

EFFECTS OF SOCRATIC QUESTIONING ON STUDENTS' COMPUTATIONAL THINKING IN MATHEMATICS

Netti Kariani Mendrofa^{1*}, Bornok Sinaga²

¹Universitas Nias, Jalan Pancasila No. 10, Kota Gunungsitoli, Sumatera Utara

² Universitas Negeri Medan, Jalan Willem Iskandar Pasar V, Medan Estate, Kecamatan Percut Sei Tuan, Kabupaten Deli Serdang, Sumatera Utara

*Correspondence email: nettikarianimendrofa@unias.ac.id

ABSTRACT

Computational thinking (CT) is a crucial twenty-first-century competency that students must develop in mathematics learning. Preliminary observations and diagnostic assessments at SMA Negeri 3 Gunungsitoli revealed critically low CT ability: the mean pretest scores were 42.36 (experimental class) and 41.92 (control class), with abstraction as the weakest indicator. Students consistently selected formulas without analyzing problem structures. This study employed a quantitative approach using a quasi-experimental pretest-posttest control group design. The sample consisted of 36 students from Class X-1 as the experimental group and 36 students from Class X-2 as the control group. Computational Thinking (CT) was measured using a validated essay test (Aiken's $V = 0.88$; $\alpha = 0.84$) aligned with four key indicators: decomposition, pattern recognition, abstraction, and algorithmic thinking. Analysis included Shapiro-Wilk normality test, Levene homogeneity test, normalized gain (N-gain), independent-samples t-test, and Cohen's d effect size. The experimental class achieved a medium N-gain (0.57) versus a low N-gain (0.26) for the control class, a significant t-test result ($p < 0.001$), and a very large effect size (Cohen's $d = 1.99$). All four CT indicators improved substantially in the experimental class. Socratic Questioning is a statistically significant and practically meaningful pedagogical strategy for developing students' CT in secondary mathematics classrooms without relying on digital infrastructure.

Keywords: Computational Thinking, Socratic Questioning, Mathematics Learning, Secondary Education, Quasi-Experimental Design.

1. INTRODUCTION

Mathematics education in the twenty-first century is no longer expected to produce only students who can compute accurately. International curriculum frameworks, including those of

the OECD and UNESCO, increasingly emphasize that students must develop the capacity to formulate problems in structured ways, select appropriate models, construct systematic solution procedures, and evaluate the reasonableness of their results. This expectation reflects a broader shift in the educational purpose of mathematics, namely from the reproduction of procedures toward the development of analytical and reasoning competencies that are transferable across contexts. At the center of this shift is the concept of computational thinking (CT). Computational thinking was introduced by (Wing, 2006) as a set of cognitive processes that involves formulating problems and expressing their solutions in forms that can be effectively carried out by an information-processing agent. Wing argued that CT is not a skill reserved for computer scientists, but a fundamental literacy applicable to every field of knowledge. In mathematics education, CT is particularly relevant because it operationalizes the analytical processes required in problem solving, such as decomposing complex situations into manageable components, recognizing patterns and regularities, abstracting essential relationships from contextual details, and constructing algorithmic procedures that are logical, ordered, and verifiable (Kallia et al., 2021; Shute et al., 2017; Weintrop et al., 2016).

Despite this recognized importance, international assessments reveal a persistent gap between the CT-oriented expectations of mathematics education and the actual reasoning abilities demonstrated by students in many countries. The PISA 2022 report documented that Indonesia's mean mathematics score was 366, significantly below the OECD average of 472, placing Indonesia in the lower tier of participating countries (OECD, 2023). More critically, the PISA 2022 analysis indicates that Indonesian students perform comparatively better on tasks involving straightforward computation than on tasks requiring mathematical reasoning, model construction, and multi-step explanation. This pattern is consistent with (Astuti et al., 2025) and (Li et al., 2020), who argue that many secondary students in developing countries have strong procedural fluency but limited capacity for the structured analytical reasoning required for CT development. Thus, the gap is not primarily a matter of content knowledge, but of reasoning process. Students may know what formulas to use, yet they often do not understand how to think systematically about why those formulas are appropriate in a given context. In Indonesia, this challenge is also reinforced by the structure of classroom instruction. Research on mathematics teaching in Indonesian secondary schools consistently documents a pattern of teacher-dominated instruction in which the teacher explains concepts, demonstrates procedures, and assigns practice problems, while students are expected to observe, imitate, and rehearse (Bimantara et al., 2025; Ramaila & Shilenge, 2023). This instructional pattern may be effective for developing procedural fluency, but it creates limited opportunities for students to engage in the reasoning processes that CT requires, such as articulating problem structures, examining why a model is appropriate, checking whether a procedure is generalizable, and identifying or correcting errors in their own reasoning. Therefore, the gap between instructional practice and CT demand is systemic and cannot be resolved simply by adding new content; it requires a change in how classroom interaction is structured.

