AKSIOMA: Jurnal Program Studi Pendidikan Matematika
Volume 14, No. 4, 2025, 1078–1096

ISSN 2089-8703 (Print)
ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v14i4.12091

ENHANCING STUDENTS CRITICAL THINKING THROUGH PROBLEM-BASED LEARNING WITH CHATGPT AND GEOGEBRA

Muhammad Rizaldi[1*], Oktaviana Ainun Ratnawati[2], Chairil Faif Pasani[3]

[1,2] Universitas Palangka Raya, Palangka Raya, Indonesia
[3] Universitas Lambung Mangkurat, Banjarmasin, Indonesia
**Corresponding author. Jl. Yos Sudarso, 73111, Palangka Raya, Indonesia.*
E-mail: *muhammad.rizaldi@math.upr.ac.id* [1*)]
Oktavianainun29@fkip.upr.ac.id [2)]
chfaifp@ulm.ac.id [3)]

Received 12 February 2025; Received in revised form 12 August 2025; Accepted 02 December 2025

Abstract

Critical thinking is increasingly important with the existence of artificial intelligence (AI) especially for college students. AI is not yet fully accurate in solving mathematical problems, so critical thinking helps students analyze and evaluate its answers. This study aims to analyzing the influence of using PBL assisted by ChatGPT and GeoGebra in improving students' critical thinking skills. The population of this study consisted of all fourth-semester students of the Mathematics Education study program. The sample was selected using cluster sampling, with the experimental and control classes randomly assigned. This quasi-experimental study employed a pretest-post-test non-equivalent group design involving 47 subjects, 24 in the experimental group and 23 in the control group. The average pre-test and post-test of experimental (47.96 and 62.13) and control (55.48 and 54.26) groups were significantly different. The ANCOVA results showed that PBL with ChatGPT and GeoGebra significantly enhanced students' critical thinking. Tukey's test result showed that the difference between the experimental and control group was 14.88 which indicated that students' critical thinking in the experimental group was higher than the control group. Furthermore, the treatment did not affect identify indicator, but had a positive effect on justify, inference, clarify, and strategy decision indicators. This demonstrates that PBL assisted by ChatGPT and GeoGebra enhances students' critical thinking. The research has the potential to support lecturers in enhancing students' critical thinking and improving innovative learning models where AI can be used in education with proper guidance and implications.

Keywords: ChatGPT; college students; critical thinking; Geogebra; problem-based learning

Abstrak

Berpikir kritis semakin penting dengan adanya kecerdasan buatan (AI) khususnya bagi mahasiswa. AI belum sepenuhnya akurat dalam menyelesaikan masalah matematika, sehingga berpikir kritis membantu mahasiswa menganalisis dan mengevaluasi jawabannya. Penelitian ini bertujuan untuk menganalisis pengaruh penggunaan PBL berbantuan ChatGPT dan GeoGebra dalam meningkatkan kemampuan berpikir kritis mahasiswa. Populasi penelitian ini adalah seluruh mahasiswa Program Studi Pendidikan Matematika pada semester 4. Sampel penelitian dipilih menggunakan teknik cluster sampling dengan penentuan kelas eksperimen dan kelas kontrol secara acak. Penelitian kuasi-eksperimen ini dengan desain pretest-post-test non-equivalent group melibatkan 47 subjek, 24 pada kelompok eksperimen dan 23 pada kelompok kontrol. Rata-rata pre-test dan post-test kelompok eksperimen (47,96 dan 62,13) dan kontrol (55,48 dan 54,26) berbeda secara signifikan. Hasil ANCOVA menunjukkan bahwa PBL dengan ChatGPT dan GeoGebra secara signifikan meningkatkan berpikir kritis mahasiswa. Hasil uji Tukey menunjukkan perbedaan antara kelompok eksperimen dan kontrol sebesar 14,88 yang mengindikasikan bahwa berpikir kritis mahasiswa pada kelompok eksperimen lebih tinggi dibandingkan kelompok kontrol. Selain itu, perlakuan tidak berpengaruh pada indikator identify, tetapi memberikan pengaruh positif pada indikator justify, inference, clarify, dan strategy decision. Hal ini menunjukkan bahwa PBL yang dibantu oleh ChatGPT dan GeoGebra meningkatkan berpikir kritis mahasiswa. Penelitian ini berpotensi mendukung dosen dalam meningkatkan berpikir kritis mahasiswa dan memperbaiki model pembelajaran inovatif di mana AI dapat digunakan dalam pendidikan dengan bimbingan dan implikasi yang tepat.

***Kata kunci**: Berpikir Kritis; ChatGPT; GeoGebra; Mahasiswa; Pembelajaran Berbasis Masalah*

ISSN 2089-8703 (Print)
ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v14i4.12091

INTRODUCTION

Critical thinking is a core educational goal, fostering independent thought and deep engagement with content (Ennis, 1996; Zakaria et al., 2025). University educators emphasize its role in preparing students for a complex world, often through active learning methods (Bezanilla et al., 2021). In the 21st-century learning paradigm, critical thinking and problem-solving are prioritized alongside. Cultivating these skills requires teaching strategies that match the learning preferences and characteristics of modern students (Alashwal & Barham, 2025; Astutik et al., 2024).

The growing presence of technologies like Artificial Intelligence (AI) highlights the rising importance of critical thinking. In 2023, Indonesia ranked third globally with 1.4 billion AI application visits, 51% of which came from users aged 17–25, many being university students. However, less critical users often accept AI-generated information without further examination, which remains a concern (Sánchez-Ruiz et al., 2023; Tlili et al., 2023). One of the AI or chatbot examples that often college students use, ChatGPT's ability to solve math problems is still below the college students' average (Chang et al., 2024). Using AI-generated information without analysis can lead to serious consequences in academic work. Critical thinking enables students to evaluate such information logically, preventing misinformation and preparing them to face technological challenges in the 21st century (van den Berg & du Plessis, 2023).

This study conducted an initial assessment of two classes of fourth-semester students to evaluate their critical thinking skills. A pre-test of two fourth-semester classes showed an average critical thinking score of 51.94/100, categorized as low. This lack of critical thinking in using AI risks unfiltered acceptance of information, which can negatively affect learning, academic performance, and intellectual growth. ((Nauman, 2017); (Rizaldi et al., 2024)).

An adaptive learning model is one of the solutions to provide student's critical thinking (Yang et al., 2014). Problem-Based Learning (PBL) can enhance critical thinking, especially when supported by tools or media (Kardoyo et al., 2020). This study integrates PBL with ChatGPT and GeoGebra, where ChatGPT serves as a comparative tool to prompt students to question and verify AI-generated answers. While ChatGPT offers accessible solutions, lecturer guidance and independent verification remain essential to ensure accuracy and foster critical thinking (Zein et al., 2024). Since ChatGPT's answers may be inaccurate, this study integrates GeoGebra as a reliable support tool for Solid Analytic Geometry. GeoGebra visualizes concepts, verifies solutions, and solves problems ChatGPT cannot (Botana & Recio, 2024). This combination exposes students to differing solutions, fostering analysis, evaluation, and creative problem-solving (Egitim, 2022; Pangestika & Faiziyah, 2022; Rizaldi et al., 2021), making both tools effective for enhancing critical thinking.

