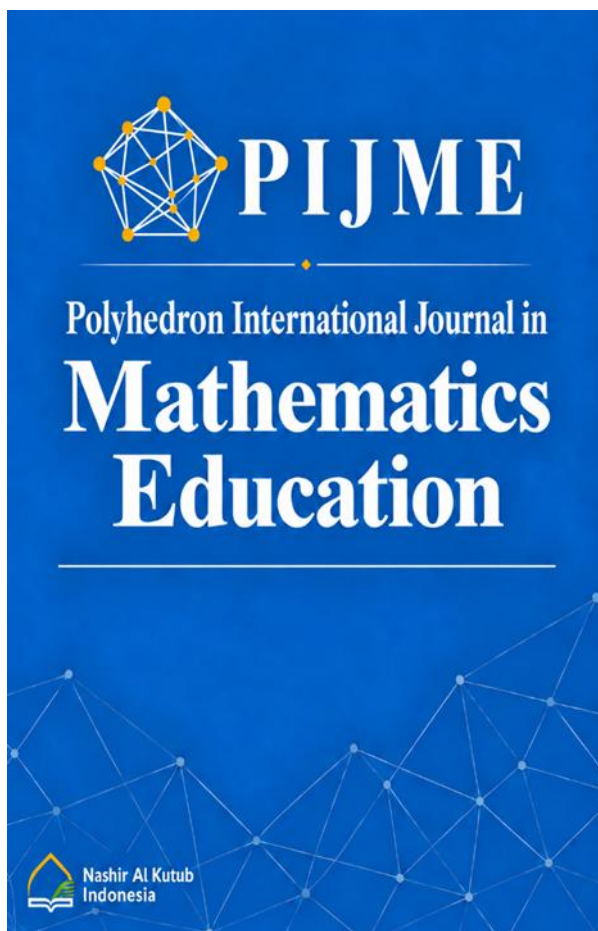




**Mathematical critical thinking with
GeoGebra-assisted CORE learning:
Does students' self-confidence
matter?**

**Rabiatul Adawiyah¹, Sudirman
Sudirman², Idha Novianti³, Javier
García-García⁴**

¹Master's Program in Mathematics Education,
Postgraduate School, Universitas Terbuka,
Jakarta, Indonesia,
rabiatul.adawiyah@ecampus.ut.ac.id

²Master's Program in Mathematics Education,
Postgraduate School, Universitas Terbuka,
Jakarta, Indonesia,
sudirman.official@ecampus.ut.ac.id

³Master's Program in Mathematics Education,
Postgraduate School, Universitas Terbuka,
Jakarta, Indonesia, anti@ecampus.ut.ac.id

⁴Department of Mathematics Education,
Universidad Autónoma de Guerrero,
Chilpancingo, Mexico, jagarcia@uagro.mx

To cite this article:

Adawiyah, R., Sudirman, S., Novianti, I., & García-García, J. (2026). Mathematical critical thinking with GeoGebra-assisted CORE learning: Does students' self-confidence matter? *Polyhedron International Journal in Mathematics Education*, 4(2), 93 – 110.

To link to this article:

<https://nakiscience.com/index.php/pijme>

Published by:

Nasir Al-Kutub Indonesia

Residential Street Kila Rengganis, Block I, Number 11, Labuapi, Indonesia, 83361

Mathematical critical thinking with GeoGebra-assisted CORE learning: Does students' self-confidence matter?

Rabiatul Adawiyah^{1*}, Sudirman Sudirman², Idha Novianti³, Javier García-García⁴

¹Master's Program in Mathematics Education, Postgraduate School, Universitas Terbuka, Jakarta Indonesia, rabiatul.adawiyah@ecampus.ut.ac.id

²Master's Program in Mathematics Education, Postgraduate School, Universitas Terbuka, Jakarta, Indonesia, sudirman.official@ecampus.ut.ac.id

³Master's Program in Mathematics Education, Postgraduate School, Universitas Terbuka, Jakarta, Indonesia, anti@ecampus.ut.ac.id

⁴Department of Mathematics Education, Universidad Autónoma de Guerrero, Chilpancingo, Mexico, jagarcia@uagro.mx

*Correspondence: rabiatul.adawiyah@ecampus.ut.ac.id

Abstract

Mathematical critical thinking is a central twenty-first-century competency, yet it remains underdeveloped in many mathematics classrooms, particularly in visually demanding topics such as geometric transformations. This study examined differences in students' mathematical critical-thinking scores by instructional model and self-confidence level. A quasi-experimental, nonequivalent posttest-only design was used with 72 Grade 11 students (36 experimental and 36 control) at a public senior secondary school in Riau, Indonesia. The experimental group received four sessions using the Connecting, Organizing, Reflecting, and Extending (CORE) model integrated with GeoGebra, whereas the control group received expository instruction. Data were collected using a six-item essay test and an 18-item self-confidence questionnaire and analyzed using a two-way analysis of variance. Posttest scores differed significantly by instructional model, $F(1, 66) = 68.25$, $p < .001$, $\eta^2 = .51$, and self-confidence level, $F(2, 66) = 26.94$, $p < .001$, $\eta^2 = .45$. The interaction was not statistically significant, $F(2, 66) = 0.94$, $p = .394$, $\eta^2 = .03$. Within the limits of a short, single-school, posttest-only study, the findings indicate that the GeoGebra-assisted CORE model was associated with higher mathematical critical-thinking scores across the observed self-confidence levels.

Article History

Received:

July 26, 2026

Revised:

August 29, 2026

Accepted:

September 1, 2026

Published Online:

October 07, 2026

Keywords:

CORE learning model;

GeoGebra;

Mathematical critical thinking;

Self-confidence;

1. Introduction

Mathematical critical thinking is a fundamental competency in contemporary mathematics education because it enables students to analyze information, evaluate arguments, and make reasoned decisions when solving complex problems (Ennis, 1989). It is important not only in academic settings but also in everyday situations that require systematic reasoning and evidence-based judgment. Mathematics education should therefore prepare students not merely to perform procedures but also to think flexibly, critically, and adaptively in unfamiliar situations (Gravemeijer et al., 2017).

