Acta Pedagogia Asiana

Research Article
Volume 4(1), 2024, 55–65
https://doi.org/10.53623/apga.v4i1.553

Comparative Analysis of the Effectiveness of Informatics Course Learning Utilizing Chatgpt

Arneitta Dwicahya Utami[1], Mia Kamayani[1]*, Estu Siduningrum[1], Nur Chalik Azhar[2]

[1]Department of Informatics Engineering, Faculty of Industrial and Informatics Technology, Universitas Muhammadiyah Prof. Dr.HAMKA, Indonesia
[2]Faculty of Industrial and Informatics Technology, Universitas Muhammadiyah Prof. Dr.HAMKA, Indonesia

*Correspondence: mia.kamayani@uhamka.ac.id

SUBMITTED: 30 November 2024; REVISED: 18 January 2025; ACCEPTED: 20 January 2025

ABSTRACT: This study examined the effectiveness of conventional teaching methods and ChatGPT in an introductory Algorithms and Programming course at the university level. ChatGPT, an AI-based NLP technology, assisted students in understanding course material through automated responses. However, its effectiveness relative to conventional methods required further evaluation, particularly concerning motivation, interaction, self-regulation, instructional structure, and the instructor's role. Using a sample of 10 students for pretest-posttest analysis, 38 respondents for the User Experience Questionnaire (UEQ), and accuracy analysis via prompt engineering, the results revealed that conventional methods better enhanced motivation and interaction. ChatGPT demonstrated strengths in attractiveness (1.982) and efficiency (2.053) but scored lower in accuracy (1.395) and novelty (1.053). Prompt engineering significantly improved response accuracy when tailored to learning modules, highlighting the importance of precise inputs. The findings suggested that while ChatGPT excelled as a supplementary tool, it was less effective as a standalone teaching method. This study contributed to the growing field of educational technology by providing insights into the integration of AI tools in learning environments.

KEYWORDS: learning effectiveness; ChatGPT; UEQ; prompt engineering

1. Introduction

The field of Natural Language Processing (NLP), a branch of artificial intelligence (AI), achieved remarkable advancements, sparking widespread interest and discussion across various domains [1]. NLP explored the interaction between humans and computers through natural language, forming the foundation of cutting-edge technologies such as generative AI. Generative AI, powered by deep learning, was capable of producing human-like responses based on user prompts, with ChatGPT being among the most popular tools utilized by students and the public [2, 3]. ChatGPT was designed as a sophisticated AI system trained on extensive datasets of internet-derived text, offering substantial assistance and support. However, its convenience also raised concerns regarding potential ethical issues and its impact on comprehension, particularly in science and technology education [4]. Despite these challenges, ChatGPT created significant opportunities for students by enhancing learning experiences and providing immediate support [5]. This development contrasted sharply with traditional learning methods, where tasks were typically accomplished through reading books, internet research, or

consulting multiple sources [6]. The divergence between these learning approaches necessitated an investigation into their comparative effectiveness.

Effectiveness in learning was defined through various indicators, including: (1) mastery of learning objectives at a class level, (2) student engagement during lessons, and (3) optimal allocation of learning time [7]. Additionally, some researchers evaluated effectiveness based on learning outcomes and active participation during educational activities [8]. Prior research, such as Shahid et al. (2022), explored the comparative effectiveness of ChatGPT and traditional teaching methods for learning English. Findings revealed that more than 60% of participants reported improved writing and speaking skills with ChatGPT. Conversely, traditional methods were deemed more effective in enhancing vocabulary for 60% of participants [4]. Aligned with the curriculum of the institution under study, introductory courses in algorithms and programming were mandatory for first-year students. Data from University X for the 2023-2024 academic year indicated that 17% of students failed this course, equating to 26 out of 156 students. To address this challenge, this research evaluated the comparative effectiveness of conventional learning methods versus those employing ChatGPT, focusing on selected algorithmic subtopics crucial for programming success. These included: (1) arrays, (2) searching algorithms, and (3) sorting algorithms, identified through discussions with course instructors.

