Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

TITLE: Beyond Pretest-Posttest: Indicator-Specific Trajectories of Critical Thinking and Problem-Solving Skills in Multiliteracy Project-Based Science Learning

Deden Agustira

dedenagustira@uskm.ac.id

Sindang Kasih University

Rani Tania Pratiwi

Universitas Kuningan

Wensley M. Reyes

University of the Philippines Manila

Nico Irawan

Thai Global Business Administration Technological College

Research Article

Keywords: critical thinking, problem solving, learning trajectories, project-based learning, multiliteracies, science education

Posted Date: November 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11508151/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

This study examined whether endpoint comparisons adequately represent changes in critical thinking and problem-solving during multiliteracy project-based science learning. We conducted a secondary analysis of repeated assessment data from 27 Grade VII students participating in a field evaluation of a project-based learning model with multiliteracy pedagogy. The reconstructed dataset comprised 2,160 observations across eight occasions and ten indicators. We summarized indicator and domain means and fitted separate generalized estimating equation models with an exchangeable working correlation and robust sandwich standard errors to estimate within-unit trends. Mean critical-thinking scores increased from 74.00 at P1 to 85.85 at P8, while problem-solving scores increased from 77.67 to 88.52. However, the piecewise models did not indicate statistically significant linear change within either instructional unit. Both domains showed a marked increase between P4 and P5, coinciding with a transition from interactions among living organisms to environmental pollution and a change in assessment context. This discontinuity cannot therefore be treated as evidence of a developmental turning point. The findings demonstrate the value of repeated indicator-level analysis for distinguishing endpoint differences from within-sequence patterns and highlight the need to examine assessment comparability when tracking complex learning outcomes.

Introduction

Critical thinking and problem solving are foundational to contemporary science education: they enable learners to engage with scientific ideas, make reasoned decisions, and respond to complex challenges in school and everyday life (Matthee & Turpin, 2019; Szabo et al., 2020; Razak et al., 2022; Puig Mauriz et al., 2023; Aposika, 2026). Project-based learning can foster higher-order and critical thinking by involving students in authentic, collaborative, and reflective work. Its outcomes, however, depend on how projects are designed, implemented, and assessed (Zhang et al., 2023; Becerra-Goicochea et al., 2026). Meta-analyses generally report beneficial associations between project-based learning and cognitive outcomes, particularly higher-order thinking, while also indicating that effects vary across contexts and methods (Nurkanti et al., 2025; Farshad & Fortin, 2026; Arsih et al., 2026).

Beyond asking whether project-based learning is effective, researchers must consider what observed gains actually signify (Liu & Pásztor, 2022; Loyens et al., 2023; Tafakur et al., 2023). A pretest–posttest design compresses an instructional sequence into two snapshots: one taken at the outset and another at the end. Although this comparison is suitable for estimating an overall difference, it may obscure temporary setbacks, delayed gains, recovery, or variation among component processes (Sasson et al., 2018). A single composite score can therefore underrepresent change in multidimensional constructs such as critical thinking and problem solving. Longitudinal educational research likewise shows that trajectories may vary by construct and context rather than conforming to one common pattern (Widlund et al., 2024).

Critical thinking and problem-solving draw on interrelated processes, including metacognition, reasoned judgment, inquiry, analysis, evaluation, application, and reflection. Problem solving additionally involves defining and representing a problem, generating and applying possible solutions, and evaluating their outcomes. These activities are shaped by self-regulation and by emotional and social conditions (Syam et al., 2020; Winne & Azevedo, 2022; Alpindo et al., 2024). In project-based science learning, students may also need to exercise cognitive flexibility, scientific reasoning, planning, evaluation, creativity, and collaboration. Because these processes are distinguishable rather than interchangeable, they warrant assessment that captures their specific contributions to learning (Miller et al., 2025; Sánchez-García & Angosto, 2026)

Repeated assessment is particularly important in project-based learning because it makes it possible to follow performance as instruction unfolds. When assessments are distributed across a learning sequence, teachers can align evidence with instructional aims, provide formative feedback, and examine the learning process rather than relying exclusively on final outcomes. Such alignment is central to the coherence of project-based learning systems (Huang et al., 2026). Performance is also shaped by the immediate learning and assessment environment, including guidance, scaffolding, opportunities for practice, collaboration, and the demands of a particular situation. Consequently, an observed score reflects both students' capabilities and the conditions under which those capabilities are elicited (Ouyang et al., 2021; Eticha et al., 2024).

This article presents a secondary analysis of repeated indicator-level data collected during the field evaluation of the Project-Based Learning model with Multiliteracy Pedagogy (PjBL-PM). The parent research-and-development project used the Dick and Carey framework to develop the model and examined its broader effects on critical thinking, problem solving, decision making, and environmental literacy (Agustira et al., 2025). The present analysis addresses a distinct question: rather than focusing only on change between baseline and endpoint, it traces ten component indicators across eight learning occasions and examines whether the apparent endpoint gains are accompanied by statistically detectable within-unit growth.

The reanalysis was guided by three questions: (1) How do the indicators of critical thinking and problem solving vary across the eight learning occasions? (2) Are the observed endpoint gains accompanied by statistically detectable longitudinal change within each instructional unit? and (3) What additional temporal variation becomes visible when indicator-level trajectories are examined instead of endpoint differences alone? These questions are descriptive and inferential, not causal, because the observations arose from a non-randomized field evaluation and the assessment context differed between the two instructional units.

Methods

Research design and setting

We undertook a secondary analysis of repeated learning-process data gathered during the field evaluation of an established PjBL-PM model. The parent project followed a research-and-development approach informed by the Dick and Carey model and included one-to-one, small-group, and field-evaluation stages (Agustira et al., 2025). The present study did not re-estimate the parent study's overall intervention effect. Instead, it treated the repeated student-by-occasion-by-indicator records as longitudinal observations.

The field evaluation took place in a Sekolah Penggerak junior high school in Kuningan Regency, West Java, Indonesia, and involved 27 Grade VII students. Instruction covered two science units: Interaction of Living Organisms and Environmental Pollution. PjBL-PM combined project-based learning with multiliteracy pedagogy and followed six instructional stages: identifying understanding, generating ideas, analyzing ideas, identifying learning experiences, developing meaningful understanding, and implementing ideas. In the parent study, these stages were referred to as the IPA-IPI syntax.

The eight occasions were split between two content units rather than forming a single, contextually uniform sequence. Occasions P1-P4 concerned Interaction of Living Organisms, whereas P5-P8 addressed Environmental Pollution. The P4-P5 boundary therefore represents more than elapsed instructional time: it also coincides with changes in content, problem context, and assessment format. This distinction informed the analytical strategy and interpretation.

Participants and measures

The analytic sample consisted of 27 Grade VII students and included ten indicators. Critical thinking was represented by basic clarification, construction skills, inference, advanced clarification, and strategies and tactics. Problem solving was represented by understanding the problem, brainstorming all possible solutions, devising a plan, carrying out the plan, and evaluating the result. In the parent study, structured assessment items were mapped to these indicators, and instrument development included expert validation and item analysis (Agustira et al., 2025).