In this study, the PBL stages with the help of ChatGPT and GeoGebra will adjust the PBL syntax (Breen & Fallon, 2005). The role of ChatGPT and GeoGebra in helping students solve the problems given. So that the PBL stages with the help of ChatGPT and GeoGebra are as follows (1) Before the

ISSN 2089-8703 (Print)
ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v14i4.12091

meeting, the lecturer orients students to the problem; (2) the lecturer organizes students in groups to learn and answer problems; (3) students in groups collecting information and solving problems using ChatGPT and GeoGebra; (4) during the meeting, students in groups presenting problem solutions and ChatGPT and GeoGebra answers; and (5) lecturers and students evaluating problem solution.

Previous studies have found that PBL improve students' critical thinking ((Magpantay & Pasia, 2022); (Marthaliakirana et al., 2022); (Sarwastuti & Purnomo, 2023)). Furthermore, PBL with the help of ChatGPT has improved students' problem solving skills, as well as their mathematical proof skills (Jinyoung & Minjeong, 2024; Mairing et al., 2024). Moreover, ChatGPT as a learning tool has a positive effect on critical thinking, although it needs to be used more carefully ((Jeong, 2023); (Rizaldi et al., 2024); (Walter, 2024); (Zhang et al., 2025)). Besides PBL and ChatGPT, PBL assisted by GeoGebra has improved mathematical ability, especially problem solving ability ((Devi et al., 2024); (Khoeriah et al., 2024)). Additionally, GeoGebra successfully enhances the critical thinking of junior and senior high school students ((Hendriana et al., 2020); (Samura & Darhim, 2023); (Sofyan et al., 2022); (Ziatdinov & Valles, 2022)).

Therefore, the purpose of this study is to enhance students' critical thinking with PBL assisted by ChatGPT and GeoGebra. In order to achieve the objectives of this study, 3 research questions will be answered, which are (1) Can the PBL model with ChatGPT and GeoGebra enhance critical thinking and its indicators? (2) How does critical thinking increase in each of its indicators? and (3) How does the PBL model process with ChatGPT and GeoGebra lead to the enhancement of critical thinking?.

METHODS

This research used quantitative methods with a quasi-experimental pre-test post-test non-equivalent control group design ((Cohen et al., 2007); (Mairing, 2017)). The improvement of students' critical thinking could be seen by conducting experiments on the PBL model assisted by ChatGPT and GeoGebra, while the indicator of critical thinking further aimed to analyze in more detail.

This research followed nine operational stages adapted from standard experimental research procedures (Lodico et al., 2010). First, the research problems were defined by identifying gaps in students' critical thinking skills through pre-test. Second, hypotheses were formulated regarding the impact of the Problem-Based Learning (PBL) model assisted by ChatGPT and GeoGebra, grounded in findings from previous studies. Third, pre-test scores were collected by administering critical thinking test to all participants before the intervention to establish baseline data. Fourth, research instruments developed and validated by experts in mathematics education especially critical thinking in solid analytical geometry. Fifth, the learning intervention was implemented over four sessions, with the experimental group receiving PBL instruction supported by ChatGPT for problem clarification and GeoGebra for visualization, while the control group received conventional instruction. Sixth, post-tests using the same critical thinking instrument were conducted immediately after the intervention to measure learning gains. Seventh,

ISSN 2089-8703 (Print)
ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v14i4.12091

students' overall critical thinking performance was analyzed using ANCOVA to compare results between groups. Eighth, each critical thinking indicator was analyzed too using ANCOVA. Finally, conclusions were drawn by summarizing the findings and relating them to the hypotheses.

The population of this study consisted of all fourth-semester students of the Mathematics Education study program. The sample was selected using cluster sampling, with the experimental and control classes randomly assigned. This quasi-experimental study employed a pretest-post-test non-equivalent group design involving 47 subjects, 24 in the experimental group and 23 in the control group. Groups were randomly selected. Data was collected through pre-test and post-test. The experimental and control group received the same pre-test and post-test. The pre-test and post-test were structured based on critical thinking indicators that were adopted from Ennis (2018) and Santos-Meneses & Drugova (2023) which were presented on Table 1. The pre-test material was plane and line, while the post-test material was sphere. The treatment period was conducted for 4 weeks, each week consisting of the four contents. The pre-test and post-test consisted of two questions. Each score was summed up for each indicator as follows

$$\mu_{1,2} = a_1 + b_1 + c_1 + d_1 + e_1 + a_2 + b_2 + c_2 + d_2 + e_2 \quad (1)$$

with μ was students' critical thinking; a, b, c, d, e were score of each indicator of critical thinking; and 1,2 were question numbers. The score scale of the pre-test and post-test were 0-100. Because there were two problems and five indicators, the maximum score for each student was $2(5 \times 10) = 100$. Moreover, the pre-test and post-test could be seen in the following links (https://s.id/26tvS) and the problems that were given to the experimental group in the following links (https://s.id/26tnK).

Table 1. Indicators of Critical Thinking

Indicator	Ability
Identify ($\boldsymbol{a}$)	Identifying implied or expli-cit information on the problem
Justify ($\boldsymbol{b}$)	Considering the credibility of information with knowledge or logic
Inference ($\boldsymbol{c}$)	Make deductive and inductive conclusions from the information obtained or knowledge
Clarify ($\boldsymbol{d}$)	Provide explanations based on the conclusions or assumptions obtained
Strategy Decision ($\boldsymbol{e}$)	Selecting possible strategies for solving the problem

The research instruments used included the pre-test, post-test, students' answers, and lesson plans. The pre-test and post-test were used to measure students' critical thinking skills before and after the treatment. Students' answers were analyzed to assess their critical thinking process in solving the problems given during learning. The lesson plans were used as a reference for implementing the PBL model integrated with ChatGPT and Geo-Gebra, ensuring consistency and alignment of the material with the research objectives. All instruments were validated by experts in mathematics

ISSN 2089-8703 (Print)
ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v14i4.12091

education and instructional development. The validation results showed that the four instruments were declared valid in terms of content, construction, and suitability with the research objectives, making them feasible for use in conducting this study.

Data analysis in this study was performed by ANCOVA (Rutherford, 2011) with the covariate as the pre-test result (critical thinking before treatment), the dependent variable as the post-test result (critical thinking after treatment), and the independent variable as the PBL with ChatGPT and GeoGebra. Subsequently, there were two hypotheses from this study, namely the first hypothesis related to students' critical thinking before treatment, which was as follows

$$H_0: \mu 1_E = \mu 1_C$$
$$H_1: \mu 1_E \neq \mu 1_C$$

where $\mu 1_E =$ average of pre-test results in the experimental group, and $\mu 1_C =$ average of pre-test results in the control group. The second hypothesis related to students' critical thinking after treatment was as follows

$$H_0: \mu 2_E = \mu 2_C$$
$$H_1: \mu 2_E \neq \mu 2_C$$

where $\mu 2_E =$ average of post-test results in the experimental group, and $\mu 2_C =$ average of post-test results in the control groups. If the ANCOVA inferred that there was a significant difference between critical thinking in the experimental and control group, then the researcher conducted further testing using the Tukey comparison test. The research used Minitab 22 to analyze the data and conclude. Furthermore, this study analyzed more details on how each critical thinking indicators was affected. Each indicator was analyzed useing ANCOVA and if there were a significant effect, the Tukey test comparison would follow.

