

Adaptive Socratic AI Scaffolding Supports Critical Thinking in Indonesian Secondary School Physics Classrooms

I Gede Purwana Edi Saputra

gedepurwana@gmail.com

State University of Malang

Punaji Setyosari

State University of Malang

Henry Praherdhiono

State University of Malang

Made Duananda Kartika Degeng

State University of Malang

Research Article

Keywords: Adaptive Scaffolding, Critical Thinking, Physics Education, Socratic AI-Based

Posted Date: October 15th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10909958/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Adaptive Socratic AI Scaffolding Supports Critical Thinking in Indonesian Secondary School Physics Classrooms

I Gede Purwana Edi Saputra^{1,2*}

¹Department of Educational Technology, Faculty of Education, State University of Malang, Malang, Indonesia.

²Department of Physics Education, Universitas Sembilanbelas November Kolaka, Kolaka, Indonesia. Orchid ID : 0000-0003-3691-5213, E-Mail : gedepurwana@gmail.com

Punaji Setyosari¹,

¹Department of Educational Technology, Faculty of Education, State University of Malang, Malang, Indonesia. Orchid ID : 0000-0003-0187-9785,, E-Mail : punaji.setyosari.fip@um.ac.id

Henry Praherdhiono¹

¹Department of Educational Technology, Faculty of Education, State University of Malang, Malang, Indonesia. Orchid ID : 0000-0002-5922-7491, E-Mail : henry.praherdhiono.fip@um.ac.id

Made Duananda Kartika Degeng^{1*}

¹Department of Educational Technology, Faculty of Education, State University of Malang, Malang, Indonesia. Orchid ID : 0000-0002-6749-6237, E-Mail : made.degeng.fip@um.ac.id

*** Corresponding Author**

Abstract: This study developed, content-validated, and evaluated the effectiveness of a Socratic AI-based adaptive scaffolding system for fostering critical thinking in secondary physics education. The system was developed through the ADDIE model and evaluated using a non-equivalent pretest-posttest control-group design. Content validity was assessed by two experts using a 20-item, four-point validation instrument analyzed with Aiken's V. The effectiveness study involved 135 Grade XI students from two secondary schools, comprising 69 students in the experimental group and 66 in the control group. Both groups received seven guided-inquiry sessions; however, only the experimental group used the Socratic AI-enhanced E-Laboratory. Posttest critical-thinking scores were analyzed using one-way ANCOVA, with pretest scores as the covariate. Expert evaluation yielded strong content validity (Aiken's $V = 0.90$). After adjustment for baseline performance, the experimental group achieved a significantly higher posttest mean than the control group, $F_{(1,132)} = 79.63$, $p < .001$, partial $\eta^2 = .374$. The adjusted means were 83.51 and 73.38, respectively, corresponding to a 10.13-point difference. These findings indicate that pedagogically constrained AI, integrating Socratic questioning and adaptive scaffolding, can support critical-thinking development by guiding reasoning rather than providing immediate solutions.

Keywords: Adaptive Scaffolding, Critical Thinking, Physics Education, Socratic AI-Based.

1. Introduction

Physics education persistently faces challenges in fostering students' critical thinking, particularly due to rote memorization, limited engagement, and difficulties in transferring conceptual understanding to real-world contexts [1], [2]. Although physics is grounded in systematic observation, experimentation, and logical reasoning, classroom instruction often prioritizes procedural fluency over epistemic understanding. As a result, students may become proficient in manipulating equations yet struggle to explain underlying assumptions, evaluate alternative explanations, or apply concepts in unfamiliar situations. Research in science education consistently shows that learners tend to retain algorithmic problem-solving routines without developing strong conceptual frameworks [3], [4], [5], [6], [7]. This gap is reflected in the dominance of formula-driven approaches that constrain conceptual reasoning, hinder knowledge transfer, and limit students' capacity for argumentation and evidence-based justification, as core dimensions of scientific reasoning.

This limitation is most evident in the development of critical thinking, which involves analyzing claims, evaluating evidence, identifying assumptions, constructing arguments, and reflecting on reasoning processes [8], [9]. In physics, these competencies underpin model construction, hypothesis testing, and theory evaluation rather than serving as supplementary skills [10], [11]. However, many instructional and assessment practices continue to privilege the correctness of final answers over the reasoning processes that produce them [12], [13], [14], [15], [16], [17]. When outcomes are prioritized over epistemic engagement, students tend to reproduce procedures rather than engage in reflective evaluation, constraining the cultivation of higher-order reasoning that physics education explicitly aims to develop.

Empirical evidence from a synthesis of 20 major studies conducted between 2021-2025 involving 3,612 Indonesia secondary school students further confirms this concern. Critical thinking performance remains at moderate to low levels, with an overall mean score of 66.6% against a 70% target, and 10 studies reporting very low achievement across key indicators. The review results are shown in Figure 1.

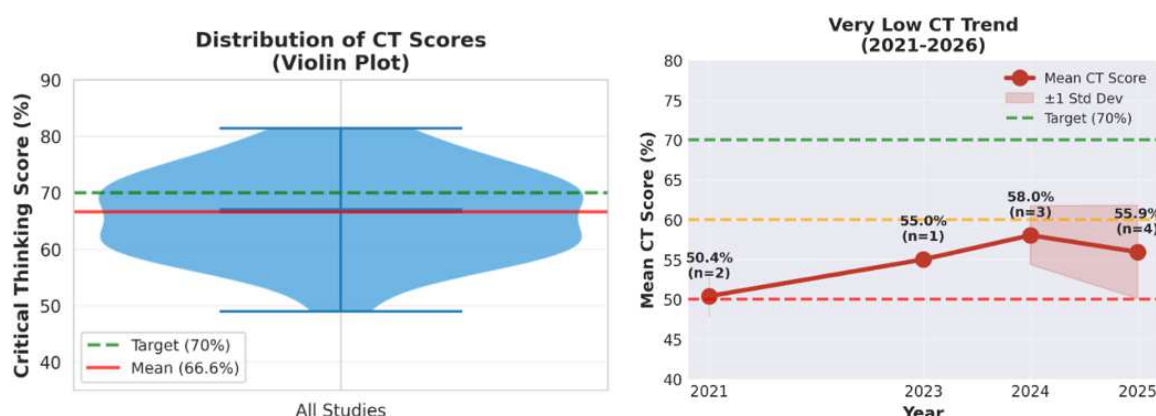


Fig. 1. Mean Score of CT and very low CT trend on physics in Indonesia secondary school.

The weakest dimensions include self-regulation, evaluation, and analytical reasoning, despite curriculum mandates emphasizing Higher Order Thinking Skills (HOTS) and the implementation of inquiry, project-based, and blended learning models [18], [19]. These findings suggest that existing instructional innovations have

improved engagement but have not consistently transformed the epistemic quality of reasoning in physics education. The review results are shown in Figure 2.

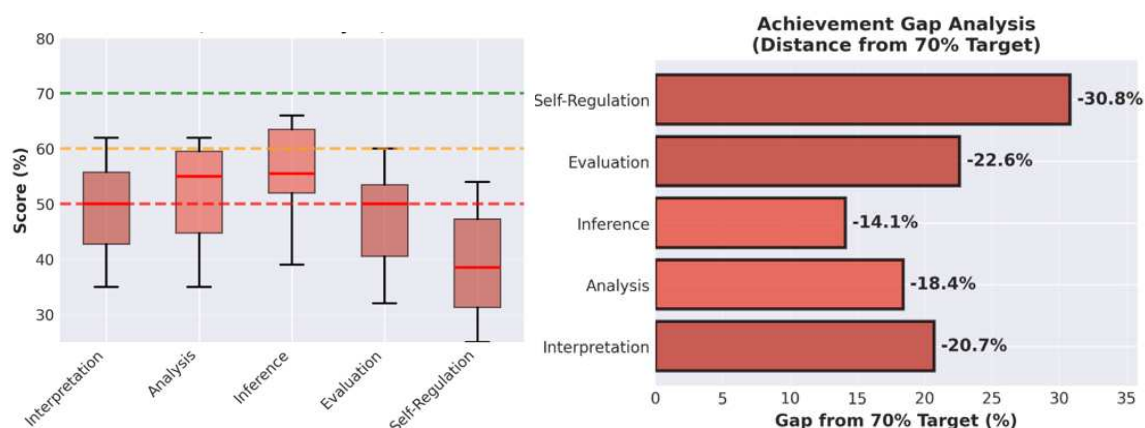


Fig. 2. Weakest score of the CT indicator on physics in Indonesia secondary school.

However, recent studies indicate various innovative instructional models have been implemented to improve CT skills, include project-based, problem-based, inquiry, and the use of technology-enhanced learning environments [20], [21], [22], [23], [24], [25]. Among these, AI-supported learning environments show particular promise due to their capacity to deliver adaptive feedback, support data-driven reasoning, and facilitate reflective problem analysis targeting higher-order cognition [26], [27], [28], [29], [30]. Previous intervention results, showed substantial increases in students' critical thinking scores across AI-supported with large effect >0.8 mean cohen's d [31], [32], [33], [34], such as intervention AI-PBL in high effectiveness (N-gain 0.91) [35]. This shift, marks AI enhanced a transition from procedural instruction to epistemically oriented learning, where students are encouraged not only to obtain correct answers but also to justify, evaluate, and reconstruct their reasoning. The review results are shown in Figure 3.

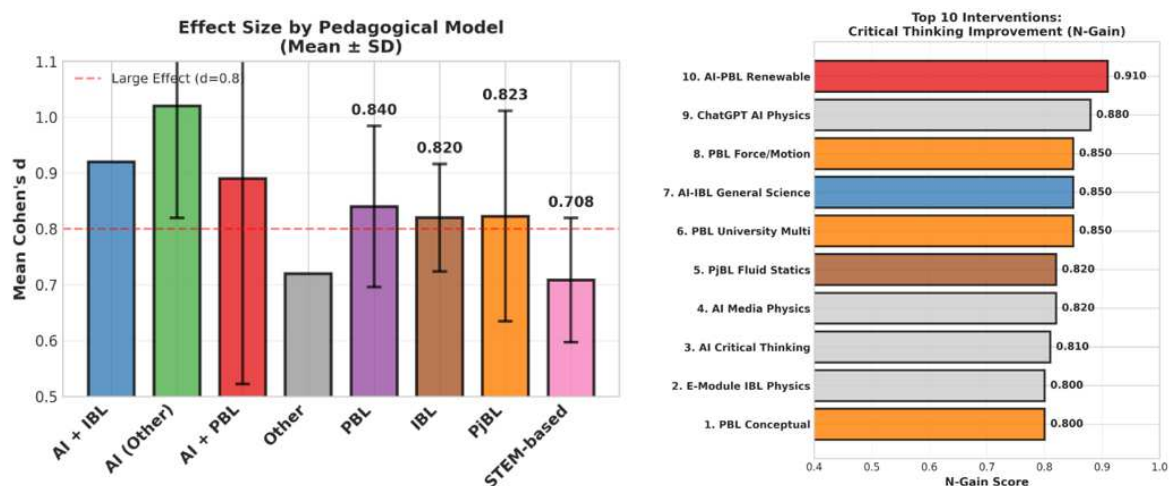


Fig. 3. The effect size and N-gain AI-supported on intervention students critical thinking.

