International Journal of Essential Competencies in Education
https://journal-center.litpam.com/index.php/ijece/index

Dec. 2025 Vol. 4, No. 2
e-ISSN: 2962-2131
pp. 100-106

When Inquiry Isn't Enough: Why Structured Reflection Can Strengthen Preservice Science Teachers' Critical Thinking

Ni Nyoman Sri Putu Verawati*, Wahyudi, Nina Nisrina

Physics Education Department, University of Mataram, Mataram, Indonesia.
*Correspondence: veyra@unram.ac.id

Article Info

Article History
Received: October 2025;
Revised: November 2025;
Published: December 2025

Keywords
Scientific inquiry;
Structured reflection;
Preservice science teachers;
Critical thinking skills;
Performance evaluation

Abstract

Prompt Inquiry-based learning is frequently treated as a ready-made route to critical thinking, but classroom inquiry often becomes procedural unless students are pushed to examine assumptions, justify claims, and revise interpretations. This commentary discusses the focal study "Emphasizing reflective processes in scientific inquiry and its impact on preservice science teachers' critical thinking skills" and argues that its strongest contribution is showing how structured reflective elements can turn inquiry activities into repeated practice in reasoning. At the same time, the evidence should be read carefully: the gains are very large, and the intervention bundles several supports (anomalies, monitoring worksheets, prompts, and feedback) that may each contribute. We outline alternative explanations, identify what the study clarifies and what it does not yet prove, and offer implications for future research designs and teacher-education practice.

https://doi.org/10.36312/waxp5a03

How to Cite
Verawati, N. N. S. P., Wahyudi, W., & Nisrina, N. (2025). When Inquiry Isn't Enough: Why Structured Reflection Can Strengthen Preservice Science Teachers' Critical Thinking. *International Journal of Essential Competencies in Education*, 4(2), 100-106. https://doi.org/10.36312/waxp5a03

INTRODUCTION

Critical thinking is typically defined as disciplined and reflective judgment about what to believe or do, involving analysis, inference, evaluation, and self-regulation (Dewey, 1933; Ennis, 2011, 2018; Facione, 2020). In teacher education, this is not just a generic graduate attribute. Preservice science teachers need to reason about evidence, evaluate competing explanations, and justify instructional decisions. Yet many teacher-prep courses treat inquiry learning as if it automatically produces these habits. That assumption is questionable: inquiry can devolve into following steps, completing worksheets, or chasing "right answers" unless the learning design explicitly demands reasoning moves and holds students accountable to standards of argument quality (Lederman & Abell, 2014; Uiterwijk-Luijk et al., 2019).

This commentary centers on the focal article Prayogi et al. (2025), *Emphasizing reflective processes in scientific inquiry and its impact on preservice science teachers' critical thinking skills*. The authors argue that inquiry "embedded as a routine" is often not robust enough to develop critical thinking, and they propose a modification: integrate structured reflective processes into the inquiry cycle. Their intervention uses anomalous phenomena to trigger curiosity, process-monitoring worksheets to make thinking visible, guided reflection prompts to require justification and revision, and performance evaluation to provide feedback aligned with critical thinking indicators (Prayogi et al., 2025). The reported quantitative gains are substantial, and lecturers' qualitative accounts emphasize attention, engagement, and perceived value.

The article offers a plausible instructional mechanism grounded in reflection theory. However, the causal story may be more complicated than "reflection caused critical thinking," because the intervention packages multiple scaffolds, the design uses intact groups without random assignment, and the outcome measure is strongly aligned to the intervention. These issues do not invalidate the work, but they affect what claims are warranted.

COMMENTARY ON THE FOCAL ARTICLE

Prayogi et al. (2025) begin from a familiar problem: preservice teachers often show weak performance in analytic reasoning and logic, and teacher educators face constraints in designing critical-thinking instruction (Bezanilla et al., 2019; Fitriani et al., 2019; Lee et al., 2021). They position inquiry learning as promising but not sufficient. This aligns with broader evidence that inquiry can improve learning outcomes, yet its impact varies depending on guidance and scaffolding (Alfieri et al., 2011; Furtak et al., 2012; Pedaste et al., 2015). The paper's move is to treat reflection not as an optional afterthought but as a designed process that repeatedly activates higher-order thinking during inquiry.

