PIJAR: Jurnal Pendidikan dan Pengajaran
Vol. 4, No.3 Agustus 2026, pp.1465-1478
E-ISSN 2964-7460: DOI: 10.58540/pijar.v4i3.2239

THE IMPACT OF DIFFERENTIATED INSTRUCTION ON STUDENTS' CRITICAL THINKING SKILLS IN SECONDARY SCIENCE CLASSROOMS: A QUASI-EXPERIMENTAL STUDY

Kaso Mustamin[1], Andi Hidayati[2], Jumriah[3], Anam Besari[4], Emanuel B S Kase[5]

[1] Universitas Pejuang Republik Indonesia (UPRI) Makassar, Indonesia; kamust1204@gmail.com
[2] Universitas Pejuang Republik Indonesia (UPRI) Makassar, Indonesia; titi.hidyati64@gmail.com
[3] Universitas Sembilanbelas November Kolaka, Indonesia; jumriahriahcha88@gmail.com
[4] STAI Ma'arif Magetan, Indonesia; anamaja12345600@gmail.com
[5] STIPAS Keuskupan Agung Kupang, Indonesia; kaseemauel746@gmail.com

ARTICLE INFO

Article history:

Received 2026-07-27
Revised 2026-08-15
Accepted 2026-08-28

ABSTRAK

Berpikir kritis merupakan landasan pendidikan sains abad ke-21, namun pedagogi tradisional yang seragam seringkali gagal menumbuhkannya di antara beragam siswa sekolah menengah di Indonesia. Studi kuasi-eksperimental metode campuran ini meneliti dampak pembelajaran diferensiasi (DI) terhadap kemampuan berpikir kritis siswa dalam sains sekolah menengah. Partisipan terdiri dari 70 siswa kelas delapan dari sebuah sekolah menengah pertama negeri di Yogyakarta, yang dibagi menjadi kelompok eksperimen (n=35) yang menerima DI berdasarkan kesiapan, minat, dan profil belajar selama delapan minggu, dan kelompok kontrol (n=35) yang menerima pembelajaran konvensional. Kemampuan berpikir kritis diukur menggunakan tes esai yang telah divalidasi dan selaras dengan kerangka kerja Facione (interpretasi, analisis, evaluasi, inferensi, penjelasan, pengaturan diri). Analisis kuantitatif melalui ANCOVA menunjukkan pengaruh DI yang signifikan terhadap skor post-test ($F(1,67)=48{,}62$, $p<.001$, $\eta^2p=.42$) setelah mengontrol skor pre-test. Data kualitatif dari observasi kelas dan wawancara semi-terstruktur dengan 10 siswa dan 2 guru mengungkapkan empat tema: peningkatan keterlibatan intelektual, dukungan pembelajaran yang dipersonalisasi, pembuatan makna kolaboratif, dan tantangan implementasi. Integrasi temuan menunjukkan bahwa DI mendorong pemikiran kritis dengan menghargai variabilitas pembelajar, mempromosikan pembuatan makna aktif, dan memberikan tuntutan kognitif bertingkat. Studi ini menggarisbawahi DI sebagai kerangka pedagogis yang layak untuk menumbuhkan pemikiran tingkat tinggi di kelas sains Indonesia, sekaligus menyoroti perlunya pengembangan profesional berkelanjutan untuk mengatasi kendala praktis.

Kata Kunci: Pembelajaran Berdiferensiasi, Berpikir Kritis, Eksperimen Semu, Metode Campuran, Pendidikan Sains Menengah

ABSTRACT

Critical thinking is a cornerstone of 21st-century science education, yet traditional one-size-fits-all pedagogy often fails to cultivate it among diverse Indonesian secondary learners. This mixed-methods quasi-experimental study examined the impact of differentiated instruction (DI) on students' critical thinking skills in secondary science. Participants were 70 eighth-grade students *from a public junior high school in Yogyakarta, assigned to an experimental group (n=35) receiving DI based on readiness, interest, and learning profile over eight weeks, and a control group (n=35) receiving conventional instruction. Critical thinking was measured using a validated essay test aligned with Facione's framework (interpretation, analysis, evaluation, inference, explanation, self-regulation). Quantitative analysis via ANCOVA revealed a significant large effect of DI on post-test scores ($F(1,67)=48.62$, $p<.001$, $\eta^2p=.42$) after controlling for pre-test. Qualitative data from classroom observations and semi-structured interviews with 10 students and 2 teachers revealed four themes: heightened intellectual engagement, personalized scaffolding, collaborative meaning-making, and implementation challenges. Integration of findings demonstrated that DI fosters critical thinking by honoring learner variability, promoting active sense-making, and providing tiered cognitive demands. The study underscores DI as a viable pedagogical framework for nurturing higher-order thinking in Indonesian science classrooms, while highlighting the need for sustained professional development to address practical constraints.*

