Journal of Pedagogy and Education Science
Vol. 5, No. 03, pp. 1146-1166
journal.iistr.org/index.php/JPES
DOI: 10.56741/IISTR.jpes.002426



Effectiveness of GeoGebra-Assisted Praxeological Learning Based on the Anthropological Theory of the Didactical in Enhancing Students' Critical Thinking

[1]Nanang Diana*, [1]Irwan Aprayadi, [1]Adi Apriadi Adiansha, [1]M.Ibnusaputra, [1]Mariamah
Corresponding Author: * nanangdiana.math@gmail.com
[1] STKIP Taman Siswa Bima, Nusa Tenggara Barat, Indonesia

ARTICLE INFO

Article history
Received 12 June 2026
Revised 12 September 2026
Accepted 13 September 2026

ABSTRACT

Research integrating GeoGebra with the Anthropological Theory of the Didactic (ATD) remains limited, particularly regarding how dynamic visualisation within a structured praxeological framework influences students' critical thinking. This study examined the effectiveness of a GeoGebra-assisted learning sequence structured around four praxeological components: task, technique, technology, and theory. A quasi-experimental non-equivalent control group design involved 63 second-semester Differential Calculus students at STKIP Taman Siswa Bima, Indonesia, comprising 32 experimental and 31 control participants. Both groups received instruction from the same lecturer across eight 100-minute sessions. Critical thinking was assessed using a ten-item essay test covering interpretation, analysis, evaluation, inference, and explanation, with satisfactory content validity (0.89) and reliability (Cronbach's $\alpha = 0.84$). Pre-test scores showed no significant difference between groups ($p = 0.784$). Following the intervention, the experimental group significantly outperformed the control group ($t(61) = 5.274$, $p < 0.001$), with a mean difference of 11.53 (95% CI: 7.16–15.90) and a large effect size (Cohen's $d = 1.33$). Normalised learning gains were moderate in the experimental group (0.53) and low in the control group (0.28). Significant improvements were observed across all five critical thinking dimensions, although explanation remained the weakest. These findings suggest that integrating GeoGebra with ATD-based praxeological organisation strengthens critical thinking by connecting dynamic visual exploration with mathematical justification and generalisation. The proposed instructional framework offers a practical approach to improving undergraduate calculus learning.

Keywords
Anthropological Theory
Critical Thinking Skills
GeoGebra
Limit Concept
Praxeological Learning

This is an open-access article under the CC–BY-SA license.

Introduction

Mathematics education in the twenty-first century is judged less by how many procedures students can execute than by how well they can reason with them. Critical thinking—the capacity to interpret information, analyse relationships, evaluate arguments, draw inferences and explain one's reasoning—has accordingly moved from the margins of the curriculum towards its centre. The evidence that Indonesian learners have not yet reached this standard is difficult to dismiss. In PISA 2022, Indonesian fifteen-year-olds averaged 366 points in mathematics against an OECD mean of 472, a decline from 379 in 2018, and only a small minority reached the proficiency levels that demand reasoning beyond routine computation [1]. Students who enter university with this profile do not shed it automatically: studies of undergraduate mathematics report that reasoning-based tasks are attempted far less successfully than computational ones, and that the gap widens as the content becomes more abstract [2], [3]. Recent Indonesian evidence also indicates that mathematical critical thinking is not an isolated academic trait but one that carries over into how students read and judge information in other domains, which raises the stakes of leaving it undeveloped [4].

Within the undergraduate curriculum, the limit concept is where this weakness becomes most visible. Limits are the first topic in which students must reason about a process that is never completed, and four decades of research have documented the mismatch between the concept image learners build from examples and the concept definition they are asked to accept [5]. Students commonly read the limit of a function at a point as the value of the function at that point, treat approaching as reaching, and cannot say why a limit may exist where the function is undefined [6]. Instruction that moves quickly from the definition to substitution rules reinforces the misreading: learners acquire a dependable computational routine and, with it, no obligation to justify anything. Because the calculus sequence is cumulative, the interpretive and inferential habits that fail to form here are unlikely to form later.

Dynamic mathematics software has been the most common response to this problem. GeoGebra links algebraic, graphical, numerical and tabular registers in a single freely available environment [7], and allows a learner to drag a point towards a target value and watch the corresponding output stabilise, so that the process of approaching becomes visible rather than merely described [8]. The empirical record is encouraging. A meta-analysis of 36 effect sizes

drawn from 29 primary studies and 2,111 students placed the overall effect of GeoGebra on students' mathematical ability at 0.96 [9], and a context-specific meta-analysis of digital tools in mathematics and science reached a comparable conclusion while adding an important qualification: outcomes depend far more on how the tool is embedded in instruction than on the tool itself [10]. Systematic reviews of the field describe the same situation from another angle. Most studies concentrate on achievement and conceptual understanding at secondary level, comparatively few examine higher-order reasoning, and fewer still specify the didactic design in enough detail for another researcher to reproduce it [11], [12]. A related strand of work adds that what matters is not the availability of a digital resource but the documentational work through which lecturers and students turn that resource into an instrument of learning [33], [34].

Design, not software, is therefore the variable that decides whether visualisation becomes reasoning, and the Anthropological Theory of the Didactical (ATD) supplies a vocabulary for that design. Developed by Chevallard, ATD models any piece of mathematical activity as a praxeology built from four components: a type of task (T), a technique (τ) for carrying it out, a technology (θ) that explains and justifies the technique, and a theory (Θ) that grounds the technology [13]. The practical block (T, τ) is what students ordinarily receive; the logos block (θ, Θ) is what is ordinarily reserved for the lecturer. The central claim of ATD is that instruction delivering the practical block alone produces exactly the procedural fluency without justification that critical thinking research keeps identifying as a deficit, and that didactic transposition should be designed so that the logos block becomes something students construct rather than receive [14]. Applications of ATD at university level have shown how praxeologies can be reconstructed and analysed for particular content domains [15], [16], and Facione's widely used account of critical thinking describes cognitive operations—interpretation, analysis, evaluation, inference and explanation—that map onto this structure with unusual directness [17].

Three gaps follow from this body of work. First, the two research programmes have run largely in parallel: GeoGebra studies rarely commit to a didactic theory beyond generic constructivism, whereas ATD studies are predominantly analytic or design-oriented and seldom report experimental outcome data [11], [15]. Second, Indonesian studies that do combine technology with structured mathematics instruction typically report a single composite score, so it remains unknown whether such designs lift every facet of critical thinking equally or only those most directly supported by visualisation [18], [19]. Third, tertiary-level evidence is thin; most GeoGebra intervention studies in Indonesia involve secondary school samples, and the few addressing calculus report achievement rather than reasoning [19], [20], [36]. Even work that examines the theoretical fit between technology and

instructional design tends to stop at the level of conceptual understanding rather than argumentation [21]. What is missing is a study that specifies the didactic design in praxeological terms, implements it with GeoGebra in a university calculus course, and reports the effect on critical thinking dimension by dimension together with an effect size that allows comparison with international evidence.