We reconstructed the repeated dataset from the student-by-indicator records presented in Appendix 23 of the dissertation. The resulting dataset contained 27 students x 8 learning occasions x 10 indicators, or 2,160 observations. We checked the reconstructed data against the indicator-level means and domain-level values reported in the source material.

The scaling procedure used here was derived during the audit; it is not explicitly identified in the dissertation's Methods section as the original scoring formula. Accordingly, it should be understood as a transparent reconstruction step rather than as a verified account of the original scoring algorithm. Agreement with published aggregate means supports numerical consistency at that level but does not independently validate every reconstructed student-level score.

Statistical analysis

The analysis addressed three complementary aims. First, we calculated each indicator's mean score at every learning occasion to describe its trajectory. Endpoint change was calculated as Δ = mean(P8)-mean(P1). These contrasts summarize the difference between the first and last observations but do not imply that change occurred steadily between them.

Second, we used generalized estimating equations (GEE) to account for the dependence among repeated observations from the same student. Separate models were fitted for critical thinking and problem solving using a Gaussian distribution, identity link, exchangeable working correlation, and robust (sandwich) covariance estimates. Piecewise slopes were estimated for P1-P4 and P5-P8 so that within-unit trends were not conflated with the marked change at the boundary between units.

Third, we examined indicator-level endpoint contrasts descriptively to identify patterns that might be concealed by domain averages. A quadratic model spanning the full sequence was considered only as a sensitivity analysis. Because the P4-P5 transition combines elapsed time with changes in content and assessment context, it was not treated as an equivalent interval within a single uninterrupted trajectory.

The significance threshold was set at α = .05, and the analyses were implemented using reproducible Python scripts. The estimated slopes were not interpreted causally, and the observed score pattern was not taken as evidence of a discrete developmental stage.

Results

Domain trajectories

Both domains had higher mean scores at P8 than at P1. Critical thinking rose from 74.00 to 85.85, an endpoint difference of 11.85 points; problem solving increased from 77.67 to 88.52, a difference of 10.85 points (Table 1). Neither series followed a consistently rising path. Across P1-P4, critical thinking changed only modestly, from 74.00 to 75.56, while problem solving declined from 77.67 to 75.89. Both domains then showed a substantial increase at P5, followed by fluctuations at the higher score level.

Table 1
Domain-level trajectories across eight learning occasions

Domain	P1	P2	P3	P4	P5	P6	P7	P8	P1-P8
Critical thinking	74.00	74.89	74.56	75.56	86.67	86.52	86.96	85.85	+ 11.85
Problem solving	77.67	75.56	75.44	75.89	88.44	86.67	89.48	88.52	+ 10.85

All five critical-thinking indicators had higher P8 than P1 means, but the magnitude of change differed. Basic clarification increased by 14.81 points and strategies and tactics by 14.44 points. Advanced clarification increased by 12.78 points, construction skills by 9.44 points, and inference by 7.78 points. The descriptive trajectories also show that the endpoint differences did not arise through a common monotonic pattern. In particular, strategies and tactics increased sharply at P5, declined at P6, and

recovered at P8. The indicator-level endpoint comparisons and occasion-by-occasion values are reported in Table 2.

Table 2
Critical-thinking endpoint and trajectory statistics

Indicator	P1	P8	Δ P1-P8	Descriptive slope	Paired p
Basic clarification	73.33	88.15	+ 14.81	2.39	.086
Construction skills	76.11	85.56	+ 9.44	2.47	.110
Inference	75.56	83.33	+ 7.78	1.59	.326
Advanced clarification	70.56	83.33	+ 12.78	2.34	.045
Strategies and tactics	74.44	88.89	+ 14.44	2.54	.008

Problem-solving indicators showed even greater differentiation. Brainstorming all possible solutions increased by 15.19 points and evaluating the result by 15.37 points. Understanding the problem and carrying out the plan each increased by 10.56 points. In contrast, devising a plan increased by only 2.59 points from P1 to P8. This endpoint result is particularly informative because the planning indicator reached 96.67 at P5 before declining to 81.48 at P8. Thus, the small endpoint gain does not indicate that planning performance remained low throughout the intervention; it indicates that a temporary peak was not sustained at the endpoint. The corresponding indicator-level endpoint comparisons and trajectory values are presented in Table 3.

Table 3
Problem-solving endpoint and trajectory statistics

Indicator	P1	P8	Δ P1-P8	Descriptive slope	Paired p
Understand the problem	77.22	87.78	+ 10.56	1.93	.200
Brainstorm all possible solutions	80.00	95.19	+ 15.19	2.50	.020
Devise a plan	78.89	81.48	+ 2.59	1.77	.705
Carry out the plan	78.33	88.89	+ 10.56	2.28	.147
Evaluate the result	73.89	89.26	+ 15.37	2.94	.004

Within-unit longitudinal growth

The piecewise GEE estimates qualify the interpretation suggested by the endpoint differences. During Unit 1, the estimated critical-thinking slope was 0.43 points per occasion (SE = 0.36, 95% CI [-0.28, 1.14], p = .232), while the problem-solving slope was − 0.54 (SE = 0.48, 95% CI [-1.48, 0.39], p = .256). Neither estimate was statistically significant. During Unit 2, the estimated slopes were − 0.20 for critical thinking (SE = 1.17, 95% CI [-2.48, 2.08], p = .864) and 0.30 for problem solving (SE = 1.11, 95% CI [-1.87, 2.48], p =

.784); these estimates were also not statistically significant. The piecewise GEE coefficients, robust standard errors, confidence intervals, and p values are reported in Table 4.

Table 4
Piecewise GEE estimates of within-unit longitudinal change

Unit	Domain	β (slope)	SE	95% CI	p
P1-P4	Critical thinking	0.433	0.363	-0.278 to 1.145	.232
P1-P4	Problem solving	-0.544	0.479	-1.484 to 0.395	.256
P5-P8	Critical thinking	-0.200	1.165	-2.484 to 2.084	.864
P5-P8	Problem solving	0.304	1.108	-1.868 to 2.475	.784

The P4-P5 discontinuity

The clearest discontinuity in the eight-occasion series occurred between P4 and P5. Mean critical-thinking scores rose by 11.11 points, from 75.56 to 86.67, while problem-solving scores increased by 12.56 points, from 75.89 to 88.44. The size of these concurrent changes makes the boundary an important feature of the observed data, but the analysis cannot determine what produced it. The transition coincided with a move from Interaction of Living Organisms to Environmental Pollution and with changes in the associated assessment context.

Accordingly, P5 is not treated as a developmental stage or a pivotal moment in cognitive development. The evidence supports a more limited conclusion: P4-P5 marks a pronounced discontinuity in observed performance that must be interpreted alongside the changes in instruction and measurement context. The indicator-specific problem-solving trajectories across P1-P8 are shown in Fig. 2, allowing the temporal patterns of the five indicators to be examined separately.