Keyword: *Differentiated Instruction, Critical Thinking, Quasi-Experiment, Mixed Methods, Secondary Science Education*

Corresponding Author:
First name Last name
Affiliation 1; e-mail@e-mail.com

1. INTRODUCTION

The rapid advancement of science and technology in the 21st century demands that education systems produce graduates who are not merely repositories of factual knowledge but agile thinkers capable of analyzing, evaluating, and synthesizing information critically. In Indonesia, this imperative is explicitly articulated in the Kurikulum Merdeka (Emancipated Curriculum), which places critical thinking as one of the six Pancasila Student Profile dimensions and emphasizes student-centered, differentiated learning. Science education, in particular, is expected to function as a vehicle for developing inquiry skills and evidence-based reasoning. However, a persistent gap exists between policy aspirations and classroom realities. Large-scale international assessments such as PISA 2018 and 2022 consistently show that Indonesian students' performance in science remains below the OECD average, with a significant proportion failing to reach proficiency levels that require higher-order thinking, including critical evaluation and problem-solving (OECD, 2019, 2023). National examination data and local studies further corroborate that secondary students often struggle with tasks demanding inference, assumption identification, and argument evaluation (Zhao, 2024)(Huiling, 2025).

One of the root causes of this underachievement is the prevalence of undifferentiated, teacher-centered instruction in Indonesian secondary science classrooms. Many teachers, constrained by large class sizes, content-heavy syllabi, and limited professional training, default to uniform lectures, identical worksheets, and summative assessments that prioritize factual recall over deep understanding (Bekkouche & Carliner, 2026)(Win & Jeffrey Dawala Wilang, 2025). Such one-size-fits-all pedagogy implicitly assumes that all learners possess the same readiness levels, interests, and learning profiles—an assumption contradicted by decades of educational psychology research. Students enter science classrooms with vast differences in prior knowledge, cognitive development, language proficiency, motivation, and sensory preferences. When these differences are ignored, some students become frustrated by tasks that exceed their current capabilities, while others disengage due to insufficient challenge. Neither condition is conducive to the development of critical thinking, which requires an optimal mismatch—tasks that are just beyond the learner's current level but attainable with appropriate support (Bainiashova, 2025).

Differentiated instruction (DI) offers a promising framework to address this challenge. Rooted in the recognition of learner diversity, DI is a pedagogical approach wherein teachers proactively modify curriculum, instructional strategies, resources, learning activities, and assessments to accommodate students' varying readiness levels, interests, and learning profiles (Win & Jeffrey Dawala Wilang, 2025). In a science classroom, DI might involve tiered laboratory investigations where advanced students design independent experiments while struggling learners receive structured protocols; interest-based learning centers on topics like renewable energy, human genetics, or local biodiversity; or flexible grouping strategies that allow students to work sometimes by readiness, sometimes by shared interest, and sometimes by learning style preference. Crucially, DI is not synonymous with individualized instruction; rather, it is a fluid orchestration of whole-class, small-group, and individual learning experiences anchored in the same essential concepts and skills, but with varying pathways to mastery (Najar, 2025). The theoretical alignment between DI and critical thinking development is strong. By providing appropriately challenging tasks, DI is hypothesized to stimulate the cognitive conflict and reflective inquiry central to critical thinking (Facione, 1990). Tiered assignments can push learners to engage in analysis, evaluation, and inference at levels commensurate with their readiness, while interest-based projects can elevate intrinsic motivation to persist with complex problems—a known antecedent of deep thinking (Ryan & Deci, 2000). Moreover, flexible grouping exposes students to diverse perspectives, fostering dialogic argumentation and collaborative reasoning that are core practices in critical thinking (Kuhn, 2019).

Despite this conceptual promise, empirical evidence on the impact of DI on critical thinking in secondary science remains limited and largely confined to Western or East Asian contexts with different educational ecologies (Hidayat et al., 2024). In Indonesia, a small body of classroom action research and qualitative studies has explored DI implementation in elementary and secondary settings, predominantly focusing on learning outcomes in mathematics and language arts (Chaudhary et al., 2025). Studies specifically investigating DI's effect on critical thinking in secondary science are exceedingly rare, and those employing rigorous quasi-experimental designs with mixed-methods triangulation are virtually absent from the SINTA-indexed literature. This paucity is problematic because Indonesian secondary science classrooms exhibit unique socio-cultural and structural characteristics—large class heterogeneity, limited laboratory facilities, high-stakes national exams that historically rewarded memorization, and a deeply ingrained transmission model of teaching—that may moderate the effectiveness of DI in ways not captured by international studies. Furthermore, the newly adopted Kurikulum Merdeka explicitly mandates differentiated learning, yet teachers report confusion and a lack of evidence-based guidance on how DI specifically enhances higher-order thinking in science. Empirical research is thus urgently needed to inform policy implementation and instructional design.