The present study addresses precisely that combination. The problem it investigates is whether organising GeoGebra-based limit instruction around the four praxeological components produces a measurably larger improvement in undergraduate students' critical thinking than conventional instruction covering the same content in the same time. Its novelty lies less in the use of GeoGebra, which is well established, than in the role assigned to the technology component of praxeology. Dragging, tracing and zooming generate observations, but in this design no observation may be generalised until students have produced a written account of why the observed behaviour occurs. The software supplies the phenomenon and the praxeological structure supplies the obligation to account for it, and it is the second element—absent from most GeoGebra studies—that is expected to convert visual exploration into argument. A second novel element is analytical rather than instructional: instead of a single composite score, the study maps each praxeological component onto the critical thinking dimension it is designed to serve and tests the five dimensions separately.

Three objectives follow. The first is to determine whether students taught through GeoGebra-assisted praxeological learning differ significantly from students taught conventionally in post-test critical thinking performance. The second is to quantify the magnitude of any difference through normalised gain and standardised effect size rather than through statistical significance alone. The third is to describe and test the pattern of improvement across the five critical thinking dimensions. Because the study relies on formal hypothesis testing, the hypotheses are stated explicitly:

- H_0: $\mu_1 = \mu_2$. There is no significant difference in post-test critical thinking scores between students who receive GeoGebra-assisted praxeological learning and students who receive conventional instruction.
- H_1: $\mu_1 > \mu_2$. Students who receive GeoGebra-assisted praxeological learning obtain significantly higher post-test critical thinking scores than students who receive conventional instruction.

Here μ_1 and μ_2 denote the population mean post-test scores of the experimental and control groups respectively, and the hypotheses were tested at $\alpha = 0.05$.

The contribution of the study is threefold. Theoretically, it moves ATD from its customary analytic use into an experimentally evaluated instructional design and proposes an explicit correspondence between praxeological components and critical thinking dimensions.

Methodologically, it reports baseline equivalence testing, effect sizes with confidence intervals and dimension-level comparisons—practices that remain uncommon in Indonesian mathematics education research and that make the findings commensurable with international evidence. Practically, it documents an eight-session sequence, including the tasks, the GeoGebra actions and the justification prompts used in each session, so that the design can be reproduced, adapted to other calculus topics, or tested in other institutions. The remainder of the paper describes the method, reports the results, discusses them against prior evidence, and closes with the limitations that qualify them.

Material and Methods

A. Research Design

The study used a quasi-experimental, non-equivalent control group design with pre-test and post-test measurement. The design was chosen because student registration at the research site assigns undergraduates to classes administratively at the start of the semester, so dissolving those classes to randomise individuals would have disrupted both the timetable and the cohort structure. Working with intact classes preserves the ecological validity of the setting, but it also means that pre-existing differences between groups cannot be ruled out by randomisation alone. The design therefore incorporates the compensating procedures recommended for non-equivalent designs: pre-test measurement, statistical testing of baseline equivalence, a single lecturer across both conditions, identical instructional time, and an explicit examination of rival explanations at the analysis stage [22], [23]. In symbolic form the design is O_1 X O_2 for the experimental group and O_1 – O_2 for the control group, where O denotes measurement and X the praxeological treatment. Fig. 1 summarises the whole procedure.

B. Participants and Group Assignment

Participants were 63 second-semester undergraduates enrolled in Differential Calculus in the Mathematics Education study programme of STKIP Taman Siswa Bima, West Nusa Tenggara, Indonesia, during the even semester of the 2025/2026 academic year. The two classes offered that semester, class A (n = 32) and class B (n = 31), together constituted the entire population of students taking the course. No sampling procedure was applied and no participant was excluded, so the group sizes reflect enrolment rather than a sampling decision; this is why the groups are of slightly unequal size. Allocation of the two intact classes to conditions was decided by lot after the pre-test had been administered and scored, which means that neither the lecturer nor the students knew at the time of baseline measurement which class would receive the treatment. Both classes met in the same computer-equipped room, at equivalent times of day, and were taught by the same lecturer throughout, so that instructor identity cannot account for any difference observed between them.

Although the sample size was fixed by enrolment, its adequacy was checked rather than assumed. An a priori power analysis conducted in G*Power 3.1 for a two-tailed independent samples t-test with $\alpha = 0.05$ and power = 0.80 indicated that 52 participants, 26 per group, are required to detect a large effect of $d = 0.80$, and that the 63 participants available yield an achieved power of approximately 0.88 for that effect size [24]. The sample is thus sufficient for detecting the medium-to-large effects typically reported in the GeoGebra literature [9], but underpowered for small effects, a constraint returned to in the limitations.

C. Instructional Procedure

The experimental treatment ran for eight weekly sessions of 100 minutes, equivalent to two fifty-minute credit hours per session and 800 minutes of instruction in total. Every session followed the same four-phase cycle derived from the praxeological components. In the task phase (approximately fifteen minutes) the lecturer posed a limit problem framed as a question about behaviour rather than about a value. In the technique phase (approximately thirty minutes) students worked in pairs with GeoGebra, dragging points, tracing values, adjusting sliders and tabulating outputs in order to establish what the function does near the point of interest. In the technology phase (approximately thirty-five minutes), which is the pivot of the design, each pair wrote an account of why the observed behaviour occurs and exchanged it with another pair for critique; no result was accepted by the class until such an account had been produced and defended. In the theory phase (approximately twenty minutes) the lecturer consolidated the accepted accounts into a formal statement, connecting them to the definitions and theorems in the course textbook. Table 1 sets out the content of the eight sessions, the GeoGebra actions used in each, and the critical thinking dimension each session was designed to exercise.

Table 1. Praxeological Design of the Eight Instructional Sessions

Session	Limit sub-topic	Praxeological activity (T, τ, θ, Θ)	Targeted indicator
1	The intuitive meaning of a limit	T: determine the value a function approaches near a given point from a table and a graph. τ: drag a point along the curve and read the corresponding output in the algebra view. θ: explain why the value approached need not equal the value of the function at that point. Θ: informal statement of the limit.	Interpretation
2	One-sided limits	T: decide whether the limit of a piecewise function exists at the junction point. τ: approach the point from each side using a slider and record both sequences of outputs. θ: justify why a two-sided limit fails when the one-sided limits differ. Θ: definition of left- and right-hand limits.	Interpretation, Analysis
3	Algebraic limits and the	T: evaluate limits that produce an indeterminate form. τ: display the original and the simplified expression in two graphic views	Analysis