For project-based science education, the principal methodological implication is that evaluations of multidimensional competencies should preserve temporal and indicator-level evidence alongside conventional effectiveness estimates. Future studies should strengthen longitudinal measurement comparability and use repeated assessments that are explicitly aligned with meaningful instructional events. Such designs can provide a more precise account of how complex cognitive performances vary during project-based science learning.

The findings therefore support a clear distinction between endpoint gain and longitudinal growth. The endpoint gains should not be interpreted as evidence of smooth or continuously accelerating cognitive development, and the P4-P5 discontinuity should not be interpreted as a discrete developmental stage or causal turning point. Indicator-level analysis adds information by showing that different cognitive processes can display different endpoint magnitudes, temporary peaks, and fluctuations that are invisible in a single gain score.

This secondary analysis offers a more fine-grained account of student performance than a simple comparison of the first and final occasions. Critical-thinking scores increased from 74.00 at P1 to 85.85 at P8, and problem-solving scores rose from 77.67 to 88.52. However, these endpoint differences were not matched by statistically significant linear trends within either instructional unit. The sharp increase between P4 and P5 occurred alongside changes in instructional content and assessment context, so it cannot be interpreted as evidence of uninterrupted cognitive growth.

Discussion

Critical thinking followed heterogeneous indicator-specific trajectories

Project-based learning (PjBL) can provide conditions for developing transferable skills, promoting deeper engagement with subject matter, and increasing learner motivation. These outcomes, however, depend on how projects are designed and enacted, including the relevance of the context, teachers' preparation, and the quality of instructional support (Belwal et al., 2021; Al-Bahadli et al., 2023; Mutanga, 2024; Chandra et al., 2025; Sánchez-García & Reyes-de-Cózar, 2025; Becerra-Goicochea et al., 2026). Well-designed PjBL typically connects authentic, learner-centred projects to explicit learning goals and combines scaffolding with assessment, collaboration, reflection, equitable participation, and appropriate technology. Together, these features can support meaningful learning and the development of important contemporary competencies (Al-Kamzari & Alias, 2025; Chandra et al., 2025; Jadhav et al., 2025; Retno et al., 2025; Tierney et al., 2025).

Considering time explicitly changes how PjBL outcomes are interpreted: rather than relying only on initial and final scores, researchers can examine the course of learning and differences in how performance develops (Zhao & Wang, 2022). Longitudinal evidence may reveal higher final scores without a consistently increasing or statistically detectable linear pattern within each instructional unit. This is one reason educational change should be examined as a process rather than inferred from two observations alone (Dowling et al., 2019).

Evidence generally associates PjBL with beneficial learning outcomes and higher-order thinking, although effects vary with context, subject matter, and implementation (Mutanga, 2024; Peng et al., 2024; Arsih et al., 2026; Becerra-Goicochea et al., 2026). Reviews and meta-analyses report promising overall effects but also identify methodological constraints and contextual heterogeneity that complicate estimates of their magnitude (Ferrero et al., 2021; Nurkanti et al., 2025; Farshad & Fortin, 2026). Moreover, a higher final score does not establish that every learner or every component skill followed a similar path. Such differences should therefore not be taken as proof of uniform change (Laven & Yang, 2025).

The piecewise GEE results reinforce this distinction. For critical thinking, the estimated within-unit slope was 0.433 points per occasion in P1-P4 (SE = 0.363, 95% CI [-0.278, 1.145], p = .232) and − 0.200 points

per occasion in P5-P8 (SE = 1.165, 95% CI [-2.484, 2.084], p = .864). For problem solving, the corresponding estimates were − 0.544 points per occasion in P1-P4 (SE = 0.479, 95% CI [-1.484, 0.395], p = .256) and 0.304 points per occasion in P5-P8 (SE = 1.108, 95% CI [-1.868, 2.475], p = .784). Thus, none of the four within-unit linear slopes provided statistically detectable evidence of systematic aggregate growth.

The main finding of this secondary longitudinal analysis is that higher endpoint scores did not provide evidence of continuous linear growth throughout the learning sequence. Critical thinking increased by 11.85 points, from 74.00 at P1 to 85.85 at P8; problem solving increased by 10.85 points, from 77.67 to 88.52. An endpoint-only interpretation might describe these differences as improvement. The repeated observations, however, show that performance fluctuated and that the largest change occurred at the boundary between two distinct instructional units.

Problem-solving trajectories reveal stronger differentiation among cognitive processes

The aggregate critical-thinking score also masks short-term reversals across the sequence. Overall performance can be higher at the final occasion even when individual components vary markedly in between. An endpoint gain therefore does not mean that every underlying cognitive operation improved simultaneously or to the same extent. Figure 1 illustrates the value of indicator-level analysis: a domain mean summarizes several distinct trajectories but cannot stand in for any single trajectory.

Project-based science tasks can give learners opportunities to exercise cognitive flexibility by working on authentic, complex problems and considering alternatives with peers. Such experiences may support problem solving and relate to gains in science and literacy, particularly when students can explore more than one idea or route to a solution (Dhimitruka & Surendran, 2023; Miller et al., 2025; Xiong et al., 2025; Honra & Monterola, 2026). Although clarification and strategy-related indicators are useful for evaluating learning, the available evidence does not justify assuming that one skill necessarily develops before another. Their expression depends on task demands, collaboration, and the wider instructional setting (Chinchilla-Rodríguez et al., 2021).

PjBL is widely recognized for fostering critical and higher-order thinking, but the broader literature documents substantial variation in how these skills are conceptualized and measured, echoing the findings referenced in query (Nurkanti et al., 2025). Differences among indicator trajectories may consequently reflect the cognitive demands of a task and the instructional circumstances, rather than the relative importance of each skill. Interpreting these patterns requires attention to both task design and assessment procedures (Estrella et al., 2020; Soysal & Soysal, 2023).

Figure 1 displays this variation across critical-thinking indicators. Basic clarification and strategies and tactics recorded relatively large endpoint increases, whereas inference changed less. Strategies and tactics also rose sharply at P5, fell at P6, and recovered at P7 before reaching 88.89 at P8. The pattern

cautions against treating the five indicators as equivalent components that change at a common, constant rate.

The critical-thinking trajectories provide a more differentiated picture than the domain mean alone. All five indicators showed higher P8 than P1, but the magnitude of change varied considerably. Basic clarification increased from 73.33 to 88.15 (+ 14.81 points), construction skills from 76.11 to 85.56 (+ 9.44), inference from 75.56 to 83.33 (+ 7.78), advanced clarification from 70.56 to 83.33 (+ 12.78), and strategies and tactics from 74.44 to 88.89 (+ 14.44). Thus, the overall gain of 11.85 points in critical thinking was not produced by a uniform shift across its component indicators.