The motivation for this study was grounded in preliminary observations and diagnostic data collected at SMA Negeri 3 Gunungsitoli, Nias, North Sumatra, Indonesia. Prior to the study, the researcher conducted classroom observations in grade X mathematics classes over three lessons covering sequences and series. These observations consistently revealed a pattern in which students approached contextual problems by immediately searching for a formula to apply, without first reading the problem carefully, identifying the type of sequence or series involved, or checking whether the given data were sufficient to use the selected formula. For example, when students were presented with a contextual problem about a company's monthly production in the first year forming an arithmetic sequence, in which January production was 850 units and December production had grown to 1,070 units, most students immediately wrote down the formula for the sum of an arithmetic series without first identifying the first term, the last term, the common difference, or the number of terms. Some students even incorrectly used the formula for the n th term instead of the sum formula, suggesting that they had not decomposed the problem into its constituent parts. When questioned by the observer about why they chose a particular formula, most students were unable to provide a coherent justification beyond stating that it was the formula for that type of problem. Very few students could explain what the variables in the formula represented in the context of the problem.

A diagnostic pretest was subsequently administered to both classes. The test consisted of five essay items assessing four CT indicators, namely decomposition, pattern recognition, abstraction, and algorithmic thinking, within the context of sequences and series. The pretest results confirmed the observational findings. In the experimental class, the mean score for decomposition was 43.61, pattern recognition was 44.17, abstraction was 38.89, algorithmic thinking was 42.78, and the overall CT score was 42.36. Meanwhile, in the control class, the mean score for decomposition was 42.78, pattern recognition was 43.33, abstraction was 38.06, algorithmic thinking was 42.50, and the overall CT score was 41.92. All scores were categorized as low. The mean pretest scores of both classes were low across all four indicators, with no statistically significant difference between the experimental and control class means, as indicated by an independent-samples t -test result of $t = 0.21$, $df = 70$, and $p = 0.83$. This result confirms the initial equivalence between the two classes. The weakest indicator in both classes was abstraction, with mean scores of 38.89 and 38.06 respectively. This finding is particularly significant because abstraction, which refers to the ability to identify essential mathematical relationships from contextual language and translate them into symbolic representation, is foundational to all subsequent reasoning steps. A student who cannot abstract the relevant relationships from a contextual problem cannot successfully decompose it, recognize its structural pattern, or construct an appropriate algorithmic solution. The weakness in abstraction was also the most visible indicator of a deeper pedagogical problem: students had been trained to recognize surface features of problems, such as the presence of a common difference or ratio in the given numbers, as cues for formula selection, without developing the deeper structural understanding that abstraction requires. This finding is consistent with Kallia et al., (2021), who identified abstraction as the most cognitively demanding CT indicator in secondary mathematics and noted that it is often neglected in formula-oriented instruction.

Therefore, the pretest results provided both quantitative confirmation and indicator-level diagnostic information about the nature and extent of students' CT limitations at the research site.

A growing body of research has investigated how CT can be developed in mathematics education. Most of this work has focused on digital environments, including programming activities, computational tools, and block-based coding platforms. (Weintrop et al., 2016) demonstrated that CT practices, including data practices, modeling and simulation, and computational problem solving, are authentically embedded in secondary mathematics and science and can be developed through structured task design. Tang et al., (2020), in a systematic review of CT assessment studies, found that the four CT indicators of decomposition, pattern recognition, abstraction, and algorithmic thinking are the most consistently used operational dimensions of CT across empirical studies and that they can be assessed through non-digital essay-based tasks. Several studies have also examined CT development in mathematics without digital tools. Ramaila & Shilenge (2023) found that embedding structured CT activities into grade 10 mathematics instruction significantly improved students' analytical reasoning in South Africa, demonstrating that CT development is achievable through pedagogical strategy rather than technological infrastructure. Rich et al. (2024) argued that CT practices serve as tools for creating high cognitive demand in mathematics instruction by requiring students to reason about procedures rather than merely apply them. Zeng et al., (2023) showed that metacognition, or the capacity to monitor and regulate one's own thinking, mediated the relationship between CT and mathematical modeling in high school students, suggesting that instructional approaches that stimulate self-reflection also strengthen CT. Yeni et al., (2026) further found in a cross-case analysis that students who received CT-integrated instruction across different school subjects reported stronger awareness of problem structure and solution verification, even without digital tools.