Session	Limit sub-topic	Praxeological activity (T, τ, θ, Θ)	Targeted indicator
	indeterminate form 0/0	and compare them. θ: explain why factoring changes the function but not the limit. Θ: removable discontinuity and the quotient law.	
4	The scope of the limit laws	T: decide whether a proposed limit law applies to a given pair of functions. τ: test the proposal on dynamic examples and search for a counterexample. θ: identify the hypothesis violated by the counterexample. Θ: formal statement of the sum, product and quotient laws.	Evaluation
5	Infinite limits and vertical asymptotes	T: describe the behaviour of a rational function near a zero of the denominator. τ: zoom and trace near the asymptote and tabulate the values obtained. θ: distinguish an infinite limit from a limit that does not exist. Θ: definition of a vertical asymptote.	Evaluation, Inference
6	Limits at infinity and horizontal asymptotes	T: compare the long-run behaviour of several rational functions. τ: vary the degrees of numerator and denominator with sliders and record the pattern. θ: justify the conjecture by comparing degrees rather than by inspecting graphs. Θ: theorem on limits of rational functions at infinity.	Inference
7	Continuity and the relation between limit and function value	T: classify the discontinuities of a given function. τ: construct examples and counterexamples in GeoGebra. θ: justify each classification against the three conditions for continuity. Θ: definition of continuity at a point.	Inference, Explanation
8	Synthesis: constructing and defending a praxeology	T: pose and solve an original limit problem. τ: build a GeoGebra applet that displays the solution. θ: write a justification of every step for a peer audience and respond to its critique. Θ: locate the problem within the limit theory developed in Sessions 1–7.	Explanation

The control group covered the same eight topics in the same eight sessions of 100 minutes with the same lecturer and the same textbook, but through conventional expository instruction: explanation of the concept, worked examples presented on the board with static graphs, individual exercises from the textbook, and a plenary discussion of the answers. GeoGebra was not used, and no structured justification phase was included. A written lesson protocol specifying the content to be covered in each session was followed in both conditions, so that the two groups differed in how the content was organised for the learner and not in what was taught or for how long.

D. Instruments, Validity and Reliability

The dependent variable was critical thinking in the context of limits, operationalised through a test of ten open-ended items, two for each of the five dimensions proposed by Facione—interpretation, analysis, evaluation, inference and explanation [17].

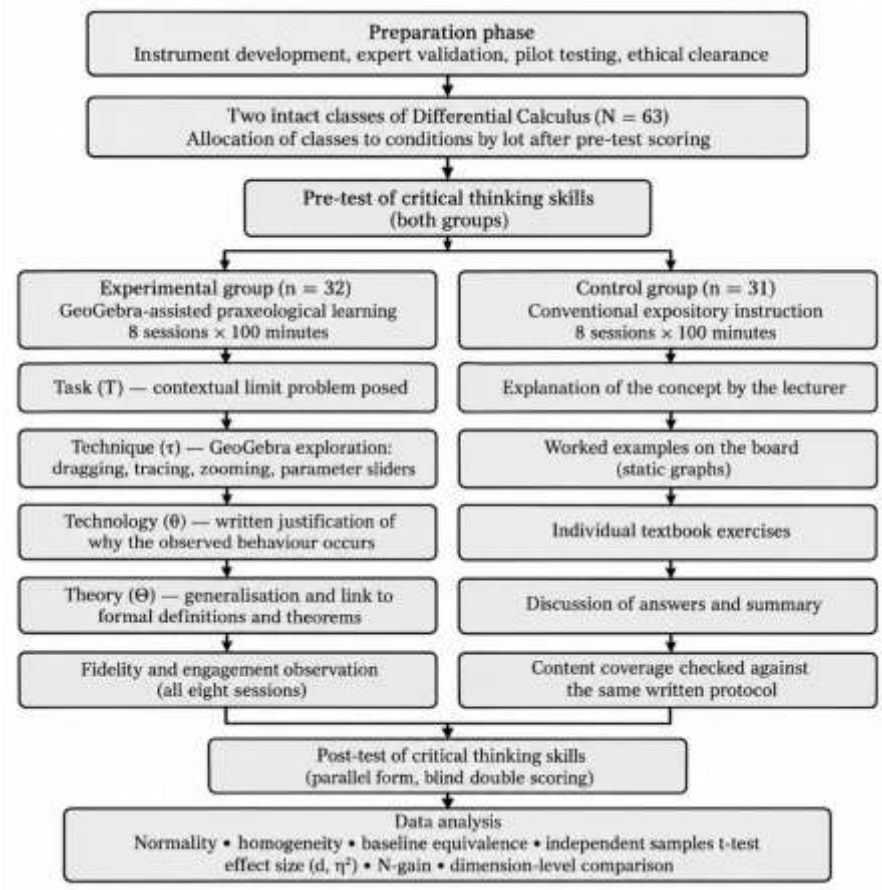


Fig. 1. Procedure of the Study

Each item required students to work with limit problems presented in graphical, tabular or symbolic form and to supply written reasoning rather than a numerical answer alone. Responses were marked with an analytic rubric on a scale of 0 to 4 per item, and the total was converted to a 0–100 scale. The pre-test and post-test were parallel forms with identical structure, item count and difficulty profile but different function families, and all scripts were scored blind to group membership. Table 2 presents the operational definition of each indicator, and Table 3 summarises the validation evidence.

Table 2. Indicators and Instruments for Measuring Critical Thinking Skills

Variable	Indicator	Operational description	Instrument	Item type
Critical thinking	Interpretation	Ability to understand and interpret limit problems presented in symbolic, tabular or graphical form	Critical thinking test (items 1–2)	Essay
Critical thinking	Analysis	Ability to identify relationships between variables and analyse the behaviour of a function near a point	Critical thinking test (items 3–4)	Essay
Critical thinking	Evaluation	Ability to assess solution strategies and judge the correctness of a given line of reasoning	Critical thinking test (items 5–6)	Essay
Critical thinking	Inference	Ability to draw logical conclusions about limit values from several representations	Critical thinking test (items 7–8)	Essay
Critical thinking	Explanation	Ability to justify reasoning and communicate a mathematical argument clearly in writing	Critical thinking test (items 9–10)	Essay

Two forms of validity were established rather than asserted. Content validity was examined by three experts—two mathematics education lecturers holding doctoral degrees and one senior calculus lecturer—who rated every item for relevance to the target dimension, clarity of wording and alignment with the course outcomes. Aiken's V was computed for each item and ranged from 0.78 to 0.94, all above the critical value of 0.75 for three raters at $\alpha = 0.05$ [25], giving a content validity index of 0.89 for the instrument as a whole. Empirical construct validity was examined in a pilot administration to 30 students from a different study programme who had already completed the course. Pearson item–total correlations ranged from 0.412 to 0.781, all exceeding the critical value of 0.361 for $n = 30$ at $\alpha = 0.05$, so all ten items were retained. Internal consistency reached a Cronbach's alpha of 0.84 for the whole test, with dimension-level coefficients between 0.71 and 0.83, and agreement between two independent scorers on 30% of the scripts yielded an intraclass correlation of 0.91.

The observation protocol used to monitor implementation contained twelve items, eight covering fidelity to the four praxeological phases and four covering student engagement during exploration, each rated on a four-point scale. Its content was validated by two experts (Aiken's V = 0.85), and agreement between the independent observer and a second rater on jointly observed sessions reached 92%.

The study protocol was reviewed and approved by the Institute for Research and Community Service (LPPM) of STKIP Taman Siswa Bima before data collection began. Students were informed of the purpose, the procedures and the voluntary nature of their participation prior to the pre-test, and all provided written informed consent; no incentive was offered, test scores were used for research purposes only and reported in aggregate, and the GeoGebra materials were made available to the control group after the post-test so that no student was disadvantaged by the condition to which their class had been assigned.