The P4-P5 discontinuity should be interpreted as a contextual transition, not a developmental pivot

The trajectory for devising a plan illustrates why a small endpoint difference should not automatically be read as limited engagement with a skill. Performance on this indicator rose notably at P5 and then declined in later observations. Its final score therefore conceals a substantial range of performance across occasions. Repeated assessment is useful precisely because it describes stability, fluctuation, and temporary peaks without requiring the researcher to make an exaggerated claim about improvement.

This interpretation aligns with science-education work that emphasizes authentic, open-ended problems as contexts for flexible knowledge use. In a meta-analysis of scientific problem-solving approaches and STEM contexts, emphasized the role of problem-solving environments in fostering scientific creativity (Pinar et al., 2025). Research on context-driven problem-solving instruction likewise highlights the relevance of guidance, scaffolding, practice, and learning conditions (Eticha et al., 2024; Lu & Xie, 2024). The present results add a temporal perspective: even over a short instructional sequence, different problem-solving components may follow distinct patterns.

Comparing brainstorming, evaluating a result, and devising a plan also reveals meaningful differences among related processes. Brainstorming and evaluation showed substantial endpoint gains, whereas planning changed little between P1 and P8 despite a pronounced intermediate high point. Generating options broadens the set of possible solutions; planning, by contrast, organizes selected options into a feasible sequence of actions. Their different trajectories reinforce the need to treat problem solving as a set of related but distinguishable processes during project work.

The devising-a-plan indicator is particularly illustrative. Its mean was 78.89 at P1, reached 96.67 at P5, and then fell to 81.48 by P8. Viewed only through the endpoint contrast, the change appears modest (+ 2.59 points); viewed across all eight occasions, it includes a marked temporary peak followed by a decline. The two summaries are not contradictory: they answer different questions about the same observations.

The problem-solving results show an even more pronounced differentiation among indicators. Understand the problem increased from 77.22 at P1 to 87.78 at P8 (+ 10.56), brainstorming all possible solutions from 80.00 to 95.19 (+ 15.19), devise a plan from 78.89 to 81.48 (+ 2.59), carry out the plan from 78.33 to 88.89 (+ 10.56), and evaluate the result from 73.89 to 89.26 (+ 15.37). The overall endpoint gain of 10.85 points therefore masks considerable variation in the degree to which different components changed.

What repeated indicator-level analysis adds beyond pretest-posttest comparisons

This distinction is central to interpreting the findings. To label P5 a developmental pivot would turn an observed score discontinuity into a mechanism that the study was not designed to test. Describing it instead as a transition associated with instructional and assessment context preserves the empirical pattern while keeping the inference within the limits of the design.

Several explanations for the P4-P5 increase remain possible. It may partly reflect learning accumulated during the first unit, but it could also be associated with familiarity with the new content, differences in task demands or assessment difficulty, context-specific performance, or the way prior knowledge interacted with the second unit. The available data cannot distinguish among these explanations. The most defensible account is therefore that P4-P5 represents an observed performance discontinuity coinciding with an instructional and assessment transition, not the emergence of a new developmental stage.

The measurement context adds a further qualification. The assessments used in P1-P4 and P5-P8 were based on different content-specific item sets. Although the reconstruction reproduced the aggregate means reported in the source material, that agreement does not establish longitudinal measurement invariance across all eight occasions. Research on response-shift bias shows that changes in how participants interpret assessment items can influence estimates of intervention effects, underscoring the need to examine measurement comparability when interpreting longitudinal results (Olivera-Aguilar & Rikoon, 2025).

Research on coherent PjBL systems similarly emphasizes that learning and assessment need to be aligned with the instructional sequence and that evidence of knowledge-in-use is shaped by the coherence of curriculum, instruction, and assessment (He et al., 2023; Li et al., 2024). The unit boundary in the present dataset should therefore be treated as analytically meaningful rather than as a neutral break in an otherwise equivalent sequence (Markula & Aksela, 2022; He et al., 2023).

The study design does not support interpreting the discontinuity as a critical cognitive turning point after four learning occasions. P1-P4 addressed Interaction of Living Organisms, whereas P5-P8 addressed Environmental Pollution. Thus, the transition involved a change not only in instructional exposure but also in subject matter, problem context, and assessment conditions. Scores consequently reflect an interaction between students capabilities and the opportunities and demands presented by each task.

Both figures show a marked shift between P4 and P5: critical thinking increased from 75.56 to 86.67, and problem solving rose from 75.89 to 88.44. Because this shift appeared concurrently in both domains and across multiple indicators, it is a salient feature of the observed data. Its cause, however, cannot be isolated from the available design.

Implications for PjBL science research and assessment

Taken together, the results support a cautious but more differentiated interpretation. Endpoint scores were higher, and indicators displayed substantial temporal variation; nevertheless, the data did not establish a uniform linear growth pattern within either instructional unit.

This distinction also helps explain why the present GEE analysis was specified piecewise rather than as a single uninterrupted linear trend. Generalized estimating equations were developed to estimate marginal relationships in longitudinal data while accounting for within-subject dependence (Liang & Zeger, 1986). In the present analysis, the piecewise specification was used to prevent the large P4-P5 discontinuity from being treated as though it represented the same temporal process as the changes occurring within P1-P4 or P5-P8.

The findings do not imply that pretest-posttest designs are unsuitable or should be abandoned. Rather, endpoint comparisons and repeated trajectory analyses address different questions. A pretest-posttest design estimates whether performance differs between two selected points, whereas repeated indicator-level analysis shows how performance varies between those points and whether an endpoint difference accompanies sustained, uneven, or discontinuous patterns.

This is consistent with the methodological argument that within-person variation can provide information that is obscured when learning is represented primarily through between-person or endpoint summaries (Saqr, 2023). In science education, the value of repeated assessment is particularly relevant to PjBL because learning opportunities are distributed across projects, investigations, artefacts, and successive applications of disciplinary ideas. Coherent PjBL research has consequently emphasized assessment systems that provide evidence across units rather than relying exclusively on a single summative outcome (He et al., 2023; Li et al., 2024).

An analysis restricted to P1 and P8 would describe the overall endpoint difference, whereas the full sequence reveals how performance varied between those points. Indicators with comparable starting and ending values can have very different intermediate profiles. The planning indicator is one example: its small endpoint contrast conceals a pronounced peak during the sequence. A more stable trajectory could produce a similar endpoint difference while suggesting a different learning process.

The reanalysis separates three conclusions that should not be conflated: mean performance was higher at the final occasion; component indicators displayed different trajectories; and the analysis did not detect statistically significant linear growth within either instructional unit. These statements can all be true because each describes a different feature of the repeated data.

Limitations and directions for future research

Evaluation of PjBL should distinguish evidence about overall effectiveness from evidence about learning processes. Comparative or meta-analytic studies can estimate whether PjBL is associated with better outcomes than alternative approaches, whereas repeated process-level measures can show how component skills behave during implementation. Noted that the generally positive direction of PjBL effects is accompanied by uncertainty about effect magnitude because of heterogeneity and methodological limitations in the existing meta-analytic literature (Farshad & Fortin, 2026). At a different analytical level, the present study demonstrates that the interpretation of one dataset changes depending on whether it is reduced to endpoint contrasts or examined as a repeated, indicator-specific sequence (Santhosh et al., 2023).