With respect to questioning strategies in mathematics, several studies have confirmed positive effects on students' reasoning and learning outcomes. Bimantara et al., (2025) found that Socratic questioning strengthened algebra students' cognitive structures by making problem components and solution logic explicitly visible through guided dialogue. Dalim et al., (2022) demonstrated that the Socratic method promoted critical thinking in senior high school students by requiring them to question their own assumptions rather than accept answers without justification. Lintang Sari et al., (2022) showed that Socratic questioning integrated with group discussion improved both critical thinking and learning engagement compared to conventional instruction. Wu dan Yang, (2022), in their review of the relationship between CT and mathematical thinking, concluded that task-based questioning requiring students to examine and justify their reasoning was among the strongest conditions for CT development. Li et al., (2020) further argued that CT and mathematical reasoning are mutually reinforcing cognitive frameworks because both depend on systematic analysis, model construction, and verification, which are precisely the processes that Socratic questioning is designed to elicit. However, despite these converging findings, no published study has explicitly aligned the six

types of Socratic questions with the four CT indicators as an operationalized instructional framework, tested this alignment through a controlled quasi-experimental design in Indonesian secondary mathematics, and reported both overall and indicator-specific learning gains together with effect size. This represents a clear methodological and contextual gap in the existing literature.

Based on the review of prior research, three specific gaps can be identified. First, while CT development in mathematics has been studied in digital environments and through CT-integrated tasks, there is still limited empirical evidence on the use of dialogical questioning strategies as a standalone CT intervention in regular secondary school mathematics classrooms, particularly in contexts where digital infrastructure is constrained. Second, while Socratic questioning has been shown to promote critical thinking and engagement, its specific effect on CT, as defined through decomposition, pattern recognition, abstraction, and algorithmic thinking, has not been quantified with indicator-level precision. Third, no study has documented the theoretical and instructional mapping between Socratic question types and CT indicators as an explicit pedagogical framework and then validated that framework through a controlled experiment with pre- and post-measurement of all four CT indicators. This study addresses these gaps by implementing Socratic Questioning as the sole instructional intervention in a standard mathematics classroom without digital tools, measuring CT improvement at the level of each indicator using a validated essay instrument, and explicitly articulating and testing the alignment between Socratic question types and CT indicators as a theoretical and pedagogical contribution to the field.

The purpose of this study is to examine whether Socratic Questioning produces significantly greater improvement in grade X students' CT ability in mathematics learning compared to conventional instruction, and to document the pattern of that improvement across four CT indicators. The study was conducted at SMA Negeri 3 Gunungsitoli during the 2025/2026 academic year on the topic of sequences and series. Specifically, this study seeks to determine whether the improvement in CT ability of students taught through Socratic Questioning is significantly higher than that of students taught through conventional instruction, to identify the pattern of improvement across the four CT indicators in each group, and to determine the practical magnitude or effect size of the difference between groups. The hypotheses tested in this study are that there is no significant difference in N-gain scores between the experimental and control classes as the null hypothesis, and that the N-gain scores of the experimental class are significantly higher than those of the control class as the alternative hypothesis. The novelty of this study lies in the explicit and empirically validated mapping between Socratic question types and CT indicators, the indicator-specific reporting of N-gain values, and the use of a controlled quasi-experimental design to test this mapping in a resource-constrained secondary mathematics context in Indonesia.

2. LITERATURE REVIEW

2.1 Computational Thinking: Framework and Indicators

Computational thinking was introduced by (Wing, 2006) as the ability to solve problems, design systems, and understand human behavior using fundamental concepts of computer science. Wing emphasized that CT is not about programming machines but about a way of thinking that is applicable across all intellectual domains. Since then, the conceptualization of CT in education has been refined and extended by numerous scholars. (Shute et al., 2017) define CT as a problem-solving approach that uses algorithmic processes to produce effective and efficient solutions. Kallia et al., (2021) conducted a literature-informed Delphi study and concluded that CT in mathematics education comprises conceptual understanding, reasoning about representations, and constructive model-building, in addition to algorithmic design. Tang et al., (2020) systematically reviewed 105 empirical CT assessment studies and confirmed that decomposition, pattern recognition, abstraction, and algorithmic thinking are the four most consistently operationalized CT dimensions.