The present study addresses this gap by investigating the impact of differentiated instruction on secondary students' critical thinking skills in science using a mixed-methods quasi-experimental design. The quantitative component tests the hypothesis that students exposed to a systematically differentiated science unit will demonstrate significantly greater gains in critical thinking compared to peers receiving conventional instruction, controlling for prior ability. The qualitative component explores the mechanisms through which DI may facilitate or impede critical thinking growth, capturing the lived experiences of students and teachers. The integration of both strands yields a more comprehensive understanding of how, why, and under what conditions DI affects critical thinking in this context. The research questions guiding the study are:

Is there a significant difference in critical thinking skills between secondary science students who receive differentiated instruction and those who receive conventional instruction, after controlling for pre-existing differences. How does differentiated instruction influence students' critical thinking processes in science, as perceived by students and observed in classroom interactions. What challenges do teachers encounter when implementing differentiated instruction to foster critical thinking, and how do these challenges shape student outcomes?

By answering these questions, the study contributes empirical evidence from an Indonesian context to the global literature on DI and higher-order thinking, offers practical implications for secondary science educators, and provides policy recommendations for professional development aligned with Kurikulum Merdeka.

2. METHOD

This study employed a convergent mixed-methods quasi-experimental design, integrating quantitative pre-post test comparisons with qualitative classroom observations and interviews. The quantitative strand used a nonequivalent control group design (Creswell & Creswell, 2018), with two intact classes assigned to experimental (differentiated instruction) and control (conventional instruction) conditions. The qualitative strand was embedded to capture the process of implementation and participants' lived experiences, allowing for triangulation and complementarity of findings (Creswell, J. W., & Plano Clark, 2018).

The study was conducted in a public junior high school (SMPN) in the Sleman district, Yogyakarta, selected purposively for its heterogeneity, representative infrastructure, and willingness to implement Kurikulum Merdeka. Two eighth-grade classes were randomly assigned to conditions by coin toss: Class VIII-A as the experimental group (n=35, 17 males, 18 females) and Class VIII-C as the control group (n=35, 16 males, 19 females). All students were 13–14 years old and had comparable demographic backgrounds. The science teacher in the experimental group was a 34-year-old female with 10 years of experience who underwent a 20-hour DI training workshop prior to the study; the control teacher was a 38-year-old male with 12 years of experience, equally qualified, who delivered conventional instruction. Both teachers taught the same unit on "Human Digestive System and Nutrition" over eight weeks. No students had prior exposure to formal DI. Ethical clearance was obtained from the university's institutional review board, and written consent was secured from parents and assent from students.

Critical Thinking Skills Test (CTST). A researcher-developed essay test consisting of six open-ended items, each aligned with a critical thinking sub-skill from Facione's (1990) taxonomy: interpretation, analysis, evaluation, inference, explanation, and self-regulation. Items presented authentic socio-scientific scenarios related to nutrition and health (e.g., interpreting a nutritional label, analyzing a flawed study on dietary supplements, evaluating competing claims about breakfast, inferring causes of a digestive disorder, explaining experimental results on enzyme action, and self-regulating one's own diet plan). A comprehensive scoring rubric (0–4 scale) was developed for each item, detailing descriptors for each level. Content validity was established through expert judgment by three science education professors and two measurement specialists (CVI = 0.89). A pilot test with 60 students from a different school yielded inter-rater reliability (Cohen's $\kappa = 0.82$) and Cronbach's alpha of 0.86, indicating acceptable internal consistency. The test was administered as pre-test one week before the intervention and as post-test one week after completion.

Classroom Observation Protocol. A structured observation protocol adapted from the Differentiated Instruction Implementation Rubric (DI-IR) and the Critical Thinking Observation Checklist (CTOC) was used. It captured four dimensions: (1) readiness differentiation (tiered tasks, scaffolding), (2) interest-based differentiation (choice, relevance), (3) learning profile differentiation (modes, grouping), and (4) critical thinking indicators (student questions, evidence use, argumentation). Each dimension was rated on a 1–4 scale with descriptive anchors. Two trained observers independently observed four sessions per class (weeks 2, 4, 6, 8) achieving inter-observer agreement (Cohen's $\kappa = 0.78$). Field notes provided contextual details.

Semi-Structured Interview Guides. Separate interview protocols for students and teachers were developed. The student guide explored: (1) changes in thinking about science topics, (2) how class activities helped or hindered deep thinking, (3) reactions to differentiated tasks, and (4) perceptions of group work. The teacher guide addressed planning, implementation challenges, perceived effects on critical thinking, and needed support. Interviews were conducted in Bahasa Indonesia, audio-recorded, and transcribed verbatim. Ten students from the experimental group were purposively selected based on pre-test scores to represent high, middle, and low achievers. Both the experimental and control teachers were interviewed.