One pedagogical tradition closely aligned with such reflective reasoning is the Socratic method. The Socratic method, characterized by dialogic questioning, represents a pedagogical approach that fosters critical thinking and reflective reasoning by positioning educators or intelligent systems as adaptive scaffolds

guiding learners to examine assumptions, evaluate alternatives, and articulate reasoning processes [36]. Advances in artificial intelligence enable the operationalization of Socratic questioning at scale through educational chatbots [37], fine-tuned large language models, and Socratic question generation systems, which provide personalized prompts, iterative feedback, and context-sensitive inquiry. Empirical studies show that such systems including Socratic tutors, Socratic Question Generation datasets, and Socratic Intelligent Conversational Agents (S-ICA) [38], significantly enhance critical thinking, reflective learning, academic achievement, and student engagement compared to conventional instructional or non-Socratic digital tools [39], [40].

The benefits are particularly evident in domains requiring analytical decision-making, where learners are encouraged to question assumptions, construct arguments, and connect prior knowledge with new insights [41]. Nevertheless, effective implementation requires pedagogical readiness, teacher expertise in questioning strategies, and contextual adaptation, while AI offers a scalable solution to extend adaptive Socratic interaction to broader learning environments. In this context, AI can operationalize Socratic questioning at scale by providing personalized prompts, iterative feedback, and context-sensitive guidance, thereby fostering analytical reasoning and argumentation in ways that mirror authentic scientific inquiry. Emerging evidence suggests that Socratic AI systems, such as LLM or structured educational chatbots and intelligent conversational agents, can enhance reflective reasoning, depth of justification, and evidence-based argumentation more effectively than unstructured digital tools, particularly when questions explicitly require explanation, evaluation of counterarguments, and articulation of evidence. The analysis result shown in Figure 4.

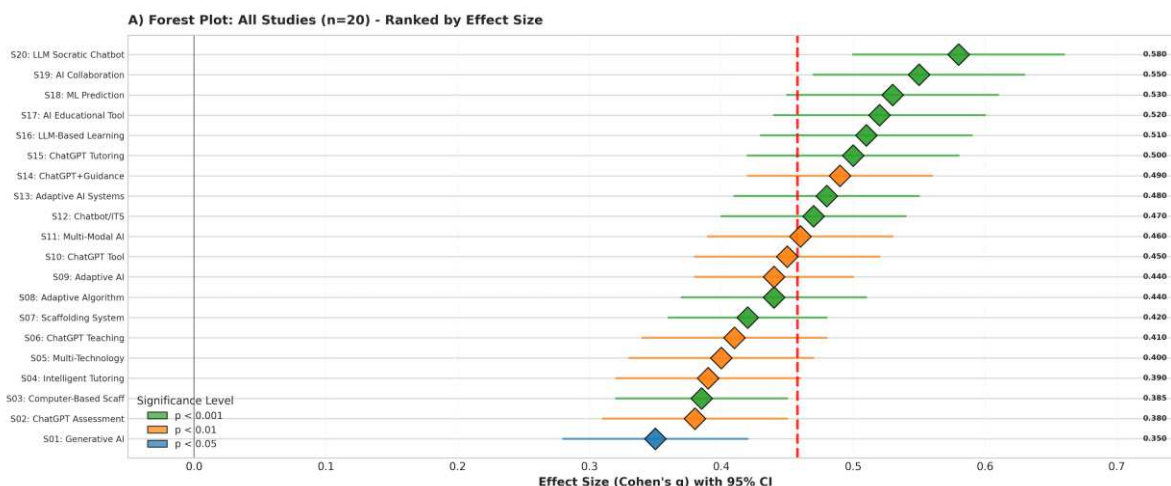


Fig. 4. Effect size of LLM Socratic Chatbot/AI on learning outcome.

From a learning theory perspective, this approach aligns with scaffolding grounded in Vygotskian theory, where learners receive temporary support from a more knowledgeable agent to perform tasks within their Zone of Proximal Development before gradually achieving independence [42], [43]. Its core elements include diagnosis of learner needs, responsive adjustment of support, and a staged handover toward independence [44], [45]. In technology-enhanced contexts, scaffolding evolves into adaptive scaffolding, in which support is dynamically

calibrated based on real-time evidence of learner performance through digital platforms, simulations, and intelligent systems [46], [47], [48], [49].

Between 2022 and 2026, research on artificial intelligence in education expanded rapidly, particularly through intelligent tutoring systems, adaptive platforms, and conversational agents that support personalized learning [50]. In physics education, these technologies can provide immediate feedback, recommend relevant resources, and adjust instructional support to students' performance and emerging needs [51]. Nevertheless, most applications remain focused on answer accuracy, procedural guidance, task completion, and instructional efficiency [52]. Far less attention has been given to AI systems that sustain reflective dialogue, challenge assumptions, and progressively develop conceptual reasoning and critical thinking [53].

This limitation is especially consequential in physics, where meaningful learning requires students to formulate problems, construct hypotheses, evaluate evidence, relate experimental findings to theory, and revise explanations when evidence is inconsistent [54]. Although AI may accelerate these processes, greater efficiency does not necessarily produce deeper conceptual understanding or stronger epistemic judgment [55].

A critical gap therefore remains between rapid technological innovation and the development of theoretically grounded, pedagogically coherent, and empirically validated instructional frameworks. Existing studies generally examine AI, Socratic dialogue, adaptive scaffolding, and critical-thinking interventions as separate components, leaving limited evidence on how they can be integrated into a unified model for sustaining inquiry and regulating instructional support [56].

To address this gap, the present study aims to develop, validate, and evaluate the effectiveness of an AI-driven Socratic adaptive scaffolding model for secondary physics education. The model integrates Socratic dialogue, guided inquiry, adaptive scaffolding, and intelligent technological support within a coherent instructional framework [37]. Rather than functioning merely as a source of answers or procedural assistance, AI is positioned as a dialogic and adaptive partner that supports students in articulating assumptions, examining evidence, refining explanations, and reflecting on the development of their reasoning [57], [58]. The study's contribution and novelty are further established through VOSviewer-based bibliometric mapping of research on Socratic questioning, artificial intelligence, adaptive scaffolding, critical thinking, intelligent tutoring, and physics education. Figure 5 illustrates the resulting network structure, emerging thematic trends, cross-cluster relationships, and the novelty position of the proposed model.

stage of reasoning, from concept activation through hypothesis testing to metacognitive reflection. A teacher-facing analytics layer keeps educators, not the system, in charge of interpreting evidence and deciding when support should shift or fade, preserving learner agency alongside AI assistance. Situated within Indonesia's Merdeka Curriculum, the model offers a pedagogically grounded route toward inquiry, autonomy, and reflective reasoning in secondary physics classrooms.

2. Methodology

Research Design

This study employed a Research and Development (R&D) methodology [59], operationalized through the ADDIE instructional design model, which comprises five iterative phases: Analysis, Design, Development, Implementation, and Evaluation [60]. The comprehensive stage shown in figure 6.

ADDIE-Based Development Procedure of the AI-Driven Socratic Adaptive Scaffolding Model

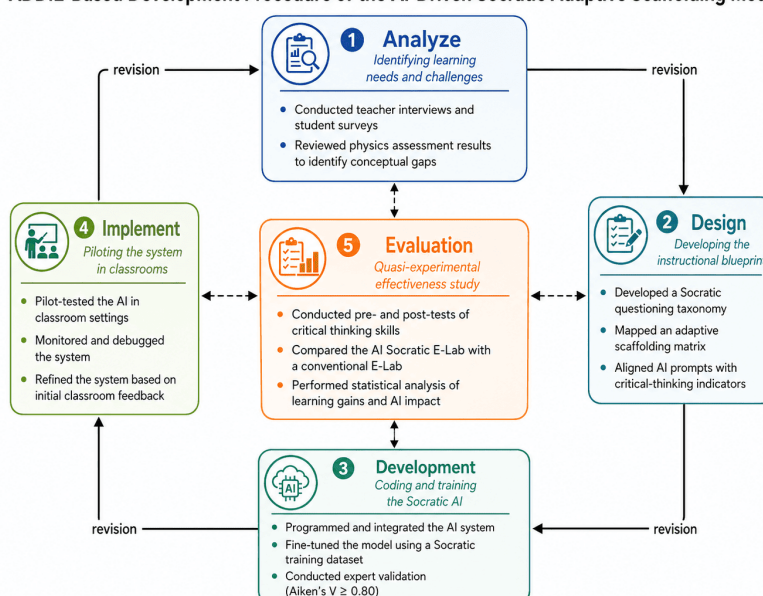


Fig. 6. ADDIE instructional design for multimedia mobile learning development procedure.

Based on figure 6, teacher interviews, student surveys, and physics assessment results in the Analysis phase were examined to identify conceptual difficulties, weak CT performance, and the need for personalized reasoning support. The Design phase established the instructional and technological blueprint, including the Socratic questioning taxonomy, adaptive scaffolding matrix, reasoning-state classification, AI guardrails, and alignment with guided inquiry and critical-thinking indicators. In the Development phase, the architecture was programmed and integrated into the e-laboratory, the AI model was trained using structured Socratic prompts, and the product was validated by physics education and educational technology experts. The Implementation phase involved classroom piloting to assess usability, system stability, adaptive functioning, and instructional consistency, followed by revisions based on initial feedback. Finally, the Evaluation phase examined effectiveness through a non-equivalent pretest–posttest control-group design [61], comparing the Socratic AI–integrated e-laboratory with a conventional e-

laboratory. Evaluation findings informed iterative revision and finalization of the system.

Research Procedure

The research was conducted in two integrated phases: (1) system development using the ADDIE model and (2) effectiveness evaluation through a quasi-experimental design [62]. This used to examine Effectiveness in student's critical thinking after the intervention of a Socratic AI-Based. The comprehensive procedure shown in figure 7.

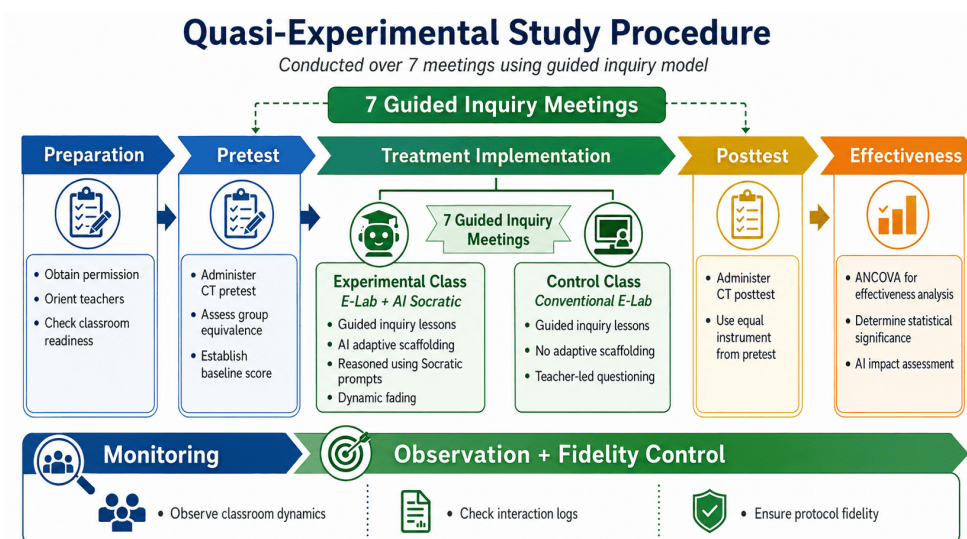


Fig. 7. Quasi-experimental study procedure integrating guided inquiry.

