

ChatGPT effects on cognitive skills of undergraduate students: Receiving instant responses from AI-based conversational large language models (LLMs)

Harry Barton Essel^a, Dimitrios Vlachopoulos^{b,*}, Albert Benjamin Essuman^c, John Opuni Amankwa^d

^a Department of Educational Innovations, Kwame Nkrumah University of Science and Technology, Kumasi, AK, 315-7530, Ghana

^b Rotterdam School of Management, Erasmus University, Burgemeester Oudlaan 50, 3062, PA, Rotterdam, the Netherlands

^c Department of Industrial Design, Kwame Nkrumah University of Science and Technology, Kumasi, AK, 315-7530, Ghana

^d Department of Communication Design, Kwame Nkrumah University of Science and Technology, Kumasi, AK, 315-7530, Ghana

ARTICLE INFO

Keywords:

ChatGPT
Large language models
Critical thinking
Creative thinking
Reflective thinking
Undergraduate students

ABSTRACT

This study investigated the impact of using ChatGPT, a state-of-the-art generative AI-based model, on the critical, creative, and reflective thinking skills of university students in Ghana. The study utilized a mixed-methods research approach, incorporating quantitative and qualitative data collection instruments, and an experimental procedure with a pretest-posttest control group. The study ultimately enlisted a sample of 125 students randomly allocated to either the experiment group (60 students) or the control group (65 students). The research was conducted in the context of a Research Methodology course, which had adopted the flipped classroom approach. The students in the experiment group engaged with ChatGPT for in-class tasks, while those in the control group used traditional databases and search engines for similar tasks. Data were collected using the Critical Thinking Scale, Creative Thinking Scale, Reflective Thinking Scale, and a student interview guide (semi-structured). The study's findings illustrated that incorporating ChatGPT influenced the students' critical, reflective, and creative thinking skills and their dimensions discernibly. As a result, the study provides suggestions for academics, instructional designers, and researchers working in educational technology.

1. Introduction

Recently, there has been an increasing fascination with artificial intelligence (AI) large language models (LLMs) and their practical application. AI has recently stimulated revolutionary innovations in several industries globally, with LLMs making influential contributions (Li et al., 2023). In education, there is a trend towards enforcing the 5th era of the Internet, also known as the Internet of Things (IoT), thus, resulting in a growing enthusiasm for integrating AI-assisted teaching and learning (Al Darayseh, 2023) with LLMs. LLMs represent a significant breakthrough in AI technology, allowing for new natural language generation and understanding opportunities. LLMs are advanced neural network models that have a vast number of parameters. These models incorporate billions of parameters and are usually trained on large amounts of text data (175 billion parameters), often gigabytes or even

terabytes in size (Mbakwe et al., 2023) obtained from the Internet through supervised and reinforced learning techniques (Kung et al., 2023). Moreover, LLMs can improve their abilities by fine-tuning, resulting in more sophisticated capabilities and human-like text outputs (Leippold, 2022 & Kasneci et al., 2023). With the revolution of generative AI, academics have begun to apply machine learning, natural language processing (NLP), and LLMs to the maturation of conversational AI interface, making their application a new study area in higher education (Essel et al., 2022). One contemporary phenomenon in the LLMs revolution is ChatGPT (Chat Generative Pre-trained Transformer), which garnered substantial awareness due to its capability to function in a myriad collection of natural language activities (Tili et al., 2023).

* Corresponding author.

E-mail addresses: hbessel.art@knust.edu.gh (H.B. Essel), vlachopoulos@rsm.nl (D. Vlachopoulos), abessuman.art@knust.edu.gh (A.B. Essuman), ajopuni@knust.edu.gh (J.O. Amankwa).

<https://doi.org/10.1016/j.caeai.2023.100198>

Received 12 July 2023; Received in revised form 17 December 2023; Accepted 20 December 2023

Available online 23 December 2023

2666-920X/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

2. Related studies

2.1. An overview of ChatGPT technology

ChatGPT is a conversational AI interface, within the LLMs category, which can respond to natural language inputs and generate human-like responses developed recently by OpenAI (OpenAI, 2023). In contrast to previous AI interfaces, which were mostly Deep Learning (DL) models that retained and identified patterns in data (Rospigliosi, 2023), ChatGPT belongs to a new class of AI algorithms. ChatGPT is trained to estimate the probability of a specific succession of words founded on the context of the words that preceded it (Liu et al., 2023; Naseem et al., 2021). ChatGPT is primarily designed for engaging in human-like conversations, but its capabilities extend much further. It excels in creating original content such as stories, poems, and novels, and can replicate various behaviors it has learned to mimic (Tili et al., 2023). This AI demonstrates remarkable proficiency in a range of natural language tasks, from crafting coherent essays and performing translations to answering questions (Rospigliosi, 2023) and creating computer code (Kasneji et al., 2023). In addition, ChatGPT is capable of interactive dialogues, where it can follow up on previous responses, acknowledge its own mistakes, challenge incorrect premises, and decline inappropriate requests (OpenAI, 2023). Additionally, ChatGPT generates ongoing conversations that lead to follow-up questions, providing distinct experiences from search engines that usually do not preserve a chronology of the progression of an answer but present a list of separate links to resources based on the relevance of specific keywords utilized as search queries (Firat, 2023). ChatGPT provides additional questions that enhance and broaden the answers, addressing any challenges the questioner presents. Also, ChatGPT can enhance academic research and constructive writing tasks, critical thinking, problem-solving, and develop research skills (Dwivedi et al., 2023; Sullivan et al., 2023). It can also suggest unexplored aspects and current research topics, giving students a better understanding and analysis of a particular topic (Kasneji et al., 2023). Notwithstanding, the discussion surrounding the application of AI-based models in higher education reached a critical crossroads following the public release of ChatGPT in November 2022.

2.2. Educational potential of ChatGPT in higher education

Recently, informal observations of ChatGPT's usage indicate proof of deductive reasoning, a sequential thought process, and the ability to maintain a long-term dependency (Kung et al., 2023). Besides, multiple preprints of academic research, blog posts, and media sources have highlighted the benefits of utilizing ChatGPT in the educational ecosystem (Tili et al., 2023). The expedience of ChatGPT in higher education has been recognized as a possible area of attraction due to its myriad applications. Some authors (Tili et al., 2023 & Mollick & Mollick, 2022) have even recommended integrating ChatGPT into instructional didactic to enhance interactive learning experience. Moreover, ChatGPT's conversational format encourages students to exchange questions and answers, promoting awareness and deeper personal reflection through scaffolding learning (Darvishi et al., 2022; Rospigliosi, 2023). ChatGPT's feature of answering follow-up questions encourages students to challenge and clarify information, facilitating synthesis with actual knowledge and promoting a more in-depth understanding of numerous meanings and notions. The utilization of ChatGPT in higher education environments demonstrates its potential to foster personalized learning approaches, significantly enhance student engagement, and encourage self-directed learning. Additionally, it shows promise in advancing cognitive competence among students (Sanusi, Olaleye, Agbo, & Chiu, 2022). Despite these advantages, the impact of ChatGPT on critical thinking, creative thinking, and reflective thinking in relation to students' learning outcomes remains an area yet to be fully explored and understood. From its inception, ChatGPT received an overwhelmingly positive reception, particularly noticeable

among academics and the general public on various social media platforms. This enthusiasm is largely attributed to the AI's perceived ability to revolutionize our understanding of professional work, cognitive processes, and the essence of human creativity (Mbakwe et al., 2023). This widespread acclaim highlights ChatGPT's potential in transforming educational methods and contributing to a deeper understanding of learning processes.

3. Research gap and questions

In the context of the present study, Ghana, most academics in higher have been concerned about the potential adverse effect that ChatGPT can pose on students' learning experiences and cognition since the use of generative AI models has not been analyzed extensively. A comprehensive Twitter sentiment analysis was conducted on the general adoption of ChatGPT as an AI model and discovered that users (including academics) have conflicting perspectives (Haque et al., 2022). Hence, the vulnerabilities academics perceive about ChatGPT have shown profound and instantaneous protective measures against its adoption in higher education as conglomerates of academic institutions in the United States of America (New York and Los Angeles) prohibited ChatGPT from academic networks due to the perceived menace of utilizing it to cheat in academic work (Shen-Berro, 2023). Like many developing countries, Ghana has no standard policy governing the integration of AI-based models into educational institutions' learning and teaching process. This lack of policy highlights the incomplete nature of the techno-centric perspective of the educational ecosystem since it does not account for the power dynamics involved in implementing such mechanisms, such as government or local governance policies, educational institution governors, respective educators, or even the students themselves (Luckin et al., 2022). Despite the potential of ChatGPT to improve cognitive skills and enhance the learning experience, some academics may still view its use in the classroom negatively, perpetuating discrimination and biases (Kooli, 2023). As a result, these academics may choose to continue leveraging conventional teaching resources and methodologies rather than incorporating ChatGPT into their teaching practices. Moreover, many educators and educational institutions may lack the knowledge or expertise to effectively incorporate new technologies into their teaching practices, especially when leveraging and integrating LLMs and may tag LLMs as a negative influence on students cognitive skills. According to Blomhøj (2011), having solid cognitive abilities means dealing with a particular challenge in an informed and thoughtful manner. These abilities would necessitate the use of 21st-century learning skills such as problem-solving (Tsankov, 2018), creative thinking (Kassymova et al., 2020), critical thinking (Biasi, Valencia, & Obregon, 2019), and reflective thinking (Chen et al., 2019). Many students, with Ghana not an exception, often need help analyzing, synthesizing, evaluating, and deeding information into new experiences (Akpur, 2020), leading to inadequate problem-solving skills and limited creativity. Despite the growing use of ChatGPT and other conversational AI models in higher education, the effects of these models on such cognitive skills, which have become the bane of academics, still need to be explored. Thus, this study aims to give an answer to the following research questions:

1. Are the scores related to critical thinking skills of undergraduate students utilizing ChatGPT significantly different from those of students using traditional lecture-based methods?
2. Do the scores reflecting creative thinking skills of undergraduate students who utilize ChatGPT significantly vary from those engaged in traditional lecture-based methods?
3. Are the scores indicative of reflective thinking skills among undergraduate students leveraging ChatGPT significantly distinct from those employing conventional lecture-based methods?
4. What are the perceptions and attitudes of undergraduate students regarding the use of ChatGPT as an educational tool?