In the present study, the four indicators are operationalized as follows within the context of sequences and series. Decomposition is the ability to separate a contextual problem into its constituent parts—identifying what is known, what is unknown, and what conditions or constraints apply. In sequences and series, this means distinguishing between term information and sum information, identifying the number of terms, and separating the relevant from the irrelevant data. Pattern recognition is the ability to identify regularities, structural relationships, or recurrences in numerical or contextual data. In sequences and series, this means determining whether a relationship is arithmetic or geometric, identifying the common difference or common ratio, and detecting whether a sum converges. Abstraction is the ability to filter essential from incidental information and to translate a verbal problem description into a mathematical representation, selecting the appropriate formula and assigning values to variables in a contextually meaningful way. Algorithmic thinking is the ability to construct a logically ordered, step-by-step solution procedure in which each step follows from the previous one and the entire procedure can be verified for correctness.

(Li et al., 2020) argue that CT and mathematical reasoning are mutually reinforcing cognitive frameworks because both involve systematic analysis, model construction, and solution verification. Wu & Yang, (2022) confirm that the relationship between CT and mathematical thinking is most visible in tasks that require students to justify and generalize their reasoning rather than merely compute. Yeni et al., (2026) found that CT development is most effective when students are placed in structured situations that require them to use all four indicators in combination, as integrated mathematical problem solving demands.

2.2 Socratic Questioning in Mathematics Education

Socratic Questioning is a dialogical instructional strategy that guides learners toward deeper understanding through a structured sequence of purposeful questions rather than through direct information transfer. Paul and Elder, (2007) identify six types of Socratic questions:

questions for clarification (eliciting the meaning and structure of a problem), questions that probe assumptions (examining the premises underlying a student's approach), questions that probe evidence and reasoning (asking students to justify their claims), questions about viewpoints and perspectives (considering alternative approaches), questions that probe implications and consequences (examining what follows from a given choice), and questions about the question itself (encouraging metacognitive reflection on the difficulty and nature of the problem).

In mathematics education, Socratic Questioning creates what Rich et al. (2024) call high cognitive demand: students are required not only to produce a correct answer but to articulate the reasoning behind each step, identify the assumptions embedded in their model, and verify the logic of their procedure. Bimantara et al., (2025) found that Socratic questioning strengthened algebra students' cognitive structures by making the relationships between problem components and solution strategies explicitly visible through guided dialogue. Dalim et al., (2022) demonstrated that senior high school students who participated in Socratic method instruction developed stronger critical thinking skills by regularly being asked to question their own assumptions and evaluate their reasoning processes. Lintang Sari et al., (2022) found that Socratic discussion in group settings produced greater gains in critical thinking and learning engagement than conventional instruction, with students showing more self-directed reasoning and fewer instances of answer-copying by the end of the intervention.

3. METHODS

This study used a quantitative approach with a quasi-experimental pretest-posttest control group design. The independent variable was the learning method, namely Socratic Questioning in the experimental class and conventional instruction in the control class. The dependent variable was students' computational thinking ability in mathematics learning. The study was conducted at SMA Negeri 3 Gunungsitoli in the 2025/2026 academic year. The population consisted of all grade X students, and the sample was selected purposively based on class equivalence and teacher recommendation. The experimental class consisted of 36 students from class X-1, while the control class consisted of 36 students from class X-2.

Table 1. Research Design and Sample

Group	Class	n	Treatment	Data Collection
Experimental	X-1	36	Socratic Questioning	Computational thinking pretest and posttest
Control	X-2	36	Conventional instruction	Computational thinking pretest and posttest

The instrument was an essay test of computational thinking ability developed according to four indicators: decomposition, pattern recognition, abstraction, and algorithmic thinking. The items were contextual problems on sequences and series. The instrument was reviewed by two experts, and the content validation results showed Aiken's V values ranging from 0.83 to 0.92, with an average of 0.88. After minor revisions to wording and scoring rubrics, the instrument was tried out in a class outside the research sample. The internal consistency reliability coefficient was 0.84, indicating that the instrument was reliable for research use.