The research procedure comprised five stages: preparation, pretesting, treatment implementation, posttesting, and effectiveness evaluation. During preparation, permissions, teacher orientation, and classroom readiness were secured. Both groups completed an identical critical-thinking pretest to establish baseline comparability. Across seven meetings, both classes followed guided inquiry, while the experimental class used the Socratic AI-enhanced E-Laboratory and the control class used the conventional E-Laboratory.

Guided inquiry was selected because its cycle of questioning, hypothesis formulation, experimentation, evidence analysis, and conclusion construction reflects the epistemic practices required in physics [63], [64], [65]. The AI layer complemented this structure by diagnosing students' reasoning, delivering adaptive Socratic prompts, and gradually fading support as competence increased. According to [NO_PRINTED_FORM] [66] this integration was intended to move students beyond procedural problem solving toward assumption testing, evidence evaluation, conceptual justification, and reflective judgment. Classroom dynamics, interaction logs, and implementation fidelity were monitored throughout, followed by an equivalent posttest to assess changes in critical-thinking performance.

Sample and Data Collection

The study population consisted of Grade XI students enrolled in physics courses at senior secondary schools (Sekolah Menengah Atas) in Kolaka, Southeast

Sulawesi, Indonesia. Two schools with comparable academic profiles, infrastructure readiness, and national curriculum implementation were selected. Using cluster sampling, four intact classes comprising 135 students were included to preserve naturally occurring classroom conditions. One class from each school was assigned to the experimental group and one to the control group, yielding 69 students in the experimental condition and 66 in the control condition. Baseline comparability was assessed using pretest scores before the intervention.

Data were collected using two complementary instruments: a media validation sheet and a CT test. The media validation sheet was used to establish the pedagogical, functional, and technical feasibility of the Socratic AI-based adaptive scaffolding system before classroom implementation. Meanwhile, the critical-thinking test was administered as an identical pretest and posttest to measure changes in students' reasoning performance following the intervention. The instruments, assessed dimensions, format, and scoring methods are summarized in Table 1.

Table 1. Overview of the data-collection instruments and assessment procedures

Instrument	Format	Assessed dimensions/indicators	Scoring and analysis
Media validation sheet	20-item expert-validation instrument administered using a four-point Likert scale.	Socratic pedagogical alignment; adaptive scaffolding and critical-thinking support; AI functional and technical quality; usability and feasibility.	Item-level and overall validity were analyzed using Aiken's V [67].
Critical-thinking test	5 essay-based physics problems administered as an identical pretest and posttest.	Clarification, assumption identification, evidence evaluation, inference, and reflective judgment based on Ennis's CT framework [68] .	Responses were scored using a four-level analytic rubric based on reasoning quality, coherence, and justification.

Because the quality and interpretability of the intervention depended on the validity of the developed system, the media validation instrument was subjected to detailed expert review prior to classroom implementation. The media validation instrument was adapted and synthesized from empirical design and evaluation criteria reported in three recent studies on Socratic conversational agents [38], explainable adaptive feedback [69], and educational AI quality [70]. These criteria were contextualized into 20 statements across four dimensions: Socratic pedagogical alignment, adaptive scaffolding and critical-thinking support, AI functional and technical quality, and usability and implementation feasibility. The indicators addressed the system's capacity to sustain reflective questioning, provide relevant and adaptive support, maintain scientifically accurate and logically comprehensible responses, prevent direct-answer generation, and operate effectively within guided-inquiry physics instructional. The item-level validation results are presented in Table 2.

Table 2. Content Validity Instrument of the Socratic AI–Based Adaptive Scaffolding

Indicator	Item Number	Validation statement
Pedagogical alignment with Socratic questioning	1	The AI prompts encourage students to clarify physics problems, concepts, and underlying assumptions.
	2	The AI questions require students to provide reasons and evidence rather than merely state final answers.
	3	The Socratic dialogue encourages students to consider alternative explanations and possible implications.
	4	The AI responses consistently maintain a dialogic and non-directive Socratic questioning pattern.
	5	The Socratic prompts are aligned with the objectives and phases of guided inquiry in physics learning.
Adaptive scaffolding and critical-thinking support	6	The system can distinguish students' fragmented, procedural, conceptual, and reflective reasoning states.
	7	The level of AI support is adjusted to students' identified reasoning needs.
	8	The system gradually reduces scaffolding as students demonstrate greater reasoning competence.
	9	The prompts support clarification, assumption identification, evidence evaluation, inference, and reflective judgment.
	10	The adaptive scaffolding promotes learner autonomy and prevents excessive dependence on AI assistance.
AI functional and technical quality	11	Each AI agent performs its phase-specific instructional function consistently.
	12	The integrated AI system provides stable and timely responses during learning activities.
	13	The embedded guardrails effectively prevent direct-answer generation and pedagogically inappropriate responses.
	14	The system accurately records interaction logs, scaffold use, and students' reasoning-state transitions.
	15	The AI-generated responses are contextually relevant and scientifically accurate within the physics content.
Usability and implementation feasibility	16	The interface, navigation, and operating instructions are clear and easy to use.
	17	The system can be operated reliably under actual classroom and e-laboratory conditions.
	18	Teacher-facing analytics provide interpretable information for monitoring students' reasoning development.
	19	The system supports teacher decision-making without replacing professional pedagogical judgment.
	20	The overall system is feasible for supporting guided inquiry and critical-thinking development in secondary physics education.

Analyzing of Data

Data analysis proceeded in three stages. First, the content validity of the developed system was evaluated using Aiken's V coefficient for each validation item. Second, descriptive statistics were used to summarize students' critical-thinking

scores before and after the intervention [71]. Third, the effectiveness of the Socratic AI-based adaptive scaffolding intervention was examined using one-way ANCOVA, with posttest critical-thinking scores as the dependent variable and pretest scores as the covariate to control for baseline differences and improve statistical precision [72]. Before conducting ANCOVA, the assumptions of residual normality, homogeneity of variance, linearity, homogeneity of regression slopes, and independence of the covariate from treatment were systematically assessed. Residual normality was examined using the Shapiro-Wilk test [73], whereas homogeneity of variance was evaluated using Levene's test [74]. Statistical significance was set at $\alpha = 0.05$, and all analyses were performed using SPSS statistical software.

3. Findings/Results

Socratic AI-Based Development and Validation

The development process began with the Analysis phase, which examined students' learning needs through diagnostic assessments and interviews with physics teachers. The findings indicated that, although many students could perform procedural problem-solving steps, they continued to experience difficulties in self-regulation, evidence evaluation, and analytical reasoning. Students frequently applied formulas mechanically without explicitly identifying assumptions, justifying conceptual relationships, or evaluating the validity of their solutions. Teachers also reported challenges in supporting heterogeneous classrooms characterized by differences in prior knowledge, learning readiness, and cognitive processing. These findings suggested that students' limited critical-thinking performance was associated not only with conceptual difficulties but also with insufficiently differentiated cognitive support. The analysis therefore established the need for adaptive Socratic scaffolding capable of identifying students' reasoning states and calibrating instructional support accordingly.

Building on the needs-analysis findings, the Design phase produced an integrated Socratic AI architecture that combines reasoning diagnosis, adaptive scaffold selection, prompt orchestration, and pedagogically controlled LLM processing. As illustrated in Figure 8a, student input is routed through the frontend and API gateway to a diagnostic engine that classifies reasoning into fragmented, procedural, conceptual, or reflective states. The orchestration server then selects context-appropriate prompt templates and Socratic scaffolds from the available repositories. The LLM alignment layer constrains response generation through Socratic role conditioning, response filtering, curriculum alignment, and reasoning-state controls. Rather than providing direct solutions, the system generates non-leading prompts that encourage clarification, evidence-based justification, conceptual connection, and reflection. Interaction data and reasoning metrics are subsequently returned to the analytics layer to refine future scaffold selection.

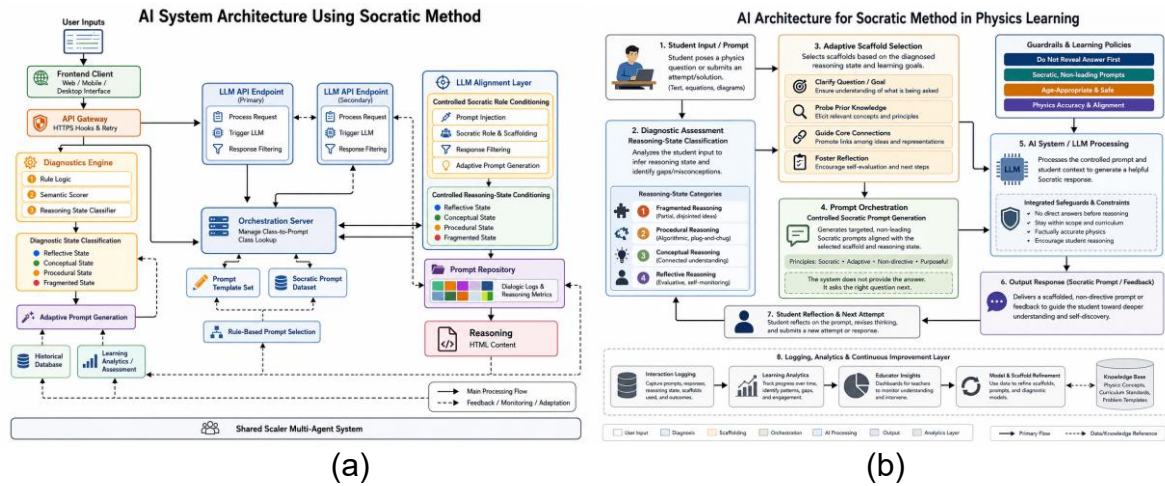


Fig. 8. Integrated Socratic AI architecture for adaptive scaffolding in physics learning: (a) system-level architecture and data flow; (b) pedagogical interaction cycle and continuous improvement process.

Figure 8b illustrates the corresponding pedagogical interaction cycle. Student responses are diagnosed, matched with an appropriate scaffold, transformed into controlled Socratic prompts, and processed by the LLM under predefined guardrails. The resulting feedback guides students to reconsider assumptions, evaluate evidence, revise explanations, and submit improved responses. This iterative cycle supports progressively calibrated assistance while preserving productive cognitive effort and learner autonomy. Within the guided-inquiry E-Laboratory, dialogue logs, scaffold use, and reasoning development are continuously recorded to support adaptation and teacher monitoring. Collectively, the two diagrams show how the technical architecture and instructional workflow operate together to support structured epistemic engagement and critical-thinking development in physics learning. Following the completion of the architectural design, the system was operationalized into a structured workflow, as illustrated in the Socratic AI flowchart in figure 9.