E. Data Analysis

Data were analysed in SPSS 27. Descriptive statistics characterised the score distributions. The assumption of normality was checked with the Kolmogorov–Smirnov test and the assumption of equal variances with Levene's test, in each case at $\alpha = 0.05$. Baseline equivalence was then tested with an independent samples t-test on the pre-test scores, because comparable means alone do not establish that two groups start from the same place.

The research hypothesis was tested with an independent samples t-test on the post-test scores. Because a p-value indicates only whether a difference is detectable and not how large it is, three complementary indices were computed: the 95% confidence interval of the mean difference, Cohen's d with the pooled standard deviation, interpreted against the conventional benchmarks of 0.20, 0.50 and 0.80 [26], and eta squared obtained from the t statistic.

Table 3. Summary of the Validity and Reliability Analysis

Instrument	Type of validity or reliability	Procedure and criterion	Result	Decision
Critical thinking test (10 essay items)	Content validity	Aiken's V from three expert raters; criterion V ≥ 0.75	V = 0.78–0.94; content validity index = 0.89	All items relevant
Critical thinking test	Empirical construct validity	Pearson item–total correlation in a pilot with 30 students; criterion r > 0.361	r = 0.412–0.781	All 10 items retained
Critical thinking test	Internal consistency	Cronbach's alpha for the whole test and by dimension	α = 0.84 (dimensions 0.71–0.83)	High reliability
Critical thinking test	Scorer reliability	Two independent blind scorers on 30% of scripts; intraclass correlation	ICC = 0.91	Excellent agreement
Observation protocol (12 items)	Content validity	Aiken's V from two expert raters	V = 0.85	Acceptable
Observation protocol	Inter-observer agreement	Percentage agreement on jointly observed sessions	92%	Acceptable

Hedges' correction for small samples was also applied [27]. The magnitude of learning improvement relative to the maximum attainable was quantified with the normalised gain of Hake and classified as low, moderate or high [28]. Finally, the five critical thinking dimensions were compared between groups with five independent samples t-tests using a Bonferroni-adjusted significance level of 0.01 to control the family-wise error rate; comparisons among the dimensions themselves were not tested inferentially and are reported descriptively.

Results

A. Descriptive Statistics

Table 4 reports the distribution of critical thinking scores before and after the intervention, and Fig. 2 displays the same information graphically together with the normalised gain.

Table 4. Descriptive Statistics of Critical Thinking Scores

Group	Test	N	Mean	Std. Deviation	Minimum	Maximum
Experimental	Pre-test	32	54.31	8.42	40	68
Experimental	Post-test	32	78.65	7.51	65	92
Control	Pre-test	31	53.74	7.98	41	67
Control	Post-test	31	67.12	8.36	52	82

The two groups entered the study at practically the same level (experimental M = 54.31, SD = 8.42; control M = 53.74, SD = 7.98), a difference of 0.57 of a point on a 100-point scale. After the intervention the experimental group had gained 24.34 points and the control group 13.38, so the experimental group improved by roughly four-fifths again as much. Two features of the dispersion are worth noting. The standard deviation of the experimental group fell from 8.42 to 7.51, indicating that the gain was distributed across the class rather than concentrated in its strongest members, whereas the control group's dispersion widened slightly from 7.98

to 8.36. The minimum score in the experimental group after instruction (65) exceeded the pre-test mean of both groups, which suggests that even the lowest-performing students in that class reached a level of reasoning that was above average before the treatment.

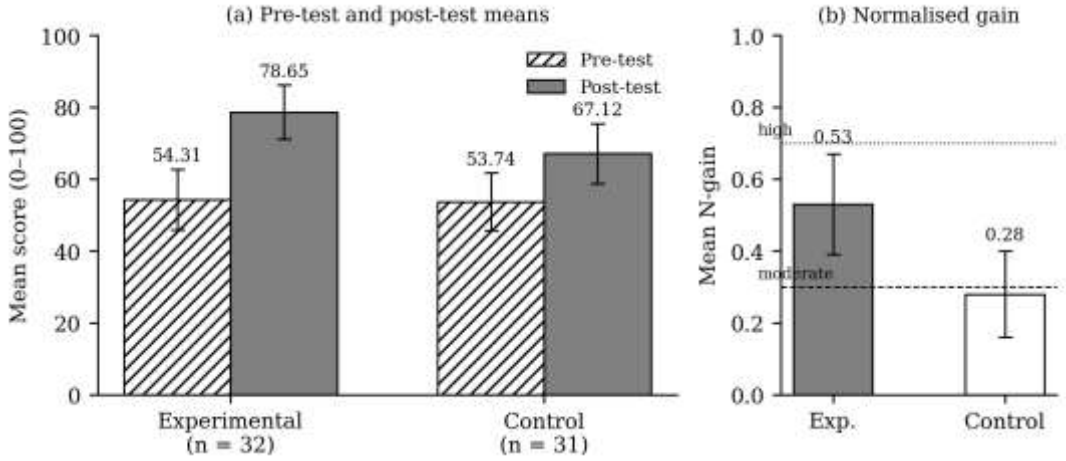


Fig. 2. Pre-test and Post-test Means and Normalised Gain by Group

B. Prerequisite Tests

Before hypothesis testing, the assumptions underlying parametric analysis were examined. Table 5 reports the Kolmogorov–Smirnov test of normality and Table 6 Levene's test of homogeneity of variance.

Table 5. Normality Test Results (Kolmogorov–Smirnov)

Group	Test	Statistic	Sig. value	Interpretation
Experimental	Pre-test	0.107	0.200	Normal
Experimental	Post-test	0.142	0.089	Normal
Control	Pre-test	0.118	0.174	Normal
Control	Post-test	0.139	0.097	Normal

Table 6. Homogeneity of Variance Test Results (Levene)

Variable	Levene statistic	df1	df2	Sig. value	Interpretation
Pre-test scores	0.412	1	61	0.523	Homogeneous
Post-test scores	1.284	1	61	0.261	Homogeneous

All significance values in Table 5 exceeded 0.05, so the null hypothesis of normality was retained for both groups on both occasions, and Levene's test was non-significant for the pre-test ($F = 0.412$, $p = 0.523$) and for the post-test ($F = 1.284$, $p = 0.261$). The assumptions of normality and equal variances were therefore satisfied, and the independent samples t-test was applicable without correction.

C. Baseline Equivalence and Hypothesis Testing

Table 7 reports the inferential comparison of the two groups on the pre-test, on the post-test and on the normalised gain, together with the associated effect sizes and confidence intervals.

Table 7. Independent Samples t-Test Results with Effect Sizes

Measure	t	df	Sig. (2-tailed)	Mean difference	95% CI	Cohen's d	η^2
Pre-test scores	0.276	61	0.784	0.57	[−3.57, 4.71]	0.07	0.001
Post-test scores	5.274	61	0.000	11.53	[7.16, 15.90]	1.33	0.313
Normalised gain	7.601	61	0.000	0.25	[0.18, 0.32]	1.92	0.487

The pre-test comparison was clearly non-significant, $t(61) = 0.276$, $p = 0.784$, mean difference = 0.57, 95% CI [−3.57, 4.71], $d = 0.07$. The confidence interval is narrow and centred close to zero, so the groups can be treated as equivalent in critical thinking before the intervention, and the post-test comparison can be interpreted as an effect of condition rather than of pre-existing advantage.

