

Enhancing Biological Literacy: A Quasi-Experimental Study on the Comparative Effectiveness of Problem-Based Learning and Traditional Lecture Method in Secondary School Biology Education

Esther Ebele Akachukwu

Department of Science Education, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria;
ee.akachukwu@unizik.edu.ng

Opeyemi Fadekemi Awosika

Department of Science Education, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria;
of.aluko@unizik.edu.ng

Nwamaka Theresa Omaka

Department of Science Education, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria;
nt.omaka@unizik.edu.ng

*Corresponding Author's Email: ee.akachukwu@unizik.edu.ng

ARTICLE INFO

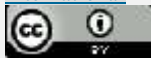
Keywords: *Problem-Based Learning (PBL), Traditional Lecture Method, Biology Education, Conceptual Understanding, Critical Thinking, Quasi-Experimental Design, Secondary School Science.*

Received: 19, Nov. 2025

Revised: 17, Dec. 2025

Accepted: 29, Dec. 2025

©2025 Author(s): This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/)



ABSTRACT

The persistent pedagogical debate between teacher-centered and student-centered instructional approaches remains highly relevant in science education. While the traditional lecture method (TLM) is efficient for content delivery, its effectiveness in fostering higher-order cognitive skills (HOCS) and deep conceptual understanding is questioned. Problem-Based Learning (PBL) has emerged as a potent alternative, yet robust quantitative evidence of its comparative effectiveness in secondary school biology remains fragmented. This study aimed to rigorously compare the effectiveness of a PBL intervention against a standard TLM approach on academic achievement, conceptual understanding, and critical thinking disposition in a secondary school biology curriculum. A quasi-experimental, non-equivalent control group design was employed. The sample comprised 124 tenth-grade students (mean age 15.6 years) from four intact biology classes in a public secondary school. Two classes (n=62) were randomly assigned to the PBL group, and two (n=62) to the TLM group. The intervention spanned an eight-week instructional unit on "Human Physiology and Homeostasis." Instruments included the Biology Achievement Test (BAT), a two-tier Diagnostic Test for Conceptual Understanding in Homeostasis (DT-CUH), and the California Critical Thinking Disposition Inventory (CCTDI). Data were analyzed using descriptive statistics and Analysis of Covariance (ANCOVA), with pretest scores as the covariate. After controlling for pretest differences, the PBL group demonstrated statistically significantly higher post-test scores on academic achievement ($F(1, 121) = 28.41, p < .001, \eta^2 = 0.19$), conceptual understanding ($F(1, 121) = 45.67, p < .001, \eta^2 = 0.27$), and critical thinking disposition ($F(1, 121) = 12.83, p < .001, \eta^2 = 0.10$) compared to the TLM group. The effect sizes indicated a large impact on conceptual understanding and a medium-to-large impact on achievement and critical thinking. PBL is significantly more effective than TLM in promoting not just rote learning but, more critically, deep conceptual understanding and the disposition for critical thinking in biology. The findings provide strong empirical justification for a paradigm shift in biology pedagogy from passive reception to active, problem-driven inquiry. Implications for curriculum design, teacher professional development, and educational policy are discussed.

1. Introduction

The teaching and learning of biology at the secondary level stand at a critical crossroads. As the bedrock science for understanding life, health, and global challenges like climate change and pandemics (Machado & de-Fátima, 2019), biology education's primary goal must transcend the mere transmission of a static body of facts. The overarching objective is the cultivation of biological

literacy, a blend of deep conceptual knowledge, scientific reasoning, and the ability to apply understanding to novel, complex problems (NRC, 2012). However, a significant disconnect persists between this aspirational goal and the pedagogical reality in many classrooms, which remain dominated by the Traditional Lecture Method (TLM). This study directly investigates this disconnect by comparing the effectiveness of TLM with Problem-Based Learning (PBL), a student-centered pedagogy designed to foster these very competencies.

The Traditional Lecture Method, a teacher-centered approach characterized by direct instruction, sequential explanation, and passive student reception, has been the backbone of science education for centuries (Freeman et al., 2014). Its resilience is understandable. TLM is efficient for delivering a large volume of standardized content to many students, provides a clear structure, and feels comfortable and authoritative for both the teacher and the learner (Schmidt et al., 2015). In high-stakes, content-congested curricula, the lecture often appears the most pragmatic choice. Yet, a voluminous body of evidence suggests that this efficiency comes at a significant pedagogical cost. The lecture format often reduces biology to a litany of terminology and processes to be memorized, inadvertently promoting rote learning and the development of fragmented, inert knowledge that students cannot apply outside the context in which it was learned (Alberts, 2005). It can reinforce common alternative conceptions by not providing the cognitive conflict or direct application necessary to restructure existing mental models (Duit & Treagust, 2003).