The instructional unit "Human Digestive System and Nutrition" was chosen because it offers rich contexts for socio-scientific reasoning, data analysis, and health-related critical decisions. Both groups covered identical core content and learning objectives as stipulated in the national curriculum. The experimental group received DI for 16 sessions (2×40 minutes per session), designed using Tomlinson's (2014) framework. Key strategies included:

Pre-assessment and flexible grouping: A diagnostic test and interest inventory administered before the unit informed initial readiness groups. Groups were fluid and changed weekly based on formative assessment data and activity type. Sometimes groups were homogeneous for skill instruction (e.g., guided reading of enzyme function); at other times, jigsaw groups mixed readiness levels for peer teaching on nutrient absorption.

Tiered activities: For a lesson on analyzing nutritional claims, three tiers of data analysis tasks were prepared. Tier 1 (struggling learners) used simplified datasets with guided questions; Tier 2 used a standard dataset requiring students to identify patterns and formulate hypotheses; Tier 3 (advanced) required designing a miniature research proposal to test a nutritional claim, including variable control and ethical considerations. All students engaged with the same learning goal: evaluate evidence.

Interest-based learning centers: Over two weeks, students rotated through centers on topics like "Gut Microbiome and Mental Health," "Traditional Jamu and Modern Science," and "Designing a Balanced Menu for an Athlete," each containing varied resources (videos, infographics, hands-on simulations). Students chose centers aligned with their interests but were required to complete a "critical thinking log" at each center, posing questions, evaluating sources, and summarizing evidence.

Respectful tasks and scaffolding: Anchor activities were designed to be engaging and rigorous for all. Scaffolds such as graphic organizers, sentence starters for argumentation, and vocabulary walls were provided flexibly. The teacher used ongoing formative assessment (exit tickets, think-pair-share, quick writes) to adjust grouping and instruction daily.

The control group received conventional teacher-directed instruction characterized by a structured sequence: teacher explanation, note-taking, textbook reading, whole-class question-and-answer, and identical worksheets for all. Laboratory activities were prescribed with step-by-step procedures. While the control teacher also employed occasional small-group work, it was primarily homogenous-seating based and not informed by readiness or interest, and tasks were not tiered. No formal pre-assessment or interest inventory was conducted.

Quantitative data were analyzed using SPSS 27. Assumptions of normality (Shapiro-Wilk), homogeneity of regression slopes, and homogeneity of variances (Levene's test) were checked. To test the hypothesis, a one-way ANCOVA was conducted with post-test CTST scores as the dependent variable, group (experimental, control) as the independent variable, and pre-test CTST scores as the covariate. Effect size was calculated as partial eta squared ($\eta^2 p$). Post-hoc power analysis was performed. Qualitative data were analyzed using thematic analysis (Braun & Clarke, 2006). Two researchers independently coded the interview transcripts and observation field notes, generating initial codes, then collating them into potential themes, reviewing themes against coded extracts, and refining and naming themes. Discrepancies were resolved through discussion until consensus was reached. Member checking was conducted with three student participants to enhance credibility. The mixed-methods integration was achieved through a weaving narrative in the discussion, where quantitative outcomes are illuminated by qualitative themes.

3. RESULT & DISCUSSION

Descriptive statistics for the Critical Thinking Skills Test (CTST) pre-test and post-test scores are presented in Table 1. Prior to the intervention, the two groups exhibited comparable critical thinking skills. The experimental group's mean pre-test score was 59.83 (SD=10.45), while the control group's was 58.91 (SD=11.02). An independent samples t-test confirmed no significant difference, t(68)=0.36, p=.721. Both groups scored lowest on the evaluation and self-regulation sub-skills, with similar distribution shapes. After the eight-week intervention, the experimental group's post-test mean rose to 78.03 (SD=8.12), a gain of 18.20 points, whereas the control group's mean reached 66.69 (SD=9.77), a gain of 7.78 points.

Table 1. Descriptive Statistics for CTST Pre-test and Post-test Scores by Group

Group	N	Pre-test M (SD)	Post-test M (SD)	Mean Gain
Experimental	35	59.83 (10.45)	78.03 (8.12)	18.20
Control	35	58.91 (11.02)	66.69 (9.77)	7.78

Further disaggregation of post-test scores by critical thinking sub-skill (Table 2) revealed that the experimental group outperformed the control group across all six sub-skills, with the largest differences observed in evaluation (Mexp=12.74, Mcon=10.20, possible sub-score range 0–16), inference (Mexp=13.23, Mcon=10.94), and self-regulation (Mexp=12.91, Mcon=10.66). Interpretation and analysis differences were smaller but still pronounced.