On the post-test the difference was significant, $t(61) = 5.274$, $p < 0.001$, mean difference = 11.53, 95% CI [7.16, 15.90]. The null hypothesis H_0 is therefore rejected and the alternative hypothesis H_1 accepted. The accompanying effect sizes show that the difference is not merely detectable but substantial: Cohen's $d = 1.33$ (Hedges' $g = 1.31$) falls well above the 0.80 benchmark for a large effect, and $\eta^2 = 0.313$ indicates that about 31% of the variance in post-test scores is associated with the instructional condition. The lower bound of the confidence interval, 7.16 points, is itself educationally meaningful, so the conclusion does not depend on the point estimate alone.

D. Normalised Gain Analysis

Normalised gain was computed for each student individually and averaged within groups; the results appear in Table 8.

Table 8. Normalised Gain Analysis Results

Group	N	Mean N-gain	SD	Minimum	Maximum	Category
Experimental	32	0.53	0.14	0.24	0.79	Moderate
Control	31	0.28	0.12	0.08	0.52	Low

The experimental group achieved a mean normalised gain of 0.53, which falls in the moderate category, against 0.28 for the control group, which falls in the low category [28]. The difference of 0.25 gain units was significant, $t(61) = 7.601$, $p < 0.001$, $d = 1.92$, 95% CI [0.18, 0.32]. The distributions are also informative: the lowest individual gain in the experimental group (0.24) was close to the control group's mean, and the highest control-group gain (0.52) barely reached the experimental mean, so the two distributions overlap only at their extremes. Interpreted as efficiency rather than as raw improvement, the experimental group converted a little over half of the improvement available to it, whereas the control group converted little more than a quarter.

E. Analysis by Critical Thinking Indicator

Table 9 compares the two groups on each of the five critical thinking dimensions, and Fig. 3 displays the same comparison. Each dimension was tested separately with a Bonferroni-adjusted significance level of 0.01.

Table 9. Between-Group Comparison Across the Five Critical Thinking Indicators

Indicator	Experimental M ± SD (%)	Control M ± SD (%)	Mean diff.	t(61)	p	Cohen's d
Interpretation	82.4 ± 9.14	71.2 ± 10.02	11.2	4.637	<0.001	1.17
Analysis	79.6 ± 9.86	66.8 ± 10.41	12.8	5.012	<0.001	1.26
Evaluation	76.3 ± 10.52	64.5 ± 10.88	11.8	4.376	<0.001	1.10
Inference	80.1 ± 9.47	68.2 ± 10.23	11.9	4.793	<0.001	1.21
Explanation	75.8 ± 10.94	63.7 ± 11.26	12.1	4.326	<0.001	1.09

The experimental group scored significantly higher on every dimension, with all five comparisons remaining significant after Bonferroni adjustment and all five effect sizes exceeding 1.00. Two patterns in Table 9 deserve separate attention because they answer different questions. In terms of attainment, the experimental group performed best on interpretation (82.4%) and inference (80.1%) and least well on explanation (75.8%), and the control group showed the same internal ordering at a lower level. In terms of the size of the group difference, however, the five dimensions were remarkably similar, ranging only from 11.2 percentage points for interpretation to 12.8 for analysis.

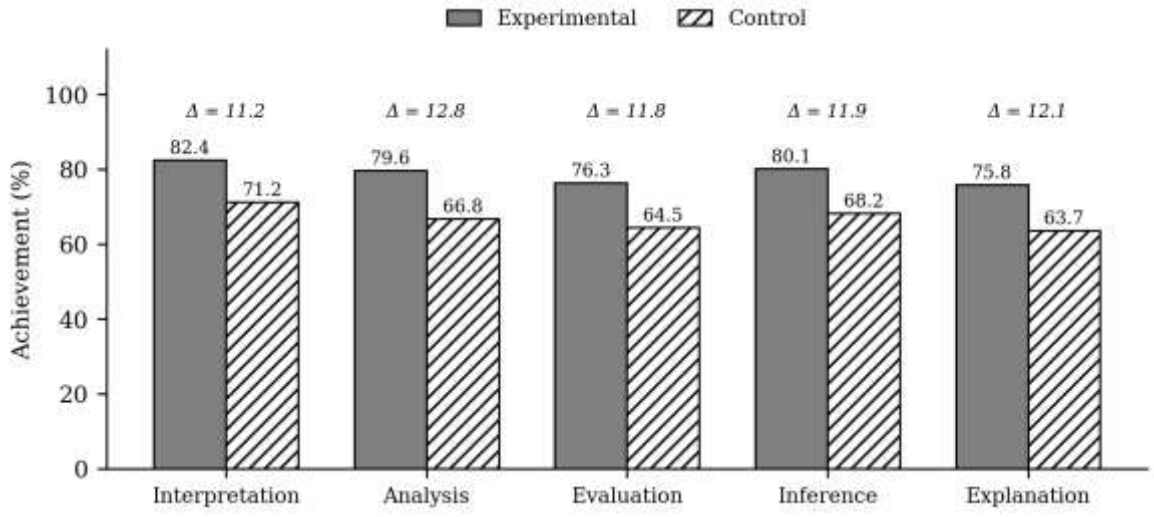


Fig. 3. Achievement Profile Across the Five Critical Thinking Indicators

The intervention therefore raised the whole profile by a broadly uniform margin rather than reshaping it. It should be emphasised that the ordering among the dimensions was not tested inferentially; the between-group comparison within each dimension is a statistical result, whereas any statement that one dimension improved more than another is a descriptive observation about the present sample.

F. Implementation Fidelity and Student Engagement

Because a treatment effect can only be attributed to a design that was actually implemented, the observation protocol was completed by the independent observer in all eight experimental sessions. Table 10 reports the results.

Table 10. Implementation Fidelity and Student Engagement

Aspect observed	Mean score (1–4)	Percentage	Category
Task phase — problems posed as specified in the protocol	3.75	93.8	Very high
Technique phase — GeoGebra exploration carried out by students	3.69	92.3	Very high
Technology phase — written justification produced and critiqued	3.50	87.5	High
Theory phase — generalisation linked to formal statements	3.44	86.0	High
Overall implementation fidelity	3.60	89.9	High
Student engagement during exploration	3.55	88.8	High

Overall implementation fidelity reached 3.60 of a possible 4.00 (89.9%), which falls in the high category. Fidelity was highest in the task and technique phases, where the activity is concrete and easy to sustain, and lowest in the technology and theory phases, where students must produce and defend justifications; the observer noted that these phases occasionally ran short when exploration overran its allotted time. Student engagement averaged 3.55 (88.8%) and did not decline across the eight sessions, the mean of the final two sessions (3.52) being indistinguishable from that of the first two (3.58). This last observation is relevant to the interpretation of the results, since a purely novelty-driven response to unfamiliar software would be expected to fade over eight weeks, whereas engagement that is sustained is more typically associated with a change in the structure of the activity itself [35].

