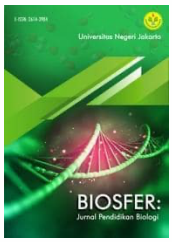
Biosfer: Jurnal Pendidikan Biologi, 19(2), 766-775 (2026)

Biosfer: Jurnal Pendidikan Biologi

Journal homepage: http://journal.unj.ac.id/unj/index.php/biosfer



Inquiry learning based on real-life problem solving (IL-RLPS) to develop critical thinking in air pollution education

Maria Gaudensia Ladja*, Nuny Nurhabibah, Ari Widodo, Sariwulan Diana

Biology Education, Faculty of Mathematics and Natural Science, Universitas Pendidikan Indonesia, Indonesia

*Corresponding author: mariagaudensialadja1@upi.edu

ARTICLE INFO

Article history
Received: 02 June 2026
Revised: 14 August 2026
Accepted: 20 September 2026

Keywords:
Air pollution education
Ceiling effect
Critical thinking skills
Inquiry-based learning
Normalized gain

ABSTRACT

Science instruction in Indonesian secondary classrooms often positions students as passive recipients, limiting opportunities for evidence-based reasoning. This gap is particularly acute for critical thinking development, which is central to biology and science education yet underexplored in the context of air-pollution-based inquiry in Indonesia. This study examined the effect of an *inquiry learning* model based on *real-life problem solving* (IL-RLPS) focused on air pollution on the critical thinking skills of 31 Grade VIII students at a public junior high school in Bandung, Indonesia. The study used a one-group pretest–posttest design conducted over two class meetings through six learning phases: stimulation, problem formulation, data collection, application, communication, and reflection, culminating in the design of an air filter prototype. Student's critical thinking skills were measured using a 10-item test, validated by two science education experts (content validity ratio = 0.89), supported by an implementation observation sheet and a seven-criterion project rubric (α = 0.83). The Wilcoxon signed-rank test showed a statistically significant improvement ($Z = -6.853$; $p < .001$) with a large effect size ($r = 0.87$). The overall N-gain was 0.16 (low), largely attributable to a ceiling effect among seven high-scoring students; excluding these students raised the N-gain to 0.21, with 17 students recording a mean gain of 0.57 (moderate). Interpretation and knowledge-transfer indicators showed the strongest gains, whereas solution evaluation and investigation design-higher-order metacognitive demands-remained comparatively underdeveloped. These findings suggest that IL-RLPS is a promising, low-cost strategy for cultivating critical thinking in biology classrooms, particularly for students with initially low-to-moderate ability, provided that future implementations extend beyond two sessions and incorporate explicit scaffolding of evaluative reasoning.

© 2026 Universitas Negeri Jakarta. This is an open-access article under the CC-BY license (https://creativecommons.org/licenses/by/4.0)

INTRODUCTION

Critical thinking is an essential competency for students to interpret evidence, evaluate information, and make reasoned decisions in science learning (Osborne, 2010). In Indonesian science classrooms, however, students' critical-thinking performance remains a concern, particularly in tasks requiring analysis, evaluation, and evidence-based reasoning (Arifah et al., 2023). One contributing factor is the persistence of teacher-centered instruction that emphasizes knowledge transmission and memorization, providing limited opportunities for students to question, analyze, and evaluate scientific information (Carvalho et al., 2015; Wati et al., 2026). Such conditions highlight the need for learning approaches that engage students in authentic problems and provide opportunities to construct and justify solutions through evidence.

Inquiry-based learning has philosophical roots in Dewey's pragmatism, which emphasizes learning through experience, investigation, and engagement with meaningful problems (Dewey, 1910, 1938). In science and biology education, inquiry provides students with opportunities to formulate questions, investigate evidence, construct explanations, and evaluate possible solutions. Empirical evidence indicates that guided inquiry, particularly when accompanied by appropriate scaffolding, can promote higher-order reasoning and critical-thinking development among secondary students (Lazonder & Harmsen, 2016; Basak Erkacmaz et al., 2023; Wati et al., 2026). Recent studies in biology education likewise indicate that inquiry-oriented learning can improve students' critical-thinking performance by engaging them in investigation and evidence-based reasoning. However, inquiry learning is often implemented primarily as a process for investigating scientific concepts, while the learning process is not always explicitly organized around an authentic real-life problem that requires students to develop and evaluate a practical solution. This limitation provides a rationale for integrating inquiry learning with real-life problem solving (RLPS) in the present study.