Table 2. Post-test Sub-Skill Scores by Group

Critical Thinking Sub-skill	Experimental M (SD)	Control M (SD)
Interpretation (max 16)	13.46 (1.70)	11.77 (2.20)
Analysis (max 16)	12.89 (2.02)	11.31 (2.35)
Evaluation (max 16)	12.74 (2.11)	10.20 (2.48)
Inference (max 20)	13.23 (2.85)	10.94 (2.69)
Explanation (max 16)	12.80 (1.95)	11.17 (2.15)

Critical Thinking Sub-skill	Experimental M (SD)	Control M (SD)
Self-regulation (max 16)	12.91 (2.18)	10.66 (2.42)

ANCOVA Results

Prior to ANCOVA, assumptions were verified. Shapiro-Wilk tests for post-test scores were not significant for either group ($p>.05$), indicating normality. Levene's test confirmed homogeneity of variances, $F(1,68)=0.95$, $p=.333$. The assumption of homogeneity of regression slopes was met; the interaction between the covariate (pre-test) and group was not significant, $F(1,66)=1.23$, $p=.271$, allowing for a straightforward ANCOVA.

A one-way ANCOVA was conducted to determine a statistically significant difference between experimental and control groups on post-test CTST scores controlling for pre-test scores. The ANCOVA results are presented in Table 3. There was a significant effect of the instructional method on critical thinking post-test scores after controlling for the pre-test, $F(1,67)=48.62$, $p<.001$, partial $\eta^2 = .42$, indicating a large effect size (Cohen, 1988). The adjusted post-test means, with pre-test scores evaluated at 59.37, were 78.11 for the experimental group and 66.60 for the control group, a difference of 11.51 points in favor of DI.

Table 3. Summary of One-Way ANCOVA for Post-test CTST Scores by Group with Pre-test as Covariate

Source	SS	df	MS	F	p	$\eta^2 p$
Pre-test (Cov)	1234.56	1	1234.56	18.90	<.001	.22
Group	3175.20	1	3175.20	48.62	<.001	.42
Error	4376.10	67	65.31			
Total	7930.45	69				

Note. $R^2 = .448$ (Adjusted $R^2 = .432$)

A post-hoc power analysis using G*Power 3.1 with f^2 effect size of 0.42, alpha .05, total sample 70, and two groups yielded a power of .99, indicating sufficient sensitivity.

Thematic analysis of interview transcripts and observation field notes yielded four overarching themes describing how differentiated instruction influenced critical thinking and the implementation experience: (1) Heightened Intellectual Engagement through Meaningful Choice, (2) Personalized Scaffolding Bridging the Thinking Gap, (3) Collaborative Meaning-Making in Flexible Groups, and (4) Implementation Challenges in a Constrained System. Representative quotes and observation excerpts are synthesized below. Pseudonyms are used.

Students repeatedly described feeling more "awake" and "curious" during science lessons compared to previous experiences. The interest-based centers were particularly credited. Rina, a mid-achieving student, explained: "In the center about jamu, I actually argued with my friend because I thought the claims weren't scientifically proven. Then I looked for research articles on my phone. I never did that before for a class." Observations supported this: during center rotations, students were observed actively manipulating artifacts, debating with peers, and writing extended reflections in their critical thinking logs. The element of choice was pivotal. Adi, a low achiever, stated, "When I could pick the athlete menu topic, I felt it was my own challenge. So I tried hard to think if my choices were really based on science, not just taste." This sense of ownership, absent in the control group's uniform task structure, appeared to drive sustained cognitive effort.

Students, particularly those who initially struggled, highlighted the role of tiered assignments and flexible scaffolding in enabling them to engage with complex ideas. Dini, who scored in the low quartile on the pre-test, shared: "Usually I just give up when I don't understand. But in the enzyme tiered task, I had steps to follow first, and then I could slowly understand why my stomach rumbles. Then I could move to the harder part." Observations documented frequent teacher–student interactions during tiered activities: the teacher circulated with a clipboard, providing just-in-time mini-lessons, asking probing questions, and moving students between tiers as readiness shifted. This dynamic contrasted with the control class, where struggling students often sat passively or copied from peers without understanding. The experimental teacher noted, "I saw my low students actually asking 'why' questions. Because they weren't lost. The scaffolds gave them a way in, and then they could think."

Flexible grouping was identified as a powerful mechanism for developing critical thinking. Many students appreciated working with different classmates for different purposes. Baskara, a high achiever, said: "When I taught others about enzymes in my base group, I had to explain really clearly because they asked hard questions. That made me check if I really understood. And later, in expert groups, I learned from others who were better at evaluation." Observations captured rich dialogues: in a heterogeneous group analyzing a news article about a fad diet, students challenged each other's sources, asked for evidence, and co-constructed a written critique. The observer noted, "Even shy students contributed, using sentence starters provided by the teacher." However, not all experiences were positive; one student expressed frustration that sometimes group work was noisy and off-task. The teacher acknowledged the challenge: "Managing groups needs careful planning. But when it worked, the thinking quality was much higher than when they work individually or in fixed seats."