Discussion

The results give a consistent answer to the question posed. Students who worked through limit content organised as praxeologies with GeoGebra outperformed their peers on every critical thinking dimension measured, and the advantage is large by conventional standards (d = 1.33) rather than merely statistically detectable. Because the two groups were equivalent at baseline, received the same content for the same number of minutes from the same lecturer, and were assessed with the same blind-scored instrument, the difference is most plausibly attributed to the way learning activity was organised rather than to content, time or teacher.

Read through ATD, the finding says something more specific than the familiar claim that technology helps. What distinguished the experimental condition was not access to dynamic graphs but the requirement that every observation produced by dragging or tracing be converted into a justification before it could be generalised—the technology component θ

that conventional calculus teaching normally withholds from students and keeps for the lecturer [13], [14]. GeoGebra makes the process of approaching perceptible, which supports interpretation; but visibility by itself obliges nobody to explain anything, and explanation is where critical thinking is actually exercised. The praxeological sequence supplies that obligation, and the fidelity data in Table 10 indicate that the obligation was in fact enforced in roughly nine sessions out of ten. On this reading the effective ingredient of the treatment is a didactic rule rather than a piece of software, which also explains why studies that add GeoGebra to otherwise unchanged instruction report smaller and more variable outcomes [10], [11].

Compared with prior evidence, the effect obtained here is at the upper end of what the literature reports. It exceeds the pooled effect of 0.96 found across a decade of GeoGebra research [9] and is consistent with the central conclusion of Hillmayr and colleagues that the instructional framing, not the tool, governs the size of the outcome [10]. The direction of the result also agrees with work on dynamic geometry environments showing that dragging supports the discernment of invariants [29], [30] and that systematic exploration can carry students from particular cases to general conjectures [31]. What the present design adds is the step those studies leave to the teacher's discretion: an explicit, timetabled demand for justification before generalisation. The size of the effect suggests that this step is not a detail. At the same time, the findings are more modest than a significance test alone would imply. A normalised gain of 0.53 is moderate, not high, and the experimental group still left roughly half of the available improvement unrealised after eight sessions.

Three aspects of the result deserve a more critical reading. First, although every dimension improved, the internal profile of the experimental group did not change: interpretation remained its strongest dimension and explanation its weakest, exactly as in the control group. The intervention lifted the whole profile by a comparable margin rather than closing the gap between the perceptually supported dimensions and the argumentative ones. Eight sessions appear sufficient to install a justification routine but not to develop written mathematical argumentation as a competence in its own right, which is consistent with evidence that evaluative and explanatory skills develop over longer instructional horizons [19], [20]. Second, the design cannot separate the contribution of GeoGebra from the contribution of the praxeological structure, because the two were introduced together. The experimental condition is a package, and the data licence conclusions about the package rather than about its components. Third, the dimension-level results are a set of five independent comparisons, not a profile analysis; the claim that the design serves interpretation particularly well is an inference from attainment levels and from the theoretical mapping in Table 1, not a tested proposition.

The contribution of the study can now be stated precisely. Theoretically, ATD is normally employed to analyse existing praxeologies or to design study-and-research paths that are evaluated qualitatively; here it is operationalised as an experimental treatment with a measured outcome, and the correspondence proposed in Table 1 between the four praxeological components and the five critical thinking dimensions provides a structure that other researchers can confirm, refine or refute. Empirically, the study supplies what is, to our knowledge, the first quasi-experimental estimate of the joint effect of GeoGebra and an explicitly praxeological design on critical thinking in Indonesian undergraduate calculus, reported with baseline equivalence testing, confidence intervals, effect sizes and dimension-level detail, so that it can be entered into future meta-analyses rather than merely cited.

For practice, the implications are more specific than a general recommendation to adopt digital tools. For lecturers, the change that carries the effect in this study is a protected phase of roughly thirty-five minutes in which students write why the behaviour they have observed occurs, before the class is permitted to generalise; Table 1 supplies eight ready-made prompts of this kind, and the same structure transfers to derivatives and integrals without modification. For programme designers, the finding argues for writing the four praxeological phases into the semester learning plan (RPS) of Differential Calculus as a required sequence rather than leaving the use of software to individual initiative, and for scheduling calculus in a room with computer access. For assessment policy, the dimension-level results show why a single composite score is insufficient: the composite conceals the fact that explanation remained the weakest dimension in both groups even after a successful intervention, so course rubrics should report the five dimensions separately and programme-level outcome mapping should treat evaluation and explanation as targets in their own right. For institutions, the cost implication is encouraging, since GeoGebra is free and the decisive element is a short staff development activity on writing justification prompts rather than procurement. Comparable findings on platform-based redesign of teaching in Indonesian higher education point in the same direction: the pedagogical protocol, not the platform, is what has to be developed [32].

Because the design is quasi-experimental, the causal reading offered above should be examined against the alternatives rather than assumed. Four rival explanations deserve consideration. The first is a novelty or Hawthorne effect: students may have responded to the unfamiliarity of the software and to the visible attention of a research study rather than to the design itself. This cannot be excluded, but two features of the data make it less likely as a complete explanation. The intervention ran for eight weeks, long enough for novelty to fade, and the observed engagement ratings did not decline across that period. A design with an active control group that uses GeoGebra without the praxeological structure would settle the question, and is recommended below.

The second is differential instructor attention or expectancy. Using one lecturer for both conditions removes instructor identity as a confounder, but it introduces the possibility that the lecturer invested more enthusiasm in the condition he had designed. This was mitigated by a written lesson protocol, identical contact time, independent observation of both conditions against the same protocol, and blind double scoring of all scripts, but it cannot be ruled out entirely; only a design in which different instructors are crossed with conditions can do that.

The third is selection. Allocation was randomised at the level of the class rather than the individual, and equivalence was verified on the outcome variable itself ($p = 0.784$). Equivalence on prior mathematics achievement, prior familiarity with GeoGebra and general academic ability was not verified, however, and with only two clusters the class-level randomisation cannot balance such characteristics reliably.

The fourth concerns measurement. Pre-test and post-test were parallel forms rather than the same instrument, which limits practice effects, and both groups sat both forms, so any residual testing effect applies equally to each. The N-gain statistic, however, assumes that the maximum attainable score is a meaningful ceiling, and its interpretation would change if the instrument were more difficult. Generalisation is also bounded: one institution, one study programme, one cohort of 63 students and one topic within calculus. The findings should be read as evidence that the design can work under these conditions, not that it will work everywhere.

Conclusion

This study set out to determine whether organising GeoGebra-based limit instruction around the four praxeological components of the Anthropological Theory of the Didactical improves undergraduate students' critical thinking more than conventional instruction covering the same content in the same time. It does. Groups that were statistically equivalent before the intervention ($p = 0.784$) differed significantly afterwards, $t(61) = 5.274$, $p < 0.001$, with a large effect size ($d = 1.33$, $\eta^2 = 0.313$), and the experimental group reached a moderate normalised gain of 0.53 against a low gain of 0.28. The advantage held on all five critical thinking dimensions, and implementation fidelity of 89.9% indicates that the design tested was the design delivered.