In the current academic landscape, the ascendance of AI technologies, particularly those exemplified by systems like ChatGPT, presents a pivotal opportunity to reassess and potentially redefine pedagogical strategies. This paper posits that it is imperative to rigorously evaluate whether and how ChatGPT can augment critical thinking skills in students - a competency that is increasingly indispensable in the digital and information-driven era.

To this end, the first research question is crafted to investigate the extent to which the utilization of ChatGPT as a pedagogical aid engenders notable variances in students' capacities to analyze and critically evaluate information. This inquiry is juxtaposed against the backdrop of conventional, lecture-based, educational methodologies, providing a comparative analysis of efficacy in fostering critical thinking.

Furthermore, the second research question delves into the realm of creative thinking. Recognizing that creativity is a cornerstone of innovation and problem-solving in a multitude of fields, this question seeks to ascertain whether the integration of ChatGPT into educational paradigms can act as a catalyst for enhanced creativity among students. By comparing the creative outputs and thought processes of students engaged with ChatGPT against those adhering to conventional educational techniques, the study aims to uncover any significant disparities in creativity levels attributable to the use of this AI tool.

The third research question pivots towards the concept of reflective thinking, a crucial component of DL and critical self-assessment. This inquiry is aimed at determining the differential impact of ChatGPT, as opposed to conventional educational methods, on students' abilities to engage in reflective thinking. Such an investigation is crucial for understanding whether ChatGPT can foster more profound self-reflection and assimilation of learned content, thus enhancing the depth and quality of learning experiences.

Lastly, given the burgeoning role of AI in educational contexts, it is imperative to comprehend student attitudes and perceptions towards such technologies. The final research question, therefore, focuses on elucidating students' viewpoints and attitudes concerning the use of ChatGPT as an educational tool. This exploration is intended to shed light on the acceptability of AI in educational settings and identify potential barriers to its effective integration.

Each of these research questions is underpinned by a fundamental objective: to thoroughly explore and harness the potential of AI, particularly tools like ChatGPT, in transforming various cognitive skills and shaping student attitudes towards learning in the digital age.

4. Methods

4.1. Research design

This study was designed as a mixed-method approach (sequential explanatory), an experiment with pretest-posttest design and a qualitative design of responses to follow-up questions. By employing such methodological approach, a holistic response to the research questions can be given by combining quantitative data (cognitive skills scores) with qualitative data (opinions of undergraduate students). This combination allows for a more effective interpretation of the results and helps to uncover the underlying factors contributing to any observed differences (Ivankova et al., 2006). The addition of qualitative data in this study can offer insights into the underlying reasons behind any differences in critical thinking, creative thinking, and reflective thinking scores. Understanding the students' perspectives, experiences, and attitudes towards the use of ChatGPT can shed light on the potential benefits or challenges associated with each approach. Finally, this design allows for a direct comparison between the cognitive skills scores of undergraduate students using ChatGPT and those using conventional, lecture-based research methods.

4.2. Research protocol & study population

The study cohort consisted of 125 undergraduate students enrolled in a "Quantitative Research Design" course during the second semester of the academic year 2022–2023. This course, which carries 2 TPC (Theory, Practical, Credit) units, is designed to provide students with a comprehensive understanding of research methodologies. Throughout the semester, students receive 1 h of theoretical instruction and engage in 3 h of hands-on training per week, amounting to a total of 48 h over 12 weeks. It's worth noting that students' participation in this study was entirely voluntary, and their decision to opt in or out had no bearing on their grades for the course. At the outset of the experiment, the research team clearly explained the purpose and goals of the study to the participating students. Prior to the commencement of the study, all students underwent a pretest that assessed their levels of Creative Thinking, Critical Thinking, and Reflective Thinking using standardized scales. Those who provided informed consent were then randomly assigned to either the experimental group (EG) or the control group (CG) using a simple random sample technique facilitated by a random number generator in Microsoft Excel. The control CG comprised 65 students, while the EG included 60 students.

In the study group, men comprised 60.8% (76) of the sample, while women comprised 39.2% (49) of the total. Students in both the CG and EG were able to gain experience concerning the learning environment and procedures leveraged throughout the course.

The average age of all the students was 21.711 (± 4.77). Their cumulative weighted average (CWA) scores were in the second upper division, averaging 68.67 (± 6.81). Most of the students in both the CG and EG were male, with 35 (58.3%) students in the EG and 41 (63.1%) students in the CG. The preceding specifications and CWA scores of the CG and EG were similar. Table 1 provides a summary of the students' descriptive traits.

As presented in Fig. 1, which illustrates the process of the intervention protocol, the Critical Thinking Scale (CTS), Creative Thinking Scale (MCTS), and Reflective Thinking Scale (RTS) were administered as a pretest to both groups of respondents. After participating in sampling method activities over 3 weeks, the respondents were re-examined utilizing the same scales. A semi-structured opinion guide for students was also implemented to garner qualitative data. This form was employed to acquire qualitative data established on the students' research objectives and, therefore, to achieve triangulation by gathering data from respondents leveraging quantitative and qualitative approaches to enhance the findings and achieve more details. The semi-structured guide for the students was utilized to determine the benefits and hindrances of ChatGPT in the activities and to expose underlying characteristics that could influence the quantitative findings.

The present study was conducted under the Helsinki Declaration (1975) and comparable ethical standards, with the approval of the

Table 1
Descriptive traits of the respondents.

		EG (n = 60) n (%)	CG (n = 65) n (%)	X; t; P- value
Gender	Men	35 (58.3)	41 (63.1)	X = 5.731 ^a P = 0.337
	Women	25 (41.7)	24 (36.9)	
Age	Mean \pm	21.45 \pm	21.79 \pm	t = 1.121 ^b P = 0.325
	SD	4.61	4.34	
Academic standings	Mean \pm	68.21 \pm	68.63 \pm	T = 0.072 ^b P = 0.673
	SD	6.14	6.53	
Prior experience with Chatbots	Yes	29 (48.3)	28 (43.1)	X = 4.504 ^a P = 0.205
	No	31 (51.7)	37 (56.9)	

\pm SD = standard deviation; a Fisher's exact test; b Independent samples t-test.

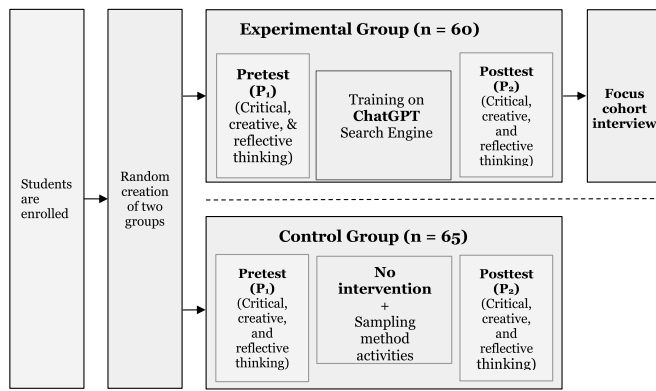


Fig. 1. The intervention protocol flowchart.

Ethics Committee of the Department of Educational Innovations in Science and Technology (EIST/REF No: January 01, 2023), Kwame Nkrumah University of Science and Technology, Kumasi, Ghana. A digital written and verbal informed consent was solicited from the students who volunteered to be part of the study.

4.3. Measures

4.3.1. Sociodemographic traits

As part of the study, sociodemographic data were garnered from the students. These traits included gender, age, CWA, previous experiences with conversational AI Chatbots such as ChatGPT or GPT-3, and type of chatbot experienced.

4.3.2. Critical thinking scale (CTS)

The study utilized an 11-item CTS developed by Sosu (2013) to assess the critical thinking skills of the participating students. This scale is designed to evaluate two key latent constructs: 'Critical Openness,' comprising seven items, and 'Reflective Skepticism,' consisting of four items. Each item in the scale was rated by the students on a 5-point Likert-type scale, where '1' corresponded to 'Strongly Disagree' and '5' denoted 'Strongly Agree.' By summing the scores of these 11 items, an overall score for each student was computed, ranging from 11 to 55 points. The scoring range was categorized into three levels: scores between 11 and 34 indicated 'Low Critical Thinking Skills,' scores from 35 to 44 represented 'Moderate Critical Thinking Skills,' and scores falling within the range of 45–55 were indicative of 'High Critical Thinking Skills.' For a deeper understanding of the critical thinking skills, the scale also allowed for a more granular assessment. Specifically, the 'Critical Openness' scores spanned from 7 to 35, with corresponding cut-off points: scores of 29–35 reflected 'High Critical Openness,' scores between 22 and 28 indicated 'Moderate Critical Openness,' and scores ranging from 7 to 21 represented 'Low Critical Openness.' On the other hand, 'Reflective Skepticism' scores ranged from 4 to 20, with cut-off values: scores of 17–20 denoted 'High Reflective Skepticism,' scores from 13 to 16 signified 'Moderate Reflective Skepticism,' and scores within the range of 4–12 indicated 'Low Reflective Skepticism.' The study's internal consistency reliability was robust, as indicated by a Cronbach's alpha coefficient (α) of 0.93 for the Critical Thinking Scale. Furthermore, the McDonald's alpha coefficient (ω) was calculated as 0.94 for this specific study. To assess the goodness of fit of the Critical Thinking Scale, Confirmatory Factor Analysis (CFA) was performed, and the results were in accordance with established standards. The fit indices included $\chi^2/sd = 2.98$, Comparative Fit Index (CFI) = 0.96, Goodness of Fit Index (GFI) = 0.94, Incremental Fit Index (IFI) = 0.95, CFI = .94, Adjusted Goodness of Fit Index (AGFI) = 0.95, Root Mean Square Error of Approximation (RMSEA) = 0.061, and Standardized Root Mean Square Residual (SRMR) = 0.049. These indices collectively indicated an acceptable fit for the CTS within the study's framework (Tabachnick

& Fidell, 2013).