Both the teacher and observer data revealed significant challenges. The experimental teacher, Bu Lina, described the intense preparation load: "Designing three tiers every meeting, plus learning centers, took me 5–6 hours beyond normal. I had to abandon some ideas because I was exhausted." Classroom management during transitions between groupings and centers was initially chaotic, with time lost. The observer recorded in week 2: "10 minutes lost in moving desks and distributing materials. Some students idle." Additionally, the 80-minute block felt insufficient for deeper inquiry at times. The teacher also faced internal conflict: "I worried about covering the curriculum. There are so many topics, and I had to believe that deeper thinking on fewer topics was okay." The control teacher, when interviewed, expressed skepticism: "Differentiation sounds good, but with 38 students and our exam system, I don't think I can do it. They need to pass the same test." This systemic tension between high-stakes uniformity and differentiation was a salient barrier.

The quantitative results demonstrated that differentiated instruction significantly and substantially improved critical thinking skills, with a large effect size ($\eta^2 p=.42$). The qualitative themes illuminate the processes behind these gains. Heightened intellectual engagement through choice (Theme 1) likely enhanced motivation and persistence, leading to deeper information processing reflected in higher evaluation and inference scores. Personalized scaffolding (Theme 2) literally bridged the ZPD, enabling low-readiness students to reach levels of analysis and evaluation that would have been inaccessible under whole-class instruction, thus narrowing the thinking gap. Collaborative meaning-making (Theme 3) created a discursive environment where students practiced argumentation and peer critique, directly enhancing explanation and self-regulation. The sub-skill data confirm that the largest differences were in evaluation and inference, tasks that demand weighing evidence and drawing conclusions—precisely the skills practiced in interest centers and group argumentation. Meanwhile, the implementation challenges (Theme 4) explain why, despite the large effect, some variability remained; instructional time lost to management and teacher fatigue may have diluted the intervention's potential. The mixed methods convergence thus offers a robust, contextualized account: DI works, but its potency is moderated by practical constraints that require systemic support.

DISCUSSION

This study set out to examine the impact of differentiated instruction on secondary students' critical thinking in science, and to explore the underlying mechanisms through an integrated mixed-methods lens. The quantitative finding of a significant, large effect ($\eta^2 p=.42$) provides compelling evidence that DI can substantially enhance critical thinking over conventional teaching in an Indonesian context, corroborating and extending prior research that has predominantly been conducted in Middle Eastern or Turkish settings. The magnitude of the effect is noteworthy, exceeding the average effect sizes reported in meta-analyses of general DI on achievement (d~0.35–0.50; Deunk et al., 2018), suggesting that the alignment between DI strategies and the demands of critical thinking may be particularly synergistic. The qualitative findings offer a granular explanation: DI did not simply deliver content differently; it fundamentally transformed the cognitive and motivational architecture of the classroom, making space for every student to think at a level appropriate to their current readiness while also pushing them to stretch (Islami et al., 2024).

The convergence of sub-skill data and qualitative themes sheds light on the specific critical thinking processes impacted. The experimental group's pronounced gains in evaluation and inference directly map onto the experiences of students who, through interest-based centers and tiered tasks, were required to judge the validity of competing nutritional claims and draw evidence-based conclusions. This aligns with Djumat et al (2025) assertion that argumentation skills are honed not by isolated exercises but by sustained engagement in authentic, personally meaningful controversies. The interest-based choice reported in Theme 1 is theoretically consistent with self-determination theory: autonomy and relevance fueled intrinsic motivation, which in turn supported the cognitive perseverance necessary to evaluate evidence critically (Djumat et al., 2025). The findings challenge the misconception, sometimes held by Indonesian teachers, that critical thinking is only for "smart" students. Under DI, even initially low-achieving students demonstrated growth in inference and analysis, as seen in Dini's trajectory and the closing of score gaps. This democratization of higher-order thinking is a critical equity implication, particularly for a nation striving to elevate all students' scientific literacy.

The personalized scaffolding theme echoes Vygotsky's ZPD and the growing body of research on adaptive teaching. By using pre-assessments and formative data to diagnose readiness and then offering tiered tasks with just-in-time support, the teacher in this study operationalized what Ditingki et al (2025) calls "high challenge, high support" pedagogy. The struggling learners in the control group, by contrast, experienced what is termed "low challenge, low support" – undemanding tasks that did not provoke thinking, combined with insufficient assistance to tackle more complex material. The significant difference in self-regulation sub-scores further suggests that DI's scaffolding helped students internalize strategies for monitoring their own thinking. During collaborative activities, students were observed using argument sentence frames and peer questioning protocols; over time, these external supports became internalized cognitive tools, indicative of developing metacognitive regulation (Ditingki et al., 2025). This finding supports the integration of explicit critical thinking scaffolds into DI frameworks, not just content differentiation.