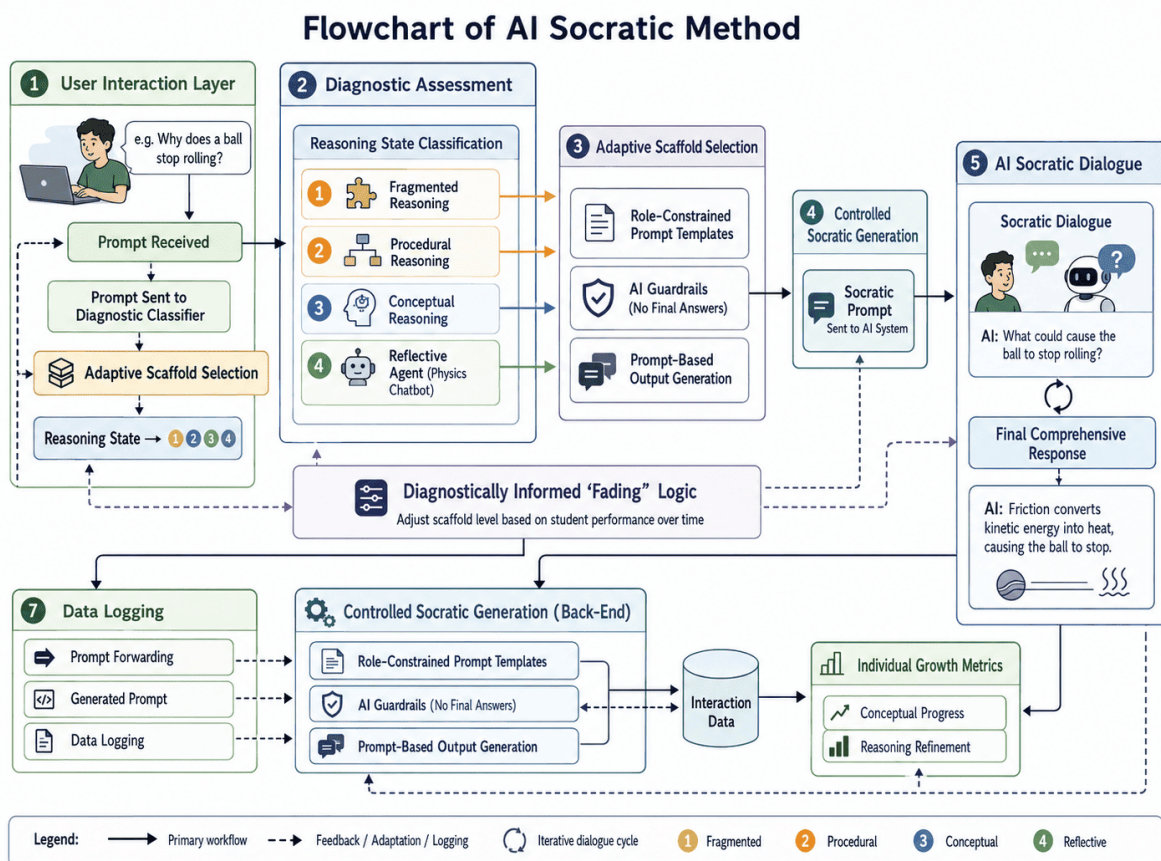


Fig. 9. Flowchart of the Socratic AI method for fostering students' critical thinking in physics learning.

Figure 9 illustrates the operational workflow of the Socratic AI system. The process begins when a student submits a question, explanation, or revised response through the E-Laboratory interface. The input is routed to the diagnostic assessment layer, where rule-based and semantic analyses classify the student's reasoning as fragmented, procedural, conceptual, or reflective. Based on this diagnosis, the adaptive scaffold-selection mechanism determines an appropriate role-constrained prompt template while applying AI guardrails that prevent premature answer provision.

The system subsequently generates a controlled, non-leading Socratic prompt that encourages students to clarify assumptions, justify evidence, examine conceptual relationships, or reconsider their conclusions. Student responses initiate an iterative diagnostic-dialogic cycle in which the reasoning state and required level of support are continuously reassessed. A diagnostically informed fading mechanism progressively reduces scaffold intensity as students demonstrate greater conceptual understanding and reasoning autonomy.

Interaction data, generated prompts, and reasoning trajectories are recorded within the analytics layer to support scaffold refinement and teacher monitoring. A comprehensive explanation is provided only after the dialogic process has sufficiently supported students' reasoning. Within the broader instructional framework, this workflow is embedded in guided inquiry, whose stages of questioning, hypothesis formulation, experimentation, evidence analysis, and conclusion construction provide a coherent context for developing critical thinking in physics.

To operationalize this diagnostic dialogic workflow across the guided inquiry sequence, four specialized AI agents were developed, each serving a distinct pedagogical function. The AI Case Study Guide supports problem orientation and critical questioning by presenting contextualized physics cases and prompting students to identify problems, variables, and initial assumptions. The AI Lab Partner operates during experimental planning and investigation, supporting hypothesis formulation, variable identification, procedural reasoning, and evidence collection. The AI Discussion and Concept Scaffolder is activated during data interpretation, discussion, and conclusion construction to examine the consistency between evidence and claims, evaluate alternative explanations, and facilitate conceptual integration. In addition, the AI Physics Chatbot provides bounded cognitive support across all inquiry phases when students require clarification or encounter cognitive difficulty.

This phase-specific orchestration ensures that adaptive scaffolding is not delivered as generic or randomly generated assistance, but is aligned with the epistemic demands of each inquiry task. Collectively, the four agents translate the system-level workflow illustrated in Figure 9 into a context-sensitive multi-agent learning environment embedded within the guided inquiry E-Laboratory. The resulting integration is presented in Figure 10.

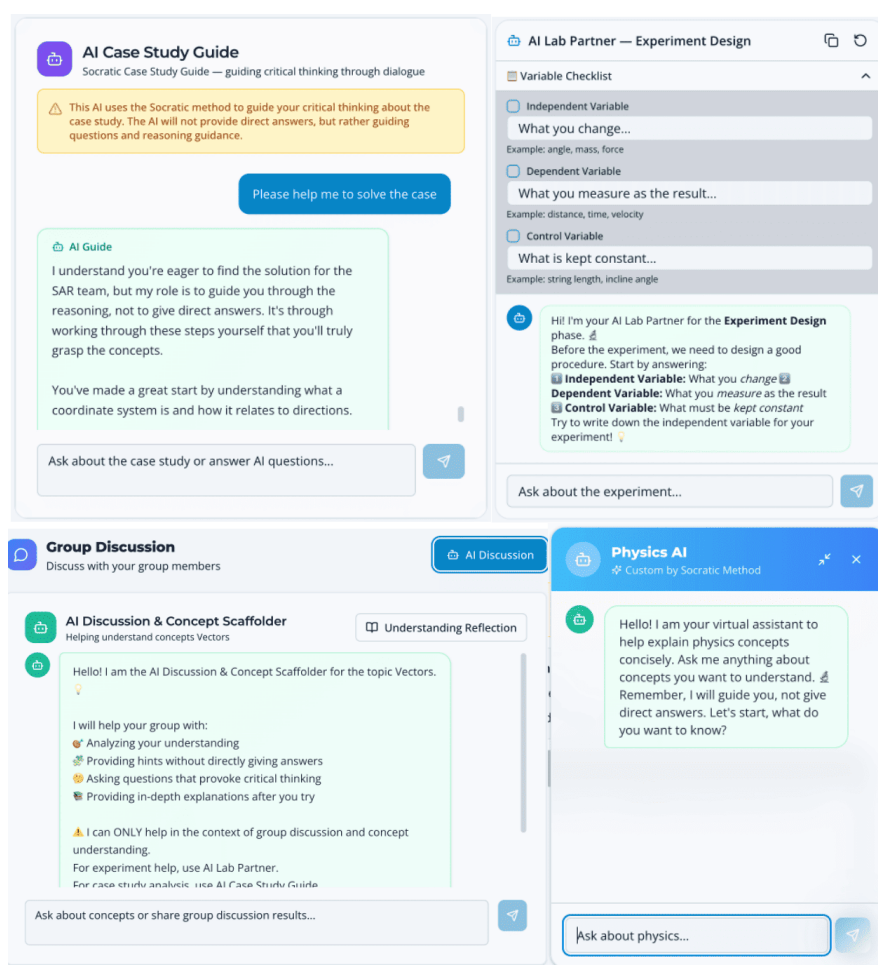


Fig. 10. Phase-aligned integration of four Socratic AI agents within the guided inquiry E-Laboratory.

Each AI agent performs a distinct yet complementary function in supporting students' critical thinking. The AI Case Study Guide operates during problem

orientation and critical questioning by presenting contextualized physics scenarios and prompting students to formulate investigable questions, identify relevant variables, and articulate initial assumptions. The AI Lab Partner supports experimental planning and investigation by guiding hypothesis construction, variable selection, outcome prediction, and evidence interpretation through Socratic prompts that examine causal reasoning and evidential adequacy.

The AI Discussion and Concept Scaffold is activated during data interpretation, discussion, and conclusion construction. It facilitates reflective dialogue by examining the consistency between evidence and claims, encouraging consideration of alternative explanations, and supporting the synthesis of physics principles. The AI Physics Chatbot, available across all inquiry phases, provides bounded clarification when students encounter conceptual difficulty or cognitive overload. Its responses are deliberately limited to preserve productive struggle and learner autonomy. Collectively, these agents constitute a phase-aligned, multi-agent Socratic scaffolding system that provides context-sensitive cognitive support throughout physics inquiry.

Before classroom implementation, the developed system was evaluated by experts in physics education and educational technology. The 20-item validation instrument covered four dimensions: Socratic pedagogical alignment, adaptive scaffolding and critical-thinking support, AI functional and technical quality, and usability and implementation feasibility. Each item was rated on a four-point scale and analyzed using Aiken's V. The item-level results are presented in Table 3.

Table 3. Aiken's V of Media Validation results

Item	Validator 1	Validator 2	$\Sigma s (r-1)$	Aiken's V	Interpretation
1	4	4	6	1.00	Very High
2	3	4	5	0.83	High
3	4	3	5	0.83	High
4	4	4	6	1.00	Very High
5	3	4	5	0.83	High
6	4	3	5	0.83	High
7	3	4	5	0.83	High
8	4	4	6	1.00	Very High
9	3	4	5	0.83	High
10	4	4	6	1.00	Very High
11	4	3	5	0.83	High
12	3	4	5	0.83	High
13	4	4	6	1.00	Very High
14	3	4	5	0.83	High
15	4	4	6	1.00	Very High
16	3	3	4	0.67	Moderate
17	4	4	6	1.00	Very High
18	3	4	5	0.83	High
19	4	4	6	1.00	Very High
20	4	4	6	1.00	Very High

Table 3 shows that the item-level Aiken's V coefficients ranged from 0.67 to 1.00, with an overall coefficient of 0.90, indicating strong content validity of the Socratic AI-based adaptive scaffolding system. The high level of agreement between validators provides evidence that the system was pedagogically coherent and appropriately aligned with Socratic questioning principles, adaptive scaffolding

mechanisms, critical-thinking indicators, and the intended physics-learning context. Nineteen items met the predetermined acceptance criterion, whereas Item 16 required minor revision to improve interface clarity and usability before classroom implementation.