4.3.3. Creative thinking scale (MCTS)

The study employed a 25-item Creative Thinking Scale (MCTS) designed Ozgenel and Çetin (2017) to gauge the creative thinking abilities of the participating students. This comprehensive scale encompasses six latent constructs, each measuring distinct aspects of creative thinking: I) 'Courage,' consisting of four items; II) 'Innovation Search,' with eight items; III) 'Inquisitive,' comprising three items; IV) 'Self-Discipline,' consisting of five items; V) 'Doubt,' including two items; and VI) 'Flexibility,' which encompasses three items. Each item in the scale was assessed by students using a five-point Likert-type scale, where '1' represented 'Strongly Disagree' and '5' indicated 'Strongly Agree.'

The MCTS was designed to yield scores ranging from a minimum of 25 points to a maximum of 125 points, with higher scores indicating a greater level of creative thinking proficiency. To ensure the scale's reliability, the study computed internal consistency reliability coefficients, with Cronbach's alpha (α) estimated at 0.89 and McDonald's alpha coefficients (ω) calculated as 0.90, underscoring the scale's robustness. To evaluate the goodness of fit of the MCTS within the study's context, CFA was conducted, and the results aligned with established criteria for model fit. The fit indices reported were as follows: $\chi^2/sd = 3.01$, CFI = 0.92, GFI = .94, IFI = .97, AGFI = 0.90, CFI = 0.93, RMSEA = .077, and SRMR = .069. These fit indices collectively indicated an acceptable fit for the MCTS within the study's framework.

4.3.4. Reflective thinking scale (RTS)

The 16-item RTS, developed by Kember et al. (1999), was employed to estimate reflective thinking of the students. The RTS is composed of four latent constructs: I) Understanding (4 items); II) Reflection (4 items); III) Critical Reflection (4 items); IV) Habitual Action (4 items). Each item in the constructs was estimated on a five-point Likert-type scale with 1 denoting "Strongly disagree" and 5 denoting "Strongly agree". The RTS had a minimum score of 16 points and a maximum score of 80 points with higher scores illustrating greater levels of reflective thinking. The internal consistency reliability coefficient (Cronbach's alpha) of the RTS was estimated as $\alpha = 0.96$ and McDonald's alpha coefficients of $\omega = 0.97$ for the current study. According to Tabachnick and Fidell (2013), the fit indices for this study's CFA of the RTS were acceptable. These indices included $\chi^2/sd = 2.91$, CFI = 0.91, GFI = 0.95, IFI = 0.94, AGFI = 0.93, CFI = 0.94, RMSEA = .082 and SRMR = 0.075, which are acceptable according to Tabachnick and Fidell (2013).

4.3.5. Semi-structured opinion guide for students

The researchers designed a semi-structured student interview guide to assess the students' opinions on leveraging ChatGPT to learn sampling methods. The focus group forum was performed in one meeting in which 15 EG members participated. Besides, the interview was held virtually a week after the experiment. The researchers wrote the questions in the interview form. Following this, two field experts in educational research evaluated the questions, and they were altered to agree with the feedback. Then, the final guide was developed (see Appendix). The form seeks to define the students' positive and negative experiences during the implementation of ChatGPT and to evaluate their readiness to persist in leveraging this mechanism.

4.4. Intervention procedure in the experimental group (EG)

ChatGPT, the study's intervention, was involved at the beginning of the semester. The EG received an intervention to improve their critical, creative, and reflective thinking skills leveraging ChatGPT. Students were assigned to the intervention group, and the intervention was delivered via the FC approach with the a Learning Management System (LMS). The intervention group was provided with educational resources such as lecture videos and reading materials. They used the videos

connected to the lesson topics via the LMS one week before each lesson. Additionally, students were provided with a series of prompts related to the topic of the week, which they were required to explore and respond to using ChatGPT. These prompts were developed to stimulate the students to think critically, creatively, and reflectively. Examples of the prompts included "What are the different types of research design?, Which one do you think is most suitable for your research question, and why?", "What are some of the challenges that you might encounter when designing a research study? How can you overcome these challenges?" and "Can you think of a real-life example of a research study that had a significant impact on society? What were the key findings, and how did they impact the field?"

During the first 30 min of the lecture, the lecturer provided a brief review of the topic and then encouraged students to ask questions and engage in discussions. Interactive methods such as discussions, problem-solving, and brainstorming were leveraged in the class to promote maximum interaction between the instructor and students. The lesson time was utilized more effectively in the EG as students had already watched the lesson videos and came prepared for the class. This approach allowed for more interactive methods like the question and answer session to cover the subjects students needed assistance with. Thus, the EG students were also encouraged to leverage ChatGPT to clarify any possible misconceptions they may have had about the lesson. The theoretical sessions were 40–50 min, while practical sessions lasted for 60–70 min, with a total duration of 120 min for the research methods courses. A topic test (based on discussion and individual written tasks) was given per session.

Students were given a deadline to submit their responses, after which they were assessed based on the differentia and profundity of their responses. The course instructor, teaching assistants, and ChatGPT assessed the responses and provided feedback and improvement directions. Students were trained on using the ChatGPT, including accessing the mechanism through a web browser and entering their prompts in the dialogue box. Besides, the students were instructed on how to interpret the responses provided by ChatGPT and use them to guide their thinking and research.

4.5. Intervention procedure in the control group (CG)

The students who were part of the study and designated to the CG ($n = 65$) attended the synchronous course with the rest of the class. The CG received the same model via the Schoology LMS but without ChatGPT. Instead, they were given the same prompts as the EG students and asked to respond to the prompts leveraging conventional research methods, such as reading textbooks, searching for articles online, and using other conventional sources of information. For example, for the topic of "Research Design," the prompts were as follows: What are the different types of research design? Which one do you think is most suitable for your research question, and why?, "What are some of the challenges that you might encounter when designing a research study? How can you overcome these challenges?" and "Can you think of a real-life example of a research study that significantly impacted society? What were the key findings, and how did they impact the field?"

Like the EG, the students in the CG received 40–50 min of theoretical sessions, and 60–70 min of practical sessions, with a total duration of 120 min for the research methods courses. Besides, a lesson test (based on discussion and individual written tasks) was given after each session. Students were given a deadline to submit their responses, after which they were assessed based on the differentia and profundity of their responses. The course instructor (lead author) and teaching assistants assessed the responses and provided feedback and improvement directions. Furthermore, the students in the CG differed from the EG in that they were not permitted to use ChatGPT or any other LLMs during their in-class question-and-answer activities and assignments.

Following the study's conclusion, the EG and CG were administered posttest assessments of the CTS, MCTS, RTS. Similarly, students in the

EG were administered a semi-structured opinion guide and requested to evaluate the effectiveness of leveraging ChatGPT for their course assignments.

4.6. Statistical analysis

Microsoft Excel 365 was utilized to compute the data. The data were inputted and examined using Jamovi 2.3.24 ([The Jamovi Project, 2021](#); [Fox & Weisberg, 2020](#); [Lenth, 2020](#)). The Skewness and Kurtosis Tests were utilized to estimate the normality of the data. The data demonstrated normal distribution (values was ± 2). Besides, a Shapiro-Wilk test was utilized to determine if the distribution of the numerical variables was normally distributed. The analysis results illustrated that the study group was normally distributed. The posttest scores of the critical, creative, and reflective thinking scales were compared after controlling for the pretest scores to address the research questions. Moreover, the posttest scores of the critical, creative, and reflective thinking measures for the CG and EG were compared by controlling the pretest scores to estimate the scales' constructs. The ANCOVA estimate was utilized in the study to compare the scales' post-test scores and their constructs by controlling the pretest scores. The pretest scores were controlled as the variance induced by the pretest scores of the two groups preceding the experimental process did not affect the result, as the EG and CG groups were not randomly formed. A significance (discernible) level of less than $p < 0.05$ and a confidence level of 95% were leveraged to estimate statistical discernibility in the analysis of all tests.

The EG stated their thoughts regarding using the ChatGPT in terms of positive, negative, and prospective features. Following the study's completion, the researchers created a raw data document in Microsoft Word by transcribing the responses. Two coders then utilized content analysis to organize the data to categorize related responses under the same topics. The [Miles and Huberman \(1994\)](#) formula was employed to estimate the validity of the codes produced by two coders. The standard codes were divided into the absolute number of codes and multiplied by 100%. The reliability ratio was discovered as 0.95 (95%) $([\text{consensus number}]/[\text{absolute consensus} + \text{number of disagreements}])$. The researchers concentrated on the 5% distinction and consented that this distinction is representative of qualitative analysis. [Table 5](#) provides an overview of the topics that came up. The findings also include a few student statements about their encounters.

5. Results

5.1. Critical thinking

To address the RQ 1, data were garnered from the respondents and tested for statistical discernibility between the critical thinking scale and its dimensions (Reflective Skepticism and Critical Openness) scores of the CG and EG. CG's critical thinking pretest score was $m = 24.9$, and the posttest score was $m = 30.6$. In contrast, the pretest score for the EG was $m = 28.4$, while their posttest score was $m = 39.2$.

The pretest scores of the respondents in both EG and CG were estimated employing the CTS and its dimensions. Afterwards, the one-way ANCOVA test was leveraged to analyze any differences between the two groups. Preliminary checks were completed to calculate the assumptions of normality, linearity, and homogeneity of variance. A Shapiro-Wilk test indicated that the critical thinking scale scores were normally distributed in the CG and EG, CTS ($W = 0.981$, $p = 0.076$), critical openness ($W = 0.984$, $p = 0.090$), and reflective skepticism ($W = 0.970$, $p = 0.081$). Levene's test demonstrated that the assumptions of normality were not violated, critical thinking scale [$F(1, 123) = 0.368$], $p = 0.545$, critical openness [$F(1, 123) = 0.469$, $p = 0.701$], and reflective skepticism [$F(1, 123) = 0.563$, $p = 0.190$]. These results provide confirmation that the scores in both the CG and the EG effectively met and adhered to the necessary assumptions required for the subsequent statistical analysis, reassuring the validity and reliability of

the data for our research. The estimated marginal means (EMM) for the construct and its dimensions are illustrated in Table 2.

