

THE EFFECTIVENESS OF USING CHATGPT ON THE CRITICAL THINKING SKILLS OF VOCATIONAL HIGH SCHOOL ACCOUNTING STUDENTS

Rahmawati¹, Nujmatul Laily²

¹Faculty of Economics and Business, State University of Malang, Indonesia, rahmawati.2204216@students.um.ac.id

²Faculty of Economics and Business, State University of Malang, Indonesia, nujmatul.laily.fe@um.ac.id

DOI

<https://doi.org/10.26740/jupe.v14n3.p297-308>

Article history

Received

7 May 2026

Revised

29 July 2026

Accepted

25 August 2026

How to cite

Rahmawati, R., & Laily, Nujmatul. (2026). The effectiveness of using ChatGPT on the critical thinking skills of Vocational High School accounting students. *Jurnal Pendidikan Ekonomi (JUPE)*, 14(3), 297-308.

<https://doi.org/10.26740/jupe.v14n3.p297-308>

Kata Kunci: Kecerdasan Buatan Generatif, ChatGPT, Kemampuan Berpikir Kritis, Siswa Akuntansi

Keywords: *Generative Artificial Intelligence, ChatGPT, Critical Thinking Skills, Accounting Student*

Corresponding author

Rahmawati

rahmawati.2204216@students.um.ac.id

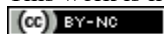
Abstrak

Penelitian ini bertujuan mengetahui efektivitas penggunaan ChatGPT terhadap kemampuan berpikir kritis siswa SMK Akuntansi pada mata pelajaran layanan perbankan. Metode penelitian menggunakan pendekatan kuantitatif dengan jenis *quasi experimental* berupa *nonequivalent control group design*. Sampel penelitiannya dilakukan di SMK Negeri 1 Malang yang terdiri dari dua kelas, yakni kelas eksperimen di kelas XI AKL (Akuntansi Keuangan dan Lembaga) 3 dan kelas kontrol di kelas XI AKL 1. Data dikumpulkan dari tes kemampuan berpikir kritis yang dibagikan sebelum dan sesudah pembelajaran. Data dianalisis dengan menerapkan uji *Independent Sample T-Test* dan *N-Gain*. Hasil temuan memperlihatkan perbedaan hasil tes di kedua kelas, di mana kelas eksperimen meningkat lebih tinggi dibanding kelas kontrol. Penggunaan ChatGPT dipembelajaran mengindikasikan dapat mendukung kemampuan berpikir kritis siswa selama tetap berada di bawah bimbingan guru yang memberikan arahan secara tepat.

Abstract

This study aims to determine the effectiveness of ChatGPT use on the critical thinking skills of vocational high school accounting students in banking services courses. The research method uses a quantitative approach with a quasi-experimental type in the form of a nonequivalent control group design. The research sample was conducted at SMK Negeri 1 Malang which consisted of two classes, namely the experimental class in class XI AKL (Financial and Institutional Accounting) 3 and the control class in class XI AKL 1. The data were analyzed using the Independent Sample T-Test and N-Gain analysis. The findings indicate that there were differences in test scores between the two classes, with the experimental class showing greater improvement than the control class. The use of ChatGPT in education suggests that it can support students' critical thinking skills as long as it is used under the guidance of a teacher who provides appropriate direction.

This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).



INTRODUCTION

In the era of 21st-century learning, the need to master 4C skills (Communication, Collaboration, Critical Thinking and Problem Solving, Creativity and Innovation) has become increasingly important in the workplace (Prayogi, 2020; Thornhill-Miller et al., 2023). A crucial skill that vocational high school students must possess is critical thinking, which refers to the ability to analyze, evaluate, and draw logical conclusions from various types of information and arguments (Afifah et al., 2025; Cui & Zhao, 2024; Dad & Akbar, 2023). In the context of education, particularly in the accounting department, critical thinking skills are of utmost importance because students must be able to evaluate financial information, make ethical judgments, and solve complex problems related to the business world (Cruz et al., 2020; Wolcott & Sargent, 2021).

However, several studies indicate that the critical thinking skills of accounting students are relatively low. Wolcott & Sargent (2021) state that the accounting curriculum tends to focus too heavily on procedures, which hinders students' analytical abilities. Meanwhile, conventional accounting teaching methods, in which teachers explain the material while students merely listen, result in students adopting a passive attitude, struggling to understand and critically evaluate the material, and being unable to apply accounting concepts independently (Risanti & Bayu, 2022; Rossouw & Steenkamp, 2025). Kurniawan et al. (2021), Rosyida & Bahtiar (2024), dan Widiastuti & Hamidi (2023) also reveal that the critical thinking skills of vocational high school students, particularly those in the Accounting program, fall into the low category, especially in the ability to make assumptions, draw conclusions, and evaluate arguments skills that can pose serious obstacles when entering the workforce (Alviasari et al., 2025). This finding is further supported by interviews with accounting teachers at SMK Negeri 1 Malang, which revealed that students' critical thinking skills are still not optimal, particularly in analyzing cases, evaluating information, and drawing conclusions independently, especially in the Banking Services course. Although some students are able to identify key issues, these abilities are not yet consistently demonstrated and still require teacher guidance to connect concepts and justify their reasoning logically. Furthermore, classroom learning activities are generally conducted through discussions, presentations, and teacher-centered explanations. While several students have independently used ChatGPT to search for information or assist with completing assignments, its integration into classroom instruction has not yet been systematically implemented as a tool to support the development of critical thinking skills. These conditions indicate that students' ability to connect concepts, evaluate information, and draw logical conclusions still needs to be strengthened through learning strategies that emphasize deeper analysis and evaluation.

With technological advancements, the emergence of Generative Artificial Intelligence (GenAI) such as ChatGPT presents new opportunities to transform learning by providing interactive and personalized learning experiences that focus on developing cognitive skills, including critical thinking (Ali et al., 2024; Chiu, 2024; Deng et al., 2025). Data from Statcounter (2026) show that ChatGPT dominates GenAI usage in Indonesia at 85%, followed by Perplexity (9.9%), Google Gemini (3.26%), Claude (1.38%), and Microsoft Copilot (0.45%). This popularity indicates a high level of student exposure to GenAI technology. Research by Essel et al. (2024) shows that the use of ChatGPT can effectively improve students' critical thinking skills compared to conventional learning.