What the study does well

The intervention is described in a way that is actionable for teacher educators. It operationalizes "reflective inquiry" as a repeatable weekly cycle: anomalous phenomenon → planning and experimentation with monitoring → structured reflection prompts → performance evaluation and feedback (Prayogi et al., 2025). This structure matters because a recurring weakness in inquiry implementation is that learners do not spontaneously engage in metacognitive monitoring, alternative explanation generation, or systematic evaluation of evidence (Miri et al., 2007; Qing et al., 2010). The authors' design directly targets that gap.

The mixed-methods structure is also a strength. The quantitative phase tests whether scores change, while lecturer interviews offer a window into feasibility and perceived mechanisms such as situational interest (Creswell & Creswell, 2018; Linnenbrink-Garcia et al., 2010). Additionally, the authors report content and construct validation of learning tools and instruments via expert review and inter-rater agreement (Chan, 2022; Emmer & Millett, 1970), which improves confidence that the intervention tools were not arbitrary.

What the results suggest

Quantitatively, Prayogi et al. (2025) report that experimental groups moved from very low baseline scores to high posttest scores on an Ennis-aligned critical thinking essay assessment, while the control group improved only modestly. Their statistical tests indicate a significant interaction between time and treatment group, and post hoc comparisons show experimental groups outperforming control at posttest (Prayogi et al., 2025). Qualitatively, lecturers attribute engagement to anomalies and collaborative activity, and they frame the reflective cycle and feedback as valuable for developing reasoning (Prayogi et al., 2025).

Interpreted cautiously, the study suggests that inquiry can become more consistently "thinking-intensive" when reflective work is explicitly prompted, documented, and evaluated rather than assumed.

CRITICAL APPRAISAL AND ALTERNATIVE INTERPRETATIONS

The focal article's overall logic is credible, but several interpretive risks matter if the goal is a strong causal claim.

The effect sizes are unusually large

The reported differences between experimental and control groups at posttest are extremely large (Prayogi et al., 2025). That is possible, but it makes alternative explanations more plausible. One possibility is assessment alignment: if reflection prompts and evaluation rubrics repeatedly train students in the same reasoning moves that the essay test rewards, students may become better at producing the kind of responses the test recognizes. That is still an educational gain, but it may reflect improvement in "argument performance in this genre" more than generalized critical thinking capacity across contexts (Ennis, 2011; Facione, 2020).

Intact groups and instructor effects

Because the study uses intact classes across institutions, the intervention may be confounded with differences in instructor style, classroom culture, and local academic expectations (Prayogi et al., 2025). The authors show pretest means are not statistically different, but equivalence on one outcome variable does not guarantee equivalence on unmeasured variables such as writing skill, motivation, prior inquiry experience, or language proficiency. A realistic alternative explanation is that lecturer implementation quality, not reflection itself, accounts for part of the gain. This is not a trivial concern in inquiry research, where teacher guidance strongly shapes outcomes (Lederman & Abell, 2014; Uiterwijk-Luijk et al., 2019).

The "active ingredient" may not be reflection alone

The intervention bundles at least four potentially powerful elements: anomalous phenomena (cognitive conflict), process monitoring (metacognitive scaffolding), guided reflection prompts (analytic rehearsal), and performance evaluation (feedback and standards). The paper tends to attribute effects to "reflective processes," but the design cannot isolate which component matters most. Cognitive conflict alone can stimulate conceptual change and deeper reasoning when properly guided (Akmam et al., 2018; Prayogi et al., 2019). Likewise, feedback aligned to explicit criteria can substantially improve reasoning quality (Procter, 2020; Larrivee, 2008). So a competing interpretation is that explicit criteria plus feedback drove much of the gain, with reflection acting as a carrier rather than the driver.