The study leveraged covariance analysis to differentiate whether there was a statistically discernible variance between the posttest scores of the CTS, reflective skepticism, and critical openness dimensions of the EG and CG. The results are recapitulated in Table 2.

After controlling for the pretest score, there was a statistically discernible effect on the posttest score of critical thinking [$F(1,122) = 36.3, p < 0.001, \eta^2p = 0.229$]. To put it differently, the posttest scores of the CTS for the EG ($m = 39.2, SD = 6.57$) were discernibly higher than the critical thinking posttest scores of the CG ($m = 30.6, SD = 7.64$). The Pretest score of critical thinking was discernibly linked to the posttest score [$F(1, 122) = 19.2, p < 0.001, \eta^2p = 0.136$] as illustrated in Table 2.

The critical openness scores of the EG and CG, a CTS dimension, were compared. Table 2 exhibits the results of one-way ANCOVA conducted on the posttest critical openness scores of the EG and CG, $F(1,122) = 43.6, p < 0.001, \eta^2p = 0.263$. To put it differently, the posttest scores of the critical openness dimension for the EG ($m = 24.7, SD = 5.38$) were discernibly higher than the critical openness dimension posttest scores of the CG ($m = 17.8, SD = 5.33$). The Pretest score of critical openness dimension was discernibly linked to the posttest score [$F(1,122) = 10.6, p = 0.001, \eta^2p = 0.080$].

Besides the reflective skepticism scores of the CG and EG, a CTS dimension, were compared. Table 2 shows the results of one-way ANCOVA performed on the posttest reflective skepticism scores of the EG and CG, $F(1,122) = 6.53, p = 0.012, \eta^2p = 0.051$. To express it differently, the posttest scores of the reflective skepticism dimension for the EG ($m = 14.5, SD = 2.62$) were discernibly higher than the reflective skepticism dimension posttest scores of the CG ($m = 12.8, SD = 3.73$). The Pretest score of critical openness dimension was discernibly linked to the posttest score [$F(1,122) = 4.92, p = 0.028, \eta^2p = 0.039$]. This suggests that the initial scores on the critical openness dimension had a discernible influence on the subsequent posttest scores.

5.2. Creative thinking

In order to address RQ2, data were collected from the respondents and tested for statistical discernible between the MCTS and its dimensions (Creative thinking, Courage, Innovative Search, Inquisitive, Self-Discipline, Doubt, and Flexible) scores of the CG and EG. CG's creative thinking pretest score was $m = 59.2$, and the posttest score was $m = 88.9$. In difference, the pretest score for the EG was $m = 57.2$, while their posttest score was $m = 92.0$.

Preliminary checks were completed to calculate the assumptions of normality, linearity, and homogeneity of variance. The pretest scores of the respondents in both CG and EG were measured utilizing the MCTS and its dimensions. Afterwards, the one-way ANCOVA test was leveraged to analyze any disparities between the two groups. A Shapiro-Wilk test showed that the creative thinking scale scores were normally spread

in the CG and EG, creative thinking scale ($W = 0.984, p = 0.145$), Courage ($W = 0.991, p = 0.059$), Innovative Search ($W = 0.982, p = 0.095$), Inquisitive ($W = 0.994, p = 0.115$), Self-Discipline ($W = 0.976, p = 0.00298$), Doubt ($W = 0.976, p = 0.127$), and Flexible ($W = 0.985, p = 0.134$). Levene's test demonstrated that the assumptions of normality were not violated, critical thinking scale [$F(1, 123) = 0.571$], $p = 0.180$], Courage [$F(1, 123) = 1.11, p = 0.293$], Innovative Search [$F(1, 123) = 0.469, p = 0.701$], Inquisitive [$F(1, 123) = 0.904, p = 0.152$], Self-Discipline [$F(1, 123) = 0.802, p = 0.372$], Doubt [$F(1, 123) = 1.90, p = 0.170$], and Flexible [$F(1, 123) = 0.743, p = 0.390$]. These results indicate that the scores met the required assumptions for the analysis. The EMM for the construct and its dimensions are illustrated in Table 3.

The study leveraged covariance analysis to measure whether there was a statistically discernible variance between the posttest scores of the MCTS, Courage, Innovative Search, Inquisitive, Self-Discipline, Doubt, and Flexible dimensions of the EG and CG. The results of the study are summarized in Table 3.

After controlling for the pretest score, there was a statistically significant effect on the posttest score of creative thinking [$F(1,122) = 9.91, p = 0.002, \eta^2p = 0.075$], as illustrated in Table 3. To state it differently, the posttest scores of the critical thinking scale for the EG ($m = 92.0, SD = 6.52$) were significantly higher than the creative thinking posttest scores of the CG ($m = 88.9, SD = 6.39$). The pretest score of creative thinking was discernibly linked to the posttest score [$F(1,122) = 7.02, p = 0.009, \eta^2p = 0.054$].

Besides, there was a statistically significant effect on the posttest score of the courage dimension [$F(1,122) = 5.54, p = 0.020, \eta^2p = 0.043$] after controlling for the pretest score of the courage dimension as witnessed in Table 3. In other words, the posttest scores of the courage dimension for the EG ($m = 13.1, SD = 2.07$) were discernibly higher than the courage dimension posttest scores of the CG ($m = 12.3, SD = 3.26$). The pretest score was discernibly linked to the posttest score [$F(1,122) = 12.28, p < 0.001, \eta^2p = 0.091$] of the courage dimension.

The innovative search dimension scores of the CG and EG were also compared. The results (Table 3) of one-way ANCOVA conducted on the posttest innovative search dimension scores of the EG and CG, $F(1,122) = 23.9, p < 0.001, \eta^2p = 0.164$. To set it differently, the posttest scores of the innovative search dimension for the EG ($m = 28.12, SD = 3.95$) were discernibly higher than the critical openness dimension posttest scores of the CG ($m = 24.63, SD = 3.36$). The pretest score of the innovative search dimension was discernibly linked to the posttest score [$F(1,122) = 11.9, p < 0.001, \eta^2p = 0.089$].

Further, the inquisitive dimension scores of the EG and CG were approximated. Table 3 exhibits the results of one-way ANCOVA executed on the posttest inquisitive dimension scores of the EG and CG, $F(1,122) = 5.56, p = 0.020, \eta^2p = 0.044$. In other words, the posttest scores of the inquisitive dimension for the EG ($m = 12.40, SD = 2.11$) were discernibly higher than the inquisitive dimension posttest scores of the CG ($m = 10.89, SD = 3.82$). The pretest score of the inquisitive dimension was discernibly linked to the posttest score [$F(1,122) = 4.47,$

Table 2
Covariate analysis of critical thinking scale and its dimensions.

	Source	Sum of squares	df	Mean square	f	p	EMM
Critical thinking	Overall model	2472	2	1236.1	7.47	<.001	EG: 40.3
	group	856	1	856.3	9.91	<.001	CG: 32.8
	Pretest	1616	1	1615.8	7.02	<.001	
	Residuals	5434	122	44.5			
Critical Openness dimension	Overall model	1441	2	720.4	7.60	<.001	EG: 25.7
	group	1159	1	1159.2	12.28	<.001	CG: 19.4
	Pretest	282	1	281.7	5.54	0.001	
	Residuals	3245	122	26.6	33.2		
Reflective skepticism dimension	Overall model	117.0	2	58.5	6.86	0.001	EG: 15.2
	group	50.3	1	50.3	4.92	0.028	CG: 13.7
	Pretest	66.7	1	66.7	6.53	0.012	
	Residuals	1246.2	122	10.2			

EMM: Estimate Marginal Means.

Table 3

Covariate analysis of creative thinking scale and its dimensions.

	Source	Sum of squares	df	Mean square	f	p	EMM
Creative thinking	Overall model	671	2	335.7	7.47	<.001	EG: 92.3
	group	393	1	392.9	9.91	0.002	CG: 88.7
	Pretest	278	1	278.5	23.20	0.009	
	Residuals	4839	122	39.7			
Courage dimension	Overall model	123.5	2	61.74	7.60	<.001	EG: 13.2
	group	38.4	1	38.40	5.54	0.020	CG: 12.1
	Pretest	85.1	1	85.09	12.28	<.001	
	Residuals	845.2	122	6.93	14.67		
Innovative Search dimension	Overall model	439	2	219.4	21.4	<.001	EG: 27.9
	group	293	1	293.1	23.9	<.001	CG: 24.8
	Pretest	146	1	145.8	11.9	<.001	
	Residuals	1496	122	12.3	7.74		
Inquisitive dimension	Overall model	95.4	2	47.68	12.92	0.003	EG: 12.3
	group	53.0	1	53.01	6.48	0.020	CG: 11.0
	Pretest	42.4	1	42.36		0.037	
	Residuals	1156.3	122	9.48			
Self-Discipline dimension	Overall model	115.9	2	58.0	4.36	0.015	EG: 19.3
	group	46.4	1	46.4	4.64	0.033	CG: 18.0
	Pretest	69.5	1	69.5	6.95	0.009	
	Residuals	1220.6	122	10.0			
Doubt dimension	Overall model	35.0	2	17.49	6.81	0.002	EG: 7.81
	group	18.6	1	18.65	6.41	0.013	CG: 8.59
	Pretest	16.3	1	16.32	5.61	0.019	
	Residuals	355.1	122	2.91			
Flexible dimension	Overall model	125.0	2	62.50	17.2	<.001	EG: 12.0
	group	59.0	1	58.98	14.5	<.001	CG: 10.6
	Pretest	66.0	1	66.02	16.2	<.001	
	Residuals	497.4	122	4.08			

EMM: Estimate Marginal Means.