The study makes two contributions. Theoretically, it operationalises ATD as an experimental instructional design rather than an analytic lens and proposes a testable correspondence between the four praxeological components and the five dimensions of critical thinking; the results support the claim that the decisive element is the technology component, the obligation to justify, and not the visualisation that precedes it. Empirically, it contributes an effect-size-based estimate for a technology-plus-design intervention in Indonesian

undergraduate calculus, reported with baseline equivalence, confidence intervals and dimension-level detail that allow the result to be compared and aggregated rather than merely cited.

For practice, the study indicates where redesign effort should be directed. The lesson of these data is not that calculus classes need software; it is that they need a protected phase in which students are required to explain why what they have observed is true, with the software used to generate something worth explaining. That phase costs nothing beyond planning, transfers directly to derivatives, integrals and series, and can be written into the semester learning plan of any calculus course. The dimension-level results add a second practical message: because explanation remained the weakest dimension in both groups even after a successful intervention, assessment in calculus should report critical thinking dimension by dimension rather than as a single score that conceals precisely the weakness most in need of attention.

Limitations and Future Research

Five limitations qualify these conclusions and should be read alongside them. The study was conducted in a single institution with 63 students drawn from one study programme, so its external validity is correspondingly narrow. Allocation was randomised at class level rather than individual level, with only two clusters, so equivalence beyond the pre-tested outcome variable cannot be guaranteed. The treatment combined GeoGebra with praxeological structure, so the design cannot attribute the effect to either component separately. Both conditions were taught by the same lecturer, which controls instructor identity at the cost of leaving experimenter expectancy uncontrolled. Finally, the outcome was measured immediately after an eight-week intervention, so nothing can be said about retention, and the sample provides adequate power only for medium-to-large effects.

These limitations translate directly into an agenda for further research. The most informative next step is a dismantling study with three conditions—GeoGebra with praxeological structure, GeoGebra without it, and conventional instruction—which would isolate the contribution of the justification phase that this study identifies as decisive. A multi-site cluster-randomised trial with several instructors crossed with conditions would address both the selection and the expectancy threats, and a delayed post-test at three and six months would establish whether the gains persist. Extending the design to derivatives, integrals and infinite series would test whether the praxeological sequence is specific to limits or general to introductory analysis. Two further lines are worth pursuing: the development of an instrument sensitive enough to grade the quality of mathematical argumentation rather than its presence, since explanation remained the weakest dimension here, and qualitative analysis of the

justifications students actually wrote, which would show how the technology component does its work rather than only that it does.

Acknowledgment

The authors thank STKIP Taman Siswa Bima for institutional support and access to the computer laboratory in which this study was conducted, the three experts who validated the critical thinking instrument, the colleague who served as independent observer across the eight experimental sessions, and above all the students who took part.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Organisation for Economic Co-operation and Development (OECD), PISA 2022 Results (Volume I): The State of Learning and Equity in Education. Paris, France: OECD Publishing, 2023, doi: 10.1787/53f23881-en.

[2] S. Radović, M. Marić, and D. Passey, "Technology enhancing mathematics learning behaviour: Shifting learning goals from reproducing to constructing knowledge," Education and Information Technologies, vol. 25, no. 1, pp. 73–92, 2020, doi: 10.1007/s10639-019-09912-1.

[3] M. D. M. López-Martín and E. Expósito-Casas, "Digital tools and critical thinking in mathematics education: A systematic review," Mathematics, vol. 10, no. 9, Art. no. 1524, 2022.

[4] Sulaiman, W. Widada, N. A. Yensy, and A. Suningsih, "The relationship between mathematical critical thinking ability and students' political literacy in senior high school," Journal of Pedagogy and Education Science, vol. 5, no. 1, pp. 388–400, 2026, doi: 10.56741/IISTR.jpes.001571.

[5] D. Tall and S. Vinner, "Concept image and concept definition in mathematics with particular reference to limits and continuity," Educational Studies in Mathematics, vol. 12, no. 2, pp. 151–169, 1981, doi: 10.1007/BF00305619.

[6] M. Ndlovu and D. Wessels, "Exploring students' understanding of limits using dynamic mathematical software," International Journal of Mathematical Education in Science and Technology, vol. 52, no. 8, pp. 1181–1198, 2021, doi: 10.1080/0020739X.2020.1717655.

[7] M. Hohenwarter and Z. Lavicza, "The strength of the community: How GeoGebra can inspire technology integration in mathematics teaching," International Journal for Technology in Mathematics Education, vol. 24, no. 3, pp. 103–112, 2017, doi: 10.1564/tme_v24.3.03.

[8] T. Kutluca and O. Birgin, "The effect of dynamic mathematics software GeoGebra on students' achievement and attitudes toward mathematics," International Journal of Mathematical Education in Science and Technology, vol. 51, no. 3, pp. 353–371, 2020, doi: 10.1080/0020739X.2019.1648898.

[9] D. Juandi, Y. S. Kusumah, M. Tamur, K. S. Perbowo, and T. T. Wijaya, "A meta-analysis of GeoGebra software decade of assisted mathematics learning: What to learn and where to go?," Heliyon, vol. 7, no. 5, Art. no. e06953, 2021, doi: 10.1016/j.heliyon.2021.e06953.

[10] D. Hillmayr, L. Ziernwald, F. Reinhold, S. I. Hofer, and K. M. Reiss, "The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis," Computers & Education, vol. 153, Art. no. 103897, 2020, doi: 10.1016/j.compedu.2020.103897.

[11] A. Bray and B. Tangney, "Technology usage in mathematics education research: A systematic review of recent trends," Computers & Education, vol. 114, pp. 255–273, 2017, doi: 10.1016/j.compedu.2017.07.004.

[12] P. Drijvers, "Digital technology in mathematics education: Why it works (or doesn't)," in Selected Regular Lectures from the 12th International Congress on Mathematical Education, S. J. Cho, Ed. Cham, Switzerland: Springer, 2015, pp. 135–151, doi: 10.1007/978-3-319-17187-6_8.