Intervention Effectiveness of Socratic AI-Based on Critical Thinking

Descriptive statistics were first examined to summarize students' critical-thinking performance before and after the intervention. Table 4 presents the mean scores and standard deviations for the experimental and control groups at pretest and posttest.

Table 4. Descriptive Statistics Results

Group	N	Pretest Mean	Pretest SD	Posttest Mean	Posttest SD
Experimental	69	60.87	2.45	83.94	3.12
Control	66	60.23	2.31	72.96	2.18

As shown in Table 4, the mean pretest difference between groups was only 0.64 points, suggesting descriptively comparable baseline critical-thinking performance. Following the seven-session intervention, the experimental group gained 23.07 points (37.9%), compared with 12.73 points (21.1%) in the control group. The experimental group therefore achieved a 10.34-point greater gain, while the between-group difference increased to 10.98 points at posttest. These descriptive results indicate greater improvement under the experimental condition. Figure 11 further presents the student-level pretest and posttest score distributions across both groups.

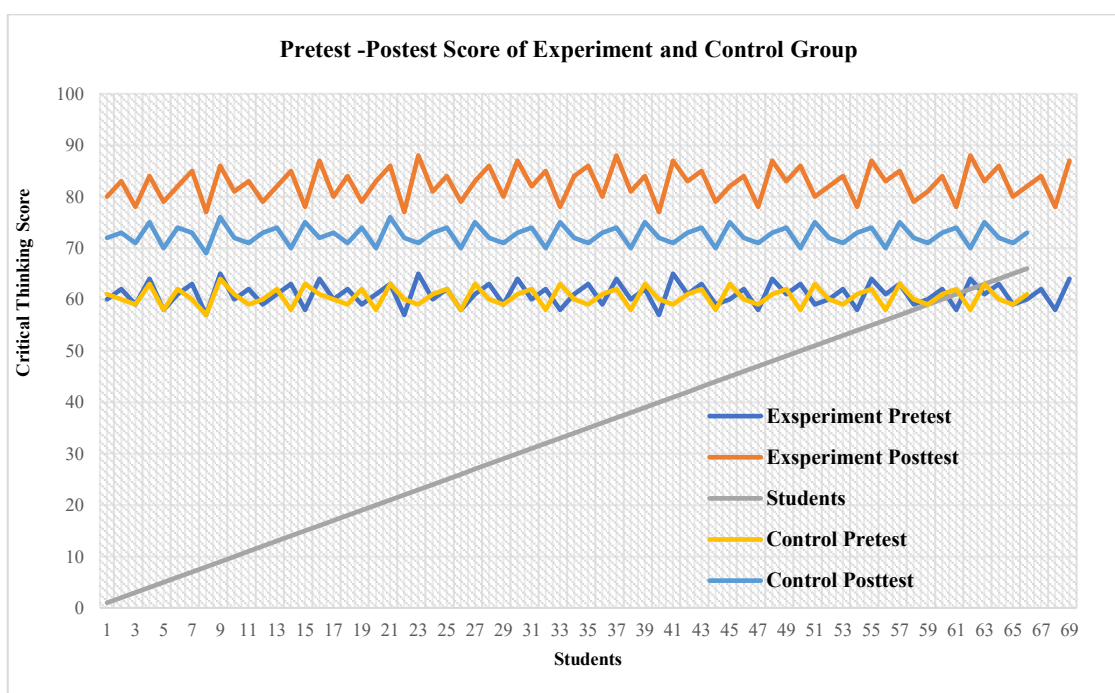


Fig. 11. Student-level pretest and posttest critical-thinking scores across the experimental and control groups.

Figure 11 shows that the experimental-group posttest scores were consistently higher than the experimental pretest and control-group scores, in line with the larger mean gain reported in Table 4. Because the figure provides descriptive evidence and the connecting lines represent sequentially ordered observations rather than longitudinal trajectories, the statistical significance of the treatment effect was subsequently examined using one-way ANCOVA after the required assumptions were verified in Table 5.

Table 5. One-Way ANCOVA Assumption Testing Summary

Assumption	Statistical Test	Sig.	Decision Rule	Interpretation
Normality of Residuals	Shapiro-Wilk	0.084	Sig. > 0.05	Residuals normally distributed
Homogeneity of Variance	Levene's Test	0.611	Sig. > 0.05	Variances homogeneous
Linearity (Covariate–DV)	Pearson Correlation	0.001	Sig. < 0.05	Significant linear relationship
Homogeneity of Regression Slopes	Interaction (Group × Pretest)	0.289	Sig. > 0.05	Slopes are parallel
Independence of Covariate	Independent t-test (Pretest)	0.480	Sig. > 0.05	Covariate independent of treatment

Based on Table 5, all statistical prerequisites for ANCOVA were satisfied. Since all assumptions were satisfied, the data met the requirements for proceeding with the one-way ANCOVA analysis, the results of which are presented in Table 6.

Table 6. One Way ANCOVA Main Effect Summary

Source	Type III SS	df	MS	F	Sig.	Partial η^2
Pretest (Covariate)	1324.882	1	1324.882	36.412	.000	.215
Group	2896.721	1	2896.721	79.628	.000	.374
Error	4729.922	132	35.833			

After controlling for pretest scores, the instructional group had a statistically significant effect on posttest critical-thinking performance, $F(1, 132) = 79.63$, $p < .001$, partial $\eta^2 = .374$. The effect size indicates that group membership accounted for 37.4% of the adjusted variance in posttest scores. Pretest performance was also a significant covariate, $F(1, 132) = 36.41$, $p < .001$, partial $\eta^2 = .215$. The estimated marginal means are reported in Table 7 to clarify the direction and magnitude of the adjusted group difference.

Table 7. Adjusted Means (Estimated Marginal Means) of One-Way ANCOVA

Group	Adjusted Mean	Std. Error	95% CI Lower	95% CI Upper
Experimental	83.51	0.72	82.09	84.93
Control	73.38	0.74	71.92	74.84

After adjustment for baseline performance, the experimental group achieved a significantly higher estimated posttest mean than the control group, with an adjusted difference of 10.13 points. The narrow confidence intervals indicate precise group estimates and are consistent with the significant treatment effect identified by

ANCOVA. Overall, the findings provide strong evidence that Socratic AI-based adaptive scaffolding was associated with substantially greater critical-thinking performance than guided inquiry using the conventional E-Laboratory. These findings warrant further examination of the pedagogical and cognitive mechanisms through which adaptive Socratic dialogue supported students' reasoning development..

4. Discussion

The first finding concerns the expert-judged content validity of the Socratic AI-based adaptive scaffolding system. The 20-item instrument yielded an overall Aiken's V coefficient of 0.90, with item-level values ranging from 0.67 to 1.00. These results provide evidence that the system adequately represented the intended pedagogical and technological constructs, particularly Socratic questioning, adaptive cognitive support, critical-thinking facilitation, and AI functionality. However, the moderate coefficient obtained for one usability-related item indicates that interface clarity and operational guidance required refinement before classroom implementation. Thus, the validation findings support the content representativeness and pedagogical coherence of the system rather than independently establishing its technical reliability or instructional effectiveness. This interpretation is consistent with recent educational AI research emphasizing scientific accuracy, logical transparency, error control, actionable feedback, and user-centered functionality as essential quality criteria [70].

The principal empirical finding was that students who received Socratic AI-based adaptive scaffolding achieved substantially higher adjusted critical-thinking scores than those who experienced guided inquiry through the conventional E-Laboratory. The adjusted difference of 10.13 points and the large group effect indicate that the experimental condition was associated with meaningful improvement beyond the gains observed under guided inquiry alone. This finding is consistent with evidence that LLM-based Socratic agents can strengthen reflective and critical-reflective thinking by encouraging learners to explain, reconsider, and refine their reasoning rather than passively receiving answers [38]. Similar improvements have been reported when Socratic dialogue is integrated with structured problem-solving support and when learners engage actively and critically with AI-generated information [75], [76], [77]. The present study extends these findings to secondary physics education by embedding adaptive Socratic interaction within a guided-inquiry E-Laboratory.

The observed effect may be explained by the epistemic design of the intervention. Unlike systems that primarily provide corrective feedback or complete solutions, the developed model constrained AI responses within a structured Socratic framework. Students were prompted to clarify assumptions, justify claims, examine evidence, identify inconsistencies, and consider alternative explanations. Such prompts may have maintained the cognitive engagement required for critical thinking while preventing premature task completion. This interpretation accords with recent design principles positioning generative AI as a provisional thinking partner that should preserve cognitive friction, metacognitive regulation, and learner epistemic agency rather than replace reasoning [56]. It is also consistent with evidence that instructional scaffolding should be calibrated to learners' prior competence, followed by structured practice and increasing opportunities for independent problem solving [78].

Adaptive support may have further contributed by matching prompt intensity to students' diagnosed reasoning states. Learners demonstrating fragmented or procedural reasoning received more explicit prompts, whereas those showing

conceptual or reflective reasoning encountered less directive support. The progressive reduction of assistance was intended to preserve productive struggle and promote reasoning autonomy. This interpretation aligns with research showing that personalized, explainable feedback can support students' monitoring, reflection, and self-regulation when recommendations are relevant and actionable [69]. Nevertheless, because scaffold trajectories were not tested as mediators of the treatment effect, their explanatory role should be regarded as theoretically plausible rather than empirically confirmed.

Guided inquiry provided the disciplinary structure within which the adaptive dialogue operated. Its sequence of problem formulation, hypothesis construction, experimentation, evidence analysis, and conclusion development required students to engage in the epistemic practices central to physics. Contemporary reviews emphasize that effective inquiry is neither minimally guided exploration nor continuous direct instruction; rather, it combines student investigation with appropriately timed instructional support [79]. Within this structure, the four AI agents delivered differentiated assistance at specific inquiry phases instead of generating generic responses. This phase-aligned design is also supported by emerging evidence that multi-agent GenAI environments can provide multidimensional assistance, personalized assessment, and knowledge elaboration more effectively than a single general-purpose chatbot [80].

The findings also highlight the importance of maintaining pedagogical control over AI assistance. Guardrails that restricted direct-answer generation and fading mechanisms that reduced support were designed to limit dependency and preserve learner agency. This feature is particularly relevant because excessive AI dependence has been associated with cognitive fatigue and lower critical-thinking performance [81]. However, the present study did not directly measure AI dependency, epistemic agency, or the mediating effects of Socratic dialogue. Moreover, the use of intact classes from two schools and a seven-session intervention limits claims regarding long-term transfer and broader generalizability. Future studies should examine dialogue logs, reasoning-state transitions, and scaffold-fading trajectories through longitudinal and multilevel analyses.

Overall, the findings suggest that the educational contribution of AI depends less on its ability to generate explanations than on the pedagogical constraints governing when, why, and how those explanations are provided. By integrating diagnostic reasoning classification, adaptive Socratic prompting, guided inquiry, and progressively reduced support, the proposed framework positioned AI as a regulated dialogic scaffold rather than an answer-generating substitute. This integration offers a theoretically grounded approach to strengthening critical thinking while maintaining productive cognitive effort and teacher-directed pedagogical oversight.