$p = 0.037$, $n^2p = 0.035$].

After controlling for the pretest score, there was a statistically significant effect on the posttest score of the self-discipline dimension [$F(1,122) = 4.64$, $p = 0.033$, $n^2p = 0.037$], as described in Table 3. To communicate it differently, the posttest scores of the self-discipline dimension for the EG ($m = 19.0$, $SD = 2.82$) were discernibly higher than the self-discipline dimension posttest scores of the CG ($m = 18.25$, $SD = 3.58$). The pretest score of the self-discipline dimension was discernibly linked to the posttest score [$F(1,122) = 6.95$, $p = 0.009$, $n^2p = 0.054$].

Besides, there was a statistically discernible effect on the posttest score of the doubt dimension [$F(1,122) = 6.81$, $p = 0.002$, $n^2p = 0.050$] after controlling for the pretest score of the doubt dimension, as seen in Table 3. In other words, the posttest scores of the doubt dimension for the EG ($m = 8.63$, $SD = 1.48$) were significantly higher than the doubt dimension posttest scores of the CG ($m = 7.77$, $SD = 1.94$). The Pretest score was discernibly linked to the posttest score [$F(1,122) = 5.61$, $p = 0.019$, $n^2p = 0.091$] of the doubt dimension.

Lastly, the flexible dimension scores of the CG and EG were approximated. Table 3 shows the results of one-way ANCOVA executed on the posttest flexible dimension scores of the EG and CG, $F(1,122) = 17.2$, $p < 0.001$, $n^2p = 0.106$. In other words, the posttest scores of the flexible dimension for the EG ($m = 12.10$, $SD = 1.66$) were discernibly higher than the flexible dimension posttest scores of the CG ($m = 10.55$, $SD = 2.50$). The pretest score of the flexible dimension was discernibly linked to the posttest score [$F(1,122) = 16.2$, $p < 0.001$, $n^2p = 0.117$].

5.3. Reflective thinking

To address research RQ 3, data were garnered from the respondents and experimented for statistical discernibility between the RTS and its dimensions (Understanding, reflection, Critical reflection, and Habitual) scores of the CG and EG. CG's reflective thinking pretest score was $m = 36.4$, and the posttest score was $m = 51.9$. In contrast, the pretest score for the EG was $m = 35.1$, while their posttest score was $m = 56.6$.

Precursory inspections were conducted to estimate the assumptions

of normality, linearity, and homogeneity of variance. The pretest scores of the respondents in both CG and EG were computed leveraging the reflective thinking scale and its dimensions. Later, the one-way ANCOVA test was leveraged to analyze any disparities between the two groups. A Shapiro-Wilk test indicated that the reflective thinking scale scores were normally spread in the CG and EG, reflective thinking scale ($W = 0.981$, $p = 0.084$, understanding ($W = 0.980$, $p = 0.079$), habitual ($W = 0.997$, $p = 0.094$), critical reflection ($W = 0.983$, $p = 0.087$), and reflection ($W = 0.987$, $p = 0.090$). Levene's test demonstrated that the assumptions of normality were not violated, reflective thinking scale [$F(1, 123) = 1.39$, $p = 0.186$], understanding [$F(1, 123) = 1.11$, $p = 0.533$], habitual [$F(1, 123) = 2.20$, $p = 0.141$], critical reflection [$F(1, 123) = 0.390$, $p = 0.081$], and reflection [$F(1, 123) = 1.84$, $p = 0.177$]. These results indicate that the scores met the required assumptions for the analysis. The EMM for the construct and its dimensions are illustrated in Table 4.

The study employed covariance analysis to calculate whether there was a statistically discernible variance between the posttest scores of the reflective thinking scale, Understanding, Habitual, Critical reflection, and Reflection dimensions of the EG and CG. The results are detailed in Table 4.

After controlling for the pretest score, there was a statistically discernible effect on the posttest scores of the RTS [$F(1,122) = 23.20$, $p < 0.001$, $n^2p = 0.160$], as described in Table 4. To describe it differently, the posttest scores of the reflective thinking for the EG ($m = 56.6$, $SD = 6.21$) were discernibly higher than the reflective thinking posttest scores of the CG ($m = 51.9$, $SD = 5.72$). The pretest score of the reflective thinking was discernibly linked to the posttest score [$F(1,122) = 6.71$, $p = 0.011$, $n^2p = 0.052$].

Additionally, there was a statistically discernible effect on the posttest score of the understanding dimension [$F(1,122) = 4.63$, $p = 0.033$, $n^2p = 0.037$] after controlling for the pretest score of the understanding dimension, as detailed in Table 4. In other words, the posttest scores of the understanding dimension for the EG ($m = 13.4$, $SD = 2.44$) were discernibly higher than the understanding dimension posttest scores of the CG ($m = 12.9$, $SD = 1.76$). The pretest score was discernibly linked

Table 4

Covariate analysis of reflective thinking scale and its dimensions.

	Source	Sum of squares	df	Mean square	f	p	EMM
Reflective thinking	Overall model	1015	2	507.3	21.10	<.001	EG: 58.1
	group	228	1	227.6	41.31	<.001	CG: 53.1
	Pretest	787	1	787.0	4.11	0045	
	Residuals	4139	122	33.9			
Understanding dimension	Overall model	135.7	2	67.83	18.29	<.001	EG: 13.5
	group	16.3	1	16.26	4.63	0.033	CG: 12.4
	Pretest	119.4	1	119.40	33.96	<.001	
	Residuals	428.9	122	3.52			
Habitual dimension	Overall model	89.73	2	44.87	14.67	<.001	EG: 15.6
	group	82.38	1	82.38	23.49	<.001	CG: 13.9
	Pretest	7.35	1	7.35	2.10	0.150	
	Residuals	427.90	122	3.51			
Critical reflection dimension	Overall model	300	2	149.78	13.2	<.001	EG: 14.7
	group	185	1	184.74	18.8	<.001	CG: 12.2
	Pretest	115	1	114.83	11.7	<.001	
	Residuals	1201	122	9.84			
Reflection dimension	Overall model	175.2	2	87.62	17.44	<.001	EG: 15.7
	group	137.4	1	137.42	20.84	<.001	CG: 13.5
	Pretest	37.8	1	37.83	5.74	0.018	
	Residuals	804.5	122	6.59			

EMM: Estimate Marginal Means.

Table 5Distribution of themes arising from students responses ($n = 15$).

Questions	Themes	n(%)
Positive aspects	Enhanced self-efficacy and confidence in research	12 (80.0)
	Heightened knowledge and performance	10 (66.7)
	Interactive, human-like, and engaging	13 (86.7)
	Pacing in learning	13 (86.7)
	Efficient in providing response	15 (100)
Negative aspects	Provided incorrect information	5 (33.3)
	Generated false citations/references	12 (80.0)
Future aspects	Should be used in all lessons/courses	11 (73.3)
	Want to use for research activities	14 (93.3)

to the posttest score [$F(1,122) = 33.96, p < 0.001, n_2p = 0.218$] of the understanding dimension.

After controlling for the pretest score, there was a statistically discernible effect on the posttest score of the habitual dimension [$F(1,122) = 23.49, p < 0.001, n_2p = 0.161$], as depicted in Table 4. To represent it differently, the posttest scores of the habitual dimension for the EG ($m = 15.2, SD = 2.08$) were discernibly higher than the habitual dimension posttest scores of the CG ($m = 13.4, SD = 1.68$). The pretest score of the habitual dimension was not discernibly linked to the posttest score [$F(1,122) = 2.10, p = 0.150$].

Further, the critical reflection dimension scores of the CG and EG were compared. Table 4 illustrates the results of one-way ANCOVA executed on the posttest critical reflection dimension scores of the CG and EG, [$F(1,122) = 13.2, p < 0.001, n_2p = 0.133$]. Similarly, the posttest scores for the EG ($m = 14.5, SD = 3.89$) were discernibly higher than the critical reflection dimension posttest scores of the CG ($m = 12.3, SD = 2.57$). The pretest score of the critical reflection dimension was discernibly linked to the posttest score [$F(1,122) = 11.7, p < 0.001, n_2p = 0.087$].

Ultimately, the flexible dimension scores of the CG and EG were approximated. Table 4 shows the results of one-way ANCOVA executed on the posttest flexible dimension scores of the EG and CG, [$F(1,122) = 20.84, p < 0.001, n_2p = 0.146$]. In other words, the posttest scores of the flexible dimension for the EG ($m = 15.9, SD = 2.78$) were discernibly higher than the flexible dimension posttest scores of the CG ($m = 13.4, SD = 2.45$). The pretest score of the flexible dimension was discernibly linked to the posttest score [$F(1,122) = 5.74, p < 0.018, n_2p = 0.045$].

5.4. Student opinions on use ChatGPT

Table 5 displays the results of the main open-ended questions that were asked to students after the intervention, inquiring about their experiences with the ChatGPT mechanism and its advantages and disadvantages. While all students were aware of ChatGPT, none of them was expecting it to be used by their instructor as a supporting tool because they have heard that other institutions have banned it. Their knowledge before this experiment, however, was limited to using it “similarly to Google search engine” or “as a live Wikipedia”.

Below there are some illustrative interview responses cited by the students in the experimental group (EG).