The use of ChatGPT in learning is closely related to constructivist theory, which views students as active learners who construct knowledge through their own experiences and thinking processes, rather than merely receiving or collecting information (Azzahra et al., 2025; Bustomi et al., 2024). In this process, students' active involvement is essential, while internal thinking and reasoning are considered more meaningful than passive information reception (Mulyadi, 2022). Thus, learning becomes meaningful when students are able to connect new information with their prior knowledge. In line with this view, John Dewey emphasized that learning experience plays a central role in knowledge construction (Loria et al., 2025). Students develop a deeper and more meaningful understanding of a subject when they are directly involved in inquiry-based activities. Therefore, constructivist learning should be implemented through models that provide opportunities for exploration, one of which is the inquiry-based learning model (Ulya, 2024). This constructivist perspective is closely aligned with the Practical Inquiry Model proposed by Garrison et al. (2001), which explains that knowledge is constructed through a sequence of inquiry processes involving exploration, integration, and resolution. These processes form the foundation of critical thinking because learners are required to investigate information, evaluate evidence, connect ideas, and formulate reasoned conclusions (Garrison et al., 2001; Loria et al., 2025; Yang & Niu, 2026).

In accounting education, the use of ChatGPT is particularly relevant because accounting learning requires not only procedural knowledge, but also the ability to interpret financial information, analyze transactions, evaluate evidence, and make logical decisions (Burney et al., 2023; Rensburg, 2024). ChatGPT facilitates the learning process by providing interactive and personalized experiences, including support in completing assignments (Hsu & Silalahi, 2024; Syahri et al., 2024). ChatGPT is widely used in educational contexts, including accounting education, as it is considered a Generative AI tool capable of generating text, performing analysis, and potentially enhancing critical thinking skills (Burney et al., 2023; Oliński et al., 2024; Rensburg, 2024). Guo & Lee (2023) show that structured interaction with ChatGPT involving analysis, evaluation, and logical inference can help students understand concepts and improve their critical thinking abilities. This finding is consistent with Tlili et al. (2023) and Nuwa et al. (2025), who state that ChatGPT can provide additional examples, re-explanations, and feedback that enable students to continuously validate their understanding and enhance their critical thinking through ongoing dialogue. ChatGPT also enables dialogic interaction that helps students test ideas, clarify concepts, and gradually construct understanding (Bannert et al., 2023). These findings indicate that the strategic integration of ChatGPT within educational settings holds significant promise for fostering the

growth of students' critical thinking competencies. This is also consistent with the findings of Elsa Sabrina et al. (2024), which demonstrate that ChatGPT can develop critical thinking skills, although variations in impact still depend on user perception. Ratnawati et al. (2024) also report that the use of ChatGPT can increase critical thinking scores by 12.63% among students.

Previous studies have demonstrated that ChatGPT can support students' critical thinking skills, learning engagement, and cognitive development across various educational contexts, particularly in higher education and general secondary education (Almazrou et al., 2025; Dedad et al., 2024; Drushlyak et al., 2025; Qawqzeh, 2024). However, empirical evidence regarding its effectiveness in vocational education, especially among vocational high school accounting students, remains limited. Furthermore, most previous studies have focused on students' perceptions, experiences, or general learning outcomes, while relatively few have examined the effect of ChatGPT on critical thinking skills through a quasi-experimental approach grounded in a specific theoretical framework. Essel et al. (2024) and Nasr et al. (2025) also recommend further investigation into how ChatGPT supports the cognitive processes underlying critical thinking across different educational contexts and disciplines. Therefore, this study addresses these gaps by examining the effectiveness of ChatGPT-assisted learning on vocational high school accounting students' critical thinking skills using a quasi-experimental design. The study is guided by constructivist theory and the Practical Inquiry Model (Garrison et al., 2001), which provide a framework for understanding how ChatGPT may support the development of critical thinking through the stages of exploration, integration, and resolution.

The effectiveness of ChatGPT in supporting critical thinking skills is also influenced by how students utilize the technology. Students who use ChatGPT passively, such as merely accepting answers without further analysis, demonstrate low levels of critical thinking ability. Conversely, students who use ChatGPT actively and reflectively are able to demonstrate better critical thinking processes, for example by asking follow-up questions, verifying the accuracy of information, connecting answers to prior knowledge, and applying information to solve real-world tasks or problems (N. R. Nasr et al., 2025). This study is also supported by the Practical Inquiry Model developed by Garrison et al. (2001), which explains the process of critical thinking through four stages, namely triggering event, exploration, integration, and resolution. In the triggering event stage, students encounter problems or confusion that stimulate curiosity (Nasr et al., 2025). In the exploration stage, students can use ChatGPT to gather information, ask follow-up questions, and compare alternative perspectives through interactive dialogue (Garrison et al., 2001). In the integration stage, students critically examine evidence by verifying ChatGPT-generated responses against prior knowledge, textbooks, or other credible sources before formulating reasoned interpretations (Garrison et al., 2001). This process reflects higher-order cognitive skills, including evaluating information, identifying potential biases, and addressing misinformation (Nasr et al., 2025). Finally, in the resolution stage, students apply the knowledge they have constructed through interaction with ChatGPT to solve contextual problems, justify decisions, and formulate evidence-based conclusions (Garrison et al., 2001).

However, the effectiveness of ChatGPT may decrease if it is not accompanied by appropriate pedagogical design (Nasr & Tu, 2025). Studies indicate that when students become overly dependent on ChatGPT, they may lose the ability to process information independently. This can lead to shallow learning, as students tend to seek direct answers without analyzing or evaluating concepts, unlike those who engage in deeper cognitive processing. Nasr et al. (2025) emphasize that ChatGPT can enhance critical thinking skills when used collaboratively, as a learning partner rather than a direct substitute, and when accompanied by reflection. Melisa et al. (2025) and Salido et al. (2025) also explain that ChatGPT can be integrated into instructional strategies to support more complex reasoning and reflection when used by teachers as a guiding tool rather than an answer provider. Therefore, when integrated within an inquiry-based learning environment and supported by the stages of the Practical Inquiry Model, ChatGPT has the potential to facilitate students' exploration of information, evaluation of evidence, and application of knowledge in problem-solving activities (Garrison et al., 2001; Nasr et al., 2025). Drawing from these theoretical and empirical premises, ChatGPT-assisted learning is expected to produce greater improvements in students' critical thinking skills than conventional learning methods. Based on the theoretical foundation and empirical studies discussed above, the conceptual framework of this study is presented in Figure 1.