RESULTS AND DISCUSSION

In this section, results presented in learning activities, statistical summaries, assumption test results, ANCOVA results, further test comparison results, ANCOVA of each critical thinking indicator, and description of each critical thinking indicator. The assumption test should be met first as a condition of conducting ANCOVA. Further comparison tests needed to be done to see in deeply of differences in critical thinking in each class. The research analyzed in more detail the differences of each critical thinking indicator. Moreover, it needed a description of students' critical thinking during the pre-test and post-test.

1. Learning Activities

This section would describe the utilization of ChatGPT and GeoGebra by the selected subjects who made good use of ChatGPT and GeoGebra. Only one solution was chosen as the representation, which best represented the students' critical thinking, rather than all the solutions generated in each meeting. This aimed to show the effectiveness of the PBL model with ChatGPT and GeoGebra in enhancing critical thinking. The following Figure 1 presented the answers of the subject, Figure 2 presented the response of ChatGPT and Figure 3 presented the graph on the solution to the problem.

AKSIOMA: Jurnal Program Studi Pendidikan Matematika
Volume 14, No. 4, 2025, 1078–1096

ISSN 2089-8703 (Print)
ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v14i4.12091

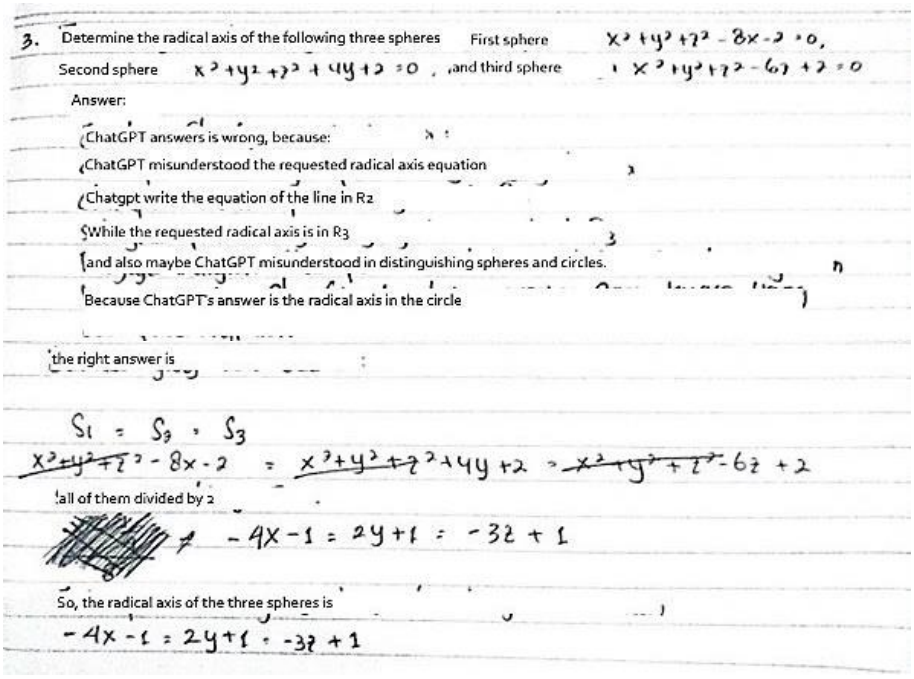

Figure 1. Subject Answers

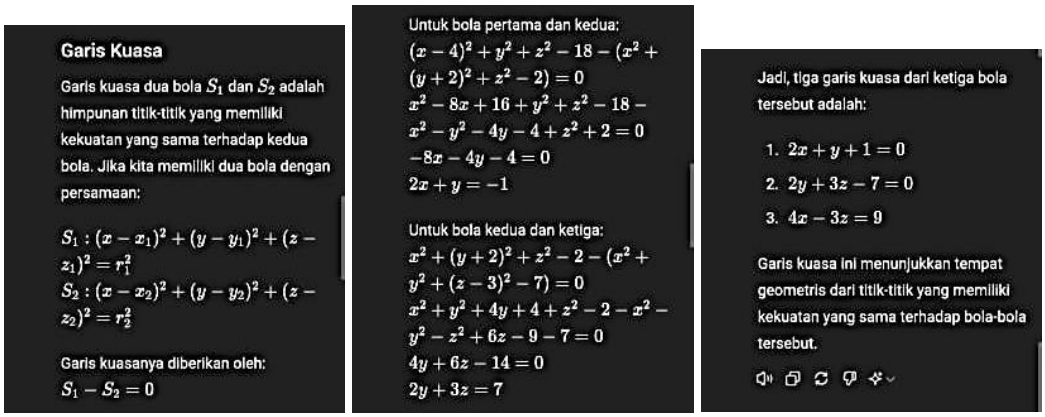

Figure 2. ChatGPT Answers

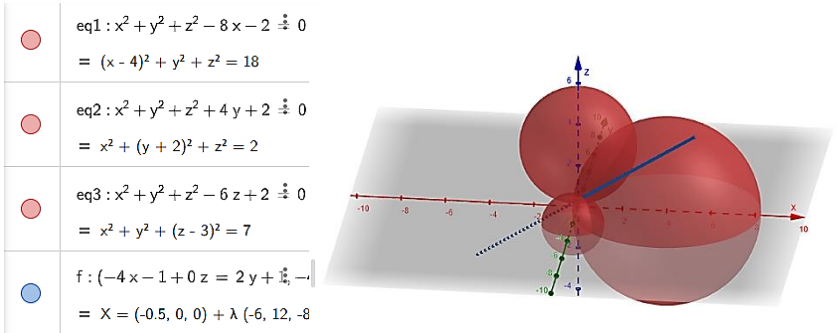

Figure 3. GeoGebra Answers

Based on Figure 1, Figure 2, and Figure 3, the subject could identify the information and consider the truth of ChatGPT's irrelevant answer. ChatGPT did not understand the problem given that the problem was about the radical

ISSN 2089-8703 (Print)
ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v14i4.12091

axis of the three spheres not the radical axis of two circles. This was corrected by the subject that the problem given was in $\mathbb{R}^3$ instead of $\mathbb{R}^2$. The subject's explanation showed that he could explain and conclude his argument against the error of ChatGPT's answer. Furthermore, the subject solved the problem correctly by making the equation of the three spheres $S_1 = S_2 = S_3$ into the equation of the requested radical axis, namely $-4x - 1 = 2y + 1 = -3z + 1$. After that, the subject examined his answer by looking at the graph in GeoGebra and was confirmed correct. It can be concluded that, despite ChatGPT's incorrect response, the subject was able to critically analyze and utilize it to solve the problem, subsequently verifying the solution using GeoGebra.

2. Critical Thinking Scores

The research showed statistical summaries to provide an overview of how students' critical thinking in each class based on the pre-test and post-test results. The result of pretest and posttest from each class can be seen in Table 2.

Table 2. Statistical Summaries of Both Group

Variable	Group	N	Mean	StDev	Min	Q1	Median	Q3	Max
Pre-test	Control	23	55.48	25.93	15.00	33.00	56.00	80.00	96.00
	Experiment	24	47.96	24.05	12.00	23.50	46.50	62.50	93.00
Post-test	Control	23	54.26	24.84	15.00	27.00	52.00	81.00	96.00
	Experiment	24	62.13	26.49	15.00	40.50	66.50	83.50	100.00

Based on Table 2, the average pre-test of the experimental (47.96) and control (55.48) groups were significantly different. Subsequently, most of the category statistical summaries and critical thinking indicators demonstrated that the control group's pre-test score was significantly greater descriptively than the experimental group. However, the average post-test of the experimental (62.13) and control (54.26) groups were also significantly different. Moreover, most of the category statistical summaries and critical thinking indicators also demonstrated that the experimental group's post-test score was significantly greater descriptively than the control group. Thus, it could be concluded descriptively that PBL with ChatGPT and GeoGebra had a positive effect on students' critical thinking.