“I think the ChatGPT activity was particularly helpful because it gave me the chance to practice pinpointing conditions where each sampling method would be appropriate. This helped me feel more confident in my ability to apply sampling methods in my own research studies.” (participant 2)

“Using ChatGPT to learn about sampling methods permitted me to learn at my own pace, which made the material less intimidating and more manageable.” (participant 3)

“I really appreciated the ChatGPT activity because it allowed me to practice applying sampling techniques to real-life scenarios. It helped me feel more confident in my understanding of the material.” (participant 4)

“ChatGPT gave intriguing suggestions to the prompts I provided. ChatGPT also helped me understand that selecting the right sampling method is critical for obtaining accurate and reliable results. I now realize that a poorly designed sampling plan can greatly undermine the credibility of a study.” (participant 8)

“At a point in time, I felt like I was communicating with my lecturer. It feels more human-like than I thought. Besides, I found the ChatGPT activity to be an effective way to learn about the different sampling methods. The interactive nature of the platform kept me engaged and made it easier to absorb the material.” (participant 9)

“ChatGPT enhanced my understanding of sampling methods. I could control and suggest how I want it to explain my prompts.” (participant 10)

“The ChatGPT activity helped make the material more engaging and interactive. I learned a lot more than I would have from a traditional lecture.” (participant 11)

"ChatGPT was really helpful in understanding complex topics like sampling methods. Next time, it would be great if it could provide more real-world examples and practical exercises in other courses too, such as historical analyses." (participant 13)

"In mathematics, ChatGPT could adapt its explanations based on the students' level, ensuring a more personalized result." (participant 15)

According to the interviews conducted with EG students, it can be concluded that they found the use of ChatGPT as a learning mechanism for sampling methods to be beneficial and influential. All of the students reported relishing the ChatGPT activity's flexibility, which allowed them to learn at their own pace and in a less intimidating mode. They also stated that the interactive nature of the platform made it easier to absorb the material and increased their confidence in their acquaintance with sampling methods. The students further asserted that the ChatGPT activity assisted them in developing critical thinking skills by enabling them to apply sampling methods to real-life scenarios. Besides, they described that the activity stimulated reflective thinking by illustrating the significance of choosing the appropriate sampling method to obtain valid and reliable results. Ultimately, the students declared that the activity stimulated creative thinking by permitting them to practice recognizing conditions where each sampling technique would be suitable. ChatGPT was suggested for diverse type of courses, such as history and mathematics.

6. Discussion

This experimental study examined how leveraging the ChatGPT large language model affects students' critical, reflective, and creative thinking skills. The study involved providing didactic assistance to the students in the EG using ChatGPT during class activities, while the CG received didactic assistance without ChatGPT. The outcomes of the study are explained and discussed below.

The first research question of the study involved examining the students' critical thinking skills in both the EG and CG. According to the study findings, there was a discernible variance in critical thinking scores between the EG and CG at both pretest and posttest. Moreover, the EG demonstrated a significant increase in critical thinking, reflective skepticism, and critical openness compared to the CG. Thus, the study's quantitative results illustrated that leveraging ChatGPT for in-class tasks effectively improved students' critical thinking skills. One feasible illustration of the significance of ChatGPT in enhancing critical thinking skills is that it furnishes students with the possibility to engage in dialogues with an AI ChatGPT model that prompts them to think critically. The ChatGPT may offer students feedback and guidance, which can help them develop a deeper understanding (Rospigliosi, 2023) of the topic and enhance their critical thinking skills (Darvishi et al., 2022; Kasneci et al., 2023). Another possible justification is that ChatGPT may provide students with a more personalized learning experience. This personalized learning procedure can benefit students who struggle with critical thinking and may require additional support to improve their skills. Unlike conventional, lecture-based classroom didactics, ChatGPT can adjust the difficulty level and pace of instruction to match individual students' needs and learning styles (Dwivedi et al., 2023; Tili et al., 2023). AI ChatGPT models have demonstrated a significant impact on critical thinking skills in a variety of empirical studies. For instance, multiple studies (Liang, 2022 & Long et al., 2016) discovered that leveraging technology-based learning intervention can improve university students' critical thinking skills. Notably, the improvement in critical thinking skills was observed not only in the overall score of the critical thinking scale but also in its two dimensions. The first dimension, critical openness, refers to the willingness to consider alternative perspectives and to be exposed to novel ideas. The second dimension, reflective skepticism, directs to the ability to question assumptions, pursue evidence, and evaluate arguments. The improvement in these

dimensions among the EG points out that ChatGPT can facilitate critical thinking skills by providing a platform for students to interact with diverse perspectives and ideas. Likewise, the study findings illustrated that the pre-critical thinking score was discernibly linked to the post-critical thinking score. This finding illustrates that students who had higher levels of critical thinking skills prior to the intervention exhibited more remarkable advancement in their critical thinking skills after leveraging ChatGPT for in-class tasks. This finding is invariant with prior studies that have illustrated a positive nexus between prior critical thinking skills and the effectiveness of interventions to improve critical thinking (Hapsari & Wu, 2022 & Goda et al., 2014). In the same line, recent studies suggest that AI Chatbot models can revolutionize education and develop new prospects for students to comprehend and develop critical thinking skills to enhance their learning experiences (Jamal et al., 2023; Tsai et al., 2023). Thus, leveraging ChatGPT can elicit critical thinking skills and problem-solving (Kasneci et al., 2023) by providing a platform for students to engage with diverse perspectives and ideas.

The study's second research question involved comparing the creative thinking skills of students who used ChatGPT for their in-class tasks and those who did not use it. The findings indicate that leveraging ChatGPT for students' in-class activities contributes to developing creative thinking skills, courage, innovative search dimension, inquisitiveness, self-discipline, doubt, and flexible skills. Thus, students in the EG increased their creative thinking due to their interactions with ChatGPT in executing in-class tasks. Literature has yet to reach a definitive conclusion regarding the influence of AI Chatbot models on students' creative thinking abilities (Tang et al., 2022). However, some studies support our findings. Chang and Yu (2015) discovered that students who engaged in an AI-driven online synchronous learning system (Chatbots) exhibited more outstanding scores regarding their creative thinking outputs according to their study. Chang (2017) conducted a study that yielded comparable findings, as it investigated the impact of AI cloud-based mobile learning tools, including Facebook and the Cubie app, on the creative thinking outputs of students. Scholars such as Gangadharbatla (2010) and Middleton (2005) have recognized creative thinking as an essential element of educational technology, for that matter, AI chatbot models. ChatGPT is implied to contribute to creative thinking via pleasure (fun), as identified by EG students interviewed, and this finding is in line with Navarrete (2013). Additionally, Root-Bernstein and Root-Bernstein (1999) based their descriptions of personal pleasure on the premise that creative thinking is closely linked with emotions and intuition. Thus, the findings in the study indicate that incorporating creative thinking in a student-centered learning approach can offer students a rewarding and enjoyable educational experience through an authentic AI chatbot model, such as ChatGPT, that facilitates profound and insightful learning (Navarrete, 2013) that advances creative thinking skills. While there may be a connection between creative thinking and intelligence, socio-constructivist perspectives or personal and relevant experiences may provide a more accurate description of the creative experience (Gangadharbatla, 2010). Although leveraging ChatGPT influenced students' creative thinking, the findings suggest that using ChatGPT, or any AI chatbot model, may only partially capture the creative experience influenced by personal and contextual factors. While there may be a connection between creative thinking and intelligence, socio-constructivist perspectives or personal and relevant experiences may provide a more accurate description of the creative experience (Gangadharbatla, 2010 & Ambrose et al., 2003). Despite this, certain studies have demonstrated that AI chatbot models could occasionally decline students' creative thinking because of their higher cognitive load demands (Tang et al., 2022). Rubino et al. (2018) discovered that using a digital platform reduced creative thinking among students. The researchers assessed the creative thinking process by independently observing the behaviors and dialogues of students in the classroom. These findings imply that since ChatGPT can promote students' creative thinking in this study, it can potentially decline

students' creative thinking. Thus, while ChatGPT may provide valuable insights and generate creative ideas, it may only partially replicate the nuances and complexity of human creativity shaped by socio-cultural factors and personal experiences. Nonetheless, certain studies have specified that the impact of AI chatbot models would rely on the scaffolding strategies employed by teachers (Chang, 2017), which aim to promote independent thinking and learning among students and reduce their dependence on teachers (Wodaj & Belay, 2021). Finally, the fact that the participants in both groups were pursuing a Quantitative Research Design course, inherently encouraged a heightened level of analytical thinking. This academic background, coupled with the students' ability to scrutinize AI-generated content, likely played a pivotal role in their readiness and inclination to engage in metacognitive examination of AI data and discussion of its prompts within their respective zones of proximal development (ZPDs). This unique characteristic of the student groups could account for the significant improvements observed in their critical, creative, and reflective thinking skills. It underscores the importance of considering the composition of the student body in interpreting the study's outcomes and highlights how their prior preparation and context were conducive to meaningful engagement with the AI-driven learning environment implemented in the course.

The third research question addressed in the study concerned the comparison of reflective thinking skills between students who utilized ChatGPT for in-class activities and those who did not use any LLMs. The study's findings demonstrated that leveraging ChatGPT for in-class tasks contributed to maturing reflective thinking skills of students. Regarding the dimension of reflective thinking, the present study reported a significant influence on the students' understanding, habitual action, critical reflection, and reflection. In line with the present study, Yilmaz (2020) discovered that there was a statistically positive significant impact on student reflective thinking skills although the study employed AI analytics. The study conducted by the author also showed that students in the EG demonstrated significantly higher levels of understanding, habitual action, critical reflection, and reflection compared to the CG. A study by Liu et al. (2023) discovered that students leveraging reflective thinking-promoting mechanisms in AI-supported conditions (in the form of a Chatbot) reported positive learning experiences and perceptions of the approach, such as reflection, engagement, augmented motivation, and feedback utilization. Moreover, the authors discovered that the AI-supported condition enhanced the students' writing skills in the EG, heightened their self-regulated learning and self-efficacy, and considerably decreased their cognitive load (Shadiev & Huang, 2020). By leveraging an AI-supported feedback model to enhance in-class tasks and elucidate writing standards, multiple studies (Allen & McNamara, 2015; Roscoe et al., 2015; Snow et al., 2015) discovered similar findings that encouraged students' self-regulated learning. In their study, Wang et al. (2023) found that providing AI-supported feedback to students resulted in a statistically significant positive impact on their self-regulated learning abilities and engagement level, and the dimensions of the self-regulating scale (planning before writing, planning during writing, attention regulation, organization, checking and correcting, and content monitoring) showed significant differences favoring the experimental group. Although AI chatbot models (automated feedback systems) can assist learning, technological limitations must be considered for pedagogical purposes (Palermo & Wilson, 2020) to foster reflective thinking successfully. These findings illustrate that when students leverage ChatGPT to assist in in-class tasks, they can observe and monitor their learning patterns, create strategies to improve them and evaluate their effectiveness. In this study, the significant difference observed in the post-reflective think scores between the EG and CG illustrated that presenting excessive information to students from different aggregators and databases may increase the effort required to perform a particular task and affect their reflection; however, leveraging ChatGPT to assist them in constructing their knowledge of this learning information can benefit students. Moreover, the findings indicated that using ChatGPT as a support system emphasized the importance of

efficient learning approaches in fostering reflection (Carless, 2019).