The primary measure of effectiveness in this study was the learning outcomes achieved by students during the instructional process. According to Eom and Ashill (2016), several factors influenced learning effectiveness, such as motivation, self-regulation, the quality of interaction between instructors and students, and course design [9]. Furthermore, this research examined the roles of user experience (UX) and prompt engineering in enhancing ChatGPT-assisted learning. UX, encompassing usability, efficiency, and satisfaction, was critical for ensuring that students felt comfortable and achieved meaningful learning outcomes [10]. Evaluating UX was therefore integral to this study. Prompt engineering, a technique for optimizing inputs to achieve desired outputs, also significantly influenced the quality of learning with ChatGPT. Crafting effective prompts led to more accurate responses, thereby enhancing learning efficiency and effectiveness [11].

This study investigated ChatGPT's role in teaching introductory Algorithms and Programming, focusing on subtopics such as arrays, searching algorithms, and sorting algorithms. These topics were chosen due to their challenging nature for first-year students. By examining pretest-posttest results, UEQ scores, and prompt engineering effectiveness, this research addressed gaps in understanding AI-assisted learning. In summary, this study aimed to evaluate the effectiveness of integrating ChatGPT into learning methodologies, particularly in the context of algorithm and programming courses, by addressing both the technological and pedagogical aspects that influenced learning outcomes.

2. Materials and Methods

This study employs a mixed-method quantitative approach with experimental, survey-based, and computational techniques.

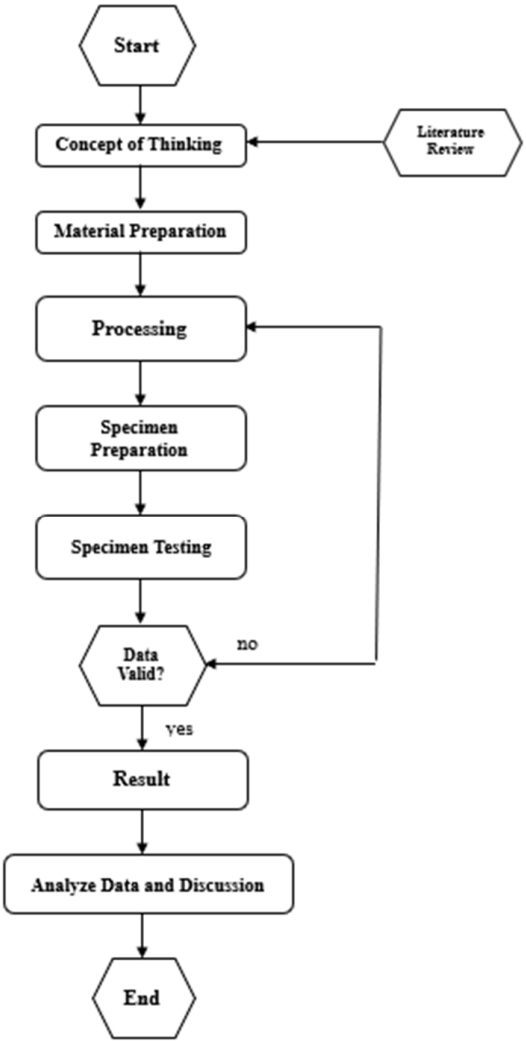

Figure 1. Research Methodology.

There were three data collection methods employed in this study: conducting experiments to determine learning effectiveness, collecting data through the UEQ, and calculating the accuracy of responses for each prompt using Python. The research design flow was illustrated in Figure 1.

2.1.Pretest-posttest analysis and questionnaire.

In this study, students were provided with learning modules based on the specified subtopics and were instructed to study the material independently. Before the experimental process began, participants took a pretest to assess their initial abilities. Ten students participated in a pretest-posttest experiment using learning modules on arrays, searching algorithms, and sorting algorithms. The experimental method, as applied in this research, was an approach to delivering learning materials that enabled students to conduct experiments and independently discover facts within the framework of the learning concept [12]. The pretest-posttest questions were related to the Algorithm and Basic Programming course. Conducting pretests and posttests before and after the research process provided a detailed framework of time and causality [13].

After completing the experiment, which included administering the pretest and posttest, participants filled out a questionnaire developed based on the study by Eom & Ashill (2016). The questions in this survey followed the framework established in their study, titled "The Determinants of Student Perceived Learning Outcomes and Satisfaction in University Online Education: An Update." The scale used in this questionnaire ranged from 1 to 3, where 1 indicated "strongly disagree," 2 indicated "neutral," and 3 indicated "agree." The categories for the average score on this scale were defined as follows [16]: Low: 1.00 - 1.66; Moderate: 1.67 - 2.33; High: 2.34 - 3.00.