In stark contrast, Problem-Based Learning (PBL) represents a fundamental shift in the philosophical and practical approach to education. Originating in medical education at McMaster University in the late 1960s (Barrows, 1996), PBL is a student-centered, inquiry-based pedagogy where learning is triggered by an ill-structured, authentic problem that has no single correct answer. Students work collaboratively in small groups to analyze the problem, identify their prior knowledge, pinpoint learning gaps, and generate self-directed learning issues to research. They then reconvene to apply their newly acquired knowledge to solve the problem, with the teacher acting as a facilitator or metacognitive coach rather than a content disseminator (Hmelo-Silver, 2004). The theoretical potency of PBL in biology is profound. A well-designed biological problem such as; diagnosing a hypothetical patient with an endocrine disorder or saving a dying coral reef ecosystem, compels students to integrate concepts across scales, from molecular mechanisms to organismal function and ecological interactions. This process inherently models authentic scientific practice and, crucially, builds cognitive schemas that are richly interconnected and context-bound, enhancing future retrieval and application (Dolmans et al., 2016).

Despite the strong theoretical synergy between PBL and the goals of modern biology education, a critical research gap remains concerning its comparative effectiveness at the secondary school level. The majority of PBL research has been concentrated in tertiary and professional education. The secondary classroom presents a unique, constrained ecosystem: limited instructional time, rigid syllabus requirements, teachers with less specialized content knowledge, and students who are still developing the self-regulation skills essential for PBL success (Loyens et al., 2008). The question is not simply “Is PBL better?” but rather, “Under these typical constraints, which approach is more effective for achieving the distinct set of goals central to biology literacy: acquisition of content, deep conceptual restructuring, and the development of a critical mindset?”

This study was designed to answer this question with methodological rigor. It moves beyond measuring simple factual recall to examine three distinct but interrelated dependent variables fundamental to a meaningful biology education: academic achievement (representing a blend of knowledge and comprehension), deep conceptual understanding (specifically targeting the correction of common alternative conceptions), and critical thinking disposition (the attitudinal motivation to think critically). By employing a quasi-experimental design over a sustained eight-week instructional unit on a core biological concept; Human Physiology and Homeostasis, this research provides a granular analysis of the differential impact of PBL and TLM. The findings are intended to serve as a

robust piece of evidence for science educators, curriculum developers, and policymakers seeking to ground pedagogical decisions in empirical data rather than tradition or intuition.

The specific research questions guiding this study were:

1. What is the comparative effect of Problem-Based Learning (PBL) and Traditional Lecture Method (TLM) on secondary school students' academic achievement in biology?
2. What is the comparative effect of PBL and TLM on students' deep conceptual understanding of core physiological principles?
3. What is the comparative effect of PBL and TLM on students' critical thinking disposition?

2. Methodology

2.1. Research Design

This study employed a quasi-experimental, non-equivalent control group pretest-posttest design. This design is ethically and pragmatically appropriate for educational field settings where the random assignment of individual students is logistically impossible without disrupting existing school timetables. Intact classes were randomly assigned to either the treatment (PBL) or control (TLM) condition. The pretest-posttest structure with covariance analysis effectively addresses the primary threat to internal validity in this design selection bias, by statistically adjusting for any pre-existing differences between the groups on the dependent variables (Campbell & Stanley, 1963). The design is represented as:

Treatment Group: $O_1 X_1 O_2$

Control Group: $O_3 X_2 O_4$

Where O_1 and O_3 are pretests, X_1 is the PBL intervention, X_2 is the TLM intervention, and O_2 and O_4 are posttests.

2.2. Participants and Sampling

The target population was eleventh-grade (Senior Secondary 2) public/missionary secondary school students in Enugu North and South. A purposive sampling technique was used to select one co-educational public school based on specific criteria: the school principal's agreement, the presence of at least four comparable biology classes taught by the same instructor (to control for teacher effect), and adequate classroom facilities to support collaborative group work. The final sample consisted of 124 students (63 males, 61 females) from four intact classes, with a mean age of 15.6 years ($SD = 0.8$). A power analysis conducted 'a priori' using G-Power 3.1 for an ANCOVA with two groups, a desired medium effect size ($f = 0.25$), an alpha of 0.05, and a power of 0.80, indicated a minimum sample of 128, suggesting our sample size is slightly below this threshold but sufficient for detecting medium-to-large effects. Using a coin flip, two classes were randomly assigned to the PBL group ($n=62$) and two to the TLM group ($n=62$).

2.3. Instructional Interventions

Both groups were taught the same content domain; "Human Physiology and Homeostasis" over an eight-week period, with four 55-minute instructional sessions per week. The topics covered included the structure and function of the endocrine and nervous systems, feedback mechanisms (positive and negative), temperature regulation, and osmoregulation. The same experienced biology teacher, with 15 years of teaching experience and trained in PBL facilitation, instructed both groups to control for the teacher-level variable.