Construct coverage and transfer remain open

The essay test captures reasoning in written form, but critical thinking is broader than essay performance and may be influenced by writing fluency and familiarity with argumentation tasks (Ennis, 2018; Dwyer et al., 2014). The study also measures immediate post-intervention outcomes. It does not yet establish whether gains persist or transfer into authentic teaching practice, such as analyzing student misconceptions or designing inquiry lessons that cultivate critical thinking in school learners (Körkkö et al., 2016; Loughran, 2002). The qualitative data is also lecturer-centered; student reflections, artifacts, and classroom observations would offer a stronger basis for mechanism claims (Braun & Clarke, 2006; Choy et al., 2017).

What a cautious conclusion looks like

Given these points, we would restate the most defensible inference as follows: a tightly scaffolded inquiry design that requires monitoring, structured reflection, and rubric-guided feedback is associated with large short-term improvements on an Ennis-aligned critical thinking essay assessment among preservice science teachers (Prayogi et al., 2025). That is a strong and useful result. But it does not yet prove that "reflection" is the sole causal mechanism or that the gains generalize beyond the measured task and context.

IMPLICATIONS FOR RESEARCH AND PRACTICE

Implications for research

Future work should test whether the reflective-inquiry package works because of reflection specifically, or because of the combination of scaffolds. One approach is a dismantling design: compare inquiry-only vs inquiry + prompts vs inquiry + feedback rubric vs inquiry + monitoring worksheet, and then the full package. This would identify which elements are necessary, which are sufficient, and whether there are interaction effects (Pedaste et al., 2012; Mäeots et al., 2016; Procter, 2020). Another priority is implementation fidelity: document what lecturers do, how consistently, and how students actually use reflection tools. Inquiry interventions often fail or succeed because of implementation variation rather than theoretical differences (Lederman & Abell, 2014).

Outcome measurement should expand beyond essays. Add performance-based tasks that resemble real teacher work: evaluating competing explanations in lab data, critiquing flawed student reasoning, designing lesson plans with embedded reasoning prompts, or microteaching episodes scored for reasoning moves (Vogt & Schmiemann, 2020; Loughran, 2002). Finally, longitudinal follow-up is essential to test durability and transfer of critical thinking habits, consistent with reflection theory's emphasis on sustained practice (Dewey, 1933; Schön, 1983; Rodgers, 2002).

Implications for practice in teacher education

Despite the cautions, the focal article offers several practical design moves that teacher educators can adopt. First, use anomalies with purpose. Anomalous phenomena should not function as entertainment. They should be framed as claims-evidence problems that force explanation revision. That matches inquiry-cycle perspectives emphasizing iterative sense-making (Pedaste et al., 2015) and evidence that anomalous data can sharpen critical thinking when connected to argument evaluation (Prayogi et al., 2019). Second, make thinking visible through monitoring. Process-monitoring tools can reduce "procedural inquiry" by forcing students to record decisions, uncertainties, and alternative explanations, supporting metacognitive control (Choy et al., 2017; Pedaste et al., 2012). Third, treat reflection as a reasoning practice tied to standards. Reflection becomes powerful when it is not just narrative ("what did you learn?") but analytic ("what evidence supports your claim?", "what alternative explanation remains plausible?", "what would change your conclusion?"). This connects directly to critical thinking indicators and to the reflective tradition in Dewey and later scholarship (Dewey, 1933; Ennis, 2018; Rodgers, 2002). Fourth, feedback matters. If performance evaluation is central, it should be implemented as feedback on reasoning quality, not merely correctness. This aligns with work on reflective assessment tasks and teacher reflective development (Larrivee, 2008; Procter, 2020; Körkkö et al., 2016).

A practical caution is workload. Over-scaffolded reflection can become compliance writing. A more sustainable approach may be fewer prompts, repeated consistently, with emphasis on quality and revision. The focal article's success may partly rely on intensity; scaling requires careful design so the cognitive demand remains high without becoming administratively heavy.

CONCLUSION

The focal study by Prayogi et al. (2025) makes a persuasive argument that inquiry alone does not reliably produce critical thinking, and it offers a clear instructional architecture for strengthening inquiry through structured reflective processes. The reported gains are large, and the intervention design is concrete enough to replicate: anomalies to trigger attention, monitoring to surface reasoning, prompts to require justification and revision, and feedback aligned to critical thinking standards.