Integrating inquiry learning with real-life problem solving (RLPS) can provide a more authentic context in which students investigate complex problems, evaluate evidence, and develop practical solutions. Real-life problems are typically open-ended and require students to connect scientific knowledge with evidence-based decision making (Carvalho et al., 2015; Fortus et al., 2005). Recent studies have also shown the potential of real-life and problem-solving-oriented learning to promote the application of scientific knowledge beyond conventional classroom tasks (Rocha et al., 2024). Air pollution provides a particularly relevant context for this integration because it is an environmental problem that students can encounter in their everyday surroundings and investigate through observable conditions and available evidence (Zagatti et al., 2020). However, existing air-pollution education studies have focused primarily on environmental knowledge, attitudes, awareness, and behavioral outcomes rather than on the development of critical thinking through an inquiry process (Isenaj et al., 2025; Kollmuss et al., 2002). Thus, the use of air pollution as an authentic problem within an inquiry learning framework remains insufficiently examined as a means of developing critical thinking among junior high school students.

Against this background, the present study examines an inquiry learning model based on real-life problem solving (IL-RLPS) in the context of air pollution. Unlike previous inquiry studies that primarily focus on investigating scientific concepts or developing critical thinking through structured inquiry activities, this study places an authentic environmental problem at the center of the inquiry process and requires students to develop a practical solution through an air-filter prototype. The study also examines changes in critical-thinking performance at both the class and individual levels, including variation in normalized gain among students with different initial performance levels. This approach provides a more detailed view of how students respond to an inquiry learning experience centered on an authentic environmental problem. Accordingly, this study addresses the following research question: How does implementing an inquiry learning model based on real-life problem solving (IL-RLPS) on the topic of air pollution affect the critical thinking skills of junior high school students at the class and individual levels?

METHODS

Research Design

This study employed a one-group pretest–posttest interventional design (Shadish et al., 2002), selected because the primary aim was to evaluate changes in students' critical thinking as a consequence of a specific pedagogical intervention in a naturalistic classroom context rather than to establish causal

inference. The absence of a control group is acknowledged as a limitation; all inferences are therefore interpretive and indicative rather than causal.

Population and Sample

Participants consisted of 31 Grade VIII students (Class VIII-E) at a public junior high school in Bandung, Indonesia, selected through purposive sampling based on teacher consent and the alignment of class characteristics with the study's instructional needs. Table 1 presents the demographic distribution of participants.

Table 1.
Participant Demographic Characteristics (n = 31)

Gender	n (%)	Prior Science Achievement	n (%)
Male	16 (51.6%)	Low (<60)	8 (25.8%)
Female	15 (48.4%)	Moderate (60–79)	16 (51.6%)
		High (≥80)	7 (22.6%)
Total	**31 (100%)**		**31 (100%)**

Note. Prior science achievement bands are based on the most recent science examination score recorded in the school's academic report.

The class spans a mixed range of prior achievement rather than a single homogeneous band, which matters for interpretation: because inquiry learning is theorized to benefit most from productive struggle at the edge of a learner's current competence (Vygotsky, 1978), a heterogeneous sample provides a more realistic test of whether IL-RLPS benefits students unevenly depending on their starting point, a question this study returns to when interpreting individual-level N-gain patterns.

Learning Procedure

Instruction was delivered across two class meetings, adapting Pedaste et al., (2015) *inquiry cycle* framework. The learning sequence comprised six phases: (1) *stimulation*, (2) *problem formulation*, (3) *data collection*, (4) *application*, (5) *communication*, and (6) *reflection*. The sequence is grounded in cognitive constructivism (Piaget, 1952) and sociocultural theory (Vygotsky, 1978), and each phase was designed to progressively activate a specific dimension of critical thinking as defined by Facione (2011). The details of each learning phase and their corresponding critical-thinking indicators are presented in Table 2, and the overall inquiry flow is illustrated in Figure 1.