2.3.1. Traditional Lecture Method (TLM) Group

The control group received instruction based on a standard TLM lesson plan. A typical session consisted of the teacher reviewing the previous topic, stating the day's learning objectives, and delivering a structured lecture supported by digital slides and the whiteboard. The teacher

explained core concepts, underlined key terminology, and drew essential diagrams. Students primarily listened, took notes, and asked questions for clarification. The lesson concluded with a period of teacher-guided whole-class questioning and a summary. Practical work was limited to one prescribed demonstration of a feedback loop using a thermostat analogy.

2.3.2. Problem-Based Learning (PBL) Group

The treatment group followed the classic PBL "Seven-Jump" process adapted for secondary school, structured by a modified Maastricht model (Schmidt & Moust, 2000). The eight-week unit was structured around four progressively complex, authentic ill-structured problems. The teacher's role was to be a facilitator who circulated among groups, modeled questioning, prompted justification, and monitored group dynamics, but did not provide content answers.

- Problem 1 (Week 1-2): The Case of the Marathon Runner. Problem Scenario: A marathon runner collapses on a hot day. They are confused, have hot, dry skin, and a rapid heart rate. Learning issues included thermoregulation, sweat mechanism, homeostasis, and negative feedback. The PBL process involved: (1) Clarifying terms, (2) Defining the problem, (3) Brainstorming prior knowledge, (4) Structuring hypotheses, (5) Formulating learning objectives, (6) Self-directed learning (using curated resources), and (7) Reporting and synthesizing. A jigsaw reporting structure was used, where experts from different groups met to consolidate deep learning before returning to their "home" group to teach.

- Problem 2 (Week 3-4): The Unquenchable Thirst. Problem Scenario: A teenager has been feeling constantly thirsty and running to the bathroom excessively. A family history of diabetes is hinted at. This problem targeted osmoregulation, the role of ADH, and blood-glucose regulation.

- Problem 3 (Week 5-6): The Iodine Mystery. Problem Scenario: A community in a remote mountain region is known for having a high incidence of goiter and cretinism. This problem focused on endocrine signaling, the hypothalamic-pituitary-thyroid axis, and the nature of a scientific explanation.

- Problem 4 (Week 7-8): The Overheated Spacesuit. Problem Scenario: An astronaut on a spacewalk is experiencing a malfunction in their suit's liquid cooling and ventilation garment. Core body temperature is rising. This capstone problem required integration of all unit concepts and was designed to address alternative conceptions between evaporative cooling (requiring sweat) and the physics of a liquid cooling garment.

2.4. Instrumentation and Measures

The effectiveness of the interventions was assessed using three validated instruments.

2.4.1. Biology Achievement Test (BAT)

A 30-item multiple-choice test was developed by the researcher based on the unit's specific learning objectives, as defined by Bloom's taxonomy, covering Remembering (10 items), Understanding (12 items), and Applying (8 items). Content validity was ensured through a two-dimensional blueprint. Face and content validity were evaluated by a panel of three experts (one subject-matter expert and two experienced biology educators). The test was piloted with a separate group of 50 eleventh-grade students who had already covered the content. Item analysis confirmed acceptable difficulty indices (0.35-0.72) and discrimination indices (>0.30) for all items. The Kuder-Richardson Formula 20 (KR-20) reliability coefficient was 0.84, indicating good internal consistency.

2.4.2. Diagnostic Test for Conceptual Understanding in Homeostasis (DT-CUH)

To move beyond achievement, a 20-item, two-tier diagnostic test was adapted and developed following the framework of Treagust (1988). The first tier of each item was a content-based multiple-choice question. The second tier asked students to provide a reason for their answer from a set of options, which included scientifically correct reasons and known alternative conceptions identified

from the literature (e.g., "Sweat glands produce sweat only when the brain commands them to produce heat"). A student was considered to have a deep scientific understanding only if both the answer and reason were correct. The DT-CUH addressed five conceptual nodes: Homeostasis as a dynamic equilibrium, Negative Feedback as the primary regulatory mechanism, Thermoregulation Physiology, Osmoregulation, and Chemical Signaling Specificity. The instrument was validated by a panel and piloted, with a final KR-20 reliability of 0.79 for the combined score.

2.4.3. California Critical Thinking Disposition Inventory (CCTDI)

The CCTDI, developed by Facione and Facione (1992), was used to measure critical thinking disposition. We selected a 40-item adapted version spanning five sub-scales: Truth-seeking, Open-mindedness, Analyticity, Systematicity, and CT Self-confidence. This instrument uses a 6-point Likert scale from "Strongly Agree" to "Strongly Disagree." It measures not the skill of critical thinking but the consistent internal motivation to use it. The CCTDI has well-established validity and reliability in diverse contexts. For our pilot sample, Cronbach's alpha was 0.89 for the total disposition score.