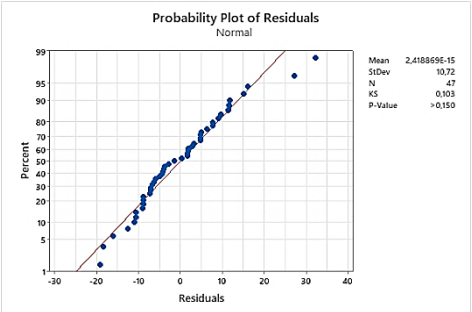

Figure 4. Result of normality tests of residuals

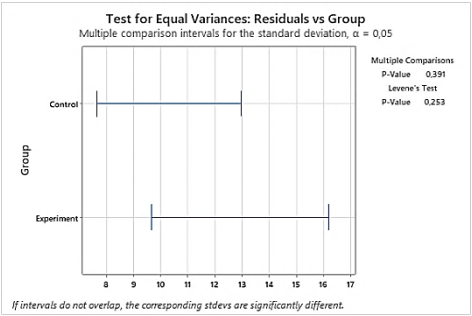

Figure 5. Result of variance homogeneity

AKSIOMA: Jurnal Program Studi Pendidikan Matematika
Volume 14, No. 4, 2025, 1078–1096
ISSN 2089-8703 (Print)
ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v14i4.12091

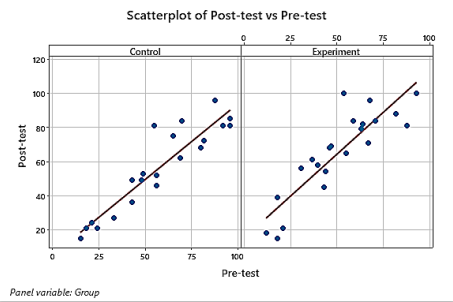

Figure 6. Result of Linearity

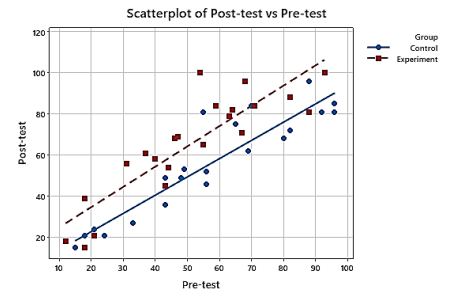

Figure 7. Result of Homogeneity of Regression Coefficients

Table 3. Assumption of ANCOVA

Assumption	p-value
Normality (Kolmogorov-Smirnov's)	>0.150
Homogeneity of Variances (Levene's test)	0.253
Linearity (Lack of fit)	0.131
Homogeneity of Regression Coefficients	0.461

The research examined the five assumptions of the ANCOVA before carrying out the analysis. The assumptions were the extraneous variable being independent of the treatments, normality of residuals, homogeneity of variance, linearity of the covariate (pre-test) and dependent variable (post-test), and homogeneity of regression coefficients. The first assumption was satisfied because the experimental class was selected randomly, so the treatment in the selected class did not affect the extraneous variable. The second assumption, based on Figure 4 and Table 3, the p-value of the Kolmogorov-Smirnov's normality test of residuals was more than $0.150 > 0.05 = \alpha$, which meant the residuals had a normal distribution with a confidence level of 95%. The third assumption, based on Figure 5 and Table 3 the p-value of the Levene's test of variances homogeneity was $0.253 > 0.05 = \alpha$, which meant the variances were homogeneous with a confidence level of 95%. The fourth assumption, based on Table 3, the p-value of the lack-of-fit was $0.131 > 0.05 = \alpha$, which meant the model accurately fit the data with a confidence level of 95%. It was indicated covariate (pre-test) and dependent variable (post-test) were linear. Furthermore, based on Figure 6, the graph showed that a randomness of most points followed a straight line in both groups. The last assumptions test, based on Table 3, the p-value of the Pre-test*Group was p-value = $0.461 > 0.05 = \alpha$, which meant regression coefficients of both groups were homogeneous with a confidence level of 95%. In addition, based on Figure 7, the graph showed that the slopes of both lines were similar. Therefore, all ANCOVA assumptions had been fulfilled.

Table 4. ANCOVA results

Source	DF	Adj SS	Adj MS	F-value	p-value
Pre-test	1	24418.4	24418.4	203.08	0.000
Group	1	2538.8	2538.8	21.11	0.000
Error	44	5290.7	120.2		
Lack-of-Fit	36	4790.2	133.1	2.13	0.131
Pure Error	8	500.5	62.6		
Total	46	30435.4			

ISSN 2089-8703 (Print)
ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v14i4.12091

Table 5. Comparisons Tukey's Test

Group	N	Mean	Grouping	
Experiment	24	65.55	A	
Control	23	50.68		B

Means that do not share a letter are significantly different.

Table 6. Difference of Group in Tukey's Test

Difference of Group Levels	Difference of Means	SE of Difference	Simultaneous 95% CI	T-Value	Adjusted p-value
Experiment - Control	14.88	3.24	(8.35; 21.40)	4.59	0.000

Individual confidence level = 95.00%

Based on Table 4, the result of the ANCOVA for the p-value of group was $0 < 0.05 = \alpha$, which meant that covariate (pre-test) affected students' critical thinking in both groups with a confidence level of 95%. This meant that the implication of PBL with ChatGPT and GeoGebra had a making difference effect on critical thinking in each class. The difference in ANCOVA showed that the results obtained would be further tested using the Tukey difference test.

Based on Table 5 and Table 6, Tukey's comparison test also showed the effect of different groups with a confidence level of 95%. The difference of means by the Tukey test was 14.88 which meant that students' critical thinking scores in the experimental group was greater than control group. Thus, there was positive effect on critical thinking by implementing PBL assisted by ChatGPT and GeoGebra. These results would be further analyzed for each indicator of critical thinking and described one of the students who represented the increase in critical thinking scores after the treatment.

3. Indicators of Critical Thinking Scores

The research analyzed more detail students' critical thinking by analyzing each indicator. It was intended to explore more deeply which indicators or abilities had increased. Thus, the positive effects of the PBL with ChatGPT and GeoGebra became more apparent. Table 7 presented a summary of each student's critical thinking indicator scores.

Table 7. Indicator Score of Critical Thinking

Variable	Group	Mean of					Total Mean
		a	***b***	***c***	***d***	***e***	
Pre-test	Control	13.70	10.26	10.17	10.22	11.13	55.48
	Experiment	14.46	7.96	7.67	8.08	9.79	47.96
Post-test	Control	14.26	9.52	10.09	9.91	10.48	54.26
	Experiment	14.25	12.46	11.58	11.29	12.54	62.13

Note: a is identify; b is justify; c is inference; d is clarify; and e is strategy decision

Based on Table 7, the average of identify indicator in pre-test (14.46) and post-test (14.25) had no significant decrease descriptively in the experimental group. However, the average of justify, inference, clarify and

ISSN 2089-8703 (Print)
ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v14i4.12091

strategy decision indicators in pre-test (7.96; 7.67; 8.08; and 9.79) and post-test (12.46; 11.58; 11.29; and 12.54) had a significant increase descriptively in the experimental group. In contrast to the control group, the average of identify indicator in pre-test (13.70) and post-test (14.26) had little increase but the average of justify, inference, clarify, and strategy decision indicator in pre-test (10.26; 10.17; 10.22; and 11.13) and post-test (9.52; 10.09; 9.91; and 10.48) had to decrease little descriptively. Thus, it could be concluded that PBL with ChatGPT and GeoGebra had no effect on identify indicator, but had a positive effect on justify, inference, clarify, and strategy decisions descriptively.