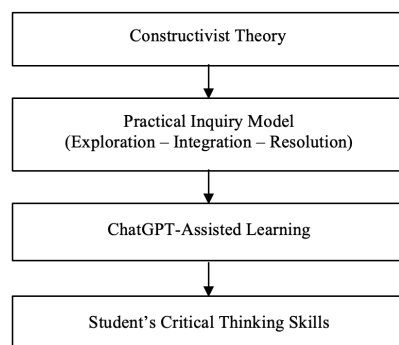


Figure 1. Conceptual Framework

Based on the conceptual framework, this study hypothesizes that there is a difference in the critical thinking skills of students between those who learn through ChatGPT-assisted learning and those who learn through conventional learning. This study aims to determine how effectively ChatGPT fosters critical thinking among vocational high school accounting students by contrasting their skill levels with those of a control group using conventional learning methods. It is expected that the findings will provide insights for teachers and schools regarding the use of ChatGPT as a learning medium that can support the strengthening of 21st-century competencies in vocational accounting education.

METHODOLOGY

This research adopts a quantitative framework utilizing a quasi-experimental design to evaluate the efficacy of ChatGPT integration in fostering critical thinking skills among vocational high school accounting students. This approach was chosen because it aligns with the characteristics of the research sample, which consists of both control and experimental groups that are pre-determined and cannot be randomly assigned (Cresswell, 2023). The experimental class received learning instruction supported by ChatGPT, while the control class was instructed using conventional instructional materials.

The population of this study consisted of 134 eleventh-grade accounting students in the second semester at SMK Negeri 1 Malang during the 2025/2026 academic year. Participants were selected through non-probability sampling, following a convenience sampling method. This technique was applied because the research could only be conducted in classes provided by the school (Creswell, 2023; Khalid, 2024). The selected sample comprised two classes, namely Class XI AKL 3 as the experimental class ($n = 27$) and Class XI AKL 1 as the control class ($n = 27$), resulting in a total sample of 54 students. The learning material used in this study focused on Banking Services in the Accounting subject. The study was conducted over three weeks, consisting of six instructional meetings in both the experimental and control classes. This duration was based on a study by Yang & Niu (2026), which implemented a three-week AI-based learning intervention to observe its effect on critical thinking skills. This research utilized a test instrument as the principal tool for data collection and subsequent analysis.

In the experimental class, learning was conducted using an inquiry-based learning model supported by ChatGPT. The learning process followed the stages of the Practical Inquiry Model, namely triggering event, exploration, integration, resolution, and reflection (Garrison et al., 2001). Students analyzed banking-related cases, explored information through interactions with ChatGPT, evaluated the accuracy of AI-generated responses using learning resources, and synthesized their findings into structured answers. Through this process, ChatGPT functioned as a learning support tool that facilitated information exploration and critical evaluation. Meanwhile, in the control class, learning followed the same inquiry-based learning stages but without the use of ChatGPT or other generative AI tools. Students explored learning materials through textbooks, modules, and classroom discussions under teacher guidance. They analyzed the given cases, reviewed concepts using conventional learning resources, and developed answers based on their own understanding. The use of the same instructional stages in both classes was intended to maintain comparability while distinguishing the presence or absence of ChatGPT as the primary treatment.

The instrument used in this study was a written test designed to measure students' critical thinking skills in the Banking Services topic, particularly the subtopics of types of banks in Indonesia and the activities of Commercial Banks and Rural Banks (BPR). The instrument consisted of five essay questions developed based on the Practical Inquiry Model proposed by Garrison et al. (2001), which includes the indicators of exploration, integration, and resolution. Three items measured the exploration indicator by requiring students to identify and analyze differences among types of banks based on their functions, ownership, and operational status. One item measured the integration indicator by asking students to connect banking concepts with real-life situations and provide logical arguments. One item measured the resolution indicator by requiring students to evaluate and determine the appropriate application of banking concepts in a given context. The same set of questions was administered as both the pretest and posttest in the experimental and control classes to ensure consistency in measuring students' critical thinking skills.

Table 1. Blueprint of the Critical Thinking Skills Instrument

Item	Indicator	Description	Cognitive Level	Form
1	Exploration	Identifying and analyzing differences in banking functions	C4	Uraian
2	Exploration	Identifying and analyzing differences in bank ownership	C4	Uraian
3	Exploration	Identifying and analyzing differences in operational status	C4	Uraian
4	Integration	Connecting concepts with real-life situations and providing logical arguments	C5	Uraian
5	Resolution	Evaluating and determining the application of banking concepts	C5	Uraian

Table 2. Instrument Score Assessment

Description	Score
The response includes all relevant concepts, presented accurately and specifically, supported by clear analysis and logical argumentation	5
The response includes most relevant concepts with reasonably clear explanations, although the analysis or argumentation is less developed	4
The response includes some correct concepts, but the analysis is incomplete and the argumentation lacks clarity	3
The response demonstrates limited conceptual understanding, inaccurate analysis, and weak argumentation	2
The response is inconsistent with the concepts, does not demonstrate analysis, and lacks argumentation	1
No response is provided or the response is entirely irrelevant	0

The instrument was administered twice, namely as a pretest before the intervention and as a posttest after the intervention, to examine changes in students' critical thinking skills. Before being used, the instrument was first tested for its feasibility through content validity, empirical validity, reliability, item discrimination, and difficulty level analyses.

Content Validity

Prior to empirical testing, the instrument underwent a content validation process to ensure its suitability for measuring students' critical thinking skills. The validation was conducted by a subject-matter expert, namely an accounting teacher responsible for the Banking Services course. The expert evaluated the instrument in terms of content relevance, item construction, language clarity, alignment with learning objectives, and consistency with the critical thinking indicators of exploration, integration, and resolution derived from the Practical Inquiry Model (Garrison et al., 2001). The content validity results were analyzed using Aiken's V coefficient (Retnawati, 2020).