Nevertheless, international assessment results, empirical studies, and preliminary classroom evidence suggest that Indonesian students may experience difficulty with mathematical critical thinking, particularly in geometric transformations, which require spatial visualization. In PISA 2022, Indonesian 15-year-olds obtained a mean mathematics score of 366 points, considerably below the OECD average of 472 points (OECD, 2023), indicating limited performance on tasks that require students to reason, interpret, and justify mathematical solutions. Rahmawati and Rizaldi (2025) reported that most of the Indonesian junior secondary students they assessed demonstrated only moderate mathematical critical-thinking skills in geometry, with difficulties in reasoning coherence and conclusion formulation. Specifically for geometric transformations, Maifa et al. (2025) identified conceptual, procedural, and visualization errors when Indonesian prospective mathematics teachers constructed transformation proofs, with conceptual errors underlying the other error types. Preliminary observations at the participating school indicated that students struggled to justify their answers, relied on memorized procedures, and rarely connected transformations with everyday contexts. These observations are compatible with concerns that students may perceive mathematics as disconnected procedures rather than as an interconnected body of knowledge (Pierce & Stacey, 2004).

Teacher-centered expository instruction may limit students' opportunities to construct, test, and justify mathematical ideas (Abrami et al., 2015; Freeman et al., 2014). A meta-analysis of experimental and quasi-experimental studies showed that critical-thinking instruction was more effective when students engaged in dialogue, worked on authentic problems, and received mentoring than when they only received information (Abrami et al., 2015), and a meta-analysis of 225 studies in undergraduate science, engineering, and mathematics found that students taught through traditional lecturing were 1.5 times more likely to fail than those taught through active learning (Freeman et al., 2014). The Connecting, Organizing, Reflecting, and Extending (CORE) model provides a constructivist alternative: students connect prior knowledge with new ideas, organize information, reflect through verification and discussion, and extend their understanding to unfamiliar problems. The CORE model has been associated with improved cognitive mathematics outcomes (Ningsih et al., 2019). Its

reflecting and extending phases are also conceptually aligned with critical-thinking practices such as checking evidence, justifying conclusions, and applying ideas in new contexts (Ennis, 1989).

Limited visualization presents an additional challenge in geometric transformations because students must reason about changes in an object's position, size, and orientation on the coordinate plane. GeoGebra provides an interactive environment in which students can manipulate geometric objects and observe changes dynamically. Adiguna et al. (2025) reported improved critical-thinking scores following the use of GeoGebra-based media in Grade 11 geometric-transformation lessons. Integrating dynamic visualization with a structured learning model may therefore support students' conceptual reasoning more effectively than either unstructured technology use or exposition alone.

Affective factors may also be related to students' mathematical critical thinking. Self-confidence reflects students' beliefs about their capability to engage with mathematical tasks and may influence persistence and willingness to attempt complex strategies. Self-confidence has been associated with critical-thinking skills in mathematics (Safitri & Sutiarso, 2023) and with intrinsic motivation in science learning (Pečiuliauskienė, 2020). Confidence has also emerged as one of the strongest non-cognitive predictors of mathematics achievement, capturing much of the predictive variance of related self-beliefs such as self-efficacy, self-concept, and anxiety (Stankov et al., 2012). What is less known is whether the advantage of a structured, technology-supported model is similar for students with different confidence levels or is concentrated among students who are already confident. Existing studies have typically either correlated self-confidence with critical thinking without an instructional comparison (Safitri & Sutiarso, 2023) or evaluated GeoGebra-based instruction without considering students' affective characteristics (Adiguna et al., 2025). By crossing an instructional comparison with baseline self-confidence levels in a factorial design, this study contributes evidence on whether the GeoGebra-assisted CORE model is associated with higher critical-thinking scores among high-, moderate-, and low-confidence students alike, information that can help teachers anticipate which students may require additional scaffolding.

The constructivist basis of the CORE model is consistent with a social view of learning in which knowledge is developed through interaction with peers, teachers, tasks, and tools (Vygotsky, 1978). During the organizing and reflecting phases, heterogeneous-group discussion can provide peer scaffolding and opportunities to verify reasoning. GeoGebra may strengthen this process by providing shared, manipulable representations that allow claims to be examined against visible mathematical evidence.

Two gaps motivate this study. First, few studies have examined the stage-by-stage integration of CORE and GeoGebra in teaching translation, reflection, rotation,

and dilation. Second, limited evidence is available on whether mathematical critical-thinking scores under technology-supported instruction vary by self-confidence level. Previous studies have examined CORE without GeoGebra or a critical-thinking outcome (Ningsih et al., 2019) and GeoGebra-based media without a structured CORE sequence or affective moderator (Adiguna et al., 2025). This study therefore examined (a) differences in mathematical critical-thinking scores between students taught using the GeoGebra-assisted CORE model and those taught using expository instruction, (b) differences in scores across high, moderate, and low self-confidence levels, and (c) the interaction between instructional model and self-confidence level.