Table 2.
IL-RLPS Learning Phases and Corresponding CT Indicators

Meeting	CT Indicator	Learning Process	Instrument	Product
1	Interpretation of phenomena	Stimulation – students observe tissues placed at different school locations for two weeks to identify differences in airborne particle levels.	Pre-test & post-test	Tissue observation sheet
	Cause-and-effect analysis	Problem Formulation–students formulate investigation questions on sources of air pollution in the school environment.	Pre-test & post-test	—
	Empirical data evaluation	Data Collection – field data collection through dust particle observation, pollutant source identification, and interpretive discussion.	Pre-test & post-test	Product design sheet
2	Inference & scientific conclusion	Application–design of an air filter prototype based on field data analysis.	Pre-test & post-test	Prototype product
	Scientific explanation & communication	Communication & Reflection – presentation of findings, evaluation of reasoning, and prototype design revision.	Pre-test & post-test	Presentation & testing

Note. Adapted from Pedaste et al., (2015). CT = critical thinking.

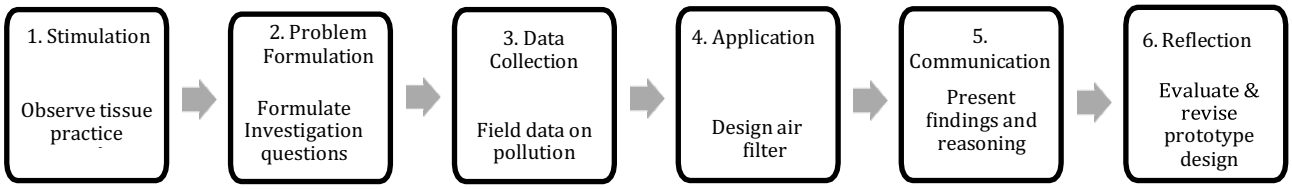


Figure 1. Six-phase inquiry flow of the IL-RLPS intervention on air pollution, adapted from Pedaste et al. (2015).

Each phase in Figure 1 is designed to target a specific critical thinking demand rather than simply structuring classroom activity: the stimulation and problem formulation phases are intended to prompt the interpretation and analysis indicators, application targets inference and explanation, and the communication–reflection phases are where evaluation is exercised through peer and self-critique of the prototype design. This mapping between the inquiry phase and cognitive demand, rather than the sequence alone, is what links the IL-RLPS procedure to the critical thinking outcomes reported in the Results section.

Research Instruments

Three instruments were employed. First, a 10-item critical-thinking test with multiple-choice reasoning items was developed based on Facione's (2011) six cognitive domains. Content validity was established through expert judgment by two science education experts (Content Validity Ratio, CVR = 0.89, exceeding Lawshe (1975) critical threshold of 0.62 for two experts), followed by revision based on their feedback. Item discrimination and difficulty indices were computed on pilot data (n = 28 from a parallel class) using classical test theory; items with a discrimination index < 0.20 were revised prior to administration. Internal consistency of the final instrument was estimated using the Kuder-Richardson 20 (KR-20) formula, yielding KR-20 = 0.74, indicating acceptable reliability for a research instrument (Kubinger, 2005). Table 3 presents the mapping between Facione's (2011) domains and the operational indicators developed for this study.

Table 3.
Mapping of Facione's (2011) Domains to Operational Indicators

Domain	Operational Definition	Item Indicator	Item No.
Interpretation	Understanding and explaining the meaning of observational data	Observation & Interpretation; Pattern Identification & Analysis	Items 1, 3
Analysis	Identifying causal relationships between variables	Cause-and-Effect Analysis; Inference & Variable Connection	Items 2, 6
Evaluation	Assessing data credibility and solution feasibility	Method Evaluation; Solution Evaluation	Items 5, 9
Inference	Drawing evidence-based conclusions and formulating hypotheses	Inference; Prediction & Hypothesis Thinking	Items 4
Explanation	Providing scientific reasoning and designing investigations	Explanation & Reasoning; Designing Investigation	Items 7, 8
Self-Regulation	Applying and transferring knowledge to new contexts	Application & Knowledge Transfer	Item 10

Note. Developed from (Facione, 2011).

Table 3 operationalizes Facione's (2011) six-domain model into the specific item indicators used in the critical thinking test, and this mapping is what allows the aggregate test score reported later to be decomposed into indicator-level results: because Interpretation and Application/Knowledge Transfer are assessed through comparatively concrete, observation-based items, whereas Evaluation and Explanation require students to justify a judgment against criteria, the indicator-level gap reported in the Results section is consistent with, rather than incidental to, the difficulty structure built into the instrument itself.