Table 3. Results of the Research Instrument

No	Rated Aspect	Validator Rating (r)	Aiken's V Index	Category
Material				
1	The test item is aligned with the learning outcomes and learning objectives	4	0,75	Moderate
2	The scope of the question and the expected answer is clearly defined	4	0,75	Moderate
3	The test item is aligned with the research objectives	5	1,00	High
Construct				
4	The test items are clearly formulated	4	0,75	Moderate
5	Clear instructions for completing the test are provided	5	1,00	High
6	A scoring rubric is provided	5	1,00	High

The results of the table show that the test instrument is in the medium and high categories so it is suitable for use as a research data collection instrument.

Empirical Validity

Test Empirical validity testing was conducted to examine each test item in order to determine the feasibility of each instrument item. An item meets the validity criteria if the Pearson correlation significance value is below 0.05 and the calculated r-value (r_{count}) is equal to or exceeds the critical r-value (r_{table}) (Retnawati, 2020). The results of the instrument validity test are presented in the following table.

Table 4. Empirical Validity Results of the Research Instrument

Item No.	R _{count}	R _{table} (N=30)	Description
Item 1	0.839	0.268	Valid
Item 2	0.817	0.268	Valid
Item 3	0.762	0.268	Valid
Item 4	0.824	0.268	Valid

Item No.	R _{count}	R _{table} (N=30)	Description
Item 5	0.697	0.268	Valid

The validity analysis reveals that all five essay questions attained R_{count} values exceeding the R_{table} threshold of 0.268. Consequently, every item in the instrument is declared valid.

Reliability Test

This serves to assess the internal consistency and dependability of the research instrument. In accordance with criteria, the instrument is deemed reliable if the Cronbach's Alpha coefficient exceeds the 0.70 threshold (Soesana et al., 2023).

Table 5. Reliability Results of the Research Instrument

Cronbach's Alpha	N of Items
.840	Not Relevant

The reliability analysis yielded a Cronbach's Alpha coefficient of 0.840 for the five-item essay instrument. Since this value exceeds the 0.70 threshold, the instrument demonstrated is deemed reliable.

Difficulty Level Test

This test is conducted to determine the level of difficulty of each test item, whether it falls into the easy, moderate, or difficult category (Sugiyono, 2023). The distribution of item difficulty levels is detailed in the table below.

Table 6. Difficulty Level Results of Research Instrument

No	Mean Score	Difficulty Index	Category
1	3.30	0.66	Moderat
2	2.74	0.55	Moderat
3	2.72	0.54	Moderat
4	3.24	0.65	Moderat
5	3.07	0.61	Moderat

Based on the results, the difficulty index ranged from 0.54 to 0.66, with all test items categorized as moderate (0.31–0.70). This indicates that the questions were neither too easy nor too difficult and were capable of facilitating students' critical thinking processes.

Item Discrimination Test

This test is conducted to determine the ability of each item's ability to distinguish between students who answer correctly and those who do not (Sugiyono, 2023). The distribution of item discrimination is detailed in the table below.

Table 7. Item Discrimination Test of Research Instrument

No	Upper Group Mean	Lower Group Mean	Discrimination Index	Category
1	4.47	2.33	0.43	Good
2	4.00	1.73	0.45	Good
3	3.67	1.47	0.44	Good
4	4.47	2.07	0.48	Good
5	3.80	1.67	0.43	Good

Based on the results, the discrimination index ranged from 0.43 to 0.48, with all test items categorized as good ($0.40 \leq DP \leq 0.70$). This indicates that each test item was able to clearly distinguish between students with high and low levels of ability.

Following the validation of the research instruments, the study proceeded with prerequisite statistical testing. The Shapiro–Wilk normality test was prioritized due to the sample size being under 50 participants, with a significance threshold of $p > 0.05$ as the criterion for a normal distribution (Sugiyono, 2023). Subsequently, Levene's test was applied to assess the equality of variances; homogeneity is confirmed if the resulting p-value exceeds the 0.05 (Chapitak et al., 2025). After these test requirements were fulfilled, the hypothesis testing analysis could then be carried out using the Independent Sample T-Test and the Normalized Gain (N-Gain) Test. The Independent Sample T-Test is pivotal in verifying whether the ChatGPT-supported framework yields a statistically significant advantage in student performance compared to the traditional instructional methods utilized in the control group (Pallant, 2020). The decision-making criterion states that if the significance value is less than 0.05, the hypothesis is accepted. Furthermore, to quantify the relative advancement in students' critical thinking skills, the Normalized Gain (N-Gain) analysis was employed. This

metric facilitates a proportional assessment of improvement by measuring the variance between pretest and posttest performance. ChatGPT is considered effective, or categorized as having a high level of effectiveness, if the gain score obtained is equal to or greater than 0.7 (Astuti & Zulaeha, 2021; Maryani et al., 2022).

RESULTS AND DISCUSSION

RESULTS

To evaluate the efficacy of the intervention, data were collected via pretest and posttest instruments from both groups. While the experimental class engaged with ChatGPT-mediated instruction, the control class proceeded without ChatGPT. The statistical breakdown of these findings is presented in the following section.

Table 8. Descriptive Statistical Analysis

Test	Class	N	Mean	Std. Deviation	Std. Error Mean
Pretest	Experimental	27	68.93	7.452	1.434
	Control	27	60.93	10.561	2.302
Posttest	Experimental	27	93.93	4.755	0.915
	Control	27	88.15	6.323	1.217

As indicated by the descriptive statistics, the sample was equally distributed; student in each class was 27. In the experimental class, the mean pretest score was 68.93 with a standard deviation of 7.452, which increased in the posttest to 93.93 with a standard deviation of 4.755. Descriptively, the data reveals an enhancement in student achievement post-intervention. Crucially, the narrower standard deviation observed in the posttest signifies a convergence in scores, suggesting that the instructional approach effectively bridged the gap between different learner levels. Furthermore, in the control class, the mean pretest score was 60.93 with a standard deviation of 10.561, which increased in the posttest to 88.15 with a standard deviation of 6.323. This indicates a descriptive improvement in students' scores as well as a more homogeneous distribution of scores after the learning process took place.