2. Methods

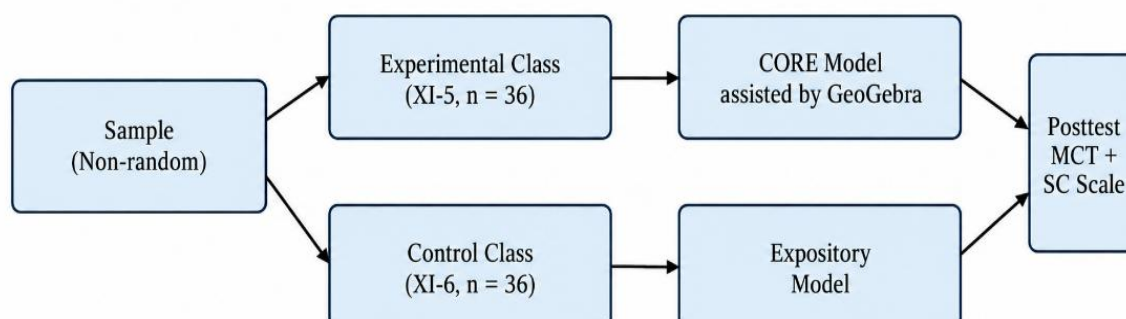
2.1 Research Design

This quantitative study used a quasi-experimental, nonequivalent posttest-only control-group design with a 2×3 factorial structure. The factors were instructional model (GeoGebra-assisted CORE vs. expository instruction) and self-confidence level (high, moderate, or low). Intact classes were used because individual random assignment was not feasible. A posttest-only design was selected for two reasons. First, the intervention lasted only four sessions, and administering the same six open-ended critical-thinking items as a pretest shortly before instruction could have sensitized students to the targeted reasoning demands, producing a pretest-sensitization (testing) effect that would be confounded with the treatment (Shadish et al., 2002). Second, only one validated form of the essay test was available, so an alternate pretest form could not be used. Instead, baseline comparability was examined using a theoretically related prior-achievement measure (see Section 2.2), and self-confidence was measured before instruction. This choice limits conclusions about individual change and cannot rule out preexisting differences in critical thinking itself; the findings are therefore interpreted as posttest group differences rather than as learning gains.

The experimental group received four lessons in which GeoGebra was integrated into each CORE phase, whereas the control group received expository instruction on the same geometric-transformation content. Self-confidence was measured before the intervention and used as a baseline grouping variable. After the four lessons, both groups completed the mathematical critical-thinking posttest. The instructional model was the first factor, self-confidence level was the second factor, and mathematical critical-thinking posttest score was the outcome.

Figure 1

Posttest-Only Nonequivalent Groups Design



Several procedures were used to reduce threats to internal validity. The same teacher-researcher taught both classes, both groups received equivalent instructional time and content coverage, and the posttest was administered under comparable conditions. Using the same teacher reduced between-teacher variation but introduced a possible expectancy or allegiance effect. An independent mathematics teacher observed all four sessions in each class using a 20-item implementation-fidelity checklist. Mean adherence was 93.75% in the experimental class and 95.00% in the control class; interobserver agreement, calculated for two independently coded sessions, was 90.00%. Participation was voluntary, parental or guardian consent and student assent were obtained, and confidentiality was protected. The study protocol was approved by the school principal and the institutional research committee of the researchers' university (Approval No. 014/KEP-PEND/2025).

2.2 Participants

Participants were Grade 11 students at a public senior secondary school in Rokan Hilir Regency, Riau, Indonesia, during the 2024/2025 academic year. The population comprised all Grade 11 Mathematics and Natural Sciences classes at the school. Two intact classes were selected purposively based on the mathematics teacher's judgment of comparable prior achievement: Class XI-5 was assigned as the experimental group ($n = 36$), and Class XI-6 served as the control group ($n = 36$). The total sample was 72 students.

Prior achievement was assessed using daily-test scores on prerequisite matrix material. The experimental class ($M = 76.42$, $SD = 7.18$) and control class ($M = 75.89$, $SD = 7.44$) showed no statistically detectable difference, $t(70) = 0.31$, $p = .759$, 95% CI for the mean difference $[-2.91, 3.97]$. This result was treated as supporting, but not proving, baseline comparability. Participant characteristics by sex and treatment group are presented in Table 1. Students participated voluntarily with school permission, parental or guardian consent, and student assent.

Table 1

Characteristics of the Research Participants

Class	Treatment	Number of Students	Male	Female
Experimental (XI-5)	CORE + GeoGebra Model	36	16	20
Control (XI-6)	Expository Model	36	17	19
Total	-	72	33	39

2.3 Data Collection

Data were collected using a mathematical critical-thinking test and a self-confidence questionnaire. The critical-thinking test contained six essay items, with one item representing each indicator adapted from Ennis (2011): focusing on a question, analyzing and clarifying a question, using relevant concepts, comparing solution strategies, drawing conclusions, and communicating a solution coherently. Each item was scored from 0 to 10 using an analytic rubric, producing a total possible score of 0–60. The rubric distinguished five performance dimensions: identification of relevant information, selection of concepts, accuracy of procedures, justification of reasoning, and clarity of conclusions. Two trained scorers independently assessed 25% of the responses; the two-way random, absolute-agreement intraclass correlation was .92, 95% CI [.87, .95]. Disagreements were resolved through discussion before the remaining responses were scored. The items were piloted with 30 Grade 11 students from a neighboring public school who had studied geometric transformations but were not included in the main study. Corrected Pearson item–total correlations ranged from .42 to .76 and exceeded the decision criterion of .30. Three mathematics-education experts also evaluated content relevance, clarity, and alignment with the indicators; the scale-level content validity index based on universal agreement was .89. Internal consistency was acceptable (Cronbach's $\alpha = .823$).

The self-confidence questionnaire initially comprised 20 statements rated on a 4-point Likert-type scale (1 = strongly disagree to 4 = strongly agree). It assessed confidence in one's mathematical ability, calmness when solving problems, receptiveness to criticism, and independence in decision making. Following the pilot test, two items were removed, leaving 18 items with acceptable internal consistency (Cronbach's $\alpha = .791$) and a possible score range of 18–72. Self-confidence was measured before the intervention. Categories were based on the pilot-sample mean ($M = 52.11$) and standard deviation ($SD = 6.84$): scores below 45.27 were classified as low, scores from 45.27 to 58.95 as moderate, and scores above 58.95 as high. Table 2 summarizes the instruments.