The research was also examined the ANCOVA to each indicator. All indicators have been examined for assumptions and satisfy all the required assumption tests. Based on Table 8 below, the ANCOVA of identify indicator was p-value 0.375 > 0.05 = α, which meant that treatment did not affect identify indicator with a confidence level of 95%. However, the ANCOVA of justify, inference, clarify and strategy decision indicators p-value respectively were 0.000; 0.000; 0.000; and 0.012 which were lower than 0.05 = α. This meant that treatment affected those four indicators with a confidence level of 95%. Subsequently, Tukey's test results of those four indicators indicated significant difference exists. In addition, it was found that justify, inference, clarify, and strategy decision indicators enhanced with the difference of mean by Tukey test respectively were 4.79; 3.74; 3.15; and 2.98.

Table 8. Result of ANCOVA and Tukey Test for Each Critical Thinking Indicator

Indicator	Group	p-value (ANCOVA)	Grouping* (Tukey Test)	Difference of Mean (Tukey Test)
Identify	Experiment Control	0.375	Not Significant	Not Significant
Justify	Experiment	0.000	A	4.79
	Control		B	
Inference	Experiment	0.000	A	3.74
	Control		B	
Clarify	Experiment	0.000	A	3.15
	Control		B	
Strategy Decision	Experiment	0.012	A	2.98
	Control		B	

**Different groupings letter indicated that significant difference exists*

4. Representation of Critical Thinking Enhancement

The researcher selected one subject in the experimental group to discuss in more detail to represent improvement in critical thinking score after the treatment. Based on the results of Table 8, PBL with ChatGPT and GeoGebra did not affect identify indicator. It was because the student's identify ability was already good beforehand. This could be seen more deeply in Figure 5 and Figure 6. In pre-test, the subject wrote the all information needed to solve the problem which were plane H_1 and H_2 also point $(1, -2, -1)$ and the objective of the problem which was to determine

ISSN 2089-8703 (Print)
ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v14i4.12091

one of the pencils of the plane. This meant that the subject wrote all the information and objective of problem. Subsequently, based on Figure 3, the subject wrote all the information in the problem on post-test. The subject also explored the center point and radius of both spheres S_1 and S_2 with precise algebraic calculations. This meant that the subject wrote the explicit and implied information. Thus, the subject also fulfilled identify indicators on pre-test and post-test.

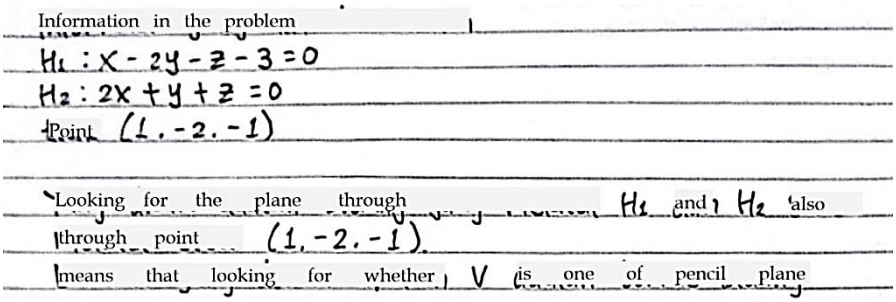

Figure 5. Subject Pre-test on Identify Indicator

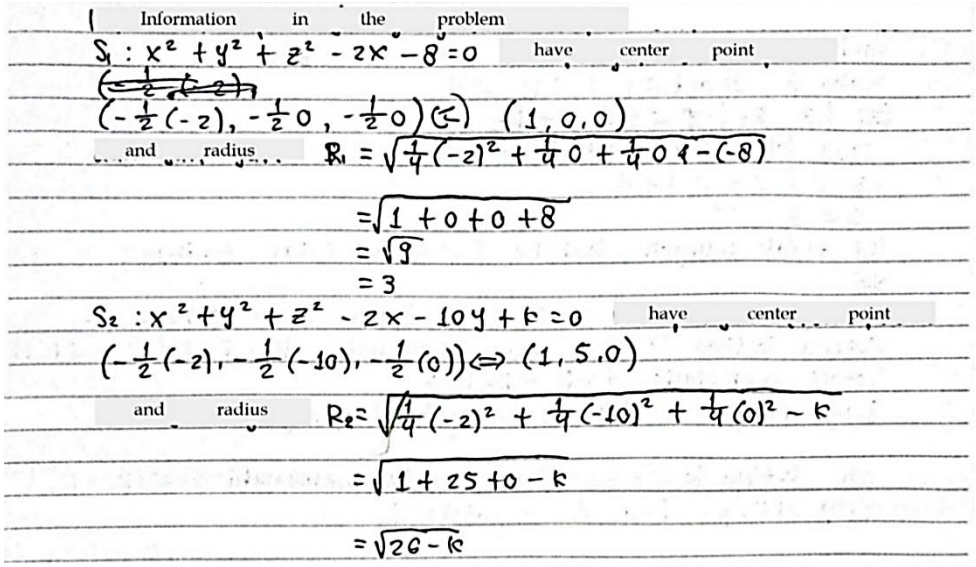

Figure 6. Subject Post-test on Identify Indicator

Before treatment was applied, the subject had irregularities when considering the credibility, interpreted, and took further steps on the information obtained. Based on Figure 7 below, the subject wrote wrong algebraic statements and calculations that plane V passed through H_1 and H_2 which was not passed through. This meant that the subject did not consider the credibility of the details in depth. This led to errors conclusions and explanations so the subject did not decide solution correctly. Thus, the subject did not fulfill justify, inference, clarify, and strategy decision indicators before treatment.

AKSIOMA: Jurnal Program Studi Pendidikan Matematika
Volume 14, No. 4, 2025, 1078–1096
ISSN 2089-8703 (Print)
ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v14i4.12091

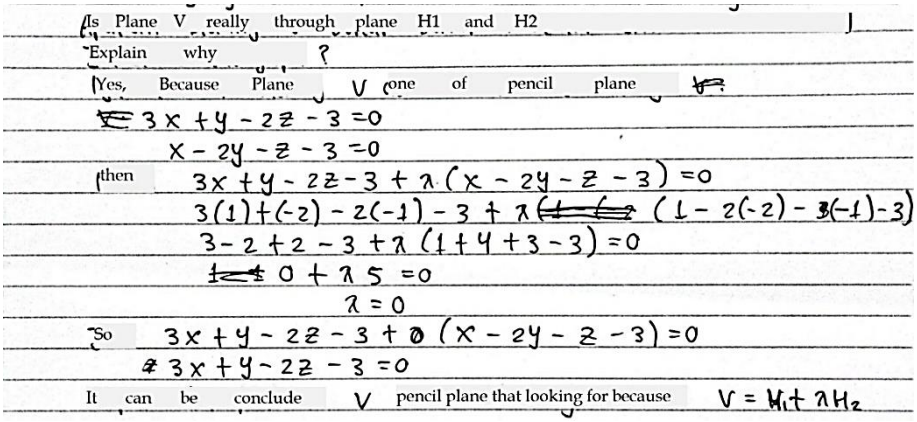

Figure 7. Subject Pre-test on Justify, Inference, Clarify and Strategy Decision Indicators