2.2. UEQ.

UEQ is a usability testing method designed to provide comprehensive and pragmatic evaluation results regarding usability and user experience. It achieves this through surveys containing subjective questions about various aspects of quality [14]. The collection and analysis of data using the UEQ tools follow several steps:

2.2.1. Data transformation.

Data collected from respondents was imported into Excel, where it was converted into positive and negative values. To minimize bias in the answers, the questionnaire responses were randomized, as illustrated in Figure 2. This figure shows the sequence of values used to reduce response bias for each item. The converted data produced individual average scores, which were grouped by their respective aspects.

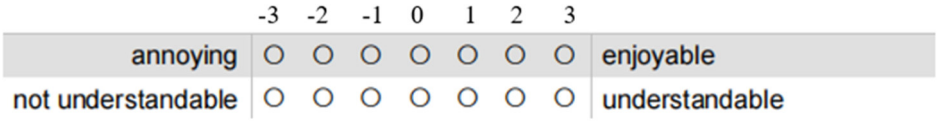

Figure 2. Data conversion (Source: UEQ instruments).

During the data transformation step, the following formula is used to calculate the mean from the data conversion:

$$\bar{x} = \frac{\sum \bar{x}[person]}{\sum item} \quad (1)$$

Where $\bar{x}$ is Individual scale mean, $\sum \bar{x}[person]$ is Total scale item score, and $\sum$item is Total number of scale items.

2.2.2. Determining UEQ results.

The average results from the data conversion are further calculated to obtain the main outcomes, which are then measured (benchmarked). The overall scale assumptions are calculated using the average results and the various mean values from the data conversion. The midpoint scale is defined as having a value between -0.8 and 0.8, representing a normal score range.

2.2.3. Establishing data benchmarks.

The UEQ data collection method employs benchmarks divided into five categories: Bad, Below Average, Average, Good, and Excellent. These categories are determined based on the results from the previously collected data.

2.3.Prompt engineering.

A prompt is a set of customized instructions provided to a large language model (LLM) to enhance its ability to generate the desired output [15]. In this study, prompt engineering was implemented by creating three prompts for each subtopic using two distinct methods. The first method involved generating responses from ChatGPT without including sample data, while the second method incorporated sample data as part of the input. The accuracy of the responses was calculated using the SequenceMatcher function in Python. This function compares the similarity between two strings by dividing the number of matching characters by the total number of characters in both texts. The resulting ratio was then multiplied by 100 to determine the percentage accuracy of the compared texts. The prompts, sample data, and responses generated by ChatGPT were provided in the Indonesian language.

3. Results and Discussion

3.1. Results of experiments.

Based on Figure 3A, which presents the results of an experiment comparing learning methods in the Algorithm and Programming Fundamentals course, it was observed that the pretest scores for both methods showed only a slight difference. The pretest results revealed that both groups started with similar levels of prior knowledge, indicating that the observed differences in posttest performance were due to the intervention methods. The pretest results for the conventional method had an average score of 52, while the ChatGPT-assisted learning method had an average score of 62. Meanwhile, the posttest results showed an average score of 88 for the conventional method and 90 for the ChatGPT-assisted learning method.

Figure 3B presents a comparison of the factors influencing student learning effectiveness. The figure illustrates that the conventional method has a greater impact on student learning effectiveness compared to the ChatGPT-assisted learning method. While the ChatGPT-assisted method achieved slightly higher posttest scores, the conventional method demonstrated a stronger influence on motivation and interaction. These results suggest that ChatGPT has limitations in replicating the instructor-led interactions that are essential for self-regulation and comprehensive learning engagement. This finding is supported by the questionnaire responses. According to the questionnaire results, ChatGPT effectively aids students by providing the answers they need while studying. However, in areas such as building motivation, facilitating interaction, and assisting with self-regulation, the conventional method proved more effective. This is likely because ChatGPT lacks a deeper understanding of human complexities, such as providing motivation and engaging in meaningful interactions, which are critical for supporting self-regulation in students. Pretest and posttest assessments were administered to evaluate students' knowledge of the *Algorithms and Basic Programming* course before and after the experimental intervention. This methodological approach allowed for a comparison of the knowledge gained by students through two distinct instructional strategies. Additionally,

factors such as motivation, instructor quality, interaction, learning structure, and self-regulation were identified as key determinants influencing student learning outcomes, encompassing both internal and external variables.