These findings matter for both research design and assessment practice. First, studies of complex outcomes in PjBL may benefit from retaining repeated observations for individual indicators instead of collapsing the evidence into a single pretest-posttest difference. This is especially pertinent to critical thinking and problem solving, which comprise multiple processes that may be elicited differently by particular tasks.

Second, curriculum, teaching, and assessment need to be aligned. In project-based science learning, assessment performance is shaped in part by the content and activities through which it is elicited. Research on coherent PjBL systems emphasizes consistency between learning objectives, instructional activity, and the evidence used to judge what learners know and can do (He et al., 2023; Li et al., 2024). When content or assessment conditions change across units, differences in scores should be interpreted with that context in view.

Third, the temporary peak in planning performance offers a question for future instructional design, not proof of a planning weakness or a need for greater support than other skills. It indicates that a final score can conceal substantial variation in planning-related performance. Future PjBL implementations might make planning more visible through structured representations, interim checkpoints, explicit criteria for comparing options, and opportunities to revise plans after testing them. Researchers could then investigate whether these supports are associated with more sustained performance rather than higher final scores alone.

Fourth, future studies should strengthen the comparability of measures across occasions. The present analysis did not establish strict longitudinal measurement invariance between the two content-specific assessment blocks. Common anchor items, linked assessment forms, or explicit longitudinal measurement models could help separate change in the intended construct from variation attributable to item difficulty or assessment context.

Conclusions

This study traced critical-thinking and problem-solving indicators across eight occasions in multiliteracy project-based science learning. Although both domains had higher mean scores at the final occasion, their trajectories were uneven and non-monotonic, with distinct fluctuations and temporary peaks. The pronounced increase from P4 to P5 coincided with changes in instructional content and assessment context and therefore cannot be attributed solely to cognitive development. In addition, the reconstructed scoring transformation has not been verified against the original scoring protocol. The findings consequently support the value of repeated indicator-level reporting as a complement to endpoint comparisons, while highlighting the need for transparent scoring documentation and stronger measurement comparability in future longitudinal studies.

Abbreviations

CT
Critical Thinking
PS
Problem-Solving
PjBL
Project-Based Learning
PjBL-PM
Project-Based Learning with Multiliteracy Pedagogy
GEE
Generalized Estimating Equations.

Declarations

Statement on the Use of Artificial Intelligence (AI)

The author acknowledges using generative AI tools, including Scopus AI and ChatGPT during manuscript preparation. The author reviewed the generated material to check that the study's context and substantive meaning were retained.

Author Contribution

Author Contributions StatementD.A. conceptualized the study, developed the research methodology, conducted the data analysis, and drafted the main manuscript. R.T.P. contributed to the interpretation of the findings and critical revision of the manuscript. W.M.R. contributed to the theoretical framing, interpretation of the results, and critical review of the manuscript. N.I. contributed to the methodological review, validation of the analysis, and revision of the manuscript. All authors reviewed and approved the final manuscript.

Acknowledgement

We thank the science teachers and participating public secondary schools in Kuningan Regency, Indonesia, for their contribution to this research.

Data Availability

Availability of data and materialsThe data supporting the findings are presented in the article. Student-level assessment records are not publicly available because access is governed by the permissions and consent arrangements established for the parent study.

References

1. Adah Miller, E., Li, T., Chen, I. C., Krajcik, J., & Codere Kelly, S. (2025). Designing for and investigating elementary students' cognitive flexibility, science, and literacy achievement in project-based science learning. *Disciplinary and Interdisciplinary Science Education Research*, *7*(1), 13. https://doi.org/10.1186/s43031-025-00131-1
2. Agustira, D., Al-Muhdhar, M. H. I., Lestari, S. R., & Pratiwi, R. T. (2025). Development of a project based learning model with multiliteracy pedagogic content to improve critical thinking skills, problem solving skills, decision making skills, and environmental literacy. *Edelweiss Applied Science and Technology*, *9*(3), 2402–2412. https://doi.org/10.55214/25768484.v9i3.5812
3. Al-Bahadli, K. H., Al-Obaydi, L. H., & Pikhart, M. (2023). The Impact of the Online Project-Based Learning on Students' Communication, Engagement, Motivation, and Academic Achievement. *PSYCHOLINGUISTICS*, *33*(2), 217–237. https://doi.org/10.31470/2309-1797-2023-33-2-217-237
4. Al-Kamzari, F., & Alias, N. (2025). A systematic literature review of project-based learning in secondary school physics: Theoretical foundations, design principles, and implementation strategies. *Humanities and Social Sciences Communications*, *12*(1), 286. https://doi.org/10.1057/s41599-025-04579-4
5. Alpindo, O., Istiyono, E., Widihastuti, & Andriyanti, E. (2024). Can Critical-thinking Skills be Measured by Analyzing Metacognition? *Journal of Teaching and Learning*, *18*(2), 194–211. https://doi.org/10.22329/jtl.v18i2.8813
6. Aposika, F. (2026). Investigating the relationship between problem-solving skills and academic performance in STEM subjects among secondary school students. *Discover Education*, *5*(1), 425. https://doi.org/10.1007/s44217-026-01487-w
7. Arsih, F., Azhar, M., Yohandri, N. A., Razak, A., Lasmana, O., Dewi, R. K., & Wati, D. D. E. (2026). Exploring the impact and evolution of project-based learning in science education: A state-of-the-art review. *International Journal of Innovation and Learning*, *40*(2), 125–150. https://doi.org/10.1504/IJIL.2026.155503