Second, an implementation observation sheet for the inquiry syntax, based on Lazonder & Harmsen's (2016) guided inquiry framework, was completed by the cooperating teacher during instruction to verify procedural fidelity. The sheet comprised 18 binary items (Yes/No) covering teacher facilitation behaviors and student inquiry activities; inter-rater agreement was not applicable as a single observer was used, but the cooperating teacher received a briefing session to standardize interpretation of each item before data collection. Overall implementation fidelity was 94.4% (17/18 items observed as fully enacted). Third, a seven-criterion project rubric assessed creativity, feasibility, problem relevance, understanding of science concepts, prototype quality, presentation, and teamwork. Rubric reliability was estimated using Krippendorff's α on scores assigned by two independent raters to a 20% random subsample of projects (α = 0.83), indicating strong inter-rater agreement (Krippendorff, 2004). The rubric is presented in Table 4.

Table 4.
Project Rubric for the Air Filter Prototype

Criterion	Score 4	Score 3	Score 2	Score 1
Creativity & Originality	Highly innovative and unique	Creative modification	Imitation with minor changes	No novelty
Feasibility & Applicability	Directly applicable, low-cost	Minor modifications needed	Major adjustments required	Unrealistic or impractical
Connection to Problem	Directly addresses tissue data	Relevant, not directly linked	General, not specific	No clear connection
Science Concept Understanding	2+ principles correctly explained	1 principle correctly explained	Principle stated, poorly explained	No scientific principle stated
Prototype Quality	Functional and neat	Adequately functional	Poorly functional	Non-functional
Presentation & Communication	Very clear; all questions answered	Clear; most questions answered	Adequately clear	Unclear
Teamwork	All members active, well-coordinated	Most members active	Some members dominate	One member dominates

The seven criteria in Table 4 were designed so that the rubric captures critical thinking as it is expressed through a physical artifact rather than a written response: feasibility and connection to problem correspond to the application or knowledge transfer indicator, while science concept understanding corresponds to explanation. This design choice allows the qualitative prototype evidence discussed later in the Results section to be interpreted as convergent, rather than merely illustrative, evidence for the quantitative test findings.

Data Analysis Techniques

Improvements in critical thinking were calculated using the normalized gain (N-gain) formula (Hake, 1998), with gains categorized as high, moderate, or low following Hake's (1998) established thresholds. Data normality was assessed using the Shapiro–Wilk test (appropriate for $n < 50$). Because data were non-normally distributed ($p < .05$), hypothesis testing used the Wilcoxon signed-rank test. Effect size was calculated via *rank-biserial correlation* ($r = Z / \sqrt{N}$; Rosenthal, 1991) with Cohen's (1988) conventions: $r > 0.5$ = large. To address ceiling effects, N-gain was additionally analyzed within the subgroup of students with pre-test scores below 80. Qualitative data from the observation sheet and project rubric were used to enrich the interpretation of quantitative findings, following a convergent mixed-methods logic (Creswell & Creswell, 2018).

RESULTS AND DISCUSSION

Significant Improvement in Critical Thinking with a Large Effect Size

Students showed an improvement in critical-thinking performance following the IL-RLPS intervention. As shown in Table 5, the mean critical-thinking score increased from 60.97 in the pretest to 75.16 in the posttest, while the standard deviation decreased from 24.41 to 17.30. This pattern indicates an overall improvement accompanied by reduced score variability among the participants.

The findings indicate that IL-RLPS was associated with a substantial improvement in students' critical-thinking performance, while the reduced variability of scores suggests more consistent outcomes across participants following the intervention. This pattern is consistent with Lazonder and

Harmsen's (2016) meta-analysis, which demonstrated that guided inquiry can produce meaningful improvements in students' scientific reasoning when appropriate instructional guidance is provided. Similarly, Sucilestari and Arizona (2020) reported that inquiry-based learning enhanced critical-thinking skills in an Indonesian science education context. The present findings extend these studies by demonstrating the potential of an inquiry approach that explicitly integrates real-life problem solving with an authentic environmental issue, namely air pollution. From a constructivist perspective, the findings support the view that critical thinking can develop through active engagement with problems, evidence, and the construction of solutions rather than through passive reception of information. The distinctive contribution of this study is therefore not merely the observed improvement in critical thinking, but the integration of guided inquiry, real-life air-pollution problems, and an air-filter prototype task as a learning sequence that connects scientific reasoning with authentic problem solving.