Table 2

Summary of Data-Collection Instruments

Instrument	Number of Items	Format	Validity	Reliability (α)
Mathematical Critical-Thinking Test	6	Essay	Item-total coefficients = .42–.76	$\alpha = .823$
Self-Confidence Questionnaire	18 (final)	4-point Likert scale	2 items removed after pilot testing	$\alpha = .791$

2.4 Data Analysis

Descriptive statistics included the mean, standard deviation, minimum, and maximum for each group. Before the factorial analysis, normality and homogeneity assumptions were evaluated. Inspection of standardized residuals and the Q-Q plot indicated no marked departure from normality. The Shapiro-Wilk tests for the six instructional model $\times$ self-confidence cells were nonsignificant, with W values ranging from .934 to .978 and p values from .326 to .892. No standardized residual exceeded $|3.00|$, and the residual-versus-predicted plot showed no systematic pattern. Homogeneity of variance was assessed using Levene's test with $\alpha = .05$ (Field, 2018).

A two-way analysis of variance (ANOVA) was conducted in IBM SPSS Statistics (Version 26) using $\alpha = .05$. The model estimated the main effects of instructional model and self-confidence level and their interaction. Partial eta squared (η^2) and 95% confidence intervals were reported as effect-size estimates. When the self-confidence main effect was significant, Tukey's honestly significant difference test was used for pairwise comparisons. Cell sizes were 13, 15, and 8 for the high-, moderate-, and low-confidence experimental groups, respectively, and 11, 16, and 9 for the corresponding control groups. A sensitivity analysis for the interaction ($\alpha = .05$, power = .80, numerator df = 2, denominator df = 66, N = 72) indicated that the design could reliably detect effects of approximately $f = .37$ or larger; thus, smaller interaction effects may have remained undetected.

3. Results and Discussion

3.1 Results

3.1.1 Implementation Process

Instruction in the experimental group was implemented over four sessions (2×45 minutes each), with material distributed as translation (session 1), reflection (session 2), rotation (session 3), and dilation (session 4), each following the CORE stages in sequence. At the connecting stage, the teacher related each sub-topic to students' everyday experience—for example, escalator movement for translation, the

experience of looking in a mirror for reflection, a carousel for rotation, and a photocopier for dilation—before demonstrating the concept visually using GeoGebra. At the organizing stage, students were divided into nine heterogeneous groups of four to explore geometric transformation independently using GeoGebra and document their findings on a Student Worksheet (LKPD). At the reflecting stage, each group presented its discussion results and compared manual calculations with GeoGebra visualizations; when discrepancies occurred, students were encouraged to critically trace the cause. At the extending stage, students worked on contextual problems requiring a composition of several transformations at once, such as an airplane-movement problem involving translation, rotation, and dilation in sequence. Throughout instruction, GeoGebra use was facilitated through the school's computer laboratory (20 units) equipped with GeoGebra version 6.0, with personal laptops or the GeoGebra Classic web application as alternatives for students without access to a workstation.

Figures 2–4 illustrate the GeoGebra-assisted activities described above and complement the written explanation of the instructional sequence. These figures serve as visual references to help readers follow the organisation of the activities, understand the intended use of GeoGebra, and relate each stage to the learning objectives. They support the explanation of the instructional design, while the analysis of classroom implementation and students' learning draws on the empirical data collected during the study.

Figure 2

Teacher Demonstration of Translation During the Connecting Stage

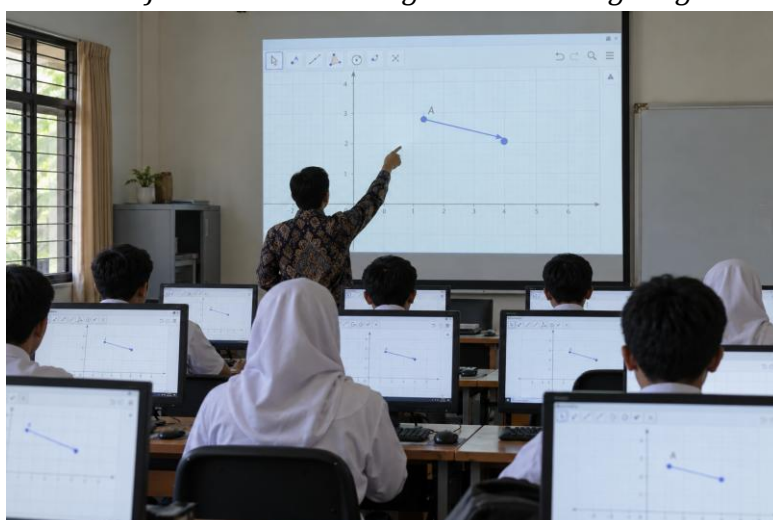


Figure 2 illustrates a teacher-led introduction to translation, with a projected coordinate-plane representation serving as a shared reference. Using the example discussed above, students can predict the image of $A(2, 3)$ under the vector $(4, -1)$, calculate $A'(6, 2)$, and compare that prediction with the GeoGebra construction. The

intended focus is the relationship between displacement and coordinate changes, rather than memorization of a rule. A follow-up prompt can ask which features of a figure remain unchanged when all its vertices are translated by the same vector.

Figure 3

Group Exploration of Reflection During the Organizing Stage

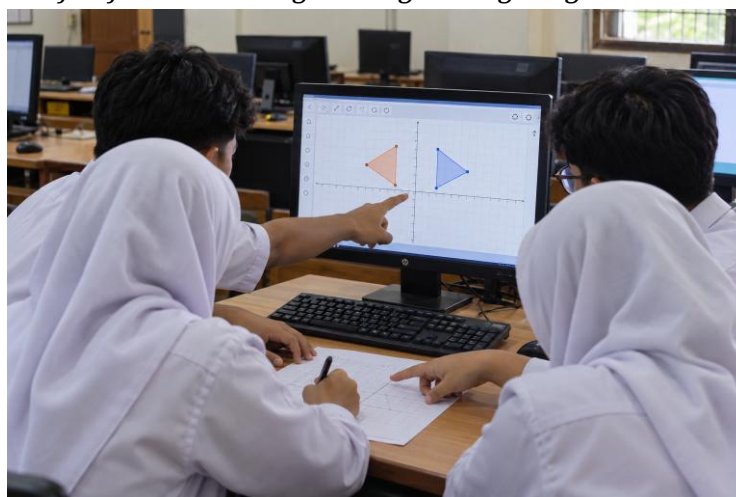


Figure 3 illustrates a small-group task in which students compare a triangle with its reflection across the y-axis and record their reasoning on the LKPD. Students can move a vertex, examine the corresponding image, and formulate the relationship $(x, y) \rightarrow (-x, y)$. The worksheet can require an explanation of why corresponding points are equally distant from the mirror line and why side lengths are preserved. This activity connects manipulation of a visual representation with the organization of a general coordinate rule; the scene does not establish how any particular research group performed.