5. Conclusion

This study developed, content-validated, and evaluated the effectiveness of a Socratic AI-based adaptive scaffolding framework embedded within a guided inquiry E-Laboratory for secondary physics education. The framework integrated diagnostic reasoning-state classification, adaptive scaffold selection, controlled Socratic prompting, and phase-aligned AI agents to support critical-thinking processes rather than direct-answer acquisition. Expert evaluation yielded an overall Aiken's V coefficient of 0.90, providing evidence of strong content validity and pedagogical coherence.

The quasi-experimental findings showed that students who received Socratic AI-supported guided inquiry achieved significantly higher adjusted posttest critical-thinking scores than those who used the conventional E-Laboratory, $F(1,132) = 79.63$, $p < .001$, partial $\eta^2 = .374$. The adjusted mean difference of 10.13 points indicates a substantial advantage associated with the experimental condition. These findings suggest that the educational value of AI derives not merely from its technological capabilities, but from the pedagogical constraints governing how support is diagnosed, generated, and progressively reduced.

The study contributes theoretically by connecting Socratic pedagogy, adaptive scaffolding, guided inquiry, and multi-agent AI within a coherent instructional architecture. Practically, it offers a structured approach to integrating AI into physics learning while preserving productive cognitive engagement, learner autonomy, and teacher oversight. Nevertheless, the findings are limited to two schools, intact classroom groups, and a seven-session intervention. Further investigation across broader contexts and longer implementation periods is therefore required to establish the framework's generalizability and sustained educational value

6. Recommendations

Future studies should evaluate the framework through longitudinal, multisite, and cross-disciplinary designs involving more diverse learners and instructional contexts. Particular attention should be given to the retention and transfer of critical-thinking gains, reasoning-state transitions, scaffold-fading trajectories, dialogue quality, and the relative contribution of each AI agent.

For practitioners, Socratic AI should be implemented as a regulated instructional scaffold rather than an unrestricted answer-generation tool. Teachers should align prompts with inquiry objectives, verify scientific accuracy, monitor student interactions, and determine when support should be increased, modified, or withdrawn. Implementation should also be preceded by teacher preparation, infrastructure assessment, small-scale piloting, and clear policies concerning privacy, ethical use, learner dependency, and human oversight..

7. Limitations

The findings should be interpreted in light of the platform's infrastructure requirements. Full functionality depends on a stable internet connection, a modern HTML5/WebGL-compatible browser, and devices capable of running real-time AI and PhET simulations. Optimal implementation requires a laptop or desktop with at least 4 GB RAM, preferably 8 GB or a tablet with at least 4 GB RAM. Although smartphones with approximately 3 GB RAM and a 6.5-inch Full HD+ display can access the platform, their smaller screens reduce the visibility and usability of simulations, data analysis, and multi-component inquiry activities; consequently, laptops, desktops, or larger tablets remain preferable. Network testing identified 3.7 Mbps as the lower access threshold, while stable simulation and AI interaction generally require at least 5 Mbps and perform optimally at 10-25 Mbps. Because the current platform does not support offline operation, its applicability remains limited in schools with weak, unstable, or unavailable internet access, potentially constraining implementation and generalizability in digitally underserved regions.

8. Ethics Approval and Consent to Participate.

This study received institutional ethical approval and research clearance through the Research Ethics Committee of Universitas Negeri Malang (Komisi Etik Penelitian Universitas Negeri Malang). Because the study involved an educational experiment evaluating the effects of a developed instructional medium on students' learning outcomes, the Committee mandated the Faculty of Education, Universitas Negeri Malang (Fakultas Ilmu Pendidikan Universitas Negeri Malang), to issue site-specific research permission letters.

Research permission for SMA Negeri 1 Kolaka was granted under Letter No. 17.3.1/UN32.1.1/KM/2026, and research permission for SMA Negeri 1 Watubangga was granted under Letter No. 17.3.2/UN32.1.1/KM/2026. Both letters were issued on 17 March 2026 and signed by Dr. Sri Wahyuni, M.Pd., in her capacity as Vice Dean I (Wakil Dekan I) of the Faculty of Education, Universitas Negeri Malang. All procedures involving human participants were conducted in accordance with the relevant institutional ethical guidelines and regulations of Universitas Negeri Malang and the principles of the Declaration of Helsinki. Permission to conduct the study was also obtained from the administrations of the participating secondary schools before data collection.

Consent to participate: Informed consent was obtained from all participating students and their parents or legal guardians before their participation in the study. Participation was voluntary, and the anonymity and confidentiality of all participants were maintained throughout the research process.

Consent for publication: Not applicable. The manuscript does not contain any identifiable personal information, images, or other participant data requiring consent for publication.

9. Funding Declaration

This research received financial support from the Indonesia Endowment Fund for Education (Lembaga Pengelola Dana Pendidikan-LPDP), Ministry of Finance of the Republic of Indonesia. The funding was provided solely for the initial data collection and analysis conducted as part of the preliminary study.

10. Clinical Trial Number

The authors declare that not applicable

11. Data Availability

The tabulated datasets supporting the conclusions of this article are available in the Mendeley Data repository (<https://doi.org/10.17632/d44w35njms.1>). The raw data underlying this study are not publicly available to protect the integrity of an ongoing dissertation research, but will be shared on reasonable request to the corresponding author.

12. Acknowledgements

I would like to express my sincere gratitude to the Lembaga Pengelola Dana Pendidikan of the Republic of Indonesia (LPDP RI) for providing financial support for this research and for funding participation in the international conference where this work was presented. This study would not have been possible without the generous support and commitment of LPDP to advancing education and research excellence.

13. Conflict of Interest

The authors declare that they have no conflict of interest.

14. Authorship Contribution Statement

Saputra: contributed to the conceptualization and design of the study, system development, data acquisition, data curation, formal and statistical analysis, interpretation of the findings, provision of technical and material support, project administration, and preparation of the original manuscript draft. Setyosari: conceptual development of the study, research design, theoretical framework, methodological supervision, interpretation of the findings, and critical revision of the manuscript for important intellectual content. Praherdhiono: development and refinement of the instructional design, research methodology, interpretation of the findings, and critical review of the manuscript. Degeng: contributed to the development of the pedagogical framework, interpretation of the data and research findings, methodological refinement, and critical revision of the manuscript.

15. Generative AI Statement

As the authors of this work, we used Scopus AI to identify relevant Scopus-indexed literature and retrieve bibliographic metadata for VOSviewer-based mapping of the state of the art and research novelty. We also used GPT-based generative AI (GPT-4) to create conceptual visual analogies of the developed prototype and Socratic AI workflow, thereby presenting complex system structures more concisely. All AI-assisted outputs were critically reviewed, verified, and revised by the authors, who take full responsibility for the accuracy and integrity of the final manuscript.

References

- [1] A. Bray and J. Williams, "Why is physics hard? Unpacking students' perceptions of physics," *J. Phys. Conf. Ser.*, vol. 1512, no. 1, p. 012002, Apr. 2020, doi: 10.1088/1742-6596/1512/1/012002.
- [2] M. Ma'ruf, A. Setiawan, A. Suhandi, and P. Siahaan, "Investigation of Student Difficulties in Basic Physics Lectures and Readiness to Implement Physics Problem Solving Assisted by Interactive Multimedia Android in Indonesia," *European Online Journal of Natural and Social Sciences*, vol. 9, no. 4, pp. 820–827, 2020, Accessed: Apr. 30, 2025. [Online]. Available: <https://european-science.com/eojnss/article/view/6113>
- [3] L. D. English and T. Lehmann, *Ways of thinking in STEM-based problem solving: Teaching and learning in a new era*. Taylor and Francis, 2024. doi: 10.4324/9781003404989.
- [4] J. G. Tupouniua, "What challenges emerge when students engage with algorithmatizing tasks?," *Journal of Pedagogical Research*, vol. 7, no. 2, pp. 93–107, 2023, doi: 10.33902/JPR.202318518.
- [5] Z. Han, W. W. M. So, and Z. H. Wan, "Challenges in assessing students' STEM-based problem solving," in *Ways of Thinking in STEM-based Problem Solving: Teaching and Learning in a New Era*, Taylor and Francis, 2024, pp. 250–262. doi: 10.4324/9781003404989-18.

- [6] S. Ardianti, D. Sulisworo, Y. Pramudya, and W. Raharjo, "The impact of the use of STEM education approach on the blended learning to improve student's critical thinking skills," *Universal Journal of Educational Research*, vol. 8, no. 3 B, pp. 24–32, 2020, doi: 10.13189/ujer.2020.081503.
- [7] D. A. Nur'aini, S. Kusairi, and M. Ali, "Impact of The STEM approach with formative assessment in PjBL on students' critical thinking skills," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, 2022. doi: 10.1088/1742-6596/2165/1/012044.
- [8] S. Vincent-Lancrin, "Critical thinking," in *Elgar Encyclopedia of Interdisciplinarity and Transdisciplinarity*, Edward Elgar Publishing Ltd., 2024, pp. 124–128. doi: 10.4337/9781035317967.ch27.
- [9] D. Hitchcock, "Critical Thinking as an Educational Ideal," in *Argumentation Library*, vol. 30, Springer Nature, 2017, pp. 477–497. doi: 10.1007/978-3-319-53562-3_30.
- [10] B. Mitrevski, "Teaching critical thinking and problem solving in physics," in *AIP Conference Proceedings*, T. M. Mishonov and A. M. Varonov, Eds., American Institute of Physics Inc., 2019. doi: 10.1063/1.5091398.
- [11] Y. Yusal, A. Suhandi, W. Setiawan, and I. Kaniawati, "Profile of pre-service physics teachers' critical thinking skills related to heat transfer," in *Journal of Physics: Conference Series*, A. G. Abdullah, M. L. Firdaus, I. Widiaty, A. B. D. Nandiyanto, A. A. Danuwijaya, and C. U. Abdullah, Eds., IOP Publishing Ltd, 2021. doi: 10.1088/1742-6596/1731/1/012075.
- [12] L. N. R. Marfu'i, "The analysis of critical thinking skills test in social-problems for physics education students with Rasch Model," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, 2019. doi: 10.1088/1742-6596/1280/5/052012.
- [13] S. B. A. Neswary, B. K. Prahani, F. C. Wibowo, and R. F. R. Uulaa, "A Profile of senior high school student's critical thinking skills in physics learning," in *Journal of Physics: Conference Series*, R. A. Firdaus, M. Satriawan, F. null, M. Khoiro, A. Realita, and U. A. Deta, Eds., Institute of Physics, 2023. doi: 10.1088/1742-6596/2623/1/012012.
- [14] A. Lumbu, I. S. Budiarti, and I. N. Asyiah, "Can higher order thinking skills to improve Physicstudents' critical thinking in the era of education 5.0?: A literature review," *Multidisciplinary Reviews*, vol. 9, no. 1, 2026, doi: 10.31893/multirev.2026032.
- [15] E. Istiyono, "Measurement of high school students' physics critical thinking skills," *Multidisciplinary Science Journal*, vol. 8, no. 6, 2026, doi: 10.31893/multiscience.2026408.
- [16] N. D. Rosyidah, S. Kusairi, A. Taufiq, and Y. Affriyenni, "Profile of students' critical thinking processes on the topics of Hydrostatic Pressure and Archimedes' principle," in *Journal of Physics: Conference Series*, S. Saputro, S. null, S. null,