Before proceeding to the hypothesis testing, the data were subjected to prerequisite analysis to verify that all underlying assumptions for parametric statistics were satisfied. The assessment of data normality and homogeneity was performed using the Shapiro-Wilk and Levene's tests, respectively.

Table 9. Prerequisite Test Results

No	Type of Test	Data Tested	Sig.	Criteria	Conclusion
1	Normalitas (Shapiro-Wilk)	Control Pretest	0.162	Sig > 0.05	Normal
		Control Posttest	0.079	Sig > 0.05	Normal
		Experimental Pretest	0.139	Sig > 0.05	Normal
		Experimental Posttest	0.294	Sig > 0.05	Normal
2	Homogeneity (Levene Test)	Data Variance	0.057	Sig > 0.05	Homogeneous

Normality was confirmed as all Shapiro-Wilk significance values were above 0.05. Furthermore, the homogeneity of variance was established via Levene's test ($p = 0.057 > 0.05$). With these assumptions met, the study continued with hypothesis testing to determine the treatment's overall effectiveness.

Table 10. Results of the Independent Sample T-Test for the Pretest

	F	Sig.	t	df	Significance (2-tailed)
Equal variances assumed	1.114	0.296	-3.216	52	0.002

Source: Pretest independent sample t-test data of the research instrument (2026)

Referring to the results of the test, the obtained value of $t = -3.216$ with a significance value of $0.02 < 0.05$ indicates that there was a difference in the initial critical thinking abilities between the two classes. This suggests that the experimental class and the control class were not entirely equivalent before the treatment was administered. Therefore, the N-Gain analysis was used to evaluate the differences more objectively by considering both the pretest and posttest scores (Astuti & Zulaeha, 2021; Maryani et al., 2022).

Table 11. Results of the Independent Sample T-Test for the Posttest

	F	Sig.	t	df	Significance (2-tailed)
Equal variances assumed	0.548	0.462	-3.795	52	< 0.001

Referring to the results of the test, the obtained value of $t = -3.795$ with a significance value of $0.001 < 0.05$, where the negative sign indicates that the mean posttest score of the control class was lower than that of the experimental class. These results highlight a notable shift in effectiveness, with the experimental group demonstrating a more

pronounced improvement in critical thinking skills. Such evidence confirms a statistically significant disparity between the two cohorts; consequently, the null hypothesis (H_0) was rejected in favor of the alternative hypothesis (H_1).

Once a significant difference was established through hypothesis testing, N-gain scores were computed for individual indicators. These scores provided a precise evaluation of the progress in students' critical thinking before and after the intervention.

Table 12. Results of the N-Gain Score for Critical Thinking Skills in the Control Class

Indicator	Pretest	Posttest	N-Gain	Description
Exploration	49.00	62.90	0.28	Low
Integration	18.00	25.20	0.48	Moderate
Resolution	16.40	25.40	0.55	Moderate

Tabel 13. Results of the N-Gain Score for Critical Thinking Skills in the Experimental Class

Indicator	Pretest	Posttest	N-Gain	Description
Exploration	45.60	69.60	0.53	Moderate
Integration	18.00	26.40	0.47	Moderate
Resolution	16.80	26.40	0.57	Moderate

As observed from the results above, the scores for each indicator in both the control and experimental classes showed differences. In the Exploration indicator, the control class obtained an N-Gain score of 0.28 (categorized as low), while the experimental class achieved an N-Gain score of 0.53 (categorized as moderate). This difference indicates that the use of ChatGPT was more effective in improving students' ability to explore information, seek alternative answers, and broaden their understanding of the given problems (Tlili et al., 2023). Furthermore, in the Integration indicator, the control class obtained an N-Gain score of 0.48, while the experimental class obtained 0.47, both categorized as moderate. These results indicate that the ability to integrate information in both classes improved relatively equally. However, in the experimental class, the information integration process was supported by ChatGPT's interactive dialogue features, enabling students to connect concepts with examples in a more systematic manner (Hsu & Silalahi, 2024; Nuwa et al., 2025). In the Resolution indicator, the control class obtained an N-Gain score of 0.55 (categorized as moderate), whereas the experimental class obtained 0.57 (categorized as moderate). Despite the marginal difference, the experimental group exhibited a more pronounced enhancement in critical thinking skills for problem-solving compared to the control class.

DISCUSSION

The Effectiveness of Using ChatGPT on the Critical Thinking Skills of Vocational High School Accounting Students

The findings demonstrate that integrating ChatGPT into the pedagogical process effectively enhances students' critical thinking skills. This advancement was evident in the post-intervention score increases across both cohorts, with the experimental group exhibiting a superior upward trend. Analysis of N-Gain scores further reveals consistent improvements across all critical thinking dimensions, specifically exploration, integration, and resolution. These results are empirically validated by the Independent Sample t-Test, which confirmed a statistically significant divergence between the two classes following the treatment.

These results suggest that incorporating GenAI-driven tools into the curriculum acts as a catalyst for student inquiry, encouraging a more profound exploration of information, test their understanding, and develop deeper reasoning compared to conventional learning methods. Theoretically, this improvement in critical thinking skills can be explained through the perspective of constructivist theory proposed by Jean Piaget (Bustomi et al., 2024; Ulya, 2024). In the experimental class treatment, the use of ChatGPT provided students with opportunities to engage in interactive dialogue. This is consistent with findings suggesting that ChatGPT can facilitate realistic conversations in the form of information exploration and reflection on students' understanding (Haleem et al., 2023; Nuwa et al., 2025). This process supports students in connecting new understanding with their prior knowledge, thereby fostering more meaningful learning. Furthermore, the improvement in students' critical thinking skills can also be explained through the stages of the thinking process in the Practical Inquiry Model, which consists of four stages of cognitive presence, namely exploration, integration, and resolution (Garrison *et al.*, 2001). Nevertheless, this study did not directly measure each stage; instead, the concept was used as a conceptual framework for interpreting the indicator data from the test items.