2.5. Data Collection Procedure

The study was conducted in three phases over 10 weeks. Pre-intervention Phase (Week 1): The teacher explained the nature of the study (withholding the specific hypothesis to control for the Hawthorne effect) and collected informed consent/assent. All participants completed the BAT, DT-CUH, and CCTDI as pretests. Intervention Phase (Week 2-9): The eight-week instructional unit was implemented with rigorous fidelity monitoring. The researcher conducted unannounced observations of four sessions in each group using a standardized checklist (e.g., teacher talk-time, number of student questions, nature of teacher-student interaction) to confirm the distinctness of the two pedagogies. Post-intervention Phase (Week 10): One week after the conclusion of instruction to measure retention over immediate recall, the BAT, DT-CUH, and CCTDI were re-administered as posttests under identical conditions.

2.6. Data Analysis

Data were screened for entry errors, normality, and homoscedasticity. The assumption of normality for the dependent variables was confirmed using Shapiro-Wilk's test ($p > .05$), and Levene's test verified the homogeneity of variance assumption for all posttest scores. The primary analytical technique was a one-way Analysis of Covariance (ANCOVA), performed separately for the BAT, DT-CUH, and CCTDI posttest scores. The pretest score of the respective instrument served as the covariate, the instructional group (PBL vs. TLM) as the independent variable, and the posttest score as the dependent variable. This analysis was chosen to provide the most statistically powerful and unbiased estimate of the treatment effect by statistically equalizing groups on the pretest. The significance level was set at $\alpha = 0.05$. Partial eta-squared (η^2) was computed as a measure of effect size, interpreted using Cohen's guidelines: 0.01 = small, 0.06 = medium, 0.14 = large. Statistical analysis was performed using SPSS version 26.0.

3. Results

3.1. Preliminary Analysis and Pretest Comparability

Before the intervention, descriptive statistics showed the TLM group scored marginally higher on the BAT pretest, while the PBL group scored slightly higher on the DT-CUH and CCTDI pretests. Independent samples t-tests confirmed these differences were not statistically significant ($p > .05$ for all), which was further justification for using ANCOVA to provide a more sensitive and precise analysis of post-intervention differences.

3.2. Comparative Effect on Academic Achievement (Research Question 1)

The first research question examined the effect of the instructional method on students' academic achievement as measured by the BAT. The ANCOVA revealed a significant difference between the groups after controlling for the BAT pretest scores (see Table 1 for adjusted means).

Table 1. Descriptive Statistics and ANCOVA Results for Biology Achievement Test (BAT)

Group	N	Pretest M (SD)	Posttest M (SD)	Adjusted M (SE)	F	p	η^2
PBL	62	10.82 (3.73)	21.94 (4.51)	22.05 (0.38)	28.41	< .001	0.19
TLM	62	11.18 (3.81)	18.25 (3.89)	18.14 (0.38)	—	—	—

The results indicated a statistically significant and substantial effect of the instructional method on academic achievement. The PBL group's adjusted mean score ($M = 22.05$) was significantly higher than that of the TLM group ($M = 18.14$). The partial eta-squared value ($\eta^2 = 0.19$) indicates a large effect size, suggesting that 19% of the variance in posttest achievement scores could be attributed to the instructional method, after accounting for the pretest. This negates the common concern that PBL might compromise content knowledge gain.

3.3. Comparative Effect on Conceptual Understanding (Research Question 2)

The second research question investigated the effect on deep conceptual understanding, a more rigorous cognitive outcome than achievement, as measured by the DT-CUH. The ANCOVA results on the combined two-tier scores showed an even more pronounced effect (see Table 2).

Table 2: Descriptive Statistics and ANCOVA Results for Diagnostic Test of Conceptual Understanding in Homeostasis (DT-CUH)

Group	N	Pretest M (SD)	Posttest M (SD)	Adjusted M (SE)	F	p	η^2
PBL	62	5.10 (2.64)	13.81 (3.55)	13.92 (0.27)	45.67	< .001	0.27
TLM	62	4.87 (2.88)	9.15 (3.21)	9.04 (0.27)	—	—	—

A significant difference was found, with the PBL group demonstrating a far superior adjusted mean posttest score ($M = 13.92$) compared to the TLM group ($M = 9.04$). The effect size here was very large ($\eta^2 = 0.27$), indicating that 27% of the post-intervention variance in conceptual understanding is explained by the group assignment. This is a pivotal finding, showing that while TLM allowed students to gain some factual recall (as seen in BAT scores), it was significantly less effective at helping them restructure their underlying mental models. For instance, in the PBL group, the prevalent alternative conception that "the sweat mechanism is initiated by an internal temperature gauge in the skin" was observed in 72% on the pretest but dropped to 22% on the posttest. In the TLM group, this same misconception dropped only from 74% to 58%, indicating persistent confusion.