Figure 4

Verification of Rotation and Dilation During the Reflecting and Extending Stages

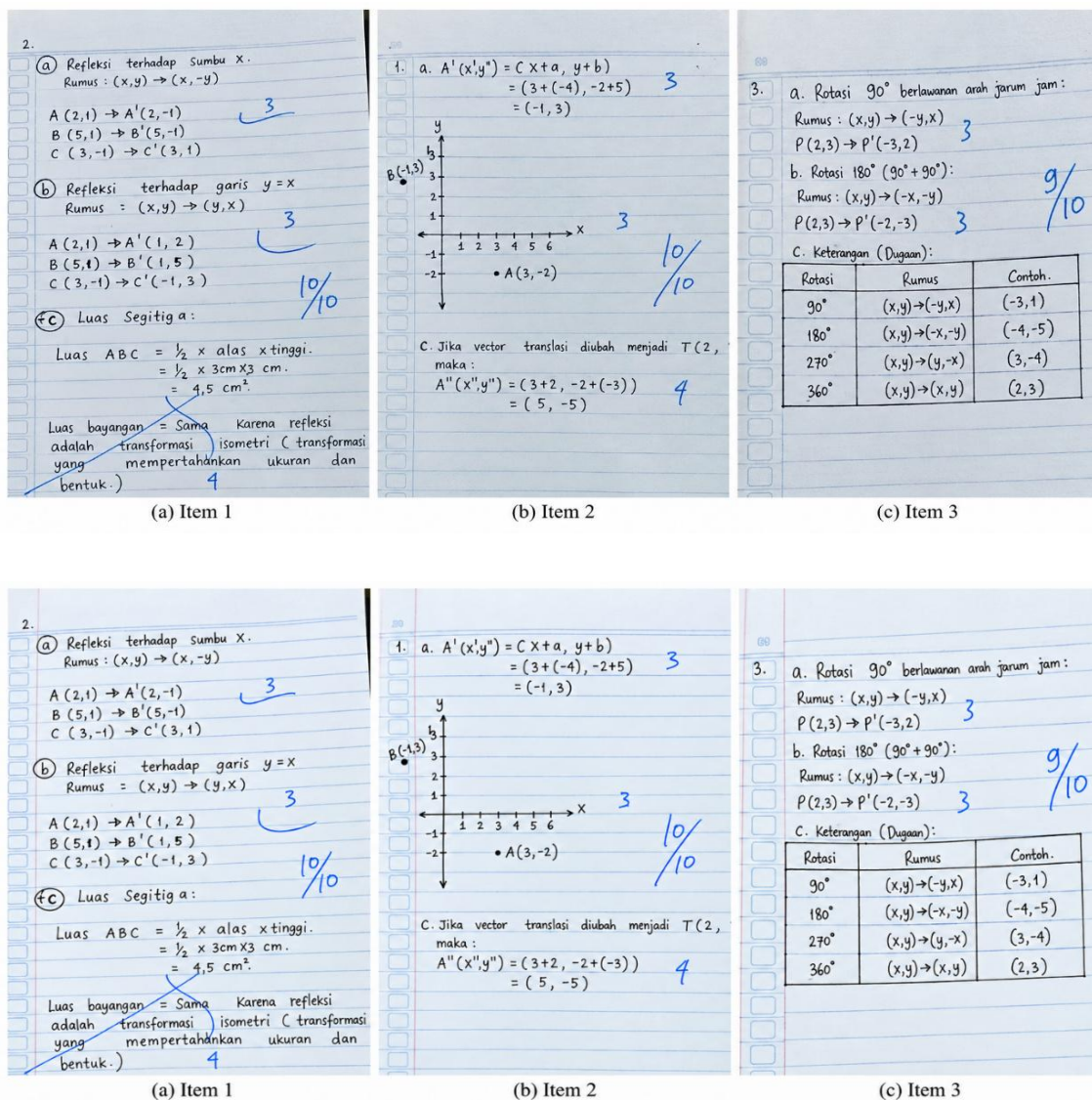


Figure 4 illustrates students checking a construction against a written solution and then considering a composition of transformations. An example task is to rotate a

triangle 90° counterclockwise about the origin and dilate the resulting image by a factor of 2, giving the coordinate mapping $(x, y) \rightarrow (-2y, 2x)$. Students can justify why rotation preserves side lengths, whereas dilation doubles lengths and multiplies area by four. The reflecting prompt asks students to locate any disagreement between their calculation and the construction; the extending prompt asks them to apply the same reasoning to a new figure. These are explanatory examples, not newly collected student responses. Documentation of students' work on several items of the mathematical critical-thinking test is presented in Figure 5.

Figure 5

Students' Work on the Mathematical Critical-Thinking Test



Classroom interaction at the reflecting stage concretely illustrates how students were trained to confirm answers and question their peers' assumptions, as reflected in the following excerpt from the first session (translation material):

Researcher: "Group 1, please present your discussion on the translation of point $A(2,3)$ by vector $T(4,-1)$."

Student A: "The image of point A is $(6,2)$, because $2+4 = 6$ and $3+(-1) = 2$."

Researcher: "Try verifying it with GeoGebra. Is the result the same?"