Data from the N-Gain analysis reveal that the experimental class achieved its highest performance gains specifically within the exploration indicator. This indicates that students' interactions with ChatGPT enabled a deeper exploration process through follow-up questions, requests for additional examples, and clarification of concepts that had not yet been understood. This exploration process is consistent with the principles of constructivist learning, which emphasize the importance of students' active participation in knowledge construction through information-seeking and reflective processes (Azzahra et al., 2025; Bustomi et al., 2024). This is also consistent with the findings of Guo & Lee (2023) showed that structured interactions with ChatGPT, which include analytical and evaluative activities, can enhance students' conceptual understanding and critical thinking skills. Furthermore, the findings of Tlili et al. (2023) and Nuwa

et al. (2025) explained that the provision of additional examples, repeated explanations, and continuous feedback from ChatGPT enables students to continuously validate their understanding through reflective dialogue.

Regarding the integration indicator, both classes demonstrated a relatively balanced improvement. This suggests that the students' ability to synthesize information is influenced by both the use of ChatGPT and the ongoing learning model. However, dialogic interaction with ChatGPT in the experimental class indicates an ability to assist students in testing ideas, clarifying concepts, and constructing understanding incrementally, as described by (Bannert et al., 2023). In terms of the resolution indicator, the experimental class exhibited slightly higher development than the control class. This indicates that the utilization of ChatGPT potentially supports students in formulating solutions and drawing conclusions based on established understanding. These findings align with Nasr et al. (2025), who posit that ChatGPT has the potential to enhance critical thinking skills when employed as a reflective tool rather than a direct substitute. This is further reinforced by the notion that ChatGPT usage can support active learning and improve educational effectiveness, provided it is directed as a cognitive aid (Grubaugh et al., 2023; Huang et al., 2024).

Furthermore, these findings indicate that ChatGPT can serve as a supportive medium for students' critical thinking processes, provided its utilization remains under appropriate teacher guidance and supervision. This underscores the enduring importance of the teacher's role in providing scaffolding or instructional support to ensure students employ technology effectively. Such a premise aligns with Technol et al. (2025), who explain that integrating ChatGPT into discussions can foster critical thinking, particularly when reinforced by scaffolding within the learning process. The results also indicate that ChatGPT's utility extends far beyond simple information retrieval; it serves as a sophisticated cognitive tool that enables students to attain a more profound comprehension of complex subject matter. This stands in contrast to previous findings suggesting that students often seek direct answers without considering or studying the underlying concepts, which leads to shallow learning.

In conclusion, the results suggest that integrating ChatGPT with inquiry-based and constructivist frameworks significantly enhances active student engagement. By facilitating stages such as information analysis, evaluation, and synthesis, this integration enables a more optimal development of critical thinking skills than traditional teaching methods (Garrison et al., 2001; Loria et al., 2025; Yang & Niu, 2026).

CONCLUSION

This study evaluates the efficacy of ChatGPT in fostering critical thinking skills within the Banking Services curriculum at an Accounting-focused Vocational High School. Empirical analysis demonstrates that the strategic integration of ChatGPT significantly bolstered critical thinking skills among students in the experimental group. This improvement is substantiated by post-treatment scores in the experimental group that were significantly superior to those of the control class. Integrating ChatGPT within the experimental framework prompted students to adopt a more proactive stance in information discovery, critical inquiry, and the self-evaluation of their conceptual understanding, while remaining under the strategic guidance of the teacher. These findings carry both practical and theoretical implications. Practically, ChatGPT serves as an instructional medium that fosters interactive learning environments and stimulates critical inquiry. From a theoretical perspective, this study contributes to the expanding body of literature on Generative Artificial Intelligence (GenAI) in pedagogy, specifically highlighting its alignment with constructivist learning frameworks and the Practical Inquiry model in developing students' critical thinking, particularly within the accounting domain.

Despite the significance of the results, it is essential to recognize the boundary conditions and limitations that shaped this investigation. First, preliminary assessment data indicated a baseline disparity in the initial competencies between the control and experimental groups, suggesting that baseline parity was not fully achieved prior to the intervention. Second, this investigation did not fully capture the qualitative nuances and intricate cognitive shifts underlying ChatGPT's influence on student critical thinking, as it relied primarily on the quantitative increase between pretest and posttest scores. Third, this study did not explicitly control for several contextual variables, such as teacher facilitation, student motivation, novelty effects arising from the introduction of AI-based technology, and differences in students' digital literacy, all of which may have influenced the learning outcomes. Lastly, since this investigation focused on a specific learning context within a singular institution, its external validity be constrained. Future inquiries should adopt more stringent protocols to guarantee initial group homogeneity, potentially through controlled randomization or pre-study matching techniques. Furthermore, utilizing data triangulation by integrating in-depth interviews, classroom observations, and teacher feedback is recommended to provide a more nuanced and multi-dimensional perspective on how ChatGPT influences the cognitive development of students. It is also consider controlling for these contextual variables to better isolate the specific effects of ChatGPT on students' critical thinking development. Lastly, future studies should broaden the demographic scope by recruiting larger, more heterogeneous samples across a variety of educational institutions, and backgrounds to bolster the external validity of the findings.

DECLARATION OF GENERATIVE AI

In preparing this article, the author used ChatGPT as an aid limited to improving grammar, sentence structure, and clarity of expression. The author bears full responsibility for all scientific content, analysis, interpretation, and conclusions.