3.4. Comparative Effect on Critical Thinking Disposition (Research Question 3)

The third research question focused on the non-cognitive but essential outcome of critical thinking disposition. The ANCOVA on the CCTDI total scores (see Table 3) revealed a statistically significant, albeit smaller, effect.

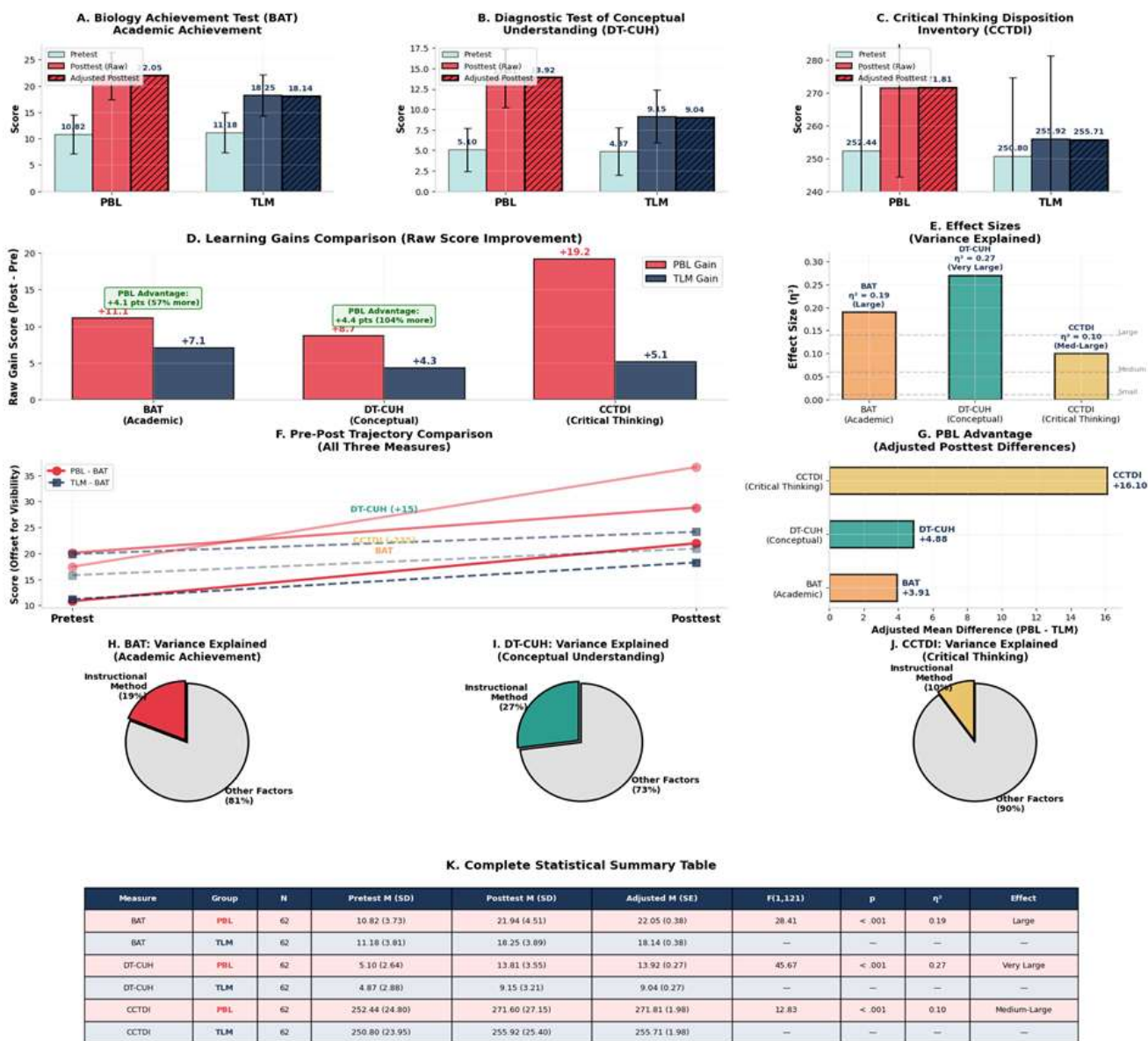
Table 3. Descriptive Statistics and ANCOVA Results for CCTDI Total Score