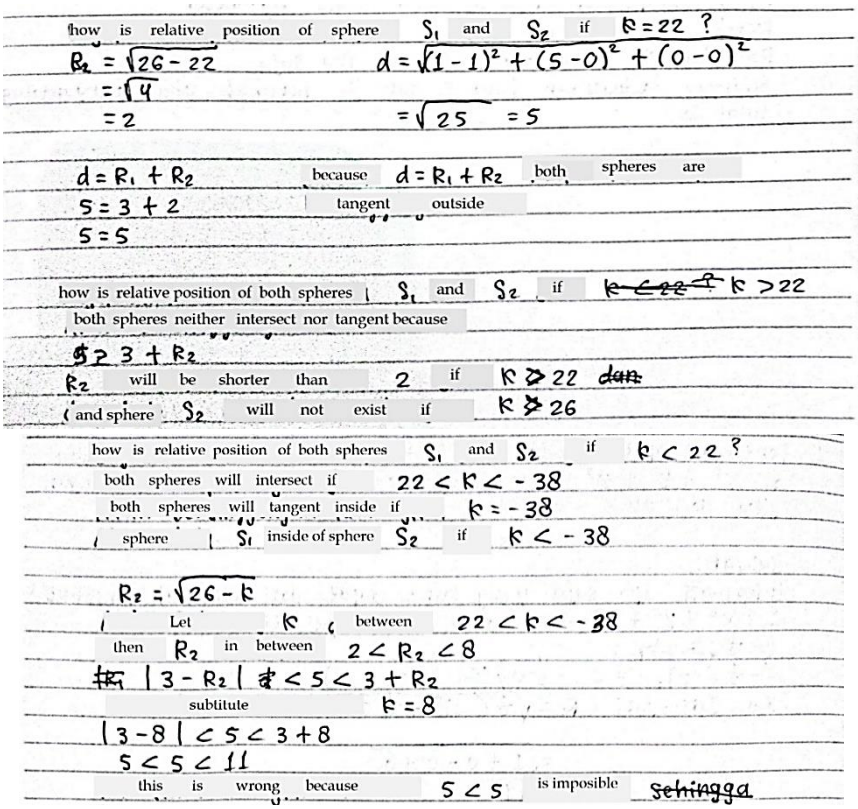

Figure 8. Subject Post-test on Justify and Strategy Decision Indicators

However, after the treatment subject's ability to consider the credibility, interpret, and make further steps to follow up on the information obtained was increased. This could be seen in Figure 8 above that the subject wrote value of k that made the relative position of both spheres change. The subject knew that the lesser value of k, the greater radius (R_2) of sphere S_1 and S_2. So that the subject precisely determined the right interval of k values with the right algebraic calculations. Subject found that the relative position of both spheres S_1 and S_2 would not always intersect when $k < 22$. The subject knew that when $k = -38$, the relative position of S_1 and S_2 did not intersect anymore and would be tangent inside. It continued to expand when

ISSN 2089-8703 (Print)
ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v14i4.12091

$k < -38$ relative position of both spheres would be sphere S_1 inside of sphere S_2. This meant that the subject considered the credibility of the details correctly and the subject was appropriate in determining the solution to the problem. Subsequently, in Figure 9 below, the proper solution led the subject to conclude and explain properly about the solution. Thus, the subject fulfilled justify, inference, clarify, and strategy decision indicators after treatment.

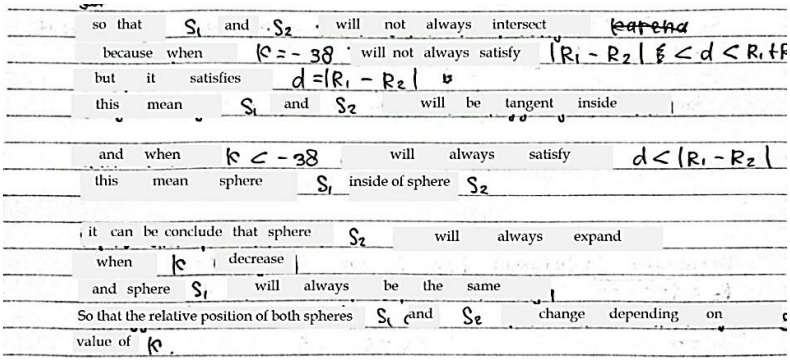

Figure 9. Subject Post-test on Inference and Clarify Indicators

5. Discussion

The findings indicate that the Problem-Based Learning (PBL) model assisted by ChatGPT and GeoGebra significantly enhanced students' critical thinking skills. This enhancement was particularly evident in the indicators of justification, inference, clarification, and strategy decision, while the identification indicator showed no significant change. The results provide a deeper narrative of how students engaged with AI-generated responses and mathematical software to develop higher-order thinking skills.

One key explanation is that ChatGPT, while offering rapid solutions, often produces incomplete or inaccurate answers. Instead of hindering the process, these inaccuracies stimulated students to question, analyze, and verify information, thereby fostering their evaluative and reasoning skills. GeoGebra, on the other hand, provided visual confirmation of mathematical concepts, enabling students to validate solutions and strengthen their logical arguments. This combination created a unique learning cycle where students were exposed to differing perspectives, encouraged to test the credibility of information, and guided to construct well-reasoned conclusions.

The main findings of this study can be summarized as follows (1) PBL integrated with ChatGPT and GeoGebra had a significant positive impact on students' overall critical thinking abilities. (2) The improvement was concentrated on higher-level indicators such as justify, inference, clarify, strategy decision, while the identify indicator remained unaffected, likely because students already possessed sufficient skills in recognizing explicit information before the intervention. (3) The learning environment encouraged students to move beyond surface-level recognition toward deeper analytical and evaluative thinking.

Several factors contributed to these results. The group-based structure of PBL required students to collaborate, compare answers, and defend reasoning, which enhanced critical discourse. ChatGPT's limitations acted as a

ISSN 2089-8703 (Print)
ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v14i4.12091

cognitive trigger, prompting students to evaluate answers critically rather than accepting them at face value. GeoGebra further strengthened students' confidence by providing a concrete visual foundation for abstract analytic geometry problems. These elements worked synergistically, making the intervention effective in enhancing critical thinking.

This research presents both strengths and limitations. Its strengths include the innovative integration of AI such as ChatGPT and mathematical software such as GeoGebra into a PBL framework, rigorous use of validated instruments, and statistical analysis with ANCOVA. However, limitations are noted that the sample size was relatively small which are only 47 students, the intervention was restricted to a single mathematical topic, and the treatment period was limited to four weeks. Furthermore, the identification indicator was unaffected, suggesting that certain aspects of critical thinking require different or longer-term interventions.