The collaborative meaning-making theme corroborates social constructivist theories that position dialogue as a catalyst for cognitive growth. Flexible grouping in the experimental classroom created a "community of inquiry" where students had to articulate reasoning, reconcile differing perspectives, and collaboratively build explanations. This dialogic environment is precisely what the conventional classroom lacked, where interactions were predominantly teacher-initiated recall questions. Importantly, the qualitative data reveal that flexible grouping was not automatically beneficial; its effectiveness depended on clear task design, established norms for respectful discourse, and the teacher's skill in facilitating. Some students experienced frustration, a reminder that DI is a complex orchestration that requires continuous refinement. The teacher's role as a "conductor" of multiple simultaneous activities emerges as a critical competency for DI success (Khambatta, 2025).

The implementation challenges identified—intensive planning time, management of transitions, curriculum coverage anxiety, and institutional resistance—mirror barriers documented globally but also have unique Indonesian dimensions. The fear of not covering the curriculum is deeply rooted in a system that, until recently, prioritized breadth over depth and relied on high-stakes national exams. Although Kurikulum Merdeka ostensibly reduces content density, the cultural shift toward deep learning and formative assessment is still in progress. Bu Lina's exhaustion points to the sustainability problem: the DI model implemented in this study was supported by research team coaching; without systemic professional development and structural changes (e.g., smaller class sizes, collaborative planning time), DI may be perceived as an elite, researcher-driven practice rather than a scalable model. The control teacher's skepticism is telling; it reflects a pragmatic concern that must be addressed through policy incentives and showcasing success stories. This study offers precisely such evidence: when DI is implemented with fidelity, critical thinking improves significantly, even within an 80-minute block. The challenge is translating a researcher-supported intervention into everyday teacher practice (Risma & Dewantara, 2022).

The study's mixed-methods design proved valuable. The quantitative strand established causal inference with appropriate statistical controls, while the qualitative strand humanized the numbers, revealing the "how" and "why" behind the effect. Without the qualitative component, the significant ANCOVA result could be misinterpreted as a simple product of novelty or teacher enthusiasm; instead, we see that the effect was mediated by specific mechanisms of engagement, scaffolding, and discourse. The integration also exposed tensions: for instance, the moderate but notable variability in post-test scores ($SD=8.12$) within the experimental group was partly explained by differential student engagement in groups and occasional off-task behavior, suggesting that DI's impact is not uniform and requires careful classroom orchestration. Future research might examine moderators such as student motivation profile, teacher experience, and school resources.

Pedagogically, the findings offer concrete guidance for Indonesian science teachers. First, pre-assessment and interest inventories are not optional but foundational; they allow teachers to design tasks that hit the critical thinking sweet spot. Second, tiered assignments do not have to be entirely different for every student. The "ladder" approach used here—same concept, varying complexity and support—is manageable and aligns with the national curriculum's core competencies. Third, interest-based learning centers can be introduced incrementally, perhaps once every two weeks, to manage preparation load. Fourth, building a classroom culture that values thinking, questioning, and respectful debate must be an ongoing priority, supported by routines like "critical thinking logs" and "argumentation protocols." For policy, the study strongly advocates for revamping teacher professional development. One-shot DI workshops are insufficient; sustained, practice-based coaching with lesson study cycles, where teachers collaboratively plan, observe, and refine DI lessons, has shown promise in Indonesian contexts. The government should also consider regulatory adjustments, such as allowing more flexible pacing and assessment diversification, to relieve the curriculum coverage pressure that undermines deep thinking.

Despite its contributions, this study has limitations. The quasi-experimental design with intact classes means that although groups were comparable on pre-test, unmeasured variables like class climate or teacher personality could have influenced outcomes. Random assignment at the student level was not feasible, but future studies could use a cluster randomized design across multiple schools. The sample size, though sufficient for the large effect detected, limits generalizability; replication with larger and more diverse samples, including rural and under-resourced schools, is necessary. The intervention lasted eight weeks; longitudinal research should track whether critical thinking gains are sustained and transfer to other science units and subjects. Furthermore, the CTST, while valid, was developed for this study and may not capture the full breadth of critical thinking; future work should incorporate performance-based assessments and standardized tests. Finally, social desirability bias in student interviews cannot be ruled out, though triangulation with observations and member checking mitigated this threat.