Yet the strongest reading should remain cautious. The study does not isolate which components are essential, it cannot fully rule out instructor and context effects, and it does not yet show long-term transfer to teaching practice. The value of the work is therefore twofold: it provides promising evidence that structured reflective inquiry can boost measured reasoning performance, and it sets an agenda for more diagnostic research that separates mechanisms and tests durability. In teacher education terms, it is a strong reminder that if we want inquiry to cultivate critical thinking, we must engineer reflection as a disciplined practice of reasoning, not assume it will appear simply because students are "doing inquiry" (Dewey, 1933; Ennis, 2018; Pedaste et al., 2015).

Author Contributions

The authors have sufficiently contributed to the study, and have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Acknowledgement

The authors would like to thank the University of Mataram for its support in the preparation of this commentary.

Declaration of Interest

The authors declare no conflict of interest.

REFERENCES

Akmam, A., Anshari, R., Amir, H., Jalinus, N., & Amran, A. (2018). Influence of Learning Strategy of Cognitive Conflict on Student Misconception in Computational Physics Course. *IOP Conference Series: Materials Science and Engineering, 335*, 012074. https://doi.org/10.1088/1757-899x/335/1/012074

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology, 103*(1), 1–18. https://doi.org/10.1037/a0021017

Bezanilla, M. J., Fernández-Nogueira, D., Poblete, M., & Galindo-Domínguez, H. (2019). Methodologies for teaching-learning critical thinking in higher education: The teacher's view. *Thinking Skills and Creativity, 33*, 100584. https://doi.org/10.1016/j.tsc.2019.100584

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Chan, P. (2022). An Empirical Study on Data Validation Methods of Delphi and General Consensus. *Data, 7*(2), 18. https://doi.org/10.3390/data7020018

Choy, S. C., Yim, J. S. C., & Tan, P. L. (2017). Reflective thinking among preservice teachers: A Malaysian perspective. *Issues in Educational Research, 27*(2), 234–251.

Creswell, J. W., & Creswell, J. D. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (5th ed.). SAGE.

Dewey, J. (1933). *How We Think: A Restatement of the Relation of Reflective Thinking to the Educative Process*. D.C. Heath.

Dwyer, C. P., Hogan, M. J., & Stewart, I. (2014). An integrated critical thinking framework for the 21st century. *Thinking Skills and Creativity, 12*, 43–52. https://doi.org/10.1016/j.tsc.2013.12.004

Emmer, E. T., & Millett, G. B. (1970). *Improving teaching through experimentation: A laboratory approach*. Prentice-Hall.

Ennis, R. H. (2011). The nature of critical thinking: An outline of critical thinking dispositions and abilities. *Inquiry: Critical Thinking Across the Disciplines, 26*(2). https://doi.org/10.5840/inquiryctnews201126214

Ennis, R. (2018). Critical Thinking Across the Curriculum: A Vision. *Topoi, 37*(1), 165–184. https://doi.org/10.1007/s11245-016-9401-4

Facione, P. A. (2020). *Critical Thinking: What It Is and Why It Counts*. Measured Reasons LLC.

Fitriani, H., Asy'ari, M., Zubaidah, S., & Mahanal, S. (2019). Exploring the Prospective Teachers' Critical Thinking and Critical Analysis Skills. *Jurnal Pendidikan IPA Indonesia, 8*(3), 379–390. https://doi.org/10.15294/jpii.v8i3.19434

Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012). Experimental and Quasi-Experimental Studies of Inquiry-Based Science Teaching: A Meta-Analysis. *Review of Educational Research, 82*(3), 300–329. https://doi.org/10.3102/0034654312457206

Guo, L. (2022). How should reflection be supported in higher education? — A meta-analysis of reflection interventions. *Reflective Practice, 23*(1), 118–146. https://doi.org/10.1080/14623943.2021.1995856

Körkkö, M., Kyrö-Ämmälä, O., & Turunen, T. (2016). Professional development through reflection in teacher education. *Teaching and Teacher Education, 55*, 198–206. https://doi.org/10.1016/j.tate.2016.01.014