When compared with previous studies, the results align with existing evidence. The result of this study was align with Magpantay and Pasia (2022), Marthaliakirana et al. (2022), and also Sarwastuti and Purnomo (2023) which showed that PBL model enhanced students' critical thinking skills. Moreover, it was also in line with Hendriana et al. (2020), Samura and Darhim (2023), Sofyan et al. (2022), and also Ziatdinov and Valles (2022) which showed that GeoGebra also enhanced students' critical thinking skills. Furthermore, it was also in line with Rizaldi et al. (2024), van den Berg and du Plessis (2023) and also Walter (2024) which showed that ChatGPT enhance students' critical thinking skills. Other research examined problem-based learning and ChatGPT to enhance critical thinking ((Jeong, 2023); (Rizaldi et al., 2024); (Walter, 2024); (Zhang et al., 2025)) that also found students' ability to analyze and evaluate increased. Additionally, the findings were consistent with those of Devi et al. (2024), Jinyoung and Minjeong (2024), Khoeriah et al., (2024), and also Mairing et al. (2024) which found that PBL, ChatGPT, and GeoGebra enhanced students' problem solving and mathematical understanding skills. Different with earlier studies that examined PBL, ChatGPT, or GeoGebra in isolation, this study demonstrates the synergistic effect of integrating all three tools to address multiple dimensions of critical thinking.

The increase in critical thinking scores resulted from the developing of student's analytical, evaluative, and decision-making abilities during the PBL learning process supported by ChatGPT and GeoGebra. This is supported by Figure 1 and Figure 2 that ChatGPT and GeoGebra was useful tools to assist PBL. This is due to the ChatGPT answers being a new perspective for students to encourage students to consider the ChatGPT answers. In accordance with Rizaldi et al. (2021) and Southworth (2022), the ability to analyze and evaluate will be encouraged by the existence of different points of view or problems, especially if one view contains aspects that are not completely true and the other view is completely true. Khalil et al. (2019) also found that GeoGebra is the best scaffolding tool for fostering mathematical thinking of analytic geometry. Students always have difficulty with abstract concepts, but through GeoGebra, most of these concepts can be concretized. GeoGebra helps students to provide visualization, conjunction of

ISSN 2089-8703 (Print)
ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v14i4.12091

algebraic and geometric thinking, and analytic and synthetic views. Thus, ChatGPT and GeoGebra are not just tools used to solve problems, but also become a new perspective to encourage students to analyze and evaluate.

On the other hand, as shown in Table 8, the integration of PBL with ChatGPT and GeoGebra positively influenced the indicators of justify, inference, clarify, and strategy decision, but did not affect the identify indicator. The issues in this study are aligned with Egitim (2022), Pangestika and Faiziyah, (2022), and also Rizaldi et al. (2021) who also found that the main challenge faced by students in higher education was not in recognizing the context, but rather in analyzing and evaluating it. It showed that in Figure 5 and Figure 6 that suggests that students generally have no difficulty in identifying explicit information provided in a problem, since such information is often directly available and requires only basic comprehension. However, greater challenges arise when students must go beyond identification and engage in higher-order thinking processes such as evaluating the credibility of information, justifying their reasoning, drawing inferences, and determining appropriate strategies. The enhancement observed in these higher-level indicators after the treatment highlights the effectiveness of PBL supported by ChatGPT and GeoGebra in fostering deeper analytical and evaluative thinking, rather than merely surface-level recognition, as illustrated in Figure 7, where the subject made errors in justification and inference during the pre-test, and in Figure 8 and Figure 9, where the subject successfully demonstrated accurate justification, inference, clarification, and strategy decision after the treatment.

CONCLUSION AND SUGGESTION

The result showed that PBL assisted by ChatGPT and GeoGebra enhances students' critical thinking. It also can be seen descriptively that the average pre-test and post-test of experimental (47.96 and 62.13) and control (55.48 and 54.26) groups were significantly different. ANCOVA confirmed that PBL with ChatGPT and GeoGebra significantly enhance students' critical thinking in the experimental group compared to the control group. Furthermore, the treatment did not affect identify indicator, but had a positive effect on justify, inference, clarify, and strategy decision indicators.

PBL model with ChatGPT and GeoGebra is more effective for students willing to critically assess AI answers, as irrelevant or unverified responses can hinder problem-solving. Future research is recommended to design strategies that motivate less-analytical students to evaluate AI responses, explore the use of AI with other adaptive tools across subjects, and conduct longitudinal studies to assess the long-term impact of such integrations on critical thinking skills along with skills such as creative thinking, problem-solving, and digital literacy.

REFERENCES

Alashwal, H. A., & Barham, A. I. (2025). Sustaining problem-based learning: A mixed-methods exploration of its long-term effects on primary students' mathematical problem solving. *Social Sciences and Humanities Open*, *12*. https://doi.org/10.1016/j.ssaho.2025.101717

Astutik, I. D., Pramasdyahsari, A. S., & Setyawati, R. D. (2024). Development of PJBL-STEM

ISSN 2089-8703 (Print)
ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v14i4.12091

based E-books Assisted by Geometry Calculator to Foster Students' Critical Thinking Ability. *Jurnal Matematika Kreatif-Inovatif*, *15*(1), 69–82. https://doi.org/10.15294/eddrkj25

Bezanilla, M. J., Galindo-Domínguez, H., & Poblete, M. (2021). Importance of teaching critical thinking in higher education and existing difficulties according to teacher's views. *Multidisciplinary Journal of Educational Research*, *11*(1), 20–48. https://doi.org/10.447/remie.2021.6159

Breen, E., & Fallon, H. (2005). Developing Student Information Literacy Skills to Support Project and Problem-based Learning. *Handbook of Enquiry & Problem Based Learning*, 179–188.

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., & Wang, Y. (2024). A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology*, *15*(3), 1–45.https://doi.org/10.1145/3641289

Cohen, L., Manion, L., & Morrison, K. (2007). *Reasearch Methods in Education*. Routledge. https://doi.org/10.4324/9781315158501-17

Devi, K., Dony, P., Ahmad, F., & Edwin, M. (2024). Pengambangan Perangkat Pembelajaran Berbasis Discovery Learning Berbantuan GeoGebra untuk Meningkatkan Kemampuan Pemecahan Masalah Matematis. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *13*(2), 379–393. https://doi.org/10.24127/ajpm.v13i2.8468

Egitim, S. (2022). Do Japanese students lack critical thinking? Addressing the misconception. *Power and Education*, *14*(3), 304–309. https://doi.org/10.1177/17577438221107203

Ennis, R. H. (1996). *Critical Thinking*. New Jersey: Printice-Hall Inc.

Ennis, R. H. (2018). Critical Thinking Across the Curriculum: A Vision. *Topoi*, *37*(1), 165–184. https://doi.org/10.1007/s11245-016-9401-4

Hendriana, H., Setiawan, W., & Aripin, U. (2020). The Effectiveness of Geogebra's Metacognitive Guidance to Develop Critical Thinking Skills. *International Journal on Emerging Mathematics Education*, *3*(2), 129. https://doi.org/10.12928/ijeme.v3i2.13806

Jeong, R. (2023). Research on Problem-Based Learning(PBL) Literature Class Method Using ChatGPT. *The Korean Society of Culture and Convergence*. https://doi.org/10.33645/cnc.2023.10.45.10.219

Jinyoung, H., & Minjeong, L. (2024). Problem-based Learning and ChatGPT : Explorative analysis of the relationship between chatgpt prompts and problem-solving skills. *The Journal of General Education.* https://doi.org/DOI:10.24173/jge.2024.01.26.4

Kardoyo, Nurkhin, A., Muhsin, & Pramusinto, H. (2020). Problem-based learning strategy: Its impact on students' critical and creative thinking skills. *European Journal of Educational Research*, *9*(3), 1141–1150.

