).

The uploaded Excel workbook was treated as the primary data source for student achievement scores. Individual-level scores from 28 students in each class were extracted, and all descriptive and inferential statistics were independently recalculated. This procedure improved analytical transparency and prevented potential errors caused by transcription inconsistencies between raw data and reported summary tables. Descriptive analyses included means, standard deviations, score differences, and normalized gain values calculated as (posttest−pretest)/(100−pretest). Assumption testing included Shapiro-Wilk tests for normality and Levene tests for homogeneity of variance. Baseline equivalence between groups was examined using an independent-samples t test. Differences in posttest scores and gain scores were analyzed using independent-samples t tests, with Hedges' g calculated as a small-sample standardized effect size because it provides a less biased estimate than Cohen's d for limited samples (Lakens, 2022). Analysis of covariance (ANCOVA) was conducted to estimate posttest differences while controlling for centered pretest scores. Before interpreting the ANCOVA, the interaction between group and pretest score was examined to evaluate the homogeneity of regression slopes assumption. Statistical decisions used an alpha level of .05. Because treatment assignment occurred through intact classes and only one class represented each condition, statistical significance was interpreted cautiously and not as equivalent to cluster-randomized causal evidence.

RESULTS AND DISCUSSION

Results

Deep learning improves critical thinking more than comparison instruction

The first finding shows a clear pattern that both instructional approaches contributed to improving students' critical thinking skills. Still, the magnitude of improvement was substantially higher in the pedagogical deep learning group. This suggests that while conventional instruction is still effective in supporting learning, the deep learning approach provides richer cognitive engagement that leads to stronger development of higher-order thinking skills.

Before the intervention, both groups demonstrated almost identical levels of critical thinking ability. The deep learning class obtained a pretest mean score of 66.12

(SD = 3.25), while the comparison class obtained a mean score of 65.85 (SD = 3.25). The difference between the two groups was only 0.27 points, which is extremely small in practical terms. Statistical testing confirmed that this difference was not significant, t(54) = 0.31, p = .759, with a very small effect size (Hedges' g = 0.08). This result is important because it establishes that both groups started from a comparable baseline, meaning that any differences observed after the intervention are unlikely to be caused by pre-existing ability differences.

To ensure the validity of further parametric analyses, assumption testing was conducted. The Shapiro-Wilk test results indicated that all score distributions (pretest, posttest, and gain scores) in both groups were normally distributed, with p-values ranging from .216 to .377. This confirms that the data met the normality assumption required for t-tests and ANCOVA. In addition, Levene's test results showed that the variances between groups were homogeneous for pretest (p = .862), posttest (p = .488), and change scores (p = .071). These results collectively confirm that the statistical conditions for comparing group means were satisfied, strengthening the reliability of subsequent inferential analyses.

After the instructional treatment, both groups showed clear improvement in critical thinking performance, indicating that learning occurred in both conditions. However, the degree of improvement differed meaningfully between groups. The deep learning group increased from 66.12 to 85.26, resulting in a mean gain of 19.14 points. In contrast, the comparison group increased from 65.85 to 82.37, resulting in a smaller gain of 16.52 points. Although both groups improved, the deep learning group consistently outperformed the comparison group in terms of absolute score increase.

Figure 1 illustrates the pattern of changes in students' critical thinking scores before and after the intervention. The figure visually compares students' performance development in the pedagogical deep learning group and the comparison instruction group, highlighting differences in improvement trends between the two instructional conditions.

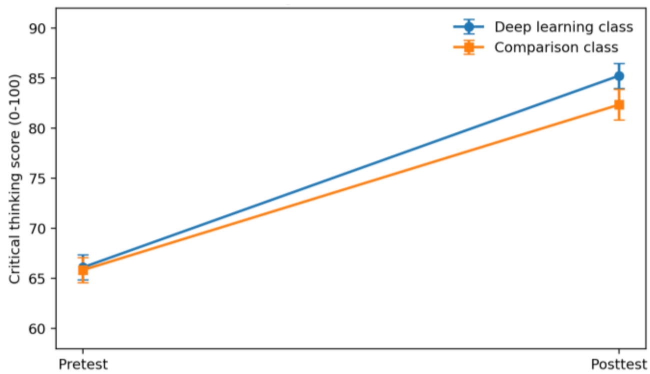

Figure 1. Changes in Students' Critical Thinking Scores from Pretest to Posttest Across Instructional Groups