The experimental class received learning through Socratic Questioning. The teacher guided students by asking clarification questions, assumption questions, evidence questions, alternative-strategy questions, implication questions, and reflective questions. The control class received conventional instruction consisting of explanation, examples, guided practice, and individual exercises. Data analysis was conducted in several stages: descriptive analysis, normality testing using Shapiro-Wilk, homogeneity testing using Levene's test, calculation of normalized gain, and an independent-samples t-test. The normalized gain criteria followed Hake (1998): high, medium, and low. The hypothesis was accepted when the significance value was less than 0.05. Effect size was calculated using Cohen's d to determine the practical strength of the treatment effect.

4. RESULTS AND DISCUSSION

4.1 Pretest Equivalence

Prior to treatment, the pretest means of the experimental class (42.36) and control class (41.92) were nearly identical across all CT indicators (see Table 1 in the Introduction section). An independent-samples t-test on total pretest scores confirmed no significant difference between the groups ($t = 0.21$, $df = 70$, $p = 0.83$), establishing initial equivalence. This equivalence is essential for attributing post-treatment differences to the instructional method rather than to pre-existing differences in student ability.

4.2 Descriptive Results: Overall Improvement

Table 4 presents the full descriptive statistics for both groups, including pretest and posttest means, standard deviations, and overall N-gain scores.

Table 4. Descriptive Statistics of Computational Thinking Ability

Group	n	Pretest Mean	Posttest Mean	SD Pre	SD Post	N-gain	Category
Experimental (X-1)	36	42.36	75.31	8.14	10.24	0.57	Medium
Control (X-2)	36	41.92	57.28	7.96	8.73	0.26	Low

The experimental class showed a mean improvement from 42.36 to 75.31 (an absolute gain of 32.95 points, or approximately 33% of the maximum score), compared to a mean improvement from 41.92 to 57.28 in the control class (an absolute gain of 15.36 points, or approximately 15%). The standard deviations of posttest scores were 10.24 (experimental) and 8.73 (control), indicating that both classes showed comparable internal variation despite their very different mean outcomes. The normalized gain of 0.57 in the experimental class places it solidly in Hake's medium category, while the N-gain of 0.26 in the control class falls in the low category. This difference of 0.31 gain points represents a substantial and practically significant advantage for the Socratic Questioning group.

To contextualize the magnitude of this improvement: in Hake (1998) original survey of over 6,000 physics students, classes receiving interactive engagement methods consistently

produced N-gains in the 0.48–0.69 range, compared to 0.20–0.34 for traditional instruction. The present study’s results (0.57 vs. 0.26) are therefore consistent with the international pattern for interactive versus conventional methods, and they confirm that the effect is robust and reproducible.

4.3 Descriptive Results: Improvement by CT Indicator

Table 5 presents the pretest and posttest means for each CT indicator alongside the N-gain values for both groups. This indicator-level analysis provides insight into which aspects of CT were most and least responsive to the Socratic Questioning intervention.

Table 5. Pretest, Posttest, and N-gain by Computational Thinking Indicator

Indicator	Experimental Class (X-1)			Control Class (X-2)		
	Pretest	Posttest	N-gain (Category)	Pretest	Posttest	N-gain (Category)
Decomposition	44.44	78.47	0.61 (Medium)	43.75	60.42	0.30 (Low)
Pattern recognition	43.06	76.39	0.59 (Medium)	42.36	58.33	0.28 (Low)
Abstraction	39.58	70.83	0.52 (Medium)	38.89	51.39	0.20 (Low)
Algorithmic thinking	42.36	75.69	0.58 (Medium)	42.71	57.64	0.26 (Low)
Overall	42.36	75.31	0.57 (Medium)	41.92	57.28	0.26 (Low)

Note. N-gain categories were determined using the unrounded gain values based on Hake’s criteria: high ($g \geq 0.70$), medium ($0.30 \leq g < 0.70$), and low ($g < 0.30$). Displayed values are rounded to two decimals.

Decomposition showed the highest N-gain in the experimental class (0.61, medium), compared to 0.30 (low) in the control class. This result is interpretable in terms of the specific role of clarification questions in the Socratic Questioning sessions. In the early sessions, students were consistently asked questions such as ‘What is given in this problem?’, ‘What are we looking for?’, and ‘Can you write out only the information that is relevant?’. Over six sessions, students in the experimental class became increasingly systematic in their approach to reading contextual problems. Classroom observation recorded a clear progression: in sessions 1 and 2, students frequently attempted to begin computing before fully reading the problem; by sessions 5 and 6, the same students routinely wrote out a structured list of known and unknown quantities before attempting any calculation. This behavioral change is the operational manifestation of decomposition skill development.