8. Becerra-Goicochea, L. J., Ballon-Tito, R. A., Feijoo-Parra, M. F. O., & Molina-Carita, P. A. (2026). *Aprendizaje basado en proyectos en educación superior: Revisión sistemática de su impacto en el desarrollo de habilidades blandas*. https://doi.org/10.5281/ZENODO.20801738
9. Belwal, R., Belwal, S., Sufian, A. B., & Al Badi, A. (2021). Project-based learning (PBL): Outcomes of students' engagement in an external consultancy project in Oman. *Education + Training*, *63*(3), 336–359. https://doi.org/10.1108/ET-01-2020-0006
10. Chandra, V., Mallet, D., Polume, P. N., Morumo, L., Mambo, A. P., & Wranga, E. (2025). Project-Based Learning in STEM. In V. Chandra, M. Lloyd, & D. Mallet (Eds.), *Integrated STEM Education in Papua New Guinea* (pp. 113–147). Springer Nature Singapore. https://doi.org/10.1007/978-981-95-0828-0_8
11. Chinchilla-Rodríguez, Z., Bu, Y., Robinson-García, N., & Sugimoto, C. R. (2021). An empirical review of the different variants of the probabilistic affinity index as applied to scientific collaboration. *Scientometrics*, *126*(2), 1775–1795. https://doi.org/10.1007/s11192-020-03815-9
12. Dhimitruka, I., & Surendran, G. (2023). Adaptation of Project-Based Learning Concepts to the Organic Chemistry I Laboratory Curriculum in a Small College Environment. *Journal of College Science Teaching*, *52*(4), 61–69. https://doi.org/10.1080/0047231X.2023.12290633
13. Dowling, N. M., Raykov, T., & Marcoulides, G. A. (2019). Examining Population Differences in Within-Person Variability in Longitudinal Designs Using Latent Variable Modeling: An Application to the Study of Cognitive Functioning of Older Adults. *Educational and Psychological Measurement*, *79*(3), 598–609. https://doi.org/10.1177/0013164418758834
14. Estrella, S., Zakaryan, D., Olfos, R., & Espinoza, G. (2020). How teachers learn to maintain the cognitive demand of tasks through Lesson Study. *Journal of Mathematics Teacher Education*, *23*(3), 293–310. https://doi.org/10.1007/s10857-018-09423-y
15. Eticha, M. D., Hunde, A. B., & Ketema, T. (2024). Designing a Context-Driven Problem-Solving Method with Metacognitive Scaffolding Experience Intervention for Biology Instruction. *Journal of Science Education and Technology*, *33*(6), 811–822. https://doi.org/10.1007/s10956-024-10107-x
16. Farshad, S., & Fortin, C. (2026). An umbrella review of meta-analyses on project-based learning: Effects on academic achievement, higher-order thinking, and 21st-century skills. *Educational Research Review*, *52*, 100809. https://doi.org/10.1016/j.edurev.2026.100809
17. Ferrero, M., Vadillo, M. A., & León, S. P. (2021). Is project-based learning effective among kindergarten and elementary students? A systematic review. *PLOS ONE*, *16*(4), e0249627. https://doi.org/10.1371/journal.pone.0249627
18. He, P., Krajcik, J., & Schneider, B. (2023). Transforming standards into classrooms for knowledge-in-use: An effective and coherent project-based learning system. *Disciplinary and Interdisciplinary Science Education Research*, *5*(1), 22. https://doi.org/10.1186/s43031-023-00088-z
19. Honra, J. R., & Monterola, S. L. C. (2026). Enhancing biology students' cognitive flexibility through problem-based learning moderated by transdisciplinary thinking. *The Journal of Educational Research*, *119*(1), 129–139. https://doi.org/10.1080/00220671.2025.2525245

20. Huang, M., Lin, X., & Zhong, B. (2026). Project-based assessment: A scoping review of concepts, implementations, and future directions. *Assessment & Evaluation in Higher Education*, 1–56. https://doi.org/10.1080/02602938.2026.2684341
21. Jadhav, S., Pawar, M., Sadhukhan, S., & Iyer, S. (2025). Exploring PBL in Real-Life Course Projects: An Analysis of Two Consecutive Years. In S. Mishra, A. Kothiyal, S. Iyer, S. Sahasrabudhe, A. Lingnau, & R. Kuo (Eds.), *Proceedings of the International Conference on Technology 4 Education 2024, Volume 2* (pp. 320–326). Springer Nature Singapore. https://doi.org/10.1007/978-981-95-1734-3_38
22. Laven, R., & Yang, D. A. (2025). Common misinterpretations of statistical significance and P-values in dairy research. *JDS Communications*, *6*(6), 721–726. https://doi.org/10.3168/jdsc.2025-0835
23. Li, T., Chen, I., Miller, A., Miller, E., Schneider, C. S., B., & Krajcik, J. (2024). The relationships between elementary students' knowledge-in-use performance and their science achievement. *Journal of Research in Science Teaching*, *61*(2), 358–418. https://doi.org/10.1002/tea.21900
24. Liang, K. Y., & Zeger, S. L. (n.d.). *Longitudinal data analysis using generalized linear models*.
25. 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*, 101069. https://doi.org/10.1016/j.tsc.2022.101069
26. 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), 39. https://doi.org/10.1007/s10648-023-09757-x
27. Lu, D., & Xie, Y. N. (2024). The application of educational technology to develop problem-solving skills: A systematic review. *Thinking Skills and Creativity*, *51*, 101454. https://doi.org/10.1016/j.tsc.2023.101454
28. Markula, A., & Aksela, M. (2022). The key characteristics of project-based learning: How teachers implement projects in K-12 science education. *Disciplinary and Interdisciplinary Science Education Research*, *4*(1), 2. https://doi.org/10.1186/s43031-021-00042-x
29. Matthee, M., & Turpin, M. (2019). Teaching critical thinking, problem solving, and design thinking: Preparing IS students for the future. *Journal of Information Systems Education*, *30*(4), 242–252. Scopus.
30. Mutanga, M. B. (2024). Students' Perspectives and Experiences in Project-Based Learning: A Qualitative Study. *Trends in Higher Education*, *3*(4), 903–911. https://doi.org/10.3390/higheredu3040052
31. Nurkanti, M., Lubis, M., Cartono, C., Hudha, A. M., Shukri, A. A. M., & Yasundari, Y. (2025). Meta-Analysis of the Effectiveness of Project-Based Learning in College Biology Education for the Development of Higher-Order Thinking Skills. *Educational Process International Journal*, *18*(1). https://doi.org/10.22521/edupij.2025.18.463
32. Olivera-Aguilar, M., & Rikoon, S. H. (2025). Intervention Effect or Measurement Artifact? Using Invariance Models to Reveal Response-Shift Bias in Experimental Studies. *Journal of Research on Educational Effectiveness*, *18*(3), 797–825. https://doi.org/10.1080/19345747.2023.2284768