The final research question had to do with an interview session where students in the EG expressed their opinions about the positives and negatives, as well as future use of ChatGPT for in-class tasks. The findings from the students' interviews suggested that leveraging ChatGPT for in-class tasks was effective in terms of enhancing critical, creative, and reflective thinking and reducing the burden of the instructor for correcting in-class tasks. Likewise, the students recognized that leveraging ChatGPT enhanced the effectiveness of learning research methodology, furnishing more opportunities for reflection as they had less information to analyze and synthesize. Lin (2019) argued that too much information could overwhelm students, but constructing their knowledge with effective learning strategies can help them benefit from the presented information. Moreover, the findings authenticated that the students were highly motivated to learn. They expressed willingness to leverage ChatGPT in other courses and recommend it to different teachers and students. These findings are inline with the study by Liu et al. (2021). These findings indicate that students perceive ChatGPT as a valuable educational tool that has the potential to enhance their learning experiences in diverse contexts. However, the study also demonstrated some adverse characteristics of ChatGPT's usage. Some students reported that ChatGPT furnished inaccurate information, while others expressed concerns that it generated false citations and references. These concerns highlight the importance of guaranteeing that AI chatbot models are accurate and reliable in higher education. Thus, McDaniel et al. (2021) stress the importance of incorporating personalized learning services for students in AI-supported learning tasks.

6.1. Conclusions

In conclusion, this experimental study has demonstrated the potential benefits of leveraging the ChatGPT to promote students' cognitive skills. Besides, the findings of this study illustrate that leveraging ChatGPT during didactic assistance in-class activities can positively impact students' critical, creative, and reflective thinking skills. Specifically, the EG exhibited significant improvement in their critical, creative, and reflective thinking scores compared to the CG, suggesting that ChatGPT can be an effective didactic mechanism for enhancing these skills. Despite the fact that this study did not assess learning outcomes, and the tests primarily measured changes in concrete skills rather than specific learning achievements, we could say leveraging ChatGPT for in-class tasks enhances the development of cognitive skills. As technology in education continues to grow, we suggest that ChatGPT can be a valuable mechanism for academics in higher education to consider. By integrating AI chatbot models in classroom tasks, academics can assist students in developing their critical, reflective, and creative thinking skills.

6.2. Implications for practice

The implications of this study for higher education are manifold. ChatGPT emerges as a valuable tool with the potential to revolutionize the teaching of critical, creative, and reflective thinking in the realm of education. Our research demonstrates that students can make substantial strides in honing these cognitive skills when guided by ChatGPT. This is particularly pertinent in the context of higher education, where fostering critical thinking is of paramount importance. The findings from this study strongly advocate for the integration of ChatGPT into higher education curricula as a means to bolster students' critical, creative, and reflective thinking abilities. Such an approach holds the promise of enhancing various courses that demand rigorous thinking, such as research methodology classes and those within the domains of social sciences, humanities, and education. Consequently, the results of this study serve as compelling evidence supporting the adoption of innovative teaching methodologies in higher education to elevate the quality of student learning outcomes.

6.3. Limitations and future studies

While the findings of this study are promising, it is crucial to acknowledge some limitations that warrant consideration. First, the study focused exclusively on third-year undergraduate students within a specific department, all of whom were enrolled in a research methodology course that employed the FC approach for one academic semester. This homogeneity of participants and course context may introduce a Hawthorne effect, wherein participants may have altered their behavior due to their awareness of being observed or participating in a novel educational experiment (McCambridge et al., 2014). To address this, future research should broaden its scope by recruiting participants from diverse departments and academic levels, including first-year, second-year, fourth-year, and postgraduate students. Additionally, exploring different teaching methods and strategies in various courses can yield more generalizable results and mitigate potential Hawthorne effects. Moreover, it is important to consider that the initial variance in critical thinking scores between the EG and CG at the pre-test stage may have influenced the observed percentage increase in scores. We acknowledge the importance of ensuring as much equivalence as possible between the two groups, particularly after the pre-test. In this study, the students were initially randomized into EG and CG to minimize potential biases. Future studies could explore methods to achieve greater equivalence between the groups post pre-test, ensuring that the results accurately reflect the impact of ChatGPT on critical thinking skills while controlling for any initial variations. Furthermore, future studies should consider conducting more extended investigations to measure the long-term impact of ChatGPT usage on students' cognitive abilities. This approach would provide valuable insights into the sustainability and durability of the observed improvements. Another limitation to consider is the reliance on self-reported data, which may be susceptible to social desirability bias. Future research could employ a mixed-methods approach, combining self-reported data with objective assessments to enhance the validity of findings. Moreover, this study focused primarily on the influence of ChatGPT on critical, creative, and reflective thinking skills. To gain a comprehensive understanding of AI's impact on education, future studies should explore its influence on other essential cognitive skills and competencies, such as metacognitive skills, problem-solving, and digital literacy. Additionally, investigating how ChatGPT supports the development of non-cognitive skills like emotional intelligence and social skills would contribute to a more holistic assessment of its educational value. It is essential to recognize that this study was conducted within the specific context of a Ghanaian university. The socio-cultural and economic backgrounds of the participants may not be representative of other settings. Therefore, future research should aim to replicate these findings in diverse educational contexts, universities, and countries to ascertain their consistency and generalizability. Ethical considerations and challenges associated with ChatGPT usage in education also warrant further exploration in future studies. Understanding the ethical implications and potential drawbacks of AI integration is crucial for responsible implementation in educational settings. Additionally, future research could delve into the impact of socio-cultural and personal factors on creative thinking compared to using ChatGPT. This could involve measuring creative thinking abilities in students exposed to different socio-cultural contexts and personal experiences and comparing their results to those who have access to ChatGPT. Lastly, an intriguing avenue for research would be to compare the effectiveness of different Language Models, such as Google Bard, in the classroom to identify the most suitable AI tools for specific educational contexts.

Ethical approval

Written informed consent was obtained from the participants of the study. This research is approved by the Department of Educational Innovations in Science and Technology Ethics Committee (EIST-EC) at

Kwame Nkrumah University of Science and Technology, Kumasi, Ghana (ref. number: EIST-EC/REF No.:/January 01, 2023).

Funding

N/A.

Declarations

No conflict of interest has been reported by the authors. All authors have read and approved the final version of the paper.

Data availability and sharing policy

The data that support the findings of this study are available from the corresponding author, [DV], upon reasonable request.

CRediT authorship contribution statement

Harry Barton Essel: Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Dimitrios Vlachopoulos:** Writing – review & editing, Supervision, Project administration, Investigation. **Albert Benjamin Essuman:** Writing – original draft, Software, Formal analysis, Data curation. **John Opuni Amankwa:** Writing – original draft, Validation, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

N/A.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.caeai.2023.100198>.

Study acronyms

LLMs	Large Language Models
ChatGPT	Chat Generative Pre-trained Transformer
AI	Artificial Intelligence
IoT	Internet of Things
NLP	Natural Language Processing
DL	Deep Learning
FC	Flipped Classroom
TPC	Theory Practical Credit
CWA	Cumulative Weighted Average
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
GFI	Goodness of Fit Index
IFI	Incremental Fit Index (IFI)
AGFI	Adjusted Goodness of Fit Index
RMSEA	Root Mean Square Error of Approximation
SRMR	Standardized Root Mean Square Residual
EMM	Estimated Marginal Means
EG	Experimental Group
CG	Control Group
CTS	Critical Thinking Scale
MCTS	Creative Thinking Scale
RTS	Reflective Thinking Scale
LMS	Learning Management System