D. A. Nugraha, and R. Perdana, Eds., Institute of Physics Publishing, 2020. doi: 10.1088/1742-6596/1511/1/012081.

- [17] L. Viennot, "Conceptual development and critical attitude in physics education: A pathway in the search for coherence," in *Upgrading Physics Education to Meet the Needs of Society*, Springer International Publishing, 2019, pp. 189–198. doi: 10.1007/978-3-319-96163-7_13.
- [18] P. Ari and M. A. H. Bunyamin, "Investigation of critical thinking skills survey research on static fluid topics," in *AIP Conference Proceedings*, N. Ishartono and M. Waluyo, Eds., American Institute of Physics, 2025. doi: 10.1063/5.0290495.
- [19] B. Jatmiko, N. R. Rahman, B. K. Prahani, F. Hasyim, H. V Saphira, and M. A. Kurtuluş, "Research analysis of critical thinking skills in physics learning in higher education," *Multidisciplinary Reviews*, vol. 9, no. 5, 2026, doi: 10.31893/multirev.2026215.
- [20] N. Nabilah and S. Suyanto, "A CIPP-based evaluation of critical thinking in science: Implementation of the Merdeka curriculum in Mataram public junior high schools," *STEM Education*, vol. 5, no. 4, pp. 587–616, 2025, doi: 10.3934/steme.2025028.
- [21] D. Fitria, A. Asrizal, and L. Lufri, "Enhancing 21st-Century Skills through Blended Problem-Based Learning with Ethnoscience Integration: A Mixed-Methods Study in Indonesian Junior High Schools," *International Journal of Learning, Teaching and Educational Research*, vol. 24, no. 1, pp. 464–480, 2025, doi: 10.26803/ijlter.24.1.23.
- [22] M. B. Panjaitan *et al.*, "THE ROLE OF EMOTIONAL REGULATION IN SHAPING STUDENTS' PERCEPTION OF PHYSICS PROBLEM-SOLVING TASKS," *TPM Test. Psychom. Methodol. Appl. Psychol.*, vol. 32, no. S2, pp. 1–10, 2025, [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-105013750538&partnerID=40&md5=8e73133672381a05437fb749ea2640d1>
- [23] A. Suparman, A. Samsudin, I. Kaniawati, and V. Lantik, "Analysis of students' critical thinking skill on optical instruments through worksheet-based inductive learning model," in *AIP Conference Proceedings*, R. Rosjanuardi, S. M. Gozali, A. Jupri, A. B. D. Nandiyanto, A. Samsudin, and L. S. Riza, Eds., American Institute of Physics Inc., 2023. doi: 10.1063/5.0155392.
- [24] N. Hermita, J. A. Alim, Z. H. Putra, D. Nasien, and H. Wijoyo, "Developing STEM autonomous learning city map application to improve critical thinking skills of primary school teacher education students," *Perspektiv Nauki i Obrazovanja*, vol. 64, no. 4, pp. 675–690, 2023, doi: 10.32744/pse.2023.4.41.
- [25] V. Serevina, D. Maulana, H. Triana, and R. Nainggolan, "Implementation of Problem Based Learning (PBL) Model on Renewable Energy Materials to Improve Students' Critical Thinking Skills," in *Journal of Physics: Conference Series*, R. Fahdiran, H. Suhendar, T. B. Prayitno, U. null, H. Nasbey, W. Indrasari, and W. Fitriani, Eds., Institute of Physics, 2024. doi: 10.1088/1742-6596/2866/1/012110.

- [26] E. Idrizi, "Exploring the Role of Explainable Artificial Intelligence(XAI) in Adaptive learning systems," in *ACM International Conference Proceeding Series*, Association for Computing Machinery, 2024, pp. 100–105. doi: 10.1145/3660853.3660877.
- [27] A. Asrifan, K. Khristianto, A. Budiman, P. I. Astuti, and A. Rossydi, "Enhancing Critical Thinking: The Role of AI in Modern Education," in *Enhancing Classroom Instruction and Student Skills With AI*, IGI Global, 2025, pp. 413–447. [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-105021761356&partnerID=40&md5=1e5ff2568bdba94eee6458e86e88c55e>
- [28] Y. Walter, "Embracing the future of Artificial Intelligence in the classroom: the relevance of AI literacy, prompt engineering, and critical thinking in modern education," *International Journal of Educational Technology in Higher Education*, vol. 21, no. 1, 2024, doi: 10.1186/s41239-024-00448-3.
- [29] L. Stuchlikova and M. Weis, "From Information to Insight: Reimagining Critical Thinking Pedagogy in the Age of Artificial Intelligence," in *ICETA 2024 - 22nd Year of International Conference on Emerging eLearning Technologies and Applications, Proceedings*, S. Fejedelem, Ed., Institute of Electrical and Electronics Engineers Inc., 2024, pp. 591–598. doi: 10.1109/ICETA63795.2024.10850787.
- [30] R. Katsenou, K. Kotsidis, A. Papadopoulou, P. Anastasiadis, and I. Deliyannis, "Beyond Assistance: Embracing AI as a Collaborative Co-Agent in Education," *Educ. Sci. (Basel)*, vol. 15, no. 8, 2025, doi: 10.3390/educsci15081006.
- [31] M. A. Khan and H. Begum, "The Impact of AI-Supported Classroom Activities on Enhancing Critical Thinking and Problem-Solving Skills," *Research Journal for Social Affairs*, vol. 3, no. 6, pp. 231–233, Oct. 2025, doi: 10.71317/rjsa.003.06.0448.
- [32] S. I. Kholida MS*, Moh. B. Tamam, S. Suprianto, M. Sumo, and Y. N. Aprilita, "Inquiry-Based Physics Learning Module with Physics Education Technology Assistance to Improve High School Students' Critical Thinking Skills and Scientific Literacy," *Jurnal IPA & Pembelajaran IPA*, vol. 9, no. 3, pp. 1003–1019, Oct. 2025, doi: 10.24815/jipi.v9i3.48745.
- [33] C. T. Anselmo, A. Bete-Liban, J. C. Ollero, D. T. Bauí, J. J. S. Pascua, and Y. A. Sario, "Role of AI in Enhancing Critical Thinking in Science Education: Challenges and Opportunities for Science Instructor," *International Journal of Multidisciplinary: Applied Business and Education Research*, vol. 6, no. 8, pp. 3877–3891, Aug. 2025, doi: 10.11594/ijmaber.06.08.12.
- [34] E. Dimitriadou and A. Lanitis, "A critical evaluation, challenges, and future perspectives of using artificial intelligence and emerging technologies in smart classrooms," *Smart Learning Environments* 2023 10:1, vol. 10, no. 1, pp. 12–, Feb. 2023, doi: 10.1186/s40561-023-00231-3.
- [35] H. Ihsan and S. Riyasni, "The Validation of Artificial Intelligence Integrated E-Module with Problem Based Learning Model on Renewable Energy to High

School Students' Critical Thinking Skills," *Physics Learning and Education*, vol. 3, no. 4, Dec. 2025, doi: 10.24036/ple.v3i4.278.

- [36] Y.-R. Ho, B.-Y. Chen, and C.-M. Li, "Thinking more wisely: using the Socratic method to develop critical thinking skills amongst healthcare students," *BMC Med. Educ.*, vol. 23, no. 1, 2023, doi: 10.1186/s12909-023-04134-2.
- [37] L. Favero, J. A. Pérez-Ortiz, T. Käser, and N. Oliver, "Enhancing Critical Thinking in Education by Means of a Socratic Chatbot," in *Communications in Computer and Information Science*, F. Bellas and O. Fontenla-Romero, Eds., Springer Science and Business Media Deutschland GmbH, 2025, pp. 17–32. doi: 10.1007/978-3-031-93409-4_2.
- [38] L. Xi, Y. Zhang, and Q. Wang, "Investigating the effects of an LLM-based Socratic conversational agent on students' academic performance and reflective thinking in higher education," *Comput. Educ.*, vol. 241, 2026, doi: 10.1016/j.compedu.2025.105494.
- [39] N. Hanafi, B. A. M. Kamil, M. Ibrahim, and S. A. Hassan, "FOSTERING CRITICAL THINKING VIA SOCRATIC DIALOGUE AMONG ENTREPRENEURIAL DEVELOPMENT STUDENTS IN A MALAYSIAN PUBLIC UNIVERSITY," *Journal of Institutional Research South East Asia*, vol. 23, no. 1, pp. 253–277, 2025, [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-105010972367&partnerID=40&md5=1485fb2427668fa3b84360d43dc9011e>
- [40] S. F. Dalim, A. S. Ishak, and L. M. Hamzah, "Promoting Students' Critical Thinking through Socratic Method: Views and Challenges," *Asian Journal of University Education*, vol. 18, no. 4, pp. 1034–1047, 2022, doi: 10.24191/ajue.v18i4.20012.
- [41] B. H. Ang, S. D. Gollapalli, and S.-K. Ng, "Socratic Question Generation: A Novel Dataset, Models, and Evaluation," in *EACL 2023 - 17th Conference of the European Chapter of the Association for Computational Linguistics, Proceedings of the Conference*, Association for Computational Linguistics (ACL), 2023, pp. 147–165. doi: 10.18653/v1/2023.eacl-main.12.
- [42] I. Isrokatun, D. Anggita, B. S. Purwono, C. Sunaengsih, and A. A. Syahid, "Scaffolding in conceptual science," in *Journal of Physics: Conference Series*, A. G. Abdullah, A. B. D. Nandiyanto, I. Widiaty, A. A. Danuwijaya, and C. U. Abdullah, Eds., IOP Publishing Ltd, 2019. doi: 10.1088/1742-6596/1402/4/044079.
- [43] C. Diaferia, V. Artoni, and D. Rollo, "Scaffolding in teacher-student interaction : an operationalization for evaluation in educational and clinical settings," in *MeMeA 2018 - 2018 IEEE International Symposium on Medical Measurements and Applications, Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2018. doi: 10.1109/MeMeA.2018.8438694.