Larrivee, B. (2008). Development of a tool to assess teachers' level of reflective practice. *Reflective Practice, 9*(3), 341–360. https://doi.org/10.1080/14623940802207451

Lederman, N. G., & Abell, S. K. (Eds.). (2014). *Handbook of Research on Science Education, Volume II*. Routledge. https://doi.org/10.4324/9780203097267

Lee, N. Y., Wang, Z., & Lim, B. (2021). The development of critical thinking: What university students have to say. *Teaching in Higher Education, 29*(1), 286–299. https://doi.org/10.1080/13562517.2021.1973412

Linnenbrink-Garcia, L., Durik, A. M., Conley, A. M., Barron, K. E., Tauer, J. M., Karabenick, S. A., & Harackiewicz, J. M. (2010). Measuring Situational Interest in Academic Domains. *Educational and Psychological Measurement, 70*(4), 647–671. https://doi.org/10.1177/0013164409355699

Loughran, J. J. (2002). Effective Reflective Practice: In Search of Meaning in Learning about Teaching. *Journal of Teacher Education, 53*(1), 33–43. https://doi.org/10.1177/0022487102053001004

Mäeots, M., Siiman, L., Kori, K., & Pedaste, M. (2016). Relation Between Students' Reflection Levels and Their Inquiry Learning Outcomes. *EDULEARN Proceedings*, 5558–5564. https://doi.org/10.21125/edulearn.2016.2324

Miri, B., David, B.-C., & Uri, Z. (2007). Purposely Teaching for the Promotion of Higher-order Thinking Skills: A Case of Critical Thinking. *Research in Science Education, 37*(4), 353–369. https://doi.org/10.1007/s11165-006-9029-2

Pedaste, M., Mäeots, M., Leijen, Ä., & Sarapuu, T. (2012). Improving Students' Inquiry Skills through Reflection and Self-Regulation Scaffolds. *Technology, Instruction, Cognition and Learning, 9*, 81–95.

Pedaste, M., Mäeots, M., Siiman, L. A., de Jong, T., van Riesen, S. A. N., 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

Prayogi, S., Muhali, M., Yuliyanti, S., Asy'ari, M., Azmi, I., & Verawati, N. N. S. P. (2019). The Effect of Presenting Anomalous Data on Improving Student's Critical Thinking Ability. *International Journal of Emerging Technologies in Learning, 14*(6), 133–137. https://doi.org/10.3991/ijet.v14i06.9717

Prayogi, S., Verawati, N. N. S. P., Bilad, M. R., Samsuri, T., Hunepi, H., Asy'ari, M., Yusup, M. Y., Azmi, I. & Ernita, N. (2025). Emphasizing reflective processes in scientific inquiry and its impact on preservice science teachers' critical thinking skills. *Social Sciences & Humanities Open, 12*, 101895. https://doi.org/10.1016/j.ssaho.2025.101895

Procter, L. (2020). Fostering critically reflective thinking with first-year university students: Early thoughts on implementing a reflective assessment task. *Reflective Practice, 21*(4), 444–457. https://doi.org/10.1080/14623943.2020.1773421

Qing, Z., Jing, G., & Yan, W. (2010). Promoting preservice teachers' critical thinking skills by inquiry-based chemical experiment. *Procedia - Social and Behavioral Sciences, 2*(2), 4597–4603. https://doi.org/10.1016/j.sbspro.2010.03.737

Rodgers, C. (2002). Defining Reflection: Another Look at John Dewey and Reflective Thinking. *Teachers College Record, 104*(4), 842–866. https://doi.org/10.1111/1467-9620.00181

Schön, D. (1983). *The Reflective Practitioner*. Temple Smith.

Uiterwijk-Luijk, L., Krüger, M., Zijlstra, B., & Volman, M. (2019). Teachers' role in stimulating students' inquiry habit of mind in primary schools. *Teaching and Teacher Education, 86*, 102894. https://doi.org/10.1016/j.tate.2019.102894

Vogt, F., & Schmiemann, P. (2020). Assessing Biology Pre-Service Teachers' Professional Vision of Teaching Scientific Inquiry. *Education Sciences, 10*(11), 332. https://doi.org/10.3390/educsci10110332