Table 5.
Descriptive Statistics of Pre-test and Post-test Critical Thinking Scores

Statistic	**Pre-test**	**Post-test**
Number of Students (n)	31.00	31.00
Mean	60.97	75.16
Standard Deviation	24.41	17.30
Minimum	20.00	40.00
Maximum	90.00	100.00

Ceiling Effect Explains Low Overall N-gain

The overall N-gain of 0.16 (low category) warrants careful interpretation, as it does not straightforwardly reflect the intervention's effectiveness. Seven students began with pre-test scores of 80 or above, leaving them with very limited room for improvement in critical thinking under the N-gain formula (Hake, 1998). When this ceiling-effect subgroup was excluded, the N-gain rose to 0.21. Table 6 and Figure 2 present individual-level patterns.

Table 6.
Individual N-gain Scores

Pattern Category	**n**	**%**	**N-gain Range**	**Interpretation**
Increased	17	54.80	0.25–1.00 (M = 0.57)	Moderate gain; predominantly low–moderate initial ability
Ceiling Effect	7	22.60	0.00	Stable high pre-test scores (≥80); mastery maintained, not stagnation
Decreased	5	16.10	−2.00 to −0.20	Attributable to non-cognitive factors (inattentiveness, limited discussion participation)
Static	2	6.50	0.00	No pre–post change

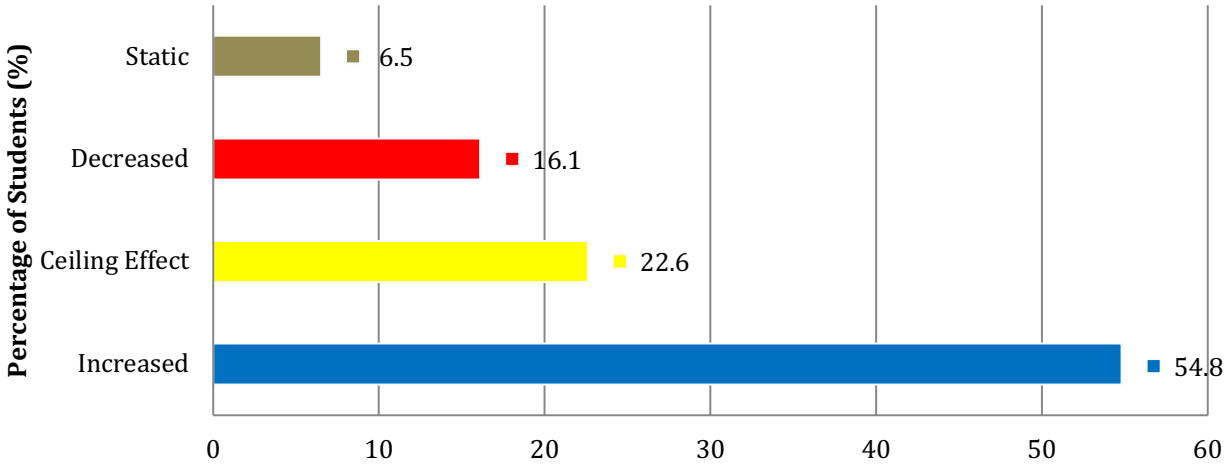


Figure 2. Distribution of Individual N-gain Patterns (Note. Blue = Increased; Yellow = Ceiling Effect; Red = Decreased; Grey = Static)

Three distinct patterns of student response emerged, indicating that the intervention did not affect all learners uniformly. Students with initially lower or moderate competence appeared to benefit most from the structured scaffolding and active engagement embedded in the guided inquiry sequence. This pattern is consistent with Vygotsky's (1978) zone of proximal development, in which learning is enhanced when instructional support enables learners to perform beyond their initial independent competence. In contrast, students who entered the intervention with high performance tended to maintain their achievement, suggesting that the intervention may have supported the consolidation of previously established competence rather than producing substantial additional gains. This interpretation is compatible with Piaget's (1952) view of cognitive development, in which learning experiences contribute to the consolidation and organization of existing cognitive structures. The small group showing decreased performance should therefore be interpreted cautiously, as the accompanying qualitative observations suggest that fluctuations were more closely associated with attentiveness and participation than with weaknesses in the inquiry design itself, consistent with the importance of learner engagement in inquiry-based science learning (Carvalho et al., 2015). Overall, the findings suggest that IL-RLPS may be particularly beneficial for students who begin with low-to-moderate levels of competence, while students with initially high performance may require more challenging and sustained inquiry tasks to promote further development.