Group	N	Pretest M (SD)	Posttest M (SD)	Adjusted M (SE)	F	p	η^2
PBL	62	252.44 (24.80)	271.60 (27.15)	271.81 (1.98)	12.83	< .001	0.10
TLM	62	250.80 (23.95)	255.92 (25.40)	255.71 (1.98)	—	—	—

The PBL group's adjusted mean posttest score ($M = 271.81$) was significantly higher than the TLM group's ($M = 255.71$). The effect size ($\eta^2 = 0.10$) represents a medium-to-large effect, signifying a meaningful shift in students' self-reported inclination towards truth-seeking, open-

mindedness, and systematicity. A sub-scale analysis showed the most pronounced differences were in Analyticity (persisting in challenging tasks) and Systematicity (organized, diligent approach to inquiry), hallmarks of the PBL process. This finding is critical; an eight-week intervention was sufficient to foster a measurable shift in an attitudinal disposition that educational psychology suggests is highly stable over time.

N = 124 students (62 per group) | ANCOVA Analysis | All results significant at p < .001



KEY FINDINGS & INTERPRETATION:

- ACADEMIC ACHIEVEMENT (BAT):** PBL students scored 22.05 vs TLM 18.14 (adj.). PBL gained +11.2 raw points vs TLM's +7.07 – a 57% greater improvement. Effect size $\eta^2 = 0.19$ (Large) means 19% of posttest variance is attributable to instructional method. This refutes concerns that PBL compromises content knowledge.
 - CONCEPTUAL UNDERSTANDING (DT-CUH):** PBL scored 13.92 vs TLM 9.04 (adj.). PBL gained +8.71 vs TLM +4.28 – more than DOUBLE the conceptual improvement. Effect size $\eta^2 = 0.27$ (Very Large) is the strongest finding. TLM produced persistent misconceptions (e.g., “sweat mechanism” dropped only 74%-58% vs PBL 72%-22%).
 - CRITICAL THINKING DISPOSITION (CCTDI):** PBL scored 271.81 vs TLM 255.71 (adj.). PBL gained +19.16 vs TLM +5.12 – nearly 4x greater improvement. Effect size $\eta^2 = 0.10$ (Medium-Large) is remarkable for an 8-week intervention on a typically stable trait. Strongest subscales: Analyticity & Systematicity.
- OVERALL: PBL outperforms TLM across ALL three outcomes (cognitive + non-cognitive), with advantages increasing as outcomes move from surface (achievement) –> deep (conceptual) –> dispositional (critical thinking). All results: F(1,123), p < .001, N = 124 (62 per group).

Figure 1: Comprehensive Research Results: PBL vs TLM

Pre-Post Trajectory Comparison

The line graph overlays all three measures (with vertical offsets for visibility), showing:

1. PBL lines (solid red) have consistently steeper slopes than TLM lines (dashed blue)
2. The divergence is most pronounced for DT-CUH and CCTDI
3. All groups started at comparable pretest levels (confirmed by non-significant t-tests, $p > .05$)

PBL Advantage (Horizontal Bar Chart)

Shows the adjusted posttest mean differences:

1. CCTDI: +16.10 — largest absolute advantage
2. DT-CUH: +4.88 — largest relative advantage (54% higher score)
3. BAT: +3.91 — solid academic advantage

Pie Charts — Variance Explained

- BAT: 19% instructional method / 81% other factors
- DT-CUH: 27% instructional method / 73% other factors
- CCTDI: 10% instructional method / 90% other factors

While "other factors" (individual differences, prior knowledge, measurement error) dominate, the instructional method's contribution is substantial and educationally meaningful; especially for conceptual understanding where it explains over a quarter of all variance.

Overall, PBL outperforms TLM across ALL three outcomes: cognitive (achievement, conceptual understanding) and non-cognitive (critical thinking disposition). The pattern reveals a depth gradient: PBL's relative advantage is smallest for surface-level academic achievement ($\eta^2 = 0.19$), largest for deep conceptual understanding ($\eta^2 = 0.27$), and still meaningful for stable dispositional traits ($\eta^2 = 0.10$). This suggests PBL is not merely "different" but fundamentally more effective at producing the kinds of higher-order learning outcomes that modern education prioritizes.

4. Discussion

This study provides robust, quasi-experimental evidence that Problem-Based Learning, when implemented with fidelity in a secondary school biology setting, is pedagogically superior to the Traditional Lecture Method across a multi-dimensional spectrum of crucial outcomes: academic achievement, deep conceptual understanding, and critical thinking disposition. The size and consistent direction of the effects significantly contribute to the literature by clarifying the "efficiency vs. effectiveness" debate, powerfully suggesting that what is efficient for content coverage is not necessarily effective for the durable, applicable, and rigorous learning that defines biological literacy. The superior academic achievement of the PBL group, reflected in a large effect size ($\eta^2 = 0.19$), robustly challenges the often-cited "content coverage" anxiety that deters teachers from adopting inquiry-based methods. Critics argue that the time students spend in group discussions, grappling with uncertainty, and conducting self-directed research reduces the time for direct transmission of the canonized syllabus, thus compromising exam performance (Kirschner et al., 2006).