References

- Akpur, U. (2020). Critical, reflective, creative thinking and their reflections on academic achievement. *Thinking Skills and Creativity*, 37, Article 100683.
- Al Darayseh, A. S. (2023). Acceptance of artificial intelligence in teaching science: Science teachers' perspective. *Computers and Education: Artificial Intelligence*, 4, Article 100132.
- Allen, L. K., & McNamara, D. S. (2015). Promoting self-regulated learning in an intelligent tutoring system for writing. In *Artificial intelligence in education: 17th international conference*. Madrid, Spain: AIED 2015. June 22-26, 2015. Proceedings 17 (pp. 827–830). Springer International Publishing.
- Ambrose, D., Cohen, L. M., & Tannenebaum, A. J. (Eds.). (2003). *Creative intelligence: Toward theoretic integration*. Cresskill N.J.: Hampton Press.
- Biasi, M. R. D., Valencia, G. E., & Obregon, L. G. (2019). A new educational thermodynamic software to promote critical thinking in youth engineering students. *Sustainability*, 12(1), 110.
- Blomhøj, M. (2011). Modelling competency: Teaching, learning and assessing competencies—Overview. *Trends in Teaching and Learning of Mathematical Modelling, ICTMA14*, 343–347.
- Carless, D. (2019). Feedback loops and the longer-term: Towards feedback spirals. *Assessment & Evaluation in Higher Education*, 44(5), 705–714.
- Chang, H. Y. (2017). How to augment the learning impact of computer simulations? The designs and effects of interactivity and scaffolding. *Interactive Learning Environments*, 25(8), 1083–1097.
- Chang, Y. S., & Yu, K. C. (2015). The relationship between perceptions of an innovative environment and creative performance in an online synchronous environment. *Computers in Human Behavior*, 49, 38–43.
- Chen, M. R. A., Hwang, G. J., & Chang, Y. Y. (2019). A reflective thinking-promoting approach to enhancing graduate students' flipped learning engagement, participation behaviors, reflective thinking and project learning outcomes. *British Journal of Educational Technology*, 50(5), 2288–2307.
- Darvishi, A., Khosravi, H., Sadiq, S., & Gasević, D. (2022). Incorporating AI and learning analytics to build trustworthy peer assessment systems. *British Journal of Educational Technology*, 53(4), 844–875.
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., ... Wright, R. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, Article 102642. <https://doi.org/10.36227/techrxiv.21789434.v1>
- Essel, H. B., Vlachopoulos, D., Tachie-Menson, A., Johnson, E. E., & Baah, P. K. (2022). The impact of a virtual teaching assistant (chatbot) on students' learning in Ghanaian higher education. *International Journal of Educational Technology in Higher Education*, 19(1), 1–19.
- Firat, M. (2023). *How chat GPT can transform autodidactic experiences and open education?*. <https://doi.org/10.31219/osf.io/9ge8m>
- Fox, J., & Weisberg, S. (2020). In *car: Companion to applied regression*. [R package]. Retrieved from. <https://cran.r-project.org/package=car>.
- Gangadharbatla, H. (2010). Technology component: A modified systems approach to creative thought. *Creativity Research Journal*, 22(2), 219–227.
- Goda, Y., Yamada, M., Matsukawa, H., Hata, K., & Yasunami, S. (2014). Conversation with a chatbot before an online EFL group discussion and the effects on critical thinking. *Journal of Information Systems Education*, 13(1), 1–7.
- Hapsari, I. P., & Wu, T. T. (2022). AI chatbots learning model in English speaking skill: Alleviating speaking anxiety, boosting enjoyment, and fostering critical thinking. In *Innovative technologies and learning: 5th international conference, ICITL 2022, virtual event, august 29–31, 2022, proceedings* (pp. 444–453). Cham: Springer International Publishing.
- Haque, M. U., Dharmadasa, I., Sworna, Z. T., Rajapakse, R. N., & Ahmad, H. (2022). *I think this is the most disruptive technology: Exploring sentiments of ChatGPT early adopters using Twitter data*. arXiv preprint arXiv:2212.05856.
- Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). Using mixed-methods sequential explanatory design: From theory to practice. *Field Methods*, 18(1), 3–20.
- Jamal, A., Solaiman, M., Alhasan, K., Tamsah, M. H., & Sayed, G. (2023). Integrating ChatGPT in medical education: Adapting curricula to cultivate competent physicians for the AI era. *Cureus*, 15(8).
- Jamovi project. (2021). In *Jamovi (2.3.24 version) [Computer Software]*. <https://www.jamovi.org>.
- Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, Article 102274.
- Kassymova, G. K., Kenzhaliyev, O. B., Koshbayeva, A. N., Triyono, B. M., & Ilmaliev, Z. B. (2020). E-learning, dilemma and cognitive competence. *Journal of Talent Development and Excellence*, 12(2s), 3689.
- Kember, D., Jones, A., Loke, A., McKay, J., Sinclair, K., Tse, H., Webb, C., Wong, F., Wong, M., & Yeung, E. (1999). Determining the level of reflective thinking from students' written journals using a coding scheme based on the work of Mezirow. *International Journal of Lifelong Education*, 18(1), 18–30.
- Kooli, C. (2023). Chatbots in education and research: A critical examination of ethical implications and solutions. *Sustainability*, 15(7), 5614.
- Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., ... Tseng, V. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digital Health*, 2(2), Article e0000198.
- Leippold, M. (2022). Thus spoke GPT-3: Interviewing a large-language model on climate finance. *Finance Research Letters*, Article 103617.
- Lenth, R. (2020). emmeans: Estimated marginal means, aka least-squares means. In [R package]. Retrieved from. <https://cran.r-project.org/package=emmeans>.
- Liang, W. (2022). *Towards a set of design principles for technology-assisted critical-thinking cultivation: A synthesis of research in English Language education*. Thinking Skills and Creativity, Article 101203.
- Lin, C. J. (2019). An online peer assessment approach to supporting mind mapping flipped learning activities for college English writing courses. *Journal of Computers in Education*, 6(3), 385–415.
- Li, Y., Sha, L., Yan, L., Lin, J., Raković, M., Galbraith, K., Lyons, K., Gašević, D., & Chen, G. (2023). Can large language models write reflectively. *Computers and Education: Artificial Intelligence*, 4, Article 100140. <https://doi.org/10.1016/j.caeai.2023.100140>
- Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(9), 1–35.
- Liu, X., Zheng, Y., Du, Z., Ding, M., Qian, Y., Yang, Z., & Tang, J. (2021). *GPT understands, too*. <https://doi.org/10.48550/arXiv.2103.10385>, arXiv.
- Long, J. D., Gannaway, P., Ford, C., Doumit, R., Zeeni, N., Sukkarieh-Haraty, O., ... Song, H. (2016). Effectiveness of a technology-based intervention to teach evidence-based practice: The EBR Tool. *Worldviews on Evidence-Based Nursing*, 13(1), 59–65.
- Luckin, R., Kukurova, M., Kent, C., & du Boulay, B. (2022). Empowering educators to be AI-ready. *Computers and Education: Artificial Intelligence*, 3, Article 100076.
- Mbakwe, A. B., Lourentzou, I., Celi, L. A., Mechanic, O. J., & Dagan, A. (2023). ChatGPT passing USMLE shines a spotlight on the flaws of medical education. *PLOS Digital Health*, 2(2), Article e0000205.
- McCambridge, J., Witton, J., & Elbourne, D. R. (2014). Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. *Journal of clinical epidemiology*, 67(3), 267–277.
- McDaniel, M. A., Einstein, G. O., & Een, E. (2021). Training college students to use learning strategies: A framework and pilot course. *Psychology Learning and Teaching*, 20(3), 364–382.
- Middleton, H. (2005). Creative thinking, values and design and technology education. *International Journal of Technology and Design Education*, 15(1), 61–71.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage.
- Mollick, E. R., & Mollick, L. (2022). New modes of learning enabled by AI chatbots: Three methods and assignments. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4300783>
- Naseem, U., Razzak, I., Khan, S. K., & Prasad, M. (2021). A comprehensive survey on word representation models: From classical to state-of-the-art word representation language models. *Transactions on Asian and Low-Resource Language Information Processing*, 20(5), 1–35.
- Navarrete, C. C. (2013). Creative thinking in digital game design and development: A case study. *Computers & Education*, 69, 320–331.
- OpenAI. (2023). *ChatGPT: Optimizing language models for dialogue*. <https://openai.com/blog/chatgpt/>.
- Özgenel, M., & Çetin, M. (2017). *Development of the marmara creative thinking dispositions scale: Validity and reliability analysis* (Vol. 46, pp. 113–132). Marmara University Atatürk Education Faculty Journal of Educational Sciences.
- Palermo, C., & Wilson, J. (2020). Implementing automated writing evaluation in different instructional contexts: A mixed-methods study. *Journal of Writing Research*, 12(1), 64–108.
- Root-Bernstein, R., & Root-Bernstein, M. (1999). *Sparks of genius: The thirteen thinking tools of creative people*. Houghton, Mifflin and Company.
- Roscoe, R. D., Snow, E. L., Allen, L. K., & McNamara, D. S. (2015). Automated detection of essay revising patterns: Applications for intelligent feedback in a writing tutor. *Grantee Submission*, 10(1), 59–79.
- Rospigliosi, P. A. (2023). Artificial intelligence in teaching and learning: What questions should we ask of ChatGPT? *Interactive Learning Environments*, 31(1), 1–3.
- Rubino, I., Barberis, C., & Malnati, G. (2018). Exploring the values of writing collaboratively through a digital storytelling platform: A mixed-methods analysis of users' participation, perspectives and practices. *Interactive Learning Environments*, 26(7), 882–894.
- Sanusi, I. T., Olaleye, S. A., Agbo, F. J., & Chiu, T. K. (2022). The role of learners' competencies in artificial intelligence education. *Computers and Education: Artificial Intelligence*, 3, Article 100098.
- Shadiev, R., & Huang, Y. M. (2020). Investigating student attention, meditation, cognitive load, and satisfaction during lectures in a foreign language supported by speech-enabled language translation. *Computer Assisted Language Learning*, 33(3), 301–326.
- Snow, E. L., Allen, L. K., Jacovina, M. E., Perret, C. A., & McNamara, D. S. (2015). You've got style: Detecting writing flexibility across time. In *Proceedings of the fifth international conference on learning analytics and knowledge* (pp. 194–202).
- Sosu, E. M. (2013). The development and psychometric validation of a critical thinking disposition scale. *Thinking Skills and Creativity*, 9, 107–119.
- Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning and Teaching*, 6(1).
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson.
- Tang, C., Mao, S., Xing, Z., & Naumann, S. (2022). Improving student creativity through digital technology products: A literature review. *Thinking Skills and Creativity*, Article 101032.
- 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), 15.

- Tsai, M. L., Ong, C. W., & Chen, C. L. (2023). Exploring the use of large language models (LLMs) in chemical engineering education: Building core course problem models with Chat-GPT. *Education for Chemical Engineers*, 44, 71–95.
- Tsankov, N. (2018). The transversal competence for problem-solving in cognitive learning. *International Journal of Cognitive Research in Science, Engineering and Education*, 6(3), 67.
- Wang, X., Liu, Q., Pang, H., Tan, S. C., Lei, J., Wallace, M., & Li, L. (2023). What matters in AI-supported learning: A study of human-AI interactions in language learning using cluster analysis and epistemic network analysis. *Computers and Education*, 194.
- Wodaj, H., & Belay, S. (2021). Effects of 7E instructional model with metacognitive scaffolding on students' conceptual understanding in biology. *Journal of Education in Science, Environment and Health*, 7(1), 26–43.
- Yılmaz, R. (2020). Enhancing community of inquiry and reflective thinking skills of undergraduates through using learning analytics-based process feedback. *Journal of Computer Assisted Learning*, 36(6), 909–921.