33. Ouyang, F., Chen, Z., Cheng, M., Tang, Z., & Su, C. Y. (2021). Exploring the effect of three scaffoldings on the collaborative problem-solving processes in China's higher education. *International Journal of Educational Technology in Higher Education*, *18*(1), 35. https://doi.org/10.1186/s41239-021-00273-y
34. Peng, J., Sun, M., Yuan, B., Lim, C. P., Van Merriënboer, J. J. G., & Wang, M. (2024). Visible thinking to support online project-based learning: Narrowing the achievement gap between high- and low-achieving students. *Education and Information Technologies*, *29*(2), 2329–2363. https://doi.org/10.1007/s10639-023-11896-1
35. Pinar, F. I. L., Panergayo, A. A. E., Sagcal, R. R., Acut, D. P., Roleda, L. S., & Prudente, M. S. (2025). Fostering scientific creativity in science education through scientific problem-solving approaches and STEM contexts: A meta-analysis. *Disciplinary and Interdisciplinary Science Education Research*, *7*(1), 18. https://doi.org/10.1186/s43031-025-00137-9
36. Puig Mauriz, B., Anaya, B., P., & Bargiela, I. M. (2023). Integrar el Pensamiento Crítico en la Educación Científica en la Era de la Post-verdad. *Revista Eureka Sobre Enseñanza y Divulgación de Las Ciencias*, *20*(3). https://doi.org/10.25267/Rev_Eureka_ensen_divulg_cienc.2023.v20.i3.3301
37. Razak, A. A., Ramdan, M. R., Mahjom, N., Zabit, M. N., Md., Muhammad, F., Hussin, M. Y. M., & Abdullah, N. L. (2022). Improving Critical Thinking Skills in Teaching through Problem-Based Learning for Students: A Scoping Review. *International Journal of Learning Teaching and Educational Research*, *21*(2), 342–362. https://doi.org/10.26803/ijlter.21.2.19
38. Retno, R. S., Purnomo, P., Hidayat, A., & Mashfufah, A. (2025). Conceptual framework design for STEM-integrated project-based learning (PjBL-STEM) for elementary schools. *Asian Education and Development Studies*, *14*(3), 579–604. https://doi.org/10.1108/AEDS-08-2024-0188
39. Sánchez-García, R., & Angosto, S. (2026). Mapping the path to effective Project-Based Learning: A systematic review and analysis of models and practices. *Humanities and Social Sciences Communications*, *13*(1), 1646. https://doi.org/10.1057/s41599-026-07870-0
40. Sánchez-García, R., & Reyes-de-Cózar, S. (2025). Enhancing Project-Based Learning: A Framework for Optimizing Structural Design and Implementation—A Systematic Review with a Sustainable Focus. *Sustainability*, *17*(11), 4978. https://doi.org/10.3390/su17114978
41. Santhosh, M., Farooqi, H., Ammar, M., Siby, N., Bhadra, J., Al-Thani, N. J., Sellami, A., Fatima, N., & Ahmad, Z. (2023). A Meta-Analysis to Gauge the Effectiveness of STEM Informal Project-Based Learning: Investigating the Potential Moderator Variables. *Journal of Science Education and Technology*, *32*(5), 671–685. https://doi.org/10.1007/s10956-023-10063-y
42. Saqr, M. (2023). Modelling within-person idiographic variance could help explain and individualize learning. *British Journal of Educational Technology*, *54*(5), 1077–1094. https://doi.org/10.1111/bjet.13309
43. Sasson, I., Yehuda, I., & Malkinson, N. (2018). Fostering the skills of critical thinking and question-posing in a project-based learning environment. *Thinking Skills and Creativity*, *29*, 203–212. https://doi.org/10.1016/j.tsc.2018.08.001

44. Soysal, Y., & Soysal, S. (2023). Relationship Between a Teacher Educator's Questions and the Development of Prospective Teachers' Critical Thinking. *ECNU Review of Education*, *6*(1), 105–140. https://doi.org/10.1177/20965311221107028
45. Syam, H., Sutawidjaja, A., Sa'dijah, C., & Abadyo, A. (2020). Junior High Students' Critical Thinking in Geometry Problem Solving. *Universal Journal of Educational Research*, *8*(11B), 5880–5887. https://doi.org/10.13189/ujer.2020.082221
46. Szabo, Z. K., Körtesi, P., Guncaga, J., Szabo, D., & Neag, R. (2020). Examples of Problem-Solving Strategies in Mathematics Education Supporting the Sustainability of 21st-Century Skills. *Sustainability*, *12*(23), 10113. https://doi.org/10.3390/su122310113
47. Tafakur, T., Retnawati, H., & Shukri, A. A. M. (2023). Effectiveness of project-based learning for enhancing students critical thinking skills: A meta-analysis. *JINoP (Jurnal Inovasi Pembelajaran)*, *9*(2), 191–209. https://doi.org/10.22219/jinop.v9i2.22142
48. Tierney, G., Urban, R., & Olabuenaga, G. (2025). Designing for Equity: Moving Project-Based Learning from Equity Adjacent to Equity Infused. *Interdisciplinary Journal of Problem-Based Learning*, *18*(1). https://doi.org/10.14434/ijpbl.v18i1.36891
49. Widlund, A., Niemivirta, M., Tuominen, H., & Korhonen, J. (2024). Growth trajectories of self-concept and interest in mathematics and language – Individual differences and cross-domain relations. *Learning and Instruction*, *91*, 101882. https://doi.org/10.1016/j.learninstruc.2024.101882
50. Winne, P. H., & Azevedo, R. (2022). Metacognition and Self-Regulated Learning. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (3rd ed., pp. 93–113). Cambridge University Press. https://doi.org/10.1017/9781108888295.007
51. Xiong, M., Chan, N. N., Wong, B. E., Xie, X., & Na, M. (2025). Investigating cognitive flexibility and innovation in interdisciplinary project-based learning: The role of openness to learning and peer feedback quality in vocational education. *Frontiers in Psychology*, *16*, 1691422. https://doi.org/10.3389/fpsyg.2025.1691422
52. Zhang, J., Wu, J., Sun, X., Yang, Y., & Zhou, M. (2023). Technology-Enabled Project-Based Learning: Let Every Child Embrace Good Learning. In H. Zeng, Z. Li, J. Guo, & Z. Zhang (Eds.), *Constructing Regional Smart Education Ecosystems in China* (pp. 93–99). Springer Nature Singapore. https://doi.org/10.1007/978-981-99-6225-9_11
53. Zhao, Y., & Wang, L. (2022). A case study of student development across project-based learning units in middle school chemistry. *Disciplinary and Interdisciplinary Science Education Research*, *4*(1), 5. https://doi.org/10.1186/s43031-021-00045-8