Figure 1 shows that both instructional groups improved in critical thinking scores after the intervention; however, the magnitude of improvement was greater in the pedagogical deep learning group. The deep learning group increased from a pretest mean score of 66.12 to a posttest mean score of 85.26, whereas the comparison group increased from 65.85 to 82.37. Although both groups started from nearly identical baseline levels, indicated by the minimal difference in pretest scores, the posttest results reveal a wider performance gap favoring the deep learning group. This pattern suggests that the pedagogical deep learning approach better supported students' critical thinking development than comparison instruction. The higher trajectory of improvement in the deep learning group indicates that instructional strategies emphasizing analysis, evaluation, problem-solving, and meaningful knowledge construction may contribute more effectively to enhancing higher-order cognitive skills. Table 1 presents the detailed descriptive statistics.

Table 1. Descriptive Statistics of Students' Critical Thinking Scores

Instructional Group	n	Pretest Mean (SD)	Posttest Mean (SD)	Mean Change (SD)	Normalized Gain Mean (SD)
Pedagogical deep learning	28	66.12 (3.25)	85.26 (3.23)	19.14 (3.36)	0.565 (0.085)
Comparison instruction	28	65.85 (3.25)	82.37 (3.91)	16.52 (4.92)	0.480 (0.121)

Table 1 shows that inferential analysis further confirmed that the difference in improvement between the two groups was statistically significant. The difference in mean gain scores was 2.63 points in favor of the deep learning group. The Welch independent-samples t-test was used due to slight differences in variance and produced a statistically significant result, $t(47.64) = 2.33$, $p = .024$. The 95% confidence interval ranged from 0.36 to 4.89, indicating that the true difference is likely positive and favors the deep learning group. The effect size (Hedges' $g = 0.61$) indicates a moderate practical effect, suggesting that the difference is not only statistically significant but also educationally meaningful.

The normalized gain analysis provides additional insight into learning efficiency relative to students' starting points. The deep learning group achieved a mean normalized gain of 0.565, while the comparison group achieved 0.480. The difference of 0.085 was statistically significant ($p = .004$), indicating that students in the deep learning condition improved more efficiently relative to their maximum possible score gain. This finding is important because it shows that deep learning not only increases final achievement but also optimizes the learning process itself.

From a pedagogical perspective, these results suggest that the deep learning model is more effective because it actively engages students in higher-order cognitive processes. During instruction, students were required to analyze ecosystem problems, identify relationships between biotic and abiotic components, evaluate evidence, construct explanations, and communicate reasoning. These activities promote deeper cognitive processing, which is essential for developing critical thinking skills. In contrast, the comparison class relied more on teacher explanation, structured examples, and routine exercises, which may support understanding but provide fewer opportunities for analytical reasoning and reflective thinking.

Deep learning effect remains significant after controlling prior ability

The superiority of pedagogical deep learning remains stable even after controlling for students' initial ability levels. This is particularly important because the study used a nonequivalent control-group design, where intact classes were assigned to different instructional conditions. In such designs, statistical control is necessary to reduce potential bias from pre-existing differences.

To address this, an analysis of covariance (ANCOVA) was conducted. Before interpreting the ANCOVA results, the assumption of homogeneity of regression slopes was tested by examining the interaction between group and pretest scores. The interaction was not statistically significant ($p = .195$), indicating that the relationship between pretest and posttest scores was consistent across both groups. This confirms that ANCOVA was appropriate and that the adjustment for baseline scores is valid.

The ANCOVA results showed that instructional condition significantly influenced posttest critical thinking scores even after controlling for pretest performance. Students in the deep learning group achieved an adjusted posttest score that was 2.82 points higher than those in the comparison group. This difference was statistically significant ($p = .004$), with a 95% confidence interval of 0.94 to 4.71. The partial eta squared value ($\eta^2p = .145$) indicates that approximately 14.5% of the variance in posttest scores can be explained by the instructional model after accounting for initial ability. In educational research, this represents a moderate-to-strong effect, suggesting that instructional design plays a meaningful role in shaping students' critical thinking outcomes. Table 2 presents a detailed comparison of all statistical outcomes.