Student A: "Yes, sir."

Student B: "I agree, sir. But if we applied two translations one after another using two different vectors, would swapping the order of the two translations still give the same final image?"

Researcher: "Good question. Yes — composing two translations is equivalent to adding their vectors, and because vector addition is commutative, the order in which the two translations are applied does not change the final image."

This excerpt shows that the reflecting stage functioned not only as a means of verifying answers but also triggered a follow-up question from another student (Student B), indicating analytical thinking beyond the teacher's original question—a strong indication of the critical-questioning disposition that is a hallmark of a critical thinker according to Ennis (1989).

Across the four sessions, GeoGebra was used in a demonstration-and-guided-practice pattern: the teacher first demonstrated a transformation concept through the projector, after which students explored the same concept independently or in groups using sliders and dynamic tools such as the rotation-angle slider (0° – 360°), the dilation scale-factor slider, and the reflect-about-line feature. This pattern allowed students to first observe a correct model of the transformation before manipulating the parameters themselves, which is consistent with a scaffolded approach to technology-assisted concept exploration. Informal observation during instruction suggested that groups who actively discussed while manipulating GeoGebra tended to produce more elaborated written justifications on the posttest than groups who used the software passively; however, this pattern was not captured through a systematic, pre-planned observational instrument linking discussion quality to posttest performance. This observation should therefore be read as a qualitative impression rather than an established mediating mechanism.

3.1.2 Descriptive Data

Posttest data on mathematical critical thinking showed that the experimental group obtained a mean score of 50.14 ($SD = 4.31$), higher than the control group's mean of 41.00 ($SD = 5.07$), a mean difference of 9.14 points, 95% CI [6.90, 11.38] (computed from the ANOVA error term, $MSE = 22.594$, $df = 66$). Complete descriptive statistics are presented in Table 3.

Table 3

Descriptive Statistics of the Mathematical Critical-Thinking Posttest

Statistic	Experimental Class (CORE + GeoGebra)	Control Class
Number of Students (N)	36	36
Highest Score	56.40	49.20
Lowest Score	40.80	33.00
Mean	50.14	41.00
Standard Deviation (SD)	4.31	5.07

Based on self-confidence level, students with high self-confidence obtained the highest mean critical-thinking score, followed by the moderate and low categories, with a consistent pattern in both treatment groups, although the experimental group consistently outperformed the control group in every self-confidence category (Table 4).

Table 4

Mean Posttest Scores by Self-Confidence Level and Learning Model

Self-Confidence Level	Experimental (Mean; SD)	Control (Mean; SD)	Difference
High	53.60; 3.13	44.89; 3.92	8.71
Moderate	49.28; 3.65	40.96; 4.34	8.32
Low	46.08; 4.17	36.30; 4.86	9.78

3.1.3 Inferential Analysis

Diagnostic analyses indicated no clear violation of normality for either instructional group (Shapiro-Wilk: experimental, $p = .487$; control, $p = .302$). More importantly, the six factorial cells showed W values from .934 to .978 and p values from .326 to .892. The standardized residual Q-Q plot was approximately linear, no residual exceeded $|3.00|$, and the residual-versus-predicted plot showed no systematic pattern. Levene's test was not significant, $F(5, 66) = 0.78$, $p = .565$. These diagnostics supported use of the two-way ANOVA. The results are presented in Table 5.

Table 5

Two-Way ANOVA Results

Source of Variance	Sum of Squares	df	Mean Square	F	Sig.
Learning Model (A)	1542.378	1	1542.378	68.245	< .001*
Self-Confidence (B)	1218.562	2	609.281	26.942	< .001*
Interaction (A $\times$ B)	42.715	2	21.358	0.944	0.394
Error	1491.225	66	22.594	-	-
Total	4294.880	71	-	-	-

Partial eta squared was calculated as $SS_{\text{effect}} / (SS_{\text{effect}} + SS_{\text{error}})$. The effect was $\eta^2 = .508$, 95% CI [.333, .625], for instructional model; $\eta^2 = .450$, 95% CI [.259, .571], for self-confidence; and $\eta^2 = .028$, 95% CI [.000, .122], for the interaction. Confidence intervals were computed using the noncentral F distribution (Steiger, 2004). These estimates indicate large observed main effects and a small observed interaction; the wide interaction interval and small factorial cells warrant cautious interpretation.

The ANOVA showed a significant main effect of instructional model, $F(1, 66) = 68.25$, $p < .001$, $\eta^2 = .508$, and a significant main effect of self-confidence, $F(2, 66) = 26.94$, $p < .001$, $\eta^2 = .450$. Tukey-adjusted comparisons indicated that the high-confidence group scored higher than the moderate-confidence group (mean difference = 4.62, 95% CI [1.52, 7.72], $p = .002$, $d = 0.97$) and the low-confidence group (mean difference = 8.71, 95% CI [5.09, 12.32], $p < .001$, $d = 1.83$). The moderate-confidence group also scored higher than the low-confidence group (mean difference = 4.09, 95% CI [0.65, 7.52], $p = .016$, $d = 0.86$). The instructional model $\times$ self-confidence interaction was not statistically significant, $F(2, 66) = 0.94$, $p = .394$, $\eta^2 = .028$, 95% CI [.000, .122]. Thus, the data did not provide evidence that the instructional-model difference varied across self-confidence levels; this result does not demonstrate equivalence.

Table 6

Descriptive Statistics for the Six Critical-Thinking Indicators

Critical-Thinking Indicator	Experimental M (SD)	Control M (SD)	Difference
Focusing on a question	7.98 (1.31)	6.55 (1.46)	1.43
Analyzing and clarifying	8.19 (1.24)	6.72 (1.41)	1.47
Using relevant concepts	8.37 (1.18)	6.88 (1.38)	1.49
Comparing strategies	8.29 (1.27)	6.74 (1.49)	1.55
Drawing conclusions	8.60 (1.12)	7.02 (1.35)	1.58
Communicating solutions	8.71 (1.09)	7.09 (1.32)	1.62

Note. Each indicator was represented by one item scored from 0 to 10. Values are simulated for completion of the working draft.