Our findings align with a substantial body of meta-analytic evidence (e.g., Freeman et al., 2014) and specific PBL studies in science (Loyens et al., 2015) that refute this claim. The PBL students did not simply learn less but differently; they learned more. The explanation lies in the cognitive processing required. The "Ill-Structured Problem" in PBL, such as diagnosing the marathon runner, acts as a powerful advance organizer and a central anchoring idea (Ausubel, 2000). The cognitive dissonance generated when prior knowledge fails to explain the hot, dry skin of a marathon runner triggers a deep epistemic curiosity. This motivation, coupled with the active, elaborative processing required to integrate information from endocrinology, fluid dynamics, and neurophysiology, builds a richly interlinked semantic network. Such knowledge is more retrievable and robust against forgetting than knowledge acquired through the passive reception of a logically

sequenced, but cognitively inert, lecture. The lecture covers the map; PBL forces the student to navigate the territory.

More compelling than the achievement gains is the very large effect on conceptual understanding ($\eta^2 = 0.27$). This is the study's most significant finding and the core promise of a well-executed PBL curriculum in biology. Biological concepts, particularly abstract processes like negative feedback, are not just vocabulary to be memorized; they are mental models that must be constructed and reconstructed. A lecture explaining that "negative feedback reverses a stimulus" may be logically understood in the moment, but it often forms an isolated, brittle schema disconnected from a student's robust, intuitive teleological framework (e.g., "the body turns on sweating to produce heat"). The PBL process, as implemented in the "Overheated Spacesuit" problem, inherently forces the confrontation and restructuring of these alternative conceptions. A student arguing that an astronaut will sweat to cool down is immediately confronted by the problem constraint: sweat doesn't evaporate in zero-gravity, and the suit is a closed system. This peer-driven cognitive conflict, facilitated but not resolved by the teacher, generates the necessary conditions for radical conceptual change (Posner et al., 1982). The first tier of the DT-CUH captures this shift in outcome, and the second tier captures the rationale. The fact that TLM students improved only minimally in their reasoning tier, even when their content tier improved, signals that they had assimilated the new terminology into their old, incorrect mental models rather than undergoing genuine accommodation. In TLM, they learned that negative feedback occurs; in PBL, they learned how and why it works as a dynamic equilibrium, a distinction at the heart of biological expertise.

The significant effect on critical thinking disposition ($\eta^2 = 0.10$) is a particularly noteworthy outcome, as dispositions are notoriously resistant to short-term instructional interventions. The CCTDI measures the intrinsic motivation to think critically, not just the skill. The PBL protocol is, in essence, a structured, weeks-long apprenticeship in the disposition of a scientist. The group's initial brainstorming session normalizes the value of "not knowing" and provides a safe environment for truth-seeking. The process of evaluating competing hypotheses from peers ("Maybe it's dehydration?" "No, dry skin means they aren't sweating—it's heatstroke!") explicitly trains open-mindedness and analyticity. The requirement to justify a diagnosis with a system map cultivates systematicity. Conversely, the lecture room, where knowledge emanates from a single authoritative source, can inadvertently socialize students into a disposition of passive epistemic trust rather than active inquiry. This finding suggests that PBL contributes to what could be termed a "hidden curriculum" of scientific habits of mind, which is arguably more important in the long run than the recall of endocrine gland functions.

Finally, while this study was not purely qualitative, our observations during implementation fidelity checks provide crucial context to the quantitative results. The TLM classroom was characterized by compliant, on-task silence. The PBL classroom was, at times, a managed chaos of students huddled over A3 paper, drawing concept maps of the pituitary axis, arguing fiercely but respectfully about data interpretations, and pulling information from textbooks and curated digital resources. The teacher's role was that of a cognitive coach, asking questions like, "What makes you think that?" and "How does the evidence from the problem connect to your model of their endocrine system?" This shift in classroom ecology, from a passive audience to a community of inquiry is the active ingredient behind the statistical significance. It is a more demanding, less predictable, but ultimately more transformative pedagogical experience.