Table 2. Between-Group Statistical Comparisons of Critical Thinking Outcomes

Outcome	Difference	95% CI	Test statistic	p	Effect size
Baseline pretest	0.27	[-1.47, 2.01]	t = 0.31	.759	g = 0.08
Posttest	2.89	[0.97, 4.82]	t = 3.02	.004	g = 0.79
Change score	2.63	[0.36, 4.89]	t = 2.33	.024	g = 0.61
ANCOVA adjusted posttest	2.82	[0.94, 4.71]	t = 3.00	.004	partial eta^2 = .145

Table 2 demonstrates a consistent advantage of the pedagogical deep learning condition across all statistical comparisons. First, the baseline pretest comparison revealed no significant difference between groups ($p = .759$; Hedges' $g = 0.08$), confirming that both classes started from an equivalent level of critical thinking ability before the intervention. Following instruction, the deep learning group achieved significantly higher posttest scores than the comparison group, with a mean difference of 2.89 points ($p = .004$; Hedges' $g = 0.79$), indicating a large standardized effect. Similarly, the analysis of change scores showed that the deep learning group improved more, with a 2.63-point advantage over the comparison group ($p = .024$; Hedges' $g = 0.61$), representing a moderate effect. Importantly, the ANCOVA analysis confirmed that this advantage remained significant after controlling for students' initial performance, with an adjusted posttest difference of 2.82 points ($p = .004$; $\eta^2p = .145$). Collectively, these findings indicate that pedagogical deep learning contributed to superior critical thinking development beyond the effect of students' initial ability levels. The convergence of results across multiple analytical approaches strengthens the conclusion that the instructional approach was associated with meaningful improvements in students' ecosystem-related critical thinking performance.

To further ensure robustness, we applied a heteroskedasticity-consistent HC3 correction. The results remained consistent ($p = .005$), indicating that the findings are stable even under more conservative statistical assumptions. This strengthens confidence in the reliability of the observed effect.

Overall, the results from multiple analytical approaches independent t-tests, change score analysis, normalized gain, and ANCOVA-converge on the same conclusion: pedagogical deep learning consistently produces higher critical thinking outcomes than comparison instruction. The consistency across methods increases the credibility of the findings.

However, because the study involved intact classroom groups and only one class per condition, the findings should be interpreted as strong evidence of instructional effectiveness within the study context rather than definitive causal proof. Despite this limitation, the convergence of statistical evidence strongly supports the conclusion that pedagogical deep learning is a more effective instructional approach for enhancing students' critical thinking in ecosystem learning.

Discussion

The present study demonstrates that pedagogical deep learning significantly enhanced students' critical thinking development in ecosystem learning compared with comparison instruction. This finding aligns with recent evidence emphasizing that meaningful, inquiry-oriented, and problem-centered learning environments promote higher-order thinking processes. Loyens et al. (2023) reported that problem- and project-based learning environments can foster critical and higher-order thinking when essential design features, such as reflection, problem analysis, and knowledge application, are deliberately implemented. Similarly, Yu and Zin (2023) highlighted that improvements in critical thinking through problem-based learning depend on the explicit integration of reasoning processes. This finding extends previous research by showing that the Indonesian pedagogical deep learning framework can serve as an effective instructional approach for strengthening critical thinking in lower-secondary ecosystem science. The result confirms that learning designs emphasizing analysis, evaluation, explanation, and reflection provide stronger cognitive experiences than predominantly teacher-directed instruction.

The superiority of the deep learning group can be explained through constructivist and active learning perspectives, which argue that learners develop deeper understanding when they actively construct knowledge through meaningful experiences. In this study, students in the deep learning condition engaged with contextual ecosystem problems, analyzed relationships among ecological components, evaluated evidence, and communicated explanations before revising their conclusions. These activities correspond with the principles of active learning described by Doolittle et al. (2023), who emphasized that active learning is valuable when students participate in cognitively demanding knowledge construction rather than simply completing activities. The finding is also consistent with Urdanivia Alarcon et al. (2023), who identified inquiry-based science learning as an effective pathway for developing scientific reasoning. Therefore, this study's theoretical contribution is demonstrating that pedagogical deep learning operationalizes critical thinking development through structured cognitive engagement within authentic science contexts.