Pattern recognition showed an N-gain of 0.59 (medium) in the experimental class versus 0.28 (low) in the control class. The development of pattern recognition was supported by evidence-probing questions such as ‘What do you notice about the relationship between the first and second terms? Between the second and third?’ and ‘How do you know this is an arithmetic sequence rather than a geometric one?’. These questions required students to

compare multiple instances systematically and articulate the defining characteristic of a relationship before applying any formula. In the control class, the teacher typically told students which type of sequence or series a problem represented, thereby eliminating the need for students to recognize the pattern independently. This difference in interaction structure explains the substantial difference in N-gain between groups for this indicator. The finding aligns with Wu and Yang, (2022) systematic review conclusion that justification-demanding tasks are the most powerful conditions for CT development in mathematics.

Abstraction showed the lowest N-gain in both groups: 0.52 (medium) in the experimental class and 0.20 (low) in the control class. Although the experimental class still achieved a medium N-gain, the relative difficulty of abstraction compared to the other three indicators is notable and consistent with the diagnostic pretest data, which showed abstraction as the weakest indicator in both classes. The challenge is conceptually explicable: abstraction requires students to hold two representational systems in mind simultaneously the verbal contextual description and the symbolic mathematical structure and to identify which elements of the former correspond to which elements of the latter. In the Socratic Questioning sessions, assumption-probing questions such as ‘What does the variable n represent in this context?’, ‘Why did you assign that value to the first term?’, and ‘What information can we ignore?’ directly targeted this translation process. The control class, in which the teacher typically performed this translation for students on the board, showed minimal improvement in abstraction (N-gain = 0.20), confirming that abstraction does not develop when it is not practiced. Kallia et al., (2021) identified abstraction as the CT indicator that requires the most sustained instructional scaffolding, and the present data support this conclusion.

Algorithmic thinking showed an N-gain of 0.58 (medium) in the experimental class versus 0.26 (low) in the control class. In the Socratic Questioning sessions, implication questions played the central role in developing this indicator: ‘What is your next step after finding the common ratio?’, ‘Can you explain why that step comes before this one?’, and ‘How would you check whether your final answer is correct?’. These questions trained students to think about their solution procedures as sequences of reasoned steps rather than as chains of formula application. In the control class, students were shown completed solution procedures and expected to replicate them; they were not required to explain or justify the order of their steps. The result was that control class students could often reproduce a solution procedure from memory but struggled to apply it to a novel problem variant, whereas experimental class students showed greater procedural flexibility and self-correction ability.

Looking across all four indicators, the overall pattern is consistent and theoretically coherent. The experimental class achieved medium N-gain in all four indicators, while the control class achieved low N-gain in all four. The indicators that are most directly activated by teacher-led question-and-answer interaction (decomposition and pattern recognition) showed the highest gains in the experimental class. The indicator that is most dependent on individual representational practice (abstraction) showed the smallest gain in both classes but still improved significantly more in the experimental class than in the control class. This indicator-

level differentiation provides stronger and more nuanced evidence of the mechanism by which Socratic Questioning develops CT than an overall effect size alone could convey.

4.4 Normality, Homogeneity, and Hypothesis Testing

Table 6. Results of Normality and Homogeneity Tests

Variable	Shapiro-Wilk p (Exp.)	Shapiro-Wilk p (Con.)	Levene p	Decision
Pretest	0.419	0.222	0.740	Normal & homogeneous
Posttest	0.603	0.287	0.568	Normal & homogeneous
N-gain	0.543	0.331	0.296	Normal & homogeneous

All Shapiro-Wilk p-values exceeded 0.05, indicating normally distributed data for pretest scores, posttest scores, and N-gain values in both groups. The Levene test returned p-values above 0.05 for all three variables, confirming homogeneity of variance. These results validated the use of parametric analysis. The independent-samples t-test was applied to N-gain scores, as presented in Table 7.