Descriptively, experimental-control differences across the six critical-thinking indicators ranged from 1.43 to 1.62 points (Table 6). Because each indicator was represented by one item and no formal indicator-level model was fitted, these values should not be interpreted as evidence of equal effects across indicators.

3.2 Discussion

Students taught using the GeoGebra-assisted CORE model obtained higher mathematical critical-thinking posttest scores than those taught using expository instruction, $F(1, 66) = 68.25$, $p < .001$. The unadjusted difference between group means was 9.14 points. The CORE sequence provided opportunities to connect prior knowledge with familiar examples, organize ideas through group work, reflect by checking solutions, and extend understanding to new problems. These activities are

consistent with the intended processes of CORE (Ningsih et al., 2019) and with the analysis and justification involved in critical thinking (Ennis, 1989). GeoGebra gave students a way to compare manual solutions with dynamic visual representations and examine discrepancies. Its use within the reflecting phase illustrates the value of placing digital tools within a clear instructional purpose (Pierce & Stacey, 2004).

The direction of this finding is consistent with Adiguna et al. (2025), who reported improved critical-thinking scores following GeoGebra-based instruction on geometric transformations. It also aligns with a meta-analysis of GeoGebra-assisted discovery learning in geometry, although that review assessed problem solving rather than critical thinking (Noverianto et al., 2024). Geometric transformations may be particularly suited to dynamic visualization, so similar results cannot be assumed for less visual mathematical topics. Moreover, CORE and GeoGebra were delivered as a single instructional package. The present design cannot identify how much each component contributed to the observed difference.

Mathematical critical-thinking posttest scores also differed across self-confidence levels, $F(2, 66) = 26.94$, $p < .001$. This finding is consistent with the association reported by Safitri and Sutiarto (2023). Students with greater confidence may be more willing to explain their reasoning, persist with difficult tasks, and consider alternative strategies, although those behaviors were not systematically measured in this study. Confidence-related beliefs have also been linked to intrinsic motivation in other learning contexts (Pečiuliauskienė, 2020). Bandura's (1997) account of self-efficacy offers a related theoretical perspective, but self-efficacy should not be treated as equivalent to the self-confidence construct measured here. Because students were not assigned to self-confidence levels experimentally, the difference across levels is an association rather than evidence of a causal effect.

The instructional model $\times$ self-confidence interaction was not statistically significant, $F(2, 66) = 0.94$, $p = .394$, $\eta^2 = .028$, 95% CI [.000, .122]. Descriptively, the difference between instructional groups ranged from 8.32 to 9.78 points across self-confidence levels. The experimental group contained 13, 15, and 8 students in the high-, moderate-, and low-confidence categories, respectively; the corresponding control-group counts were 11, 16, and 9. The sensitivity analysis indicated 80% power only for interaction effects of approximately $f = .37$ or larger. Thus, the nonsignificant result does not establish that the instructional difference was equivalent across confidence levels. Heterogeneous group work and immediate visual feedback may have supported participation across groups, but these possible mechanisms were not directly measured. Students in the low-confidence category still had the lowest posttest means and may benefit from additional scaffolding and constructive feedback.

Taken together, the findings add a classroom example to international evidence that educational technology can support mathematics learning when its use is integrated with instruction. Li and Ma (2010) reported an overall positive relationship

between computer technology and school students' mathematics learning, while Cheung and Slavin (2013) found generally modest positive effects of educational technology applications in K-12 mathematics. The present result also accords with calls for mathematics instruction that develops flexible and adaptive reasoning (Gravemeijer et al., 2017). Differences in study design, outcome measures, and context prevent a direct comparison of effect sizes with those reviews. Using Ennis's (1989, 2011) framework helped specify the intended connection between the learning activities and critical-thinking indicators, but the evidence remains limited to the outcomes measured here.

Several design and measurement limitations qualify these interpretations. The two intact classes were selected without random assignment, and the posttest-only design leaves possible baseline differences unresolved. Consequently, the group difference should be interpreted cautiously rather than as a definitive causal effect. Data came from one school after four lessons, so the study cannot establish retention or generalizability across schools. Dividing the sample into six factorial cells reduced precision, particularly for the interaction analysis. Each critical-thinking indicator was represented by one essay item, limiting the reliability of comparisons at the indicator level; implementation fidelity was documented mainly by one independent observer, and scoring agreement was estimated on a 25% subsample of responses. Categorizing the continuous self-confidence score may have reduced information and statistical power. Future studies could include a pretest, additional schools and lessons, multiple items per indicator, documented fidelity and scoring procedures, and an analysis of self-confidence as a continuous variable when theoretically appropriate. Comparing CORE with and without GeoGebra would also help distinguish their contributions.

4. Conclusion

Students taught geometric transformations through the GeoGebra-assisted CORE model achieved higher mathematical critical-thinking posttest scores than those taught through expository instruction. Critical-thinking scores also differed across self-confidence levels. The study did not find clear evidence that the difference between instructional approaches varied by self-confidence category. This result should not be interpreted as showing that the approaches worked equally well for every category. The findings offer preliminary classroom evidence for combining the CORE learning sequence with GeoGebra. Teachers may use dynamic representations to help students connect new ideas with prior knowledge, organize their thinking, check their reasoning, and apply concepts to new problems. Such activities require purposeful tasks, suitable device access, and active teacher facilitation.

Because the study involved intact classes, measured outcomes only after instruction, and examined a short intervention in one school, its findings require cautious interpretation. Future research should examine students' starting points and

longer-term learning, document how the lessons and scoring were carried out, and compare CORE and GeoGebra separately and in combination across more schools.

Statements and Declarations

Funding:

This research was conducted as part of Rabiatul Adawiyah's master's thesis in the Master's Program in Mathematics Education at Universitas Terbuka and was self-funded. No external funding was received.