Higher-Order Indicators Lag Behind Lower-Order Gains

Figure 3 reveals a consistent hierarchical pattern in indicator achievement: interpretation (84%) and application/transfer (84%) dominated, while solution evaluation (32%) and investigation design (39%) lagged considerably.

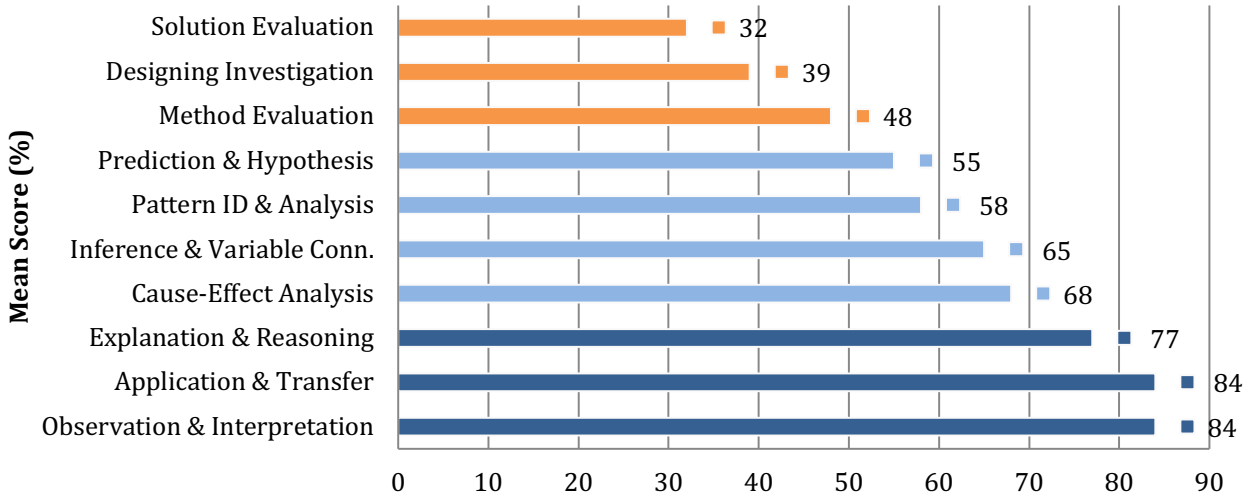


Figure 3. Critical Thinking Indicator Achievement Scores (%). Note. Dark blue = High (≥70%); Light blue = Moderate (50–69%); Orange = Low (<50%).

This pattern is explicable through cognitive developmental theory. Piaget's (1952) account of the concrete-to-formal operational transition suggests that 12–15-year-old students are more proficient at processing information through direct, hands-on experience than through abstract evaluative judgment. The stimulation and data collection phases positioned students in concrete observational contexts, placing tissues in real school settings and measuring particulate differences, thereby directly activating interpretation and causal-analysis capacities (Dewey, 1938; Facione, 2011). By contrast, solution evaluation and investigation design require *metacognitive judgment* - assessing the quality of one's own reasoning process, a higher-order function that Facione (2011) explicitly notes requires sustained, recurring practice through evidence-based discussion and reflection to develop. The finding that two-session interventions reliably raise lower-order indicators while leaving higher-order ones underdeveloped is consistent with Osborne et al.'s (2004) argument that scientific argumentation, the cognitive core of evaluative ability, requires iterative classroom practice over extended periods, not a single instructional episode.

Product Quality Reflects Critical Thinking Development

A positive pattern was observed between individual critical thinking ability and group product quality, providing convergent qualitative evidence for the quantitative findings. Air-filter prototypes developed by the five groups, illustrating their distinct filtration strategies. The five groups pursued three distinct filtration strategies, each reflecting a different level of scientific reasoning about particle removal. One group (score 92) built a cardboard housing with a fan and a mask as the filtering medium, correctly reasoning that the mask would retain fine dust particles and act as a porous barrier. Three groups (scores 70–72) used water-filled bottle systems, explaining filtration through particle entrapment in water, and in one case, linking a longer air pathway to a higher probability of pollutant capture, a relationship close to bona fide diffusion reasoning. The remaining group (score 88) used charcoal as an adsorptive medium, correctly identifying its pollutant-capturing function even without using the term "adsorption." Across all five groups, product quality and score tracked most closely with how explicitly each group could justify its choice of filter medium in causal, mechanism-based terms rather than by trial and error, consistent with the *National Research Council's* (2012) emphasis on constructing explanations as a core science practice.