The observed improvement is also consistent with recent empirical studies investigating problem-based and inquiry-based science instruction. The deep learning group achieved a larger increase in critical thinking scores (19.14 points) compared with the comparison group (16.52 points), with a statistically significant difference in gain scores and a moderate effect size. This pattern supports findings from Mutiara et al. (2024), who synthesized Indonesian research showing that problem-based learning frequently improves critical thinking outcomes in science-related subjects. Likewise, Arifin et al. (2025) reported positive effects of inquiry-based learning on critical thinking while noting that effectiveness depends on implementation quality and learning context. The current research contributes by moving beyond established instructional labels such as inquiry-based or problem-based learning and examining the broader pedagogical deep learning framework. This distinction is important because it highlights not only instructional strategies but also the quality of students' cognitive engagement throughout the learning process.

The ANCOVA findings provide additional evidence that initial differences between groups did not simply cause the instructional advantage of pedagogical deep learning. After controlling for pretest performance, the deep learning group maintained a significantly higher adjusted posttest score, with an effect size indicating meaningful educational impact. This result is consistent with methodological recommendations that quasi-experimental educational studies should statistically control baseline differences when random assignment is unavailable (Creswell & Creswell, 2023). Furthermore, the finding supports Wijnia et al. (2024), who argued that student-centered approaches often produce positive outcomes but that effects depend strongly on instructional implementation. The present study adds evidence that deep learning, when translated into structured classroom practices involving problem analysis, evidence evaluation, and reflection, can generate measurable improvements in critical thinking. The practical implication is that teachers should design learning experiences around reasoning processes rather than merely increasing student activity.

From a practical perspective, this study provides implications for science teachers, curriculum developers, and educational policymakers. The findings suggest that ecosystem instruction should incorporate authentic environmental problems because ecological concepts naturally require students to analyze complex relationships and evaluate alternative solutions. This recommendation is supported by Bossér (2024), who emphasized that socio-scientific and contextual science learning can strengthen scientific literacy by connecting classroom knowledge with real-world decision-making. Similarly, Sugrah et al. (2023) found that socio-scientific issue-based learning encourages critical thinking by requiring students to consider evidence and multiple perspectives. The novelty of the present research lies in connecting these principles with Indonesia's emerging pedagogical deep learning framework at the junior secondary level. Unlike previous studies that focus mainly on specific models such as PBL or inquiry learning, this research shows how a broader learning philosophy can translate into measurable improvements in students' reasoning abilities.

This study also contributes methodologically by improving analytical transparency in evaluating instructional interventions. Rather than relying exclusively on reported summary statistics, the analysis used individual-level achievement data and included multiple statistical approaches, including gain analysis, independent comparisons, and ANCOVA. Together, these analyses consistently showed that pedagogical deep learning produced higher critical thinking outcomes than comparison instruction. This approach addresses concerns raised in educational research regarding inconsistent measurement and interpretation of intervention effects. García-Carmona (2025) emphasized that critical thinking in science requires systematic evaluation, reasoning, and evidence-based judgment; therefore, assessment and analysis procedures should reflect these intellectual processes. This study not only identifies an effective instructional approach but also demonstrates a more rigorous analytical pathway for evaluating classroom-based innovations.

Despite its contributions, the findings should be interpreted within the limitations of the research design. Because the study used intact classes with one class representing each instructional condition, the results indicate strong evidence of instructional effectiveness within the research context rather than definitive causal relationships. Future studies should replicate this investigation across multiple schools, larger samples, and different educational contexts to examine the generalizability of the findings. Nevertheless, the current study provides important evidence that pedagogical deep learning can serve as a promising framework for developing critical thinking in science education. Its broader impact lies in supporting a shift from knowledge transmission to reasoning-oriented learning, where students can interpret evidence, construct explanations, and make informed decisions about complex ecological issues.