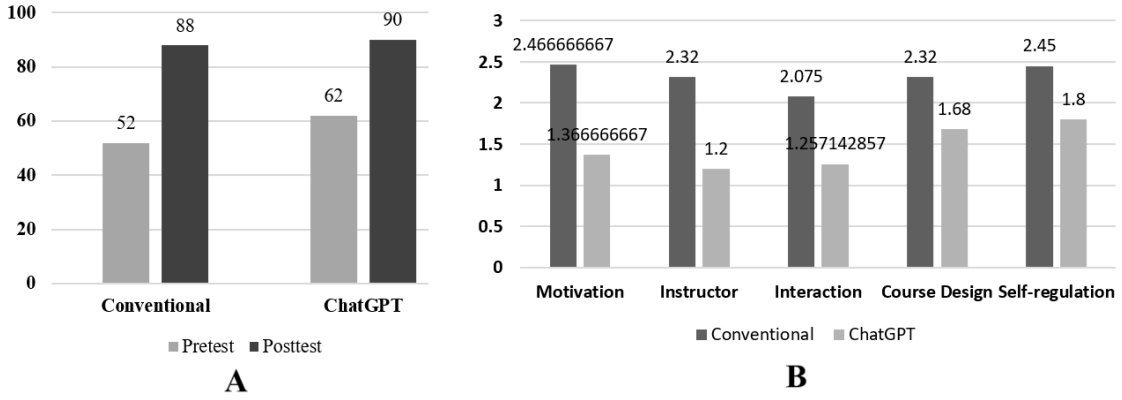

Figure 3. (A) Comparison of pretest and posttest results between conventional learning methods and ChatGPT; (B) Comparison of results on the effectiveness factors of students in conventional learning vs. ChatGPT method.

Figure 3B presents a comparison of the factors influencing student learning effectiveness. The figure illustrates that the conventional method has a greater impact on student learning effectiveness compared to the ChatGPT-assisted learning method. While the ChatGPT-assisted method achieved slightly higher posttest scores, the conventional method demonstrated a stronger influence on motivation and interaction. These results suggest that ChatGPT has limitations in replicating the instructor-led interactions that are essential for self-regulation and comprehensive learning engagement. This finding is supported by the questionnaire responses. According to the questionnaire results, ChatGPT effectively aids students by providing the answers they need while studying. However, in areas such as building motivation, facilitating interaction, and assisting with self-regulation, the conventional method proved more effective. This is likely because ChatGPT lacks a deeper understanding of human complexities, such as providing motivation and engaging in meaningful interactions, which are critical for supporting self-regulation in students. Pretest and posttest assessments were administered to evaluate students' knowledge of the *Algorithms and Basic Programming* course before and after the experimental intervention. This methodological approach allowed for a comparison of the knowledge gained by students through two distinct instructional strategies. Additionally, factors such as motivation, instructor quality, interaction, learning structure, and self-regulation were identified as key determinants influencing student learning outcomes, encompassing both internal and external variables.

3.2.User Experience with ChatGPT in learning activities.

The analysis of the overall UEQ calculations provided the average value, variance, and standard deviation for 26 items. Based on Figure 4, the attractiveness scale yielded an average positive value of 1.982, indicating that respondents enjoyed using ChatGPT to assist in their learning activities. For the clarity scale, a positive average value of 1.822 was obtained, suggesting that respondents found ChatGPT easy to understand and useful as a supportive tool for learning activities.

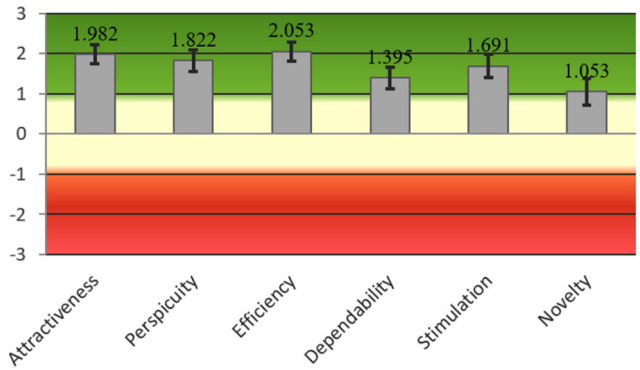

Figure 4. The average graph of impressions across the 6 scales.