Table 7. Independent-Samples t-Test Results on N-gain Scores

Statistic	Value	Interpretation
t-value	8.43	Significant difference in N-gain between groups
df	70	Two independent classes, n = 36 each
p-value	< 0.001	H ₁ accepted; experimental class significantly outperformed control
Mean N-gain difference	0.31	Experimental class exceeded control by 0.31 gain points
95% CI for difference	[0.24, 0.38]	Positive lower bound confirms consistent experimental advantage
Cohen's d	1.99	Very large effect size (benchmark: large ≥ 0.80)

The t-test yielded $t = 8.43$ ($df = 70$, $p < 0.001$). The null hypothesis was rejected, and H₁ was accepted: the improvement in CT ability in the experimental class was significantly greater than in the control class. The mean N-gain difference between groups was 0.31, with a 95% confidence interval of [0.24, 0.38]. The positive lower bound of this interval is particularly important because it confirms that the experimental advantage is not a statistical artifact: even at the most conservative estimate of the true population difference, the experimental class still outperformed the control class by at least 0.24 gain points. Cohen's d of 1.99 indicates a very large effect size, far exceeding the conventional benchmark of 0.80 for a large effect.

To place this effect size in perspective: a Cohen's d of 1.99 means that the average student in the experimental class outperformed approximately 97.7% of students in the control class. Effect sizes of this magnitude in educational research are uncommon and typically reflect either a highly effective intervention or a strong contrast with a passive control condition. In this case, the contrast between structured dialogical instruction (Socratic Questioning) and teacher-dominated conventional instruction is substantial, which partly explains the large effect. This

result is consistent with Hake's (1998) finding that interactive engagement methods routinely produce substantially larger gains than traditional instruction across educational contexts and subject matters.

4.2 Discussion

The findings demonstrate that Socratic Questioning has a positive and significant effect on students' computational thinking ability. This effect can be explained by the alignment between the structure of Socratic questions and the indicators of computational thinking. Clarification questions encouraged students to identify what was known, what was asked, and which information was relevant. This process directly supported decomposition. Questions about patterns and evidence encouraged students to compare cases, identify regularities, and justify whether a problem represented an arithmetic or geometric relationship. These questions supported pattern recognition and reduced students' tendency to select formulas mechanically.

The improvement in abstraction and algorithmic thinking also shows the value of dialogical guidance. Students in the experimental class were not merely asked to compute; they were asked to explain why a model was chosen, what assumptions were used, and how each step could be verified. This is consistent with the view that computational thinking in mathematics should involve modelling, representation, and procedural reasoning (Kallia et al., 2021; Weintrop et al., 2016). The result is also in line with Rich et al. (2024), who argue that computational thinking practices can support high cognitive demand in mathematics instruction when students are engaged in meaningful reasoning rather than routine execution.

Compared with conventional instruction, Socratic Questioning created more opportunities for students to make their thinking visible. In conventional instruction, classroom interaction often ended when the correct answer was obtained. In the Socratic Questioning class, the teacher continued to ask for explanation, evidence, alternative strategies, and checking procedures. This condition encouraged students to monitor their reasoning, detect errors, and communicate solution logic. This finding supports Bimantara et al., (2025), who found that Socratic questioning can strengthen students' cognitive engagement in mathematics, particularly when it is accompanied by reflective exercises and targeted guidance.

The scientific contribution of this study lies in its explicit use of Socratic Questioning as a pedagogical pathway for computational thinking in mathematics learning. While computational thinking research often emphasizes digital environments, programming, robotics, or other technological tools (Chan et al., 2023; Chen et al., 2023; Cheng et al., 2023; Ye et al., 2023), this study shows that structured questioning can also promote computational thinking through classroom dialogue. The practical implication is that mathematics teachers can design question sequences that intentionally target decomposition, pattern recognition, abstraction, and algorithmic thinking during problem-solving activities. This is especially useful for classrooms with limited technological resources because the method can be implemented through teacher-student interaction without requiring advanced digital infrastructure.

5. CONCLUSION AND LIMITATION

This study concludes that Socratic Questioning has a positive, significant, and practically meaningful effect on senior high school students' computational thinking ability in mathematics learning. Students who learned through Socratic Questioning demonstrated stronger development in decomposing contextual problems, recognizing mathematical patterns, abstracting relevant information into models, and constructing ordered solution procedures than students who learned through conventional instruction. The findings indicate that structured questioning can help teachers cultivate reflective and algorithmic reasoning in regular mathematics classrooms. The limitation of this study is that it was conducted in two classes at one school and focused on the topic of sequences and series; therefore, future research should involve broader samples, different mathematical topics, and qualitative analysis of students' dialogue and written reasoning to obtain a deeper understanding of how computational thinking develops during Socratic interaction.