CONCLUSION

This study concludes that pedagogical deep learning meaningfully contributes to improving seventh-grade students' critical thinking skills in ecosystem science learning. The most important finding is that students who experienced a learning sequence emphasizing contextual problems, inquiry, evidence evaluation, collaborative reasoning, and reflection developed critical thinking more than students receiving comparison instruction. The deep learning group demonstrated higher learning gains. It maintained a significant advantage after controlling for initial ability, suggesting that meaningful and reflective learning experiences can strengthen students' capacity to analyze, evaluate, and construct scientific explanations. The main scholarly contribution of this study is extending existing research on problem-based and inquiry-based learning by providing empirical evidence for implementing Indonesia's emerging pedagogical deep learning framework at the lower-secondary level. Methodologically, the study contributes by reanalyzing individual-level data and using multiple statistical approaches to strengthen analytical transparency. However, the findings should be interpreted in light of several limitations, including the use of intact classes, one class per instructional condition, limited research duration, and the absence of delayed assessment to measure long-term retention and transfer. Future research should involve larger multi-school samples, cluster-based experimental designs, longitudinal evaluation, implementation fidelity analysis, and further validation of critical thinking instruments to determine the broader effectiveness and sustainability of pedagogical deep learning across diverse educational contexts.

REFERENCES

Abu Khurma, O., & El Zein, F. (2024). Inquiry Skills Teaching and Its Relationship with UAE Secondary School Students' Critical Thinking: Systematic Review of Science Teachers' Perspectives. *Eurasia Journal of Mathematics, Science and Technology Education, 20*(2), Article em2397. https://doi.org/10.29333/ejmste/14155

AERA, APA, & NCME. (2024). *Standards for Educational and Psychological Testing*. American Educational Research Association.

Aisy, M. R., Trisnowati, E., & Siswanto, S. (2024). The Effect of the Problem-Based Learning (PBL) Model in the Context of Socio-Scientific Issues (SSI) on Critical Thinking Ability on Digestive System Material. *Jurnal Inovasi Pendidikan IPA, 10*(2), 185–195. https://doi.org/10.21831/jipi.v10i2.75231

Arifin, Z., Sukarmin, Saputro, S., & Kamari, A. (2025). The Effect of Inquiry-Based Learning on Students' Critical Thinking Skills in Science Education: A Systematic Review and Meta-Analysis. *Eurasia Journal of Mathematics, Science and Technology Education, 21*(3), Article em2592. https://doi.org/10.29333/ejmste/15988

Bossér, U. (2024). Transformation of School Science Practices to Promote Functional Scientific Literacy. *Research in Science Education, 54*(2), 265–281. https://doi.org/10.1007/s11165-023-10138-1

Creswell, J. W., & Creswell, J. D. (2023). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (6th ed.). SAGE Publications.

Doolittle, P., Wojdak, K., & Walters, A. (2023). Defining Active Learning: A Restricted Systematic Review. *Teaching & Learning Inquiry, 11*. https://doi.org/10.20343/teachlearninqu.11.25

Fauziah, A. N. M., Wasis, Hendratmoko, A. F., Mahdiannur, M. A., Ermawan, M. Z. F., Suwandi, E., & Ratri, S. Y. (2024). The Relationship Between Critical Thinking and Scientific Argumentation in Science Learning. *Jurnal Pendidikan IPA Indonesia, 13*(2), 239–254. https://doi.org/10.15294/jpii.v13i2.2585

Felton, M., Levin, D. M., De La Paz, S., & Butler, C. (2022). Scientific Argumentation and Responsive Teaching: Using Dialog to Teach Science in Three Middle-School Classrooms. *Science Education, 106*, 1354–1374. https://doi.org/10.1002/sce.21740

García-Carmona, A. (2025). Scientific Thinking and Critical Thinking in Science Education: Two Distinct but Symbiotically Related Intellectual Processes. *Science & Education, 34*, 227–245. https://doi.org/10.1007/s11191-023-00460-5

Lakens, D. (2022). Sample Size Justification. *Collabra: Psychology, 8*(1), Article 33267. https://doi.org/10.1525/collabra.33267

Liu, Y., & Pásztor, A. (2022). Effects of Problem-Based Learning Instructional Intervention on Critical Thinking in Higher Education: A Meta-Analysis. *Thinking Skills and Creativity, 45*, Article 101069. https://doi.org/10.1016/j.tsc.2022.101069