ISSN 2089-8703 (Print)
ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v14i4.12091

https://doi.org/10.12973/EU-JER.9.3.1141

Khalil, M., Khalil, U., & ul Haq, Z. (2019). Geogebra as a Scaffolding Tool for Exploring Analytic Geometry Structure and Developing Mathematical Thinking of Diverse Achievers. *International Electronic Journal of Mathematics Education*, *14*(2), 427–434. https://doi.org/10.29333/iejme/5746

Khoeriah, N., Mahmudi, A., & Sudrajat, S. (2024). Pengembangan LKS Berbasis PBL Berbantuan GeoGebra Untuk Meningkatkan Kemampuan Pemecahan Masalah Dan Kesadaran Metakognitif Siswa. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *13*(1), 65. https://doi.org/10.24127/ajpm.v13i1.8534

Lodico, M. G., Spaulding, D. T., & Voegtle, K. H. (2010). *Methods in Educational Research: From Theory to Practice, 2nd Edition* (Issue 3). Jossey-Bass.

Magpantay, I. C. D., & Pasia, A. E. (2022). Problem-Based Learning Materials in Upskilling Mathematics Critical Thinking Skills. *International Journal of Science, Technology, Engineering and Mathematics*, *2*(4), 74–91. https://doi.org/10.53378/352940

Mairing, J. P. (2017). *Statistika Pendidikan: Konsep & Penerapannya Menggunakan Minitab dan Microsoft Excel*. Yogyakarta: Penerbit Andi.

Mairing, J. P., Rizaldi, M., Pandiangan, P., & Lada, E. Y. (2024). Development of Problem , YouTube , and ChatGPT Learning Model to Improve Students ' Proving Ability in Real Analysis. *Jurnal Matematika Kreatif-Inovatif*, *15*(1), 169–184. https://doi.org/10.15294/sgz61469

Marthaliakirana, A. D., Suwono, H., Saefi, M., & Gofur, A. (2022). Problem-based learning with metacognitive prompts for enhancing argumentation and critical thinking of secondary school students. *Eurasia Journal of Mathematics, Science and Technology Education*, *18*(9). https://doi.org/10.29333/ejmste/12304

Nauman, S. (2017). Lack of critical thinking skills leading to research crisis in developing countries: A case of Pakistan. *Learned Publishing*, *30*(3), 233–236. https://doi.org/10.1002/leap.1091

Pangestika, W. A., & Faiziyah, N. (2022). Students' Mathematics Power Viewed From the Student'S Critical Thinking Skills. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *11*(3), 1897. https://doi.org/10.24127/ajpm.v11i3.5322

Rizaldi, M., Hidayanto, E., & Rahardi, R. (2021). Berpikir Kritis Siswa Melalui Aktivitas Problem Posing dengan Konteks Masalah yang Tidak Masuk Akal. *Jurnal Pendidikan: Teori, Penelitian, Dan Pengembangan*, *6*(2), 191–198. https://doi.org/10.17977/jptpp.v6i2.14444

Rizaldi, M., Sasalia, P., & Pancarita. (2024). Efektivitas ChatGPT Untuk Mendorong Berpikir Kritis Mahasiswa Dalam Menyelesaikan Masalah Aljabar. *EDU-MAT: Jurnal Pendidikan Matematika*, *12*, 168–181.

ISSN 2089-8703 (Print)
ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v14i4.12091

https://doi.org/10.20527/edumat.v12i1.18668

Rutherford, A. (2011). *ANOVA and ANCOVA: A GLM Appraoch*. John Wiley & Sons.

Samura, A. O., & Darhim. (2023). Improving Mathematics Critical Thinking Skills of Junior High School Students Using Blended Learning Model (BLM) in GeoGebra Assisted Mathematics Learning. *International Journal of Interactive Mobile Technologies*, *17*(2), 101–117. https://doi.org/10.3991/ijim.v17i02.36097

Sánchez-Ruiz, L. M., Moll-López, S., Nuñez-Pérez, A., Moraño-Fernández, J. A., & Vega-Fleitas, E. (2023). ChatGPT Challenges Blended Learning Methodologies in Engineering Education: A Case Study in Mathematics. *Applied Sciences (Switzerland)*, *13*(10). https://doi.org/10.3390/app13106039

Santos-Meneses, L. F., & Drugova, E. A. (2023). Trends in critical thinking instruction in 21st-century research and practice: Upgrading instruction in digital environments. *Thinking Skills and Creativity*, *49*, 101383. https://doi.org/https://doi.org/10.1016/j.tsc.2023.101383

Sarwastuti, H. T., & Purnomo, Y. W. (2023). Pengaruh Problem Based Learning Terhadap Keterampilan Berpikir Kritis Matematika Materi Lingkaran. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *12*(1), 473. https://doi.org/10.24127/ajpm.v12i1.6172

Sofyan, D., Sugandi, A. I., Linda, L., Sari, D. R., & Bernard, M. (2022). Penerapan Strategi React Berbantuan Geogebra Dalam Meningkatkan Kemampuan Berpikir Kritis Matematis. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *11*(3), 1668. https://doi.org/10.24127/ajpm.v11i3.4995

Southworth, J. (2022). Bridging critical thinking and transformative learning: The role of perspective-taking. *Theory and Research in Education*, *20*(1), 44–63. https://doi.org/10.1177/14778785221090853

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, *10*(1). https://doi.org/10.1186/s40561-023-00237-x

van den Berg, G., & du Plessis, E. (2023). ChatGPT and Generative AI: Possibilities for Its Contribution to Lesson Planning, Critical Thinking and Openness in Teacher Education. *Education Sciences*, *13*(10). https://doi.org/10.3390/educsci13100998

Walter, Y. (2024). Embracing the future of Artificial Intelligence in the classroom: the relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education*, *21*(1). https://doi.org/10.1186/s41239-024-00448-3

Yang, Y. T. C., Gamble, J. H., Hung, Y. W., & Lin, T. Y. (2014). An online adaptive learning

ISSN 2089-8703 (Print)
ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v14i4.12091

environment for critical-thinking-infused English literacy instruction. *British Journal of Educational Technology*, *45*(4), 723–747. https://doi.org/10.1111/bjet.12080

Zakaria, M. I., Hanri, C., Noer, S. H., Triana, M., & Widyastuti. (2025). 4C skills teaching activities for mathematics teachers: application of modified nominal group technique. *Journal of Education and Learning*, *19*(2), 626–633. https://doi.org/10.11591/edulearn.v19i2.22337

Zein, S., Yaniawati, R. P., & Mudrikah, A. (2024). The Role and Evaluation of ChatGPT as a Virtual Tutor in Improving Students ' Creative and Critical Abilities Reviewed from Probing-Prompting Abilities. *Jurnal Matematika Kreatif-Inovatif*, *15*(2), 501–517. https://journal.unnes.ac.id/journals/kreano/article/view/14343

Zhang, D., Wijaya, T. T., Wang, Y., Su, M., Li, X., & Damayanti, N. W. (2025). Exploring the relationship between AI literacy, AI trust, AI dependency, and 21st century skills in preservice mathematics teachers. *Scientific Reports*, *15*(1), 1–15. https://doi.org/10.1038/s41598-025-99127-0

Ziatdinov, R., & Valles, J. R. (2022). Synthesis of Modeling, Visualization, and Programming in GeoGebra as an Effective Approach for Teaching and Learning STEM Topics. *Mathematics*, *10*(3), 398. https://doi.org/10.3390/math10030398