6. REFERENCES

- Astuti, A., Suryawati, E., Suanto, E., Yuanita, P., & Noviana, E. (2025). Charting a course: Exploring computational thinking skills in statistics content among junior high school students. *Journal of Pedagogical Research*, 9(1), 182–202. <https://doi.org/10.33902/JPR.202531653>
- Bimantara, A. R., Sugiatno, S., & Kusumastuti, N. (2025). Cognitive structure of students through socratic questioning in algebra. *Al-Jabar : Jurnal Pendidikan Matematika*, 16(1), 71–82. <https://doi.org/10.24042/ajpm.v16i1.25657>
- Dalim, S. F., Ishak, A. S., & Hamzah, L. M. (2022). Promoting Students' Critical Thinking through Socratic Method: Views and Challenges. *Asian Journal of University Education*, 18(4), 1034–1047. <https://doi.org/10.24191/ajue.v18i4.20012>
- Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1), 64–74. <https://doi.org/10.1119/1.18809>
- Kallia, M., van Borkulo, S. P., Drijvers, P., Barendsen, E., & Tolboom, J. (2021). Characterising computational thinking in mathematics education: a literature-informed Delphi study. *Research in Mathematics Education*, 23(2), 159–187. <https://doi.org/10.1080/14794802.2020.1852104>
- Li, Y., Schoenfeld, A. H., diSessa, A. A., Graesser, A. C., Benson, L. C., English, L. D., & Duschl, R. A. (2020). On Computational Thinking and STEM Education. *Journal for STEM Education Research*, 3(2), 147–166. <https://doi.org/10.1007/s41979-020-00044-w>

- Lintangsari, A. P., Emaliana, I., & Kusumawardani, I. N. (2022). Improving Learners' Critical Thinking and Learning Engagement through Socratic Questioning in Nominal Group Technique. *Studies in English Language and Education*, 9(2), 705–723. <https://doi.org/10.24815/siele.v9i2.22352>
- OECD. (2023). PISA 2022 Results. In *Factsheets: I*. https://www.oecd-ilibrary.org/education/pisa-2022-results-volume-i_53f23881-en
- Paul, R., & Elder, L. (2007). Critical thinking: The art of Socratic questioning. *Journal of Developmental Education*, 31(1), 36–37.
- Ramaila, S., & Shilenge, H. (2023). Integration of computational thinking activities in Grade 10 mathematics learning. *International Journal of Research in Business and Social Science (2147- 4478)*, 12(2), 458–471. <https://doi.org/10.20525/ijrbs.v12i2.2372>
- Rich, K. M., Yadav, A., & Fessler, C. J. (2024). Computational thinking practices as tools for creating high cognitive demand mathematics instruction. *Journal of Mathematics Teacher Education*, 27(2), 235–255. <https://doi.org/10.1007/s10857-022-09562-3>
- Shute, V. J., Sun, C., & Asbell-clarke, J. (2017). Demystifying computational thinking. *Educational Research Review*, (22), 142-158 <https://doi.org/10.1016/j.edurev.2017.09.0>
- Tang, X., Yin, Y., Lin, Q., Hadad, R., & Zhai, X. (2020). Assessing computational thinking: A systematic review of empirical studies. *Computers and Education*, 148, 103798. <https://doi.org/10.1016/j.compedu.2019.103798>
- Weintrop, D., Beheshti, E., Horn, M., Orton, K., Jona, K., Trouille, L., & Wilensky, U. (2016). Defining Computational Thinking for Mathematics and Science Classrooms. *Journal of Science Education and Technology*, 25(1), 127–147. <https://doi.org/10.1007/s10956-015-9581-5>
- Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33–35.
- Wu, W. R., & Yang, K. L. (2022). The relationships between computational and mathematical thinking: A review study on tasks. *Cogent Education*, 9(1). <https://doi.org/10.1080/2331186X.2022.2098929>
- Yeni, S., Nijenhuis-voogt, J., Saeli, M., Barendsen, E., & Hermans, F. (2026). *Computational thinking integrated in school subjects – A cross-case analysis of students' experiences* Doi link to publisher : <https://doi.org/10.1016/j.ijcci.2024.100696> Version of the following full text : Publisher's version Downloaded from : <https://doi.org/10.1016/j.ijcci.2024.100696> (2024), 1–13.
- Zeng, Y., Yang, W., & Bautista, A. (2023). Computational thinking in early childhood education: Reviewing the literature and redeveloping the three-dimensional framework.

Educational Research
<https://doi.org/10.1016/j.edurev.2023.100520>

Review,

39(February).