[13] Y. Chevallard, "Introducing the anthropological theory of the didactic: An attempt at a principled approach," Hiroshima Journal of Mathematics Education, vol. 12, pp. 71–114, 2019, doi: 10.24529/hjme.1205.
[14] Y. Chevallard and M. Bosch, "Didactic transposition in mathematics education," in Encyclopedia of Mathematics Education, 2nd ed., S. Lerman, Ed. Cham, Switzerland: Springer, 2020, pp. 214–218, doi: 10.1007/978-3-030-15789-0_48.
[15] C. Winsløw, B. Barquero, M. De Vleeschouwer, and N. Hardy, "An institutional approach to university mathematics education: From dual vector spaces to questioning the world," Research in Mathematics Education, vol. 16, no. 2, pp. 95–111, 2014, doi: 10.1080/14794802.2014.918345.
[16] A. P. Adiredja and M. Zandieh, "The lived experience of linear algebra: A counter-story about women of color in mathematics," Educational Studies in Mathematics, vol. 104, no. 2, pp. 239–260, 2020, doi: 10.1007/s10649-020-09954-3.
[17] P. A. Facione, Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction. Millbrae, CA, USA: California Academic Press, 1990.
[18] R. A. Pratama and H. Retnawati, "The effectiveness of GeoGebra-assisted learning in improving students' mathematical reasoning," Jurnal Riset Pendidikan Matematika, vol. 5, no. 1, pp. 12–23, 2018, doi: 10.21831/jrpm.v5i1.13532.
[19] N. Nurhayati, "The use of GeoGebra in improving students' understanding of calculus concepts," Jurnal Pendidikan Matematika Indonesia, vol. 6, no. 2, pp. 85–92, 2021. [OJS/DOI belum terverifikasi]
[20] D. P. Sari and M. Mulyono, "The effect of inquiry-based learning on students' higher-order thinking skills," Jurnal Pendidikan dan Pembelajaran, vol. 27, no. 3, pp. 145–153, 2020.
[21] N. N. Parwati, I. G. P. Suharta, and I. G. W. Sudatha, "TPACK theoretical correspondence with learning variables and its application effectiveness in learning," Journal of Pedagogy and Education Science, vol. 2, no. 3, pp. 244–256, 2023, doi: 10.56741/jpes.v2i03.402.
[22] W. R. Shadish, T. D. Cook, and D. T. Campbell, Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston, MA, USA: Houghton Mifflin, 2002.
[23] J. W. Creswell and J. D. Creswell, Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, 5th ed. Thousand Oaks, CA, USA: SAGE Publications, 2018.
[24] F. Faul, E. Erdfelder, A.-G. Lang, and A. Buchner, "G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences," Behavior Research Methods, vol. 39, no. 2, pp. 175–191, 2007, doi: 10.3758/BF03193146.
[25] L. R. Aiken, "Three coefficients for analyzing the reliability and validity of ratings," Educational and Psychological Measurement, vol. 45, no. 1, pp. 131–142, 1985, doi: 10.1177/0013164485451012.
[26] J. Cohen, Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Hillsdale, NJ, USA: Lawrence Erlbaum Associates, 1988.
[27] L. V. Hedges, "Distribution theory for Glass's estimator of effect size and related estimators," Journal of Educational Statistics, vol. 6, no. 2, pp. 107–128, 1981, doi: 10.3102/10769986006002107.
[28] R. R. Hake, "Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses," American Journal of Physics, vol. 66, no. 1, pp. 64–74, 1998, doi: 10.1119/1.18809.
[29] F. Arzarello, A. Baccaglini-Frank, and F. Ferrara, "Mathematical visualization and digital technologies: New perspectives for teaching and learning," ZDM Mathematics Education, vol. 52, no. 7, pp. 1285–1297, 2020, doi: 10.1007/s11858-020-01172-2.
[30] A. Leung, A. Baccaglini-Frank, and M. A. Mariotti, "Discernment of invariants in dynamic geometry environments," Educational Studies in Mathematics, vol. 101, no. 2, pp. 177–197, 2019.
[31] M. Santos-Trigo and M. Camacho-Machín, "Problem solving and the use of digital technologies within the mathematics curriculum," International Journal of Mathematical Education in Science and Technology, vol. 50, no. 6, pp. 815–831, 2019.
[32] Sarwa, R. Putra, A. Zulfikar, L. Atika, M. D. Putri, and R. Azahar, "Enhancing vocational higher education lecturers' pedagogical competence through a digital platform-based flipped classroom," Journal of Pedagogy and Education Science, vol. 5, no. 2, pp. 586–600, 2026, doi: 10.56741/IISTR.jpes.001912.

[33] R. Nemirovsky, F. Ferrara, G. Ferrari, and N. Adamuz-Povedano, "Body motion, early algebra, and the colours of abstraction," Educational Studies in Mathematics, vol. 104, no. 2, pp. 261–283, 2020, doi: 10.1007/s10649-020-09955-2.
[34] B. Pepin, J. Choppin, K. Ruthven, and N. Sinclair, "Digital curriculum resources in mathematics education: Foundations for change," ZDM Mathematics Education, vol. 49, no. 5, pp. 645–661, 2017, doi: 10.1007/s11858-017-0879-1.
[35] T. Hidayat and A. Widodo, "Technology integration in mathematics learning and its impact on student engagement," Jurnal Pendidikan Matematika, vol. 15, no. 2, pp. 101–110, 2021.
[36] L. Rahmawati and S. Sutarto, "Guided inquiry learning to improve students' critical thinking skills," Jurnal Inovasi Pendidikan, vol. 13, no. 1, pp. 45–54, 2019.
[37] N. H. M. Sari, M. H. Ayusman, M. Agus, and A. L. Nahdi, "Pemanfaatan GeoGebra dalam pembelajaran kalkulus: Systematic literature review," SANTIKA: Seminar Nasional Tadris Matematika. [Online]. Available: https://proceeding.uingusdur.ac.id/index.php/santika/article/view/santika42434.
[38] A. Hidayat and P. Firmanti, "Navigating the tech frontier: A systematic review of technology integration in mathematics education," Cogent Education, vol. 11, no. 1, 2024, doi: 10.1080/2331186X.2024.2373559.

Authors



Nanang Diana is a historian and a full-time lecturer in the Master's Program in Pedagogy at the Teacher Training and Education College, STKIP Taman Siswa Bima, West Nusa Tenggara, Indonesia. He is widely recognized for his expertise in Didactic Design, Geometry, Real Analysis, and Statistics, which are the primary focus of his academic and research career. (email: nanangdiana.math@gmail.com).



Irwan Aprayadi is a postgraduate student in the Pedagogy program at STKIP Taman Siswa and a civil servant teacher at SMAN 2 Madapangga. He specializes in science and chemistry teaching. He has conducted research in chemistry and is currently writing several books on character education and other subjects. (email: irwan.aprayadi.syakira@gmail.com).



Adi Apriadi Adiansha is an educator and a full-time lecturer in the Elementary School Teacher Education Program (PGSD) at the Teacher Training and Education College, STKIP Taman Siswa Bima, West Nusa Tenggara, Indonesia. He is widely recognized for his expertise in Elementary Education, Learning Innovation, and Educational Research, which are the primary focus of his academic and research career. (email: adiapriadiadiansha@gmail.com).



M. Ibnusaputra is a lecturer in the Elementary School Teacher Education Study Program at STKIP Taman Siswa Bima. He holds a Master's degree in Physics Education from the University of Mataram. Throughout his academic career, several works have been published in national and international journals, with research interests focusing on education, teaching and learning, and modern technology. One of his main areas of expertise is research methodology, and he is proficient in statistical data analysis, including the use of SPSS, Ministep (Rasch Model), and NVivo. (email: ibnusaputratsb@gamil.com).



Mariamah is a full-time lecturer in the Master's Program in Pedagogy at the Teacher Training and Education College (STKIP) Taman Siswa Bima, West Nusa Tenggara, Indonesia. She is widely recognized for her expertise in Ethnomathematics and Numeracy, which have become the primary focus of her academic, teaching, and research activities to date. (email: mariamahmariamah85@yahoo.co.id).