REFERENCES

- Afifah, S. N., Dahlan, D., & Kurniawati, S. (2025). A critical thinking research trends: A bibliometric analysis (2020–2025). *Journal of Educational Sciences*, 9(6), 5102–5115. doi:10.31258/jes.9.6.p.5102-5115
- Ali, D., Fatemi, Y., Boskabadi, E., Nikfar, M., Ugwuoke, J., & Ali, H. (2024). ChatGPT in teaching and learning: A systematic review. *Education Sciences*, 14(6), 643. doi:10.3390/educsci14060643
- Almazrou, S., Alanezi, F., Almutairi, S. A., Aboalsamh, H. M., Alsedrah, I. T., Arif, W. M., Alsadhan, A. A., Alsanad, D. S., & Alqahtani, N. S. (2025). Enhancing medical students' critical thinking skills through ChatGPT: An empirical study with medical students. *Nutrition and Health*, 31(3), 1023–1033. doi:10.1177/02601060241273627
- Alviasari, A., Azizah, I., Anggraeni, M. D., & Darpa, M. M. (2025). Analysis of curriculum development and accounting learning at SMK Indonesia Raya. *Hipkin Journal of Educational Research*, 2(1), 123–138. doi:10.64014/hipkin-jer.v2i1.41
- Astuti, D., & Zulaeha, I. (2021). The effectiveness of metacognition strategy instruments for writing in French at level A2 with blended learning. *International Journal of Innovative Science and Research Technology*, 6(8), 560–566.
- Azzahra, N. T., Ali, S. N. L., & Bakar, M. Y. A. (2025). Teori konstruktivisme dalam dunia pembelajaran. *Jurnal Ilmiah Research Student*, 2(2), 64–75. doi:10.61722/jirs.v2i2.4762
- Bannert, M., Kasneci, E., Sessler, K., Stefan, K., Fischer, F., Gasser, U., Groh, G., Stephan, G., Poquet, O., Sailer, M., Schmidt, A., & Seidel, T. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 102274. doi:10.1016/j.lindif.2023.102274
- Burney, L., Church, K., Akpan, M., & Dell, S. (2023). ChatGPT and AI in accounting education and research. *Technology & Analytics*, 1–29.
- Bustomi, Sukardi, I., & Astuti, M. (2024). Pemikiran konstruktivisme dalam teori pendidikan kognitif Jean Piaget dan Lev Vygotsky. *Jurnal Review Pendidikan dan Pengajaran (JRPP)*, 7(4), 16376–16383. Retrieved from <http://journal.universitaspahlawan.ac.id/index.php/jrpp>
- Chiu, T. K. F. (2024). The impact of generative AI (GenAI) on practices, policies and research direction in education: A case of ChatGPT and Midjourney. *Interactive Learning Environments*, 32(10), 6187–6203. doi:10.1080/10494820.2023.2253861
- Cruz, G., Payan-Carreira, R., Dominguez, C., Silva, H., & Morais, F. (2020). What critical thinking skills and dispositions do new graduates need for professional life? Views from Portuguese employers in different fields. *Higher Education Research & Development*, 39(7), 1–17. doi:10.1080/07294360.2020.178540
- Cui, R., & Zhao, L. (2024). Assessing students' critical thinking in dialogue. *Journal of Intelligence*, 12(11), 1–16. doi:10.3390/jintelligence12110106
- Dad, D. H., & Akbar, D. R. A. (2023). Critical thinking as a twenty-first century skill. *Journal of Educational Research & Social Sciences Review (JERSSR)*, 3(1), 8–15.
- Dedad, M., Akastangga, F., Harmonis, S., Alaudin, R., & Hafidz, A. (2024). The impact of ChatGPT on the critical thinking ability of UIN Sunan Kalijaga students. *Manajemen Teknologi dan Informatika*, 13(3), 2023. doi:10.31940/matrix.v13i3.157-165
- Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, 227, 105224. doi:10.1016/j.compedu.2024.105224
- Drushlyak, M., Lukashova, T., & Semenikhina, O. (2025). ChatGPT-based simulation helps to develop the pre-service mathematics teachers' critical thinking. *International Journal of Instruction*, 18(1), 153–172. doi:10.29333/iji.2025.1819a
- Elsa Sabrina, Yulianti, R., Ambiyar, & Rizal, F. (2024). Mengeksplor dampak interaksi siswa dengan ChatGPT terhadap berpikir kritis dan pemecahan masalah. *Indonesian Journal of Computer Science*, 13(3), 4875–4883. doi:10.33022/ijcs.v13i3.4073
- Essel, H. B., Vlachopoulos, D., Essuman, A. B., & Amankwa, J. O. (2024). ChatGPT effects on cognitive skills of undergraduate students: Receiving instant responses from AI-based conversational large language models (LLMs). *Computers and Education: Artificial Intelligence*, 6, 100198. doi:10.1016/j.caeai.2023.100198
- Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. *The American Journal of Distance Education*, 15(1), 7–23. doi:10.1080/08923640109527071
- Grubaugh, S., Levitt, G., & Deever, D. (2023). Harnessing AI to power constructivist learning: An evolution in educational methodologies. *Journal of Effective Teaching Methods (JETM)*, 1(3), 81–83.
- Guo, Y., & Lee, D. (2023). Leveraging ChatGPT for enhancing critical thinking skills. *Journal of Chemical Education*. doi:10.1021/acs.jchemed.3c00505
- Haleem, A., Javaid, M., & Pratap, R. (2023). An era of ChatGPT as a significant futuristic support tool: A study on features, abilities, and challenges. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, 2(4), 100089. doi:10.1016/j.tbench.2023.100089