Groups with more structured prototypes and more coherent scientific argumentation were predominantly composed of students who showed greater N-gain in analysis and explanation indicators. This confirms that individual critical-thinking capacity contributes to the quality of collective reasoning in group work (Hmelo-Silver et al., 2007). The integration of *scientific practices* and *engineering design thinking* in the prototype activity, requiring students to define a problem, develop a solution from observational data, and test and explain their design, aligns with the *Next Generation Science Standards* vision of authentic science-engineering learning (National Research Council, 2012; Putra et al., 2023; Yu et al., 2020). This contributes to biology education by showing that even a minimal two-session inquiry intervention can produce design artifacts of varying scientific quality that are meaningfully correlated with students' development of reasoning.

The reflection phase also proved to be a critical juncture. Students recognized errors in their initial assumptions, such as the belief that all school areas have equivalent pollution levels, and revised their claims based on observational data. Several students identified limitations in their prototype designs and proposed improvements, indicating that self-regulation is emerging as a core component of critical thinking (Facione, 2011). Hmelo-Silver et al. (2007) and Tai et al. (2018) confirm that reflective practices in inquiry-based contexts contribute significantly to evaluative judgment, the capacity to assess evidence quality and argument strength - a finding this study corroborates at the junior high school level in Indonesia.

Pedagogical Implications

Three key pedagogical implications emerge from these findings. First, IL-RLPS demonstrates the greatest potential for students with initially low or moderate critical-thinking ability; high-ability students require additional enrichment to avoid ceiling effects. Second, evaluating intervention effectiveness under a single-group design must go beyond the mean N-gain to include subgroup analyses, identification of ceiling effects, and reporting of effect sizes for accurate interpretation. Third, developing evaluation and investigation design indicators requires higher-frequency exposure through recurring, sustained evidence-based argumentation practices systematically embedded across science learning at the junior high school level.

CONCLUSION

This study demonstrates that an inquiry learning model based on real-life problem solving (IL-RLPS), implemented through the authentic context of air pollution, can significantly support the development of junior high school students' critical thinking in biology and science education. Addressing the study's research question, the model produced a statistically significant, large-effect improvement in critical thinking, with the clearest and most consistent gains occurring among students who began with low-to-moderate ability, the group that benefited most from the structured scaffolding embedded in the six-phase inquiry sequence. A ceiling effect among already high-performing students depressed the overall normalized gain, underscoring the importance of reporting effect size and individual-level patterns alongside class-level averages when evaluating single-group interventions. Indicator-level analysis further revealed that interpretation and knowledge transfer developed most

readily, while solution evaluation and investigation design, the higher-order metacognitive demands of critical thinking, require longer interventions and explicit, iterative scaffolding of evaluative reasoning to develop fully. These findings collectively suggest that air-pollution-based IL-RLPS is a promising, ecologically valid, and low-cost strategy for biology and science classrooms in Indonesia, and that its effectiveness can be meaningfully enhanced by extending implementation beyond two sessions and deliberately embedding argumentation and reflective-evaluation activities across the curriculum.

ACKNOWLEDGMENT

The authors thank the participating teachers and students for their willingness to take part in this study, and the expert validators for their valuable feedback on the research instruments.

REFERENCES

Arifah, F., Suprapto, N., & Setiawan, B. (2023). Profile of critical thinking skills in science learning class at junior high school on additive materials. *Studies in Learning and Teaching, 4*(3), 601–607. https://doi.org/10.46627/silet.v4i3.288

Bao, L., Cai, T., Koenig, K., Fang, K., Han, J., Wang, J., Liu, Q., & Ding, L. (2009). Learning and scientific reasoning. *Science, 323*(5914), 586–587. https://doi.org/10.1126/science.1167740

Basak Erkacmaz, K., Bakirci, H., & Kara, Y. (2023). Effect of inquiry-based laboratory approach on scientific process skills, critical thinking skills, and opinions of ninth grade students: Cell unit example. *Jurnal Penelitian Dan Pembelajaran IPA, 9*(2), 139. https://doi.org/10.30870/jppi.v9i2.19733

Billah, A., Masykuri, M., Sarwanto, & Sajidan. (2021). Analysis of critical thinking in junior high school students through science learning in Indonesia: A systematic review. *Journal of Physics: Conference Series, 1796*(1), 012013. https://doi.org/10.1088/1742-6596/1796/1/012013

Carvalho, C., Fíuza, E., Conboy, J., Fonseca, J., Santos, J., Gama, A. P., & Salema, M. H. (2015). Critical thinking, real life problems and feedback in the sciences classroom. *Journal of Turkish Science Education, 12*(2), 21–31. https://doi.org/10.36681/

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

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.