Figures

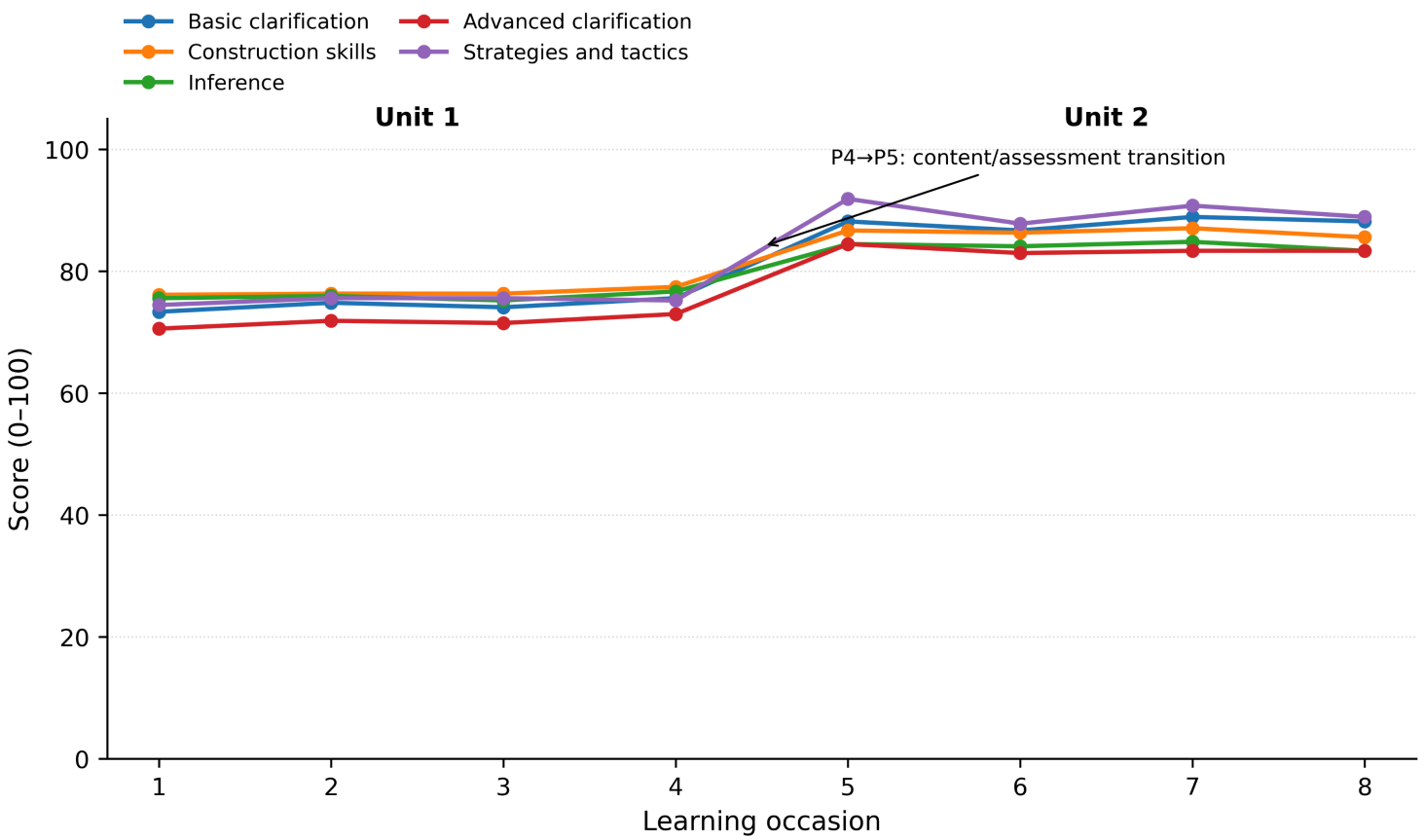


Figure 1

Indicator-specific trajectories of critical thinking across eight learning occasions.

Note: Each line represents the mean score of one critical-thinking indicator across P1-P8. Unit 1 comprises P1-P4, whereas Unit 2 comprises P5-P8. The P4-P5 transition coincided with a change in instructional content and assessment context and should not be interpreted as a pure temporal growth point.

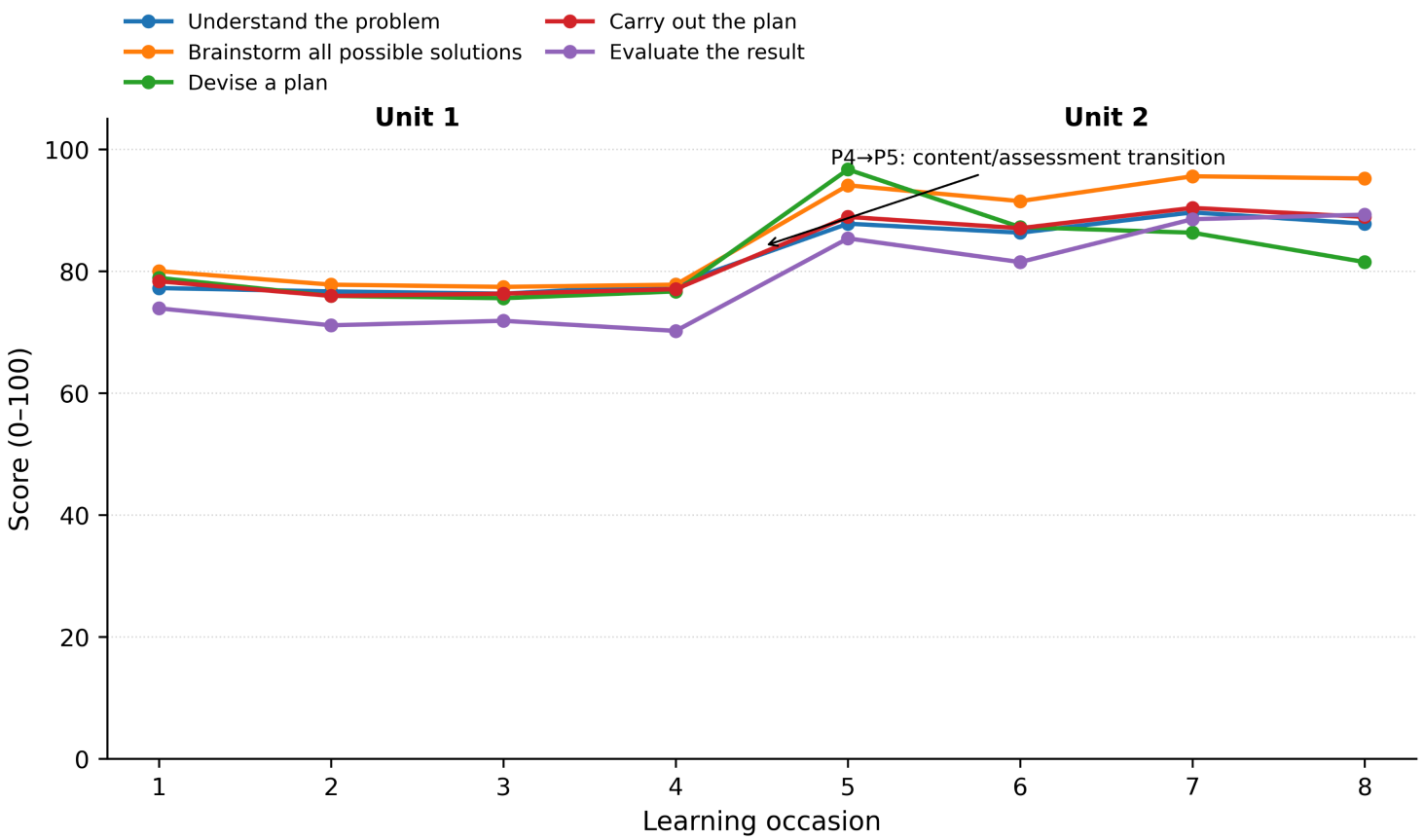


Figure 2

Indicator-specific trajectories of problem solving across eight learning occasions.

Note: Each line represents the mean score of one problem-solving indicator across P1-P8. Unit 1 comprises P1-P4, whereas Unit 2 comprises P5-P8. The P4-P5 transition coincided with a change in instructional content and assessment context; consequently, the discontinuity should not be attributed exclusively to instructional exposure or learning growth.