- Hsu, W., & Silalahi, A. D. K. (2024). Exploring the paradoxical use of ChatGPT in education: Analyzing benefits, risks, and coping strategies through integrated UTAUT and PMT theories using a hybrid approach of SEM and fsQCA. *Computers and Education: Artificial Intelligence*, 7, 100329. doi:10.1016/j.caeai.2024.100329
- Huang, W., Jiang, J., King, R. B., & Fryer, L. K. (2025). Chatbots and student motivation: A scoping review. *International Journal of Educational Technology in Higher Education*, 22(1). doi:10.1186/s41239-025-00524-2
- Huang, Z., Mao, Y., & Zhang, J. (2024). The influence of artificial intelligence technology on college students' learning effectiveness from the perspective of constructivism—Taking ChatGPT as an example. *Journal of Education, Humanities and Social Sciences*, 30, 40–46.
- Kurniawan, N. A., Hidayah, N., & Rahman, D. H. (2021). Analisis kemampuan berpikir kritis siswa SMK. *Jurnal Pendidikan: Teori, Penelitian, dan Pengembangan*, 6(3), 334. doi:10.17977/jptpp.v6i3.14579
- Loria, Junika, C., Natalia, N., Yuliana, A., Wahyuti, S., Studi, P., Agama, P., Keguruan, F., Kristen, P., Agama, I., Negeri, K., Raya, P., & Raya, K. P. (2025). Pembelajaran inkuiri: Konsep, prinsip, dan penerapannya. *Jurnal Ilmiah Multidisiplin*, 1(6), 4315–4321. doi:10.63822/mabbps58
- Maryani, Putri, N. B. H., & Supriadi, B. (2022). Development of physics REACT module assisted with PhET simulations on harmonic vibration materials. *Jurnal Ilmu Pendidikan Fisika*, 7(2), 157–167. doi:10.26737/jipf.v7i2.2357
- Melisa, R., Ashadi, A., Triastuti, A., Hidayati, S., Salido, A., Luansi Ero, P. E., Marlina, C., Zefrin, Z., & Al Fuad, Z. (2025). Critical thinking in the age of AI: A systematic review of AI's effects on higher education. *Educational Process: International Journal*, 14. doi:10.22521/edupij.2025.14.31
- Mulyadi. (2022). Teori belajar konstruktivisme dengan model pembelajaran (inquiry). *Jurnal Keislaman, Sosial, Hukum dan Pendidikan*, 7(1), 174–187.
- Nasr, N. R., Tu, C., Werner, J., Bauer, T., Yen, C., & Sujo-Montes, L. (2025). Exploring the impact of generative AI ChatGPT on critical thinking in higher education: Passive AI-directed use or human–AI supported collaboration? *Education Sciences*, 15, 1–28.
- Nuwa, M. T., Enstein, J., & Sogen, M. M. B. (2025). Pengaruh penggunaan ChatGPT terhadap kemampuan berpikir kritis siswa kelas X dalam pembelajaran informatika di SMA Kristen Citra Bangsa Mandiri Kupang. *Jurnal Ilmiah Pendidikan Dasar*, 10(4), 487–497.
- Oliński, M., Krukowski, K., & Sieciński, K. (2024). Bibliometric overview of ChatGPT: New perspectives in social sciences. *Publications*, 12(1), 9. doi:10.3390/publications12010009
- Prayogi, R. D. (2020). Kecakapan abad 21: Kompetensi digital pendidik masa depan. *Manajemen Pendidikan*, 14(2), 144–151. doi:10.23917/jmp.v14i2.9486
- Qawqzeh, Y. (2024). Exploring the influence of student interaction with ChatGPT on critical thinking, problem solving, and creativity. *International Journal of Information and Education Technology*, 14(4), 596–601. doi:10.18178/ijiet.2024.14.4.2082
- Ratnawati, O. A., Rizaldi, M., Hamdani, M., & Artuti, E. (2024). Penggunaan ChatGPT terhadap berpikir kritis mahasiswa dalam menyelesaikan masalah geometri analitik ruang. *Jurnal Ilmiah Pendidikan Matematika*, 7, 105–118.
- Janse van Rensburg, J. (2024). Artificial human thinking: ChatGPT's capacity to be a model for critical thinking when prompted with problem-based writing activities. *Discover Education*, 3, 42. doi:10.1007/s44217-024-00113-x
- Retnawati, H. (2020). *Validitas, reliabilitas, & karakteristik butir: Panduan untuk peneliti, mahasiswa, dan psikometrian*. Parama Publishing.
- Risanti, N., & Bayu, K. (2022). Analysis of students' critical thinking ability in accounting. *International Journal of Innovation, Creativity and Change*, 16(2), 15–24.
- Rossouw, M., & Steenkamp, G. (2025). Developing the critical thinking skills of first year accounting students with an active learning intervention. *The International Journal of Management Education*, 23(1), 101086. doi:10.1016/j.ijme.2024.101086
- Rosyida, L., & Bahtiar, M. D. (2024). Pengaruh kecerdasan emosional, self-confidence, dan literasi digital terhadap kemampuan berpikir kritis siswa kelas X Akuntansi SMK Negeri 4 Surabaya. *Jurnal Pendidikan Ekonomi (JURKAMI)*, 9(2), 382–395. doi:10.31932/jpe.v9i2.3603
- Salido, A., Syarif, I., Sari, M., Rias, P., Taufika, R., & Melisa, R. (2025). Integrating critical thinking and artificial intelligence in higher education: A bibliometric and systematic review of skills and strategies. *Social Sciences & Humanities Open*, 12, [article number needs verification].
- Statcounter. (2026). *AI chatbot market share Indonesia*. GlobalStats. Retrieved from <https://gs.statcounter.com/ai-chatbot-market-share/all/indonesia>
- Sugiyono. (2023). *Metode penelitian kuantitatif, kualitatif, dan R&D* (Edisi ke-2). Alfabeta.
- Syahri, A., Efriyanti, L., Zakir, S., & Imamuddin, M. (2024). Pengaruh penggunaan ChatGPT terhadap pola pikir mahasiswa dalam mata kuliah metodologi penelitian: Studi penelitian kuantitatif. *Jurnal Inovasi Pendidikan dan Teknologi*, 5(1), 135–143.
- Thornhill-Miller, B., Camarda, A., Mercier, M., Burkhardt, J. M., Morisseau, T., Bourgeois-Bougrine, S., Vinchon, F., El Hayek, S., Augereau-Landais, M., Mourey, F., Feybesse, C., Sundquist, D., & Lubart, T. (2023). Creativity, critical thinking, communication, and collaboration: Assessment, certification, and promotion of 21st century skills for the future of work and education. *Journal of Intelligence*, 11(3), 54. doi:10.3390/jintelligence11030054

- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., & Hickey, D. T. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10, 15. doi:10.1186/s40561-023-00237-x
- Ulya, Z. (2024). Penerapan teori konstruktivisme menurut Jean Piaget dan teori neuroscience dalam pendidikan. *Journal of Education*, 7(1), 12–23. Retrieved from <http://e-journal.staima-alhikam.ac.id/index.php/al-mudarris>
- Widiastuti, M., & Hamidi, N. (2023). Analisis keterampilan berpikir kritis siswa sekolah menengah kejuruan menggunakan tes W-GCTA. *Jurnal Pendidikan Akuntansi Indonesia*, 23(2), 22–36. Retrieved from <https://journal.uny.ac.id/index.php/jpakun/article/view/57747/19824>
- Wolcott, S. K., & Sargent, M. J. (2021). Critical thinking in accounting education: Status and call to action. *Journal of Accounting Education*, 56, 100731. doi:10.1016/j.jaccedu.2021.100731
- Yang, J., & Niu, T. (2026). Exploring the impact of AI feedback on college students' critical thinking: An intervention experiment based on design discipline. *Frontiers in Psychology*. Retrieved from <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2026.1806913/abstract>