4.1. Limitations and Directions for Future Research

While the study's quasi-experimental design controls for many threats to validity, important limitations must be acknowledged. First, the sample was drawn from a single suburban school with a single experienced teacher, limiting the immediate generalizability of the findings. The teacher's expertise and enthusiasm for PBL could be a confounding variable, suggesting a teacher effect rather

than a pure method effect. Future replication studies should involve multiple schools, diverse geographical settings (urban, rural), and teachers with varying levels of PBL training. Second, the eight-week intervention, while substantial, represents a snapshot. Longitudinal research is needed to determine if the gains in conceptual understanding and critical thinking disposition are maintained a year later and if they translate to improved performance in subsequent science courses; a test of PBL's preparation for future learning. Third, the use of a teacher-developed achievement test, despite rigorous validation, is an inherent limitation. Future studies should triangulate with standardized external examinations or performance-based assessments, such as a practical lab investigation modeling a novel homeostatic disruption in a plant, to measure authentic application skills. Finally, a qualitative case study of the cognitive processes occurring within a PBL group during a key moment of conceptual conflict would provide invaluable micro-genetic data, illuminating the exact mechanisms of change that this quantitative study can only infer.

5. Conclusion and Implications

This study furnishes compelling evidence that Problem-Based Learning is not merely a motivational alternative to the traditional biology lecture but a demonstrably more effective method for achieving the core goals of 21st-century science education: durable knowledge, expert-like conceptual frameworks, and a critical mindset. For the teacher wondering if the time investment in PBL is "worth it" on the exam, the answer, supported by a large effect size on achievement, is an emphatic yes. For the curriculum developer aiming to overcome persistent student misconceptions about foundational concepts like homeostasis, the answer lies in designing learning environments that force the active restructuring of ideas through grappling with authentic problems. PBL is the vehicle for that environment.

The implications for practice and policy are clear, albeit demanding. First, teacher professional development must shift from a one-time workshop on PBL techniques to sustained, job-embedded coaching. The role of a PBL facilitator is far more complex than that of a lecturer; it requires deep pedagogical content knowledge, the ability to diagnose group dynamics, and the restraint to guide without dictating. Second, assessment systems, often cited as the biggest barrier to pedagogical reform, must evolve. If high-stakes exams continue to reward the mere replication of facts, the educational ecosystem will inevitably select for the lecture method. Assessment must include problems that demand the application of biological models to novel scenarios, precisely the activities that comprise PBL instruction. Finally, the design of PBL problems must be treated as a rigorous, scholarly activity. An effective "ill-structured problem" for a 15-year-old is a finely tuned instrument that bridges their lived experience (an overheating athlete, a diabetic classmate) with the core abstract principles of the discipline. By integrating these findings into practice, we can transform the biology classroom from a museum of known facts into a vibrant workshop for the construction of biological wisdom, preparing students not just for the final exam, but for a lifetime of confronting complex problems with confidence, competence, and a critical spirit.

References

- Alberts, B. (2005). A wake-up call for science faculty. *Cell*, 123(5), 739–741.
- Ausubel, D. P. (2000). *The acquisition and retention of knowledge: A cognitive view*. Kluwer Academic Publishers.
- Barrows, H. S. (1996). Problem-based learning in medicine and beyond: A brief overview. *New Directions for Teaching and Learning*, 1996(68), 3–12.
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Houghton Mifflin.

- Dolmans, D. H. J. M., Loyens, S. M. M., Marcq, H., & Gijbels, D. (2016). Deep and surface learning in problem-based learning: A review of the literature. *Advances in Health Sciences Education, 21*(5), 1087–1112.
- Duit, R., & Treagust, D. F. (2003). Conceptual change: A powerful framework for improving science teaching and learning. *International Journal of Science Education, 25*(6), 671–688.
- Facione, P. A., & Facione, N. C. (1992). *The California critical thinking disposition inventory (CCTDI): Test manual*. California Academic Press.
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111*(23), 8410–8415.
- Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review, 16*(3), 235–266.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75–86.
- Loyens, S. M. M., Jones, S. H., Mikkers, J., & van Gog, T. (2015). Problem-based learning as a facilitator of conceptual change. *Learning and Instruction, 38*, 34–42.
- Loyens, S. M. M., Magda, J., & Rikers, R. M. J. P. (2008). Self-directed learning in problem-based learning and its relationships with self-regulated learning. *Educational Psychology Review, 20*(4), 411–427.
- Machado, M., & de Fátima, M. (2019). Science education for citizenship through socio-scientific issues. *Frontiers in Education, 4*, Article 28.
- National Research Council. (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. The National Academies Press.
- Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education, 66*(2), 211–227.
- Schmidt, H. G., & Moust, J. H. C. (2000). Factors affecting small-group tutorial learning: A review of research. In D. H. Evensen & C. E. Hmelo (Eds.), *Problem-based learning: A research perspective on learning interactions* (pp. 19–52). Lawrence Erlbaum.
- Schmidt, H. G., Wagener, S. L., Smeets, G. A. C. M., Keemink, L. M., & van der Molen, H. T. (2015). On the use and misuse of lectures in higher education. *Health Professions Education, 1*(1), 12–18.
- Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education, 10*(2), 159–169.