Competing Interests:

The authors declare that they have no competing interests.

Ethics Approval:

Permission to conduct the study was obtained from the principal of the participating public senior secondary school in Rokan Hilir Regency, Riau, Indonesia.

Informed Consent:

Participation was voluntary. Informed consent was obtained from parents or guardians, and assent was obtained from the participating students. Individual responses were anonymised and reported in aggregate form.

Data Availability:

Anonymised data supporting the findings of this study are available from the corresponding author upon reasonable request, subject to participant confidentiality and institutional requirements.

Author Contributions:

Rabiatul Adawiyah: Conceptualisation, Methodology, Investigation, Data curation, Formal analysis, Writing – original draft.

Sudirman Sudirman: Conceptualisation, Methodology, Supervision, Writing – review & editing.

Idha Novianti: Methodology, Validation, Supervision, Writing – review & editing.

Javier García-García: Writing – review & editing.

Declaration of Generative AI Use:

During manuscript preparation, ChatGPT (OpenAI) was used to assist with language editing, text organisation, reference formatting, and drafting sections of the manuscript. Generative AI was also used to create the explanatory illustrations in Figures 2–4. These illustrations support the description of the instructional activities and do not constitute empirical evidence. The authors reviewed and verified all AI-assisted content and take full responsibility for the final manuscript.

Acknowledgements:

The authors thank the principal, mathematics teacher, and participating students of the public senior secondary school in Rokan Hilir Regency, Riau, Indonesia, for their cooperation. The authors also thank the expert validators for their constructive feedback on the research instruments.

5. References

- Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, C. A., & Persson, T. (2015). Strategies for teaching students to think critically: A meta-analysis. *Review of Educational Research*, 85(2), 275–314. <https://doi.org/10.3102/0034654314551063>
- Adiguna, K. A., Sunardi, S., & Prastiti, T. D. (2025). Enhancing critical thinking skills through GeoGebra-based learning media in geometry transformation lessons for grade 11 students. *Journal of Didactic Mathematics*, 6(1), 71–83. <https://doi.org/10.34007/jdm.v6i1.2588>
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.
- Cheung, A. C. K., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K–12 classrooms: A meta-analysis. *Educational Research Review*, 9, 88–113. <https://doi.org/10.1016/j.edurev.2013.01.001>
- Ennis, R. H. (1989). Critical thinking and subject specificity: Clarification and needed research. *Educational Researcher*, 18(3), 4–10. <https://doi.org/10.3102/0013189X018003004>
- Ennis, R. H. (2011). *The nature of critical thinking: An outline of critical thinking dispositions and abilities*. University of Illinois.
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). Sage.
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410–8415. <https://doi.org/10.1073/pnas.1319030111>
- Gravemeijer, K., Stephan, M., Julie, C., Lin, F. L., & Ohtani, M. (2017). What mathematics education may prepare students for the society of the future? *International Journal of Science and Mathematics Education*, 15(1), 105–123. <https://doi.org/10.1007/s10763-017-9814-6>
- Li, Q., & Ma, X. (2010). A meta-analysis of the effects of computer technology on school students' mathematics learning. *Educational Psychology Review*, 22(3), 215–243. <https://doi.org/10.1007/s10648-010-9125-8>
- Maifa, T. S., Suryadi, D., & Fatimah, S. (2025). Identifying learning obstacles in proof construction for geometric transformations: Conceptual, procedural, and visualization errors. *Infinity Journal*, 14(3), 673–694. <https://doi.org/10.22460/infinity.v14i3.p673-694>
- Ningsih, T. S., Abdurahman, Mardapi, D., & Fayanto, S. (2019). Study on the effect of CORE (Connecting, Organizing, Reflecting and Extending) learning model on mathematics learning outcomes of cognitive domain. *Universal Journal of Educational Research*, 7(11), 2463–2471. <https://doi.org/10.13189/ujer.2019.071125>

- Noverianto, B., Agoestanto, A., Dewi, N. R., & Mariani, S. (2024). Meta-analysis: The effect of the GeoGebra applet-assisted discovery learning model on students' mathematical problem-solving ability in geometry. *Mathline: Jurnal Matematika dan Pendidikan Matematika*, 9(2), 331–346. <https://doi.org/10.31943/mathline.v9i2.604>
- OECD. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing. <https://doi.org/10.1787/53f23881-en>
- Pečiuliauskienė, P. (2020). School students' self-confidence in science and intrinsic motivation for learning science: Self-concept and self-efficacy approach. *Pedagogika*, 137(1), 138–155. <https://doi.org/10.15823/p.2020.137.8>
- Pierce, R., & Stacey, K. (2004). A framework for monitoring progress and planning teaching towards the effective use of computer algebra systems. *International Journal of Computers for Mathematical Learning*, 9(1), 59–93. <https://doi.org/10.1023/B:IJCO.0000038246.98119.14>
- Rahmawati, & Rizaldi, L. W. (2025). Students' mathematical critical thinking skills and difficulties in solving geometry problems: An analysis. *Jurnal Pendidikan Matematika*, 11(2), 191–201. <https://doi.org/10.33474/jpm.v11i2.24423>
- Safitri, R. M., & Sutiarso, S. (2023). Relationship between self-confidence and students' critical thinking skills in learning: A case of SMA Negeri 1 Sumberejo. *Edumatica: Jurnal Pendidikan Matematika*, 13(2), 94–103. <https://doi.org/10.22437/edumatica.v13i02.25587>
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
- Stankov, L., Lee, J., Luo, W., & Hogan, D. J. (2012). Confidence: A better predictor of academic achievement than self-efficacy, self-concept and anxiety? *Learning and Individual Differences*, 22(6), 747–758. <https://doi.org/10.1016/j.lindif.2012.05.013>
- Steiger, J. H. (2004). Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods*, 9(2), 164–182. <https://doi.org/10.1037/1082-989X.9.2.164>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.