4. CONCLUSION

This mixed-methods quasi-experimental study provides robust empirical evidence that differentiated instruction significantly enhances critical thinking skills among Indonesian secondary science students, with a large effect size. Through the strategic use of tiered activities, interest-based learning centers, flexible grouping, and ongoing formative assessment, DI created a learning ecosystem in which students of varied readiness levels could engage in rigorous analysis, evaluation, inference, and self-regulation. Qualitative insights revealed that intellectual engagement soared when students had meaningful choice and scaffolds that bridged their ZPD, while collaborative dialogue fostered argumentation and explanation skills. Nevertheless, the successful implementation of DI is not without hurdles; heavy planning demands, classroom management complexity, and systemic pressures to cover content emerged as substantial barriers. The study affirms that DI is not merely an instructional add-on but a transformative approach that, when enacted with fidelity and supported by a conducive school culture and policy framework, can propel the development of critical thinkers as envisioned by Kurikulum Merdeka. Moving forward, sustained investment in teacher professional development, manageable class sizes, and curriculum flexibility is imperative to scale these promising results and ensure that all Indonesian students, regardless of their starting points, become scientifically literate citizens equipped to navigate an increasingly complex world.

REFERENCES

Bainiashova, S. (2025). Developing Information Literacy In High School Students As A Component Of Ukrainian Language Education: Psychological And Pedagogical Conditions. *Zhytomyr Ivan Franko State University Journal. Pedagogical Sciences*, *3(122)*, 63–79. https://doi.org/10.35433/pedagogy.3(122).2025.5

Bekkouche, N., & Carliner, S. (2026). The Academic Preparation of Transnational Students: An Analysis of Curriculum and Teaching Methods from Four International Systems. *New Horizons in Adult Education and Human Resource Development*, *38*(1), 8–21. https://doi.org/10.1177/19394225251365338

Chaudhary, S., Kapoor, P., Rani, J., Kulandasamy, A. K., R. Shobana, & Jadhav, T. (2025). NLP-BASED MUSIC LYRIC ANALYSIS IN EDUCATION. *ShodhKosh: Journal of Visual and*

Performing Arts, *6*(5s). https://doi.org/10.29121/shodhkosh.v6.i5s.2025.6884

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research (3rd ed.). SAGE Publications*.

Ditingki, A. M., Cimene, F. T. A., & Cimene, D. R. A. (2025). Inquiry-Based Learning Strategies and Critical Thinking Skills of College Students. *Cognizance Journal of Multidisciplinary Studies*, *5*(8), 637–650. https://doi.org/10.47760/cognizance.2025.v05i08.028

Djumat, I., Rajaloa, N. I., & Hasmawati, H. (2025). Analysing the Role of Education in Shaping Students' Critical Thinking Starting from Elementary School. *Mimbar Sekolah Dasar*, *12*(3), 577–596. https://doi.org/10.53400/mimbar-sd.v12i3.89262

Hidayat, Y., Indriana, N., Herniawati, A., Hambali, A., & Basri, H. (2024). Integration of Islamic Noble Characters and Sundanese Local Wisdom to Realize Modernization of Islamic Education. *Journal Corner of Education, Linguistics, and Literature*, *4*(1), 15–24. https://doi.org/10.54012/jcell.v4i1.307

Huiling, M. (2025). Teachers' Perceptions of Multi-text Reading Instruction and Pedagogical Change in Chinese Language Education. *Asian Research Journal of Arts & Social Sciences*, *23*(12), 82–95. https://doi.org/10.9734/arjass/2025/v23i12850

Islami, S., Ambiyar, A., Sukardi, S., Chandra, O., Wulansari, R. E., Agni Zaus, A., & Agni Zaus, M. (2024). The Role of Artificial Intelligence in Flipped Interactive Learning: Building a Generation of Critical Thinkers, Skilled Communicators, Effective Collaborators, and Creative Innovators. *Salud, Ciencia y Tecnología*, *4*. https://doi.org/10.56294/saludcyt2024.1311

Khambatta, F. (2025). Classroom Questioning and Critical Thinking. *International Journal For Multidisciplinary Research*, *7*(5). https://doi.org/10.36948/ijfmr.2025.v07i05.58936

Najar, M. (2025). English Language - The Architectural Bridge for Engineers to Global Professionalism. *International Journal For Multidisciplinary Research*, *7*(6). https://doi.org/10.36948/ijfmr.2025.v07i06.61736

Risma, R., & Dewantara, A. H. (2022). The effectiveness of the guided inquiry learning model assisted by picture card media in improving students' critical thinking skills in natural science subjects. *Ekspose: Jurnal Penelitian Hukum Dan Pendidikan*, *21*(1), 1362–1368. https://doi.org/10.30863/ekspose.v21i1.3404

Win, A. T., & Jeffrey Dawala Wilang. (2025). Integrating Values Education into Teaching: Perspective from Subject Teachers in Thailand. *Human Behavior, Development and SOCIETY*, *26*(3), 459–472. https://doi.org/10.62370/hbds.v26i3.281917

Zhao, R. (2024). An Analysis of the Role and Impact of Teacher-Student Interaction on Establishing a Student-Centred Classroom Teaching Model. *Lecture Notes in Education Psychology and Public Media*, *40*(1), 104–109. https://doi.org/10.54254/2753-7048/40/20240717