Dewey, J. (1910). *How we think*. D.C. Heath & Company.

Dewey, J. (1938). *Experience and education*. Macmillan.

Facione, P. A. (2011). *Critical thinking: What it is and why it counts*. Insight Assessment.

Fortus, D., Krajcik, J., Dershimer, R. C., Marx, R. W., & Mamlok-Naaman, R. (2005). Design-based science and real-world problem-solving. *International Journal of Science Education, 27*(7), 855–879. https://doi.org/10.1080/09500690500038165

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data. *American Journal of Physics, 66*(1), 64–74. https://doi.org/10.1119/1.18809

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42*(2), 99–107. https://doi.org/10.1080/00461520701263368

Isenaj, Z. S., Moshammer, H., Berisha, M., & Weitensfelder, L. (2025). Effect of an educational intervention on pupil's knowledge, attitudes, perceptions, and behavior on air pollution in public schools in Pristina. *European Journal of Investigation in Health, Psychology and Education*, *15*(5), 69. https://doi.org/10.3390/ejihpe15050069

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Prentice Hall.

Kollmuss, A., & Agyeman, J. (2002). Mind the gap: Why do people act environmentally and what are the barriers to pro-environmental behavior? *Environmental Education Research, 8*(3), 239–260. https://doi.org/10.1080/13504620220145401

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Sage Publications.

Kubinger, K. D. (2005). Psychological test calibration using the Rasch model: Some critical suggestions on traditional approaches. *International Journal of Testing, 5*(4), 377–394.

https://doi.org/10.1207/S15327574IJT0504_3
Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.
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
Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Review of Educational Research, 86*(3), 681–718. https://doi.org/10.3102/0034654315627366
National Research Council. (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas.* National Academies Press.
Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science, 328*(5977), 463–466. https://doi.org/10.1126/science.1183944
Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching, 41*(10), 994–1020. https://doi.org/10.1002/tea.20035
Pedaste, M., Mäeots, M., Siiman, L. A., de Jong, T., van Riesen, S., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review, 14*, 47–61. https://doi.org/10.1016/j.edurev.2015.02.003
Piaget, J. (1952). *The origins of intelligence in children*. International Universities Press.
Putra, P. D. A., Sulaeman, N. F., & Supeno. (2023). Exploring students' critical thinking skills using the engineering design process in a physics classroom. *Asia-Pacific Education Researcher, 32*, 141–149. https://doi.org/10.1007/s40299-021-00640-3
Rocha, H., Viseu, F., & Matos, S. (2024). Problem-solving in a real-life context: An approach during the learning of inequalities. *European Journal of Science and Mathematics Education, 12*(1), 21–37. https://doi.org/10.30935/scimath/13828
Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Rev. ed.). Sage Publications.
Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Houghton Mifflin.
Sucilestari, R., & Arizona, K. (2020). The impact of inquiry-based learning on students' critical thinking skills. *Advances in Social Science, Education and Humanities Research, 408*, 172–175. https://doi.org/10.2991/assehr.k.200220.031
Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education, 76*, 467–481. https://doi.org/10.1007/s10734-017-0220-3
Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
Wati, I. K., Rosana, D., & Wilujeng, I. (2026). How science teachers foster critical thinking in junior secondary classrooms: A narrative inquiry from a constructivist perspective. International *Journal of Educational Reform. Advance online publication*. https://doi.org/10.1177/10567879261458338
Yu, K.-C., Wu, P.-H., & Fan, S.-C. (2020). Structural relationships among high school students' scientific knowledge, critical thinking, engineering design process, and design product. *International Journal of Science and Mathematics Education, 18*, 1001–1022. https://doi.org/10.1007/s10763-019-10007-2
Zagatti, E., Russo, M., & Pietrogrande, M. C. (2020). On-site monitoring indoor air quality in schools: A real-world investigation to engage high school science students. *Journal of Chemical Education, 97*(11), 4069–4072. https://doi.org/10.1021/acs.jchemed.0c00065