Lonergan, R., Cumming, T. M., & O'Neill, S. C. (2022). Exploring the Efficacy of Problem-Based Learning in Diverse Secondary School Classrooms: Characteristics and Goals of Problem-Based Learning. *International Journal of Educational Research, 112*, Article 101945. https://doi.org/10.1016/j.ijer.2022.101945

Loyens, S. M. M., van Meerten, J. E., Schaap, L., & Wijnia, L. (2023). Situating Higher-Order, Critical, and Critical-Analytic Thinking in Problem- and Project-Based Learning Environments: A Systematic Review. *Educational Psychology Review, 35*(2), Article 39. https://doi.org/10.1007/s10648-023-09757-x

Mutiara, E., Suyanto, S., B, N. K. L., Laksita, G. D., & Zamzami, Z. (2024). Improving Critical Thinking Skills Using Problem-Based Learning: Systematic Literature Review. *Jurnal Penelitian Pendidikan IPA, 10*(12), 988–995. https://doi.org/10.29303/jppipa.v10i12.7872

Osborne, J., & Pimentel, D. (2023). Science Education in an Age of Misinformation. *Science Education, 107*(3), 553–571. https://doi.org/10.1002/sce.21790

Rahmawati, Y., Mu'ti, A., Suyanto, S., & Herianingtyas, N. L. R. (2025). Pembelajaran Mendalam: Transformasi Pembelajaran Menuju Pendidikan Bermutu. *Jurnal Penelitian Kebijakan Pendidikan, 18*(1). https://doi.org/10.24832/jpkp.v18i1.1281

Rehman, N., Huang, X., Mahmood, A., AlGerafi, M. A. M., & Javed, S. (2024). Project-Based Learning as a Catalyst for 21st-Century Skills and Student Engagement in the Math Classroom. *Heliyon, 10*(23), Article e39988. https://doi.org/10.1016/j.heliyon.2024.e39988

Rehman, N., Zhang, W., Mahmood, A., Fareed, M. Z., & Batool, S. (2023). Fostering Twenty-First Century Skills Among Primary School Students Through Math Project-Based Learning. *Humanities and Social Sciences Communications, 10*, Article 424. https://doi.org/10.1057/s41599-023-01914-5

Strunk, K. O., & Mwavita, M. (2022). *Designing and Conducting Research in Education*. SAGE Publications.

Sugrah, N., Suyanta, & Wiyarsi, A. (2023). Promoting Students' Critical Thinking and Scientific Attitudes Through Socio-Scientific Issues-Based Flipped Classroom. *LUMAT: International Journal on Math, Science and Technology Education, 11*(1), 140–165. https://doi.org/10.31129/LUMAT.11.1.1856

Suastrawan, K. E., Suardana, I. N., & Sudiatmika, A. A. I. A. R. (2021). The Effectiveness of Science E-Modules for Class VII Junior High Schools Based on Socio-scientific Issues to Improve Students' Critical Thinking Skills. *Journal of Science Education Research, 5*(2), 1–9. https://doi.org/10.21831/jser.v5i2.42877

Urdanivia Alarcon, D. A., Talavera-Mendoza, F., Rucano Paucar, F. H., Cayani Caceres, K. S., & Machaca Viza, R. (2023). Science and Inquiry-Based Teaching and Learning: A Systematic Review. *Frontiers in Education, 8*, Article 1170487. https://doi.org/10.3389/feduc.2023.1170487

Wijnia, L., Noordzij, G., Arends, L. R., Rikers, R. M. J. P., & Loyens, S. M. M. (2024). The Effects of Problem-Based, Project-Based, and Case-Based Learning on Students' Motivation: A Meta-Analysis. *Educational Psychology Review, 36*(1), Article 29. https://doi.org/10.1007/s10648-024-09864-3

Yu, L., & Zin, Z. M. (2023). The Critical Thinking-Oriented Adaptations of Problem-Based Learning Models: A Systematic Review. *Frontiers in Education, 8*, Article 1139987. https://doi.org/10.3389/feduc.2023.1139987