- [44] I. Hidayah, T. B. Adji, and N. A. Setiawan, "Development and evaluation of adaptive metacognitive scaffolding for algorithm-learning system," *IET Software*, vol. 13, no. 4, pp. 305–312, 2019, doi: 10.1049/iet-sen.2018.5072.
- [45] A. Munshi, G. Biswas, R. Baker, J. Ocumpaugh, S. Hutt, and L. Paquette, "Analysing adaptive scaffolds that help students develop self-regulated learning behaviours," *J. Comput. Assist. Learn.*, vol. 39, no. 2, pp. 351–368, 2023, doi: 10.1111/jcal.12761.
- [46] N. F. Jumaat and Z. Tasir, "Instructional scaffolding in online learning environment: A meta-analysis," in *Proceedings - 2014 International Conference on Teaching and Learning in Computing and Engineering, LATICE 2014*, IEEE Computer Society, 2014, pp. 74–77. doi: 10.1109/LaTiCE.2014.22.
- [47] T. J. E. Faber, M. E. W. Dankbaar, R. Kickert, W. W. van den Broek, and J. J. G. van Merriënboer, "Identifying indicators to guide adaptive scaffolding in games," *Learn. Instr.*, vol. 83, 2022, doi: 10.1016/j.learninstruc.2022.101666.
- [48] M. Zuo, S. Kong, Y. Ma, Y. Hu, and M. Xiao, "The Effects of Using Scaffolding in Online Learning: A Meta-Analysis," *Educ. Sci. (Basel)*, vol. 13, no. 7, 2023, doi: 10.3390/educsci13070705.
- [49] J. Melero, D. Hernández-Leo, and J. Blat, "A review of scaffolding approaches in game-based learning environments," in *Proceedings of the European Conference on Games-based Learning*, M. Meimaris and D. Gouscos, Eds., Dechema e.V., 2011, pp. 717–724. [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84938603560&partnerID=40&md5=ce439eb02200bc6840b733cf1765fd5d>
- [50] H. Hariyanto, X. D. K. Francisca, and M. Rizqona, "Artificial Intelligence in Adaptive Education: A Systematic Review of Techniques for Personalized Learning," *Discover Education 2025 4:1*, vol. 4, no. 1, pp. 458–, Oct. 2025, doi: 10.1007/S44217-025-00908-6.
- [51] A. Meyer, T. Bleckmann, and G. Friege, "Automatic feedback on physics tasks using open-source generative artificial intelligence," *Int. J. Sci. Educ.*, May 2025, doi: 10.1080/09500693.2025.2499220.
- [52] S. Ba, L. Yang, Z. Yan, C. K. Looi, and D. Gašević, "Unraveling the mechanisms and effectiveness of AI-assisted feedback in education: A systematic literature review," *Computers and Education Open*, vol. 9, p. 100284, Dec. 2025, doi: 10.1016/J.CAEO.2025.100284.
- [53] A. Salido *et al.*, "Integrating critical thinking and artificial intelligence in higher education: A bibliometric and systematic review of skills and strategies," *Social Sciences & Humanities Open*, vol. 12, p. 101924, Jan. 2025, doi: 10.1016/J.SSAHO.2025.101924.
- [54] L. Bao, K. Koenig, Y. Xiao, J. Fritchman, S. Zhou, and C. Chen, "Theoretical model and quantitative assessment of scientific thinking and reasoning," *Phys. Rev. Phys. Educ. Res.*, vol. 18, no. 1, p. 010115, Feb. 2022, doi: 10.1103/PhysRevPhysEducRes.18.010115.

- [55] B. Jose, J. Cherian, A. M. Verghis, S. M. Varghise, M. S, and S. Joseph, "The cognitive paradox of AI in education: between enhancement and erosion," *Front. Psychol.*, vol. 16, p. 1550621, 2025, doi: 10.3389/FPSYG.2025.1550621.
- [56] M. Vendrell and S. K. Johnston, "Scaffolding critical thinking with generative AI: Design principles for integrating large language models in higher education," *Computers and Education: Artificial Intelligence*, vol. 10, p. 100572, Jun. 2026, doi: 10.1016/J.CAEAI.2026.100572.
- [57] A. Santos-Guevara, O. Aquines-Gutiérrez, H. Martínez-Huerta, W. X. Chavarría-Garza, and J. A. Azuela, "Fostering Student Engagement and Learning Perception Through Socratic Dialogue with ChatGPT: A Case Study in Physics Education," *Education Sciences 2026*, Vol. 16, Page 184, vol. 16, no. 2, p. 184, Jan. 2026, doi: 10.3390/EDUCSCI16020184.
- [58] K. S. Tang and G. B. S. Putra, "Generative AI as a Dialogic Partner: Enhancing Multiple Perspectives, Reasoning, and Argumentation in Science Education with Customized Chatbots," *Journal of Science Education and Technology 2025* 35:1, vol. 35, no. 1, pp. 128–140, Jul. 2025, doi: 10.1007/S10956-025-10240-1.
- [59] P. Setyosari, *Metode Penelitian Pendidikan & Pengembangan*, 4th ed., vol. 6. Jakarta: PRENADAMEDIA GROUP, 2020. Accessed: Dec. 22, 2025. [Online]. Available: <https://books.google.co.id/books?id=SnADwAAQBAJ&lpg=PA1&ots=6HNCNqEQD7&dq=H.%20P.%20Setyosari%2C%20Metode%20penelitian%20pendidikan%20%26%20pengembangan.%20Prenada%20%20Media%2C%202016.%20Accessed%3A&lr&pg=PP1#v=onepage&q&f=false>
- [60] R. M. Branch, *Instructional design: The ADDIE approach*. Springer US, 2010. doi: 10.1007/978-0-387-09506-6.
- [61] M. Denny, S. Denieffe, and K. O'Sullivan, "Non-equivalent Control Group Pretest–Posttest Design in Social and Behavioral Research," *The Cambridge Handbook of Research Methods and Statistics for the Social and Behavioral Sciences*, pp. 314–332, May 2023, doi: 10.1017/9781009010054.016.
- [62] E. M. Furtak, T. Seidel, H. Iverson, and D. C. Briggs, "Experimental and Quasi-Experimental Studies of Inquiry-Based Science Teaching," *Rev. Educ. Res.*, vol. 82, no. 3, pp. 300–329, 2012, doi: 10.3102/0034654312457206.
- [63] I. Yusuf, P. Setyosari, D. Kuswandi, and S. Ulfa, "Development of Integrated Physics Learning Tools in Virtual Laboratory Platform: Its Implementation through the POGIL Strategy in Indonesian Frontier Areas," *TEM Journal*, vol. 13, no. 3, pp. 2568–2580, 2024, Accessed: Dec. 23, 2025. [Online]. Available: <https://www.ceeol.com/search/article-detail?id=1262101>
- [64] S. Sholikhan, I. N. S. Degeng, P. Setyosari, and S. K. Handayanto, "Understanding the physics concept between guided inquiry or open inquiry," *Int. J. Learn. Chang.*, vol. 12, no. 2, pp. 113–123, 2020, doi: 10.1504/IJLC.2020.106784;JOURNAL:JOURNAL:IJLC;PAGEGROUP:STRING: PUBLICATION.

- [65] S. E. Dewi, P. Setyosari, Y. Hanafi, and Supriyono, "Effect of Inquiry-based Learning on Students' Critical Thinking: A Meta-analysis.," *Journal of Institutional Research South East Asia*, vol. 23, no. 3, p. 18, Sep. 2025, Accessed: Mar. 30, 2026. [Online]. Available: <https://openurl.ebsco.com/contentitem/gcd:190651189?sid=ebsco:plink:crawler&id=ebsco:gcd:190651189>
- [66] I. G. P. E. Saputra, P. Setyosari, H. Praherdhiono, and M. D. K. Degeng, "The Framework of Smart Learning Environments Based on E-Laboratory Integrated Artificial Narrow Intelligence to Enhance the Effectiveness of Physics Learning," pp. 114–133, 2026, doi: 10.1007/978-981-95-8824-4_9.
- [67] N. R. A. Nabil, I. Wulandari, S. Yamtinah, S. R. D. Ariani, and M. Ulfa, "Aiken Index Analysis to Determine The Contents Validity of The Minimum Competency Assessment Instrument Based on Chemical Science Context," *Jurnal Penelitian Pendidikan*, vol. 25, no. 2, pp. 184–191, 2022, doi: 10.20961/paedagogia.v25i2.64566.
- [68] R. Ennis, "Critical Thinking," *Inquiry: Critical Thinking Across the Disciplines*, vol. 26, no. 1, pp. 4–18, 2011, doi: 10.5840/INQUIRYCTNEWS20112613.
- [69] M. Afzaal, A. Zia, J. Nouri, and U. Fors, "Informative Feedback and Explainable AI-Based Recommendations to Support Students' Self-regulation," *Technology, Knowledge and Learning 2023 29:1*, vol. 29, no. 1, pp. 331–354, Apr. 2023, doi: 10.1007/S10758-023-09650-0.
- [70] J. von Garrel and J. Mayer, "Which features of AI-based tools are important for students? A choice-based conjoint analysis," *Computers and Education: Artificial Intelligence*, vol. 7, p. 100311, Dec. 2024, doi: 10.1016/J.CAEAI.2024.100311.
- [71] S. Loeb, S. Dynarski, D. McFarland, P. Morris, S. Reardon, and S. Reber, "Descriptive Analysis in Education: A Guide for Researchers. NCEE 2017-4023.," *National Center for Education Evaluation and Regional Assistance*, Mar. 2017, Accessed: Dec. 22, 2025. [Online]. Available: <http://ies.ed.gov/ncee/>.
- [72] T. R. Belin and S. L. T. Normand, "The role of ANCOVA in analyzing experimental data," *Psychiatr. Ann.*, vol. 39, no. 7, pp. 753–759, Jul. 2009, doi: 10.3928/00485713-20090625-01;PAGE:STRING:ARTICLE/CHAPTER.
- [73] E. González-Estrada, J. A. Villaseñor, and R. Acosta-Pech, "Shapiro-Wilk test for multivariate skew-normality," *Comput. Stat.*, vol. 37, no. 4, pp. 1985–2001, Sep. 2022, doi: 10.1007/S00180-021-01188-Y/TABLES/6.
- [74] M. E. O'Neill and K. L. Mathews, "Levene Tests of Homogeneity of Variance for General Block and Treatment Designs," *Biometrics*, vol. 58, no. 1, pp. 216–224, Mar. 2002, doi: 10.1111/J.0006-341X.2002.00216.X.
- [75] A. Yusuf, S. Bello, N. Pervin, and A. K. Tukur, "Implementing a proposed framework for enhancing critical thinking skills in synthesizing AI-generated texts," *Think. Skills Creat.*, vol. 53, p. 101619, Sep. 2024, doi: 10.1016/J.TSC.2024.101619.

- [76] R. Suriano, A. Plebe, A. Acciai, and R. A. Fabio, "Student interaction with ChatGPT can promote complex critical thinking skills," *Learn. Instr.*, vol. 95, p. 102011, Feb. 2025, doi: 10.1016/J.LEARNINSTRUC.2024.102011.
- [77] B. C. Kuo, Z. E. Bai, and C. H. Lin, "Developing an AI learning companion for mathematics problem solving in elementary schools," *Comput. Educ.*, vol. 240, p. 105463, Jan. 2026, doi: 10.1016/J.COMPEDU.2025.105463.
- [78] A. J. Martin *et al.*, "Integrating generative AI and load reduction instruction to individualize and optimize students' learning," *Learn. Individ. Differ.*, vol. 121, p. 102723, Jul. 2025, doi: 10.1016/J.LINDIF.2025.102723.
- [79] T. de Jong *et al.*, "Beyond inquiry or direct instruction: Pressing issues for designing impactful science learning opportunities," *Educ. Res. Rev.*, vol. 44, p. 100623, Aug. 2024, doi: 10.1016/J.EDUREV.2024.100623.
- [80] L. Zheng, Z. Shi, and L. Gao, "A generative artificial intelligence-enhanced multiagent approach to empowering collaborative problem solving across different learning domains," *Comput. Educ.*, vol. 241, p. 105489, Feb. 2026, doi: 10.1016/J.COMPEDU.2025.105489.
- [81] J. Tian and R. Zhang, "Learners' AI dependence and critical thinking: The psychological mechanism of fatigue and the social buffering role of AI literacy," *Acta Psychol. (Amst)*, vol. 260, Oct. 2025, doi: 10.1016/J.ACTPSY.2025.105725.