Next, the efficiency scale recorded the highest average value of 2.053, indicating that respondents could easily obtain answers through ChatGPT without unnecessary effort. On the accuracy scale, a positive average value of 1.395 was achieved, suggesting that respondents found ChatGPT to be a reliable tool for assisting with their learning activities. The stimulation scale yielded an average value of 1.691, reflecting that respondents felt motivated when using ChatGPT as a medium to support their learning. Finally, the novelty scale had the lowest average value of 1.053, indicating that while ChatGPT aligns well with current technological standards, its innovative appeal was perceived as relatively limited.

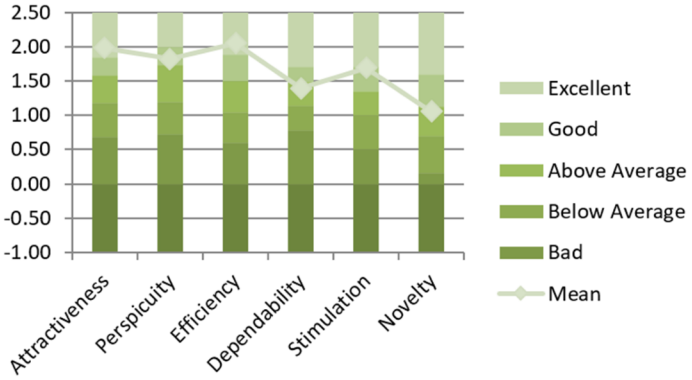

Figure 5. Standardization of the benchmark.

Based on the results of this study, it can be concluded that the evaluation for all aspects of the item scales yielded positive values, with all aspects of UX scoring above 0.8. This conclusion is drawn from the UEQ analysis, where values between -0.8 and 0.8 are considered normal, values greater than 0.8 are positive, and values less than -0.8 are negative. As shown in Figure 5, the standardization obtained in this study produced above-average results. According to the results displayed in Figure 4, the efficiency item scale received the highest rating, while the novelty item scale received the lowest rating. Therefore, it can be concluded that the novelty aspect requires further development to enhance its UX.

3.3.Comparison of results from various prompts in generating answers on ChatGPT

In addition to user experience and factors affecting student learning effectiveness, prompts also played a significant role in determining learning outcomes for students. When writing prompts, specific techniques had to be used to ensure that the generated responses aligned with the desired goals and requirements. In this study, to determine how different prompts affected the

quality of answers, the responses were compared for similarity using the SequenceMatcher library in the Python programming language. The accuracy of the similarity was compared with modules that had been validated by the course instructors as sample data. To make the comparison, two prompting conditions were used: the first condition involved not attaching a module as sample data to ChatGPT, but accuracy was still calculated based on the modules or sample data. The second condition involved attaching sample data to ChatGPT, so the responses would be generated according to the attachment. Three different prompts were used for each sub-learning topic, with the same meaning but with specific added words in each prompt. The responses generated by ChatGPT were then copied, and unnecessary characters or punctuation were removed to refine the answers. After that, the cleaned responses were converted into PDF format for comparison using SequenceMatcher.

3.3.1. Accuracy of ChatGPT responses without sample data.

Based on Table 1, there were differences in accuracy for each prompt. An accuracy level of 1% was considered low, while higher accuracy values indicated better alignment with the sample data. In the sub-learning topic of arrays, the accuracy improved with the third prompt, indicating that the response had a high similarity to the sample data. For the sub-learning topic of searching algorithms, there was also an improvement between prompt 1 and prompt 3, although the increase in accuracy was not highly significant. Similarly, for the sub-learning topic of sorting algorithms, there was an improvement between prompt 1 and prompt 2. Therefore, it can be concluded that prompt engineering had an impact on the effectiveness of student learning, allowing the responses generated by ChatGPT to better align with the requirements.

Table 1. Accuracy results table for answers without attaching sample data.

No.	Sub-Learning Topic	Prompt	Accuracy Level
1.	Array	Definition of array	1.81%
		The definition of an array and its program example	1.80%
		Explain the definition of an array and its algorithm with the declaration of contents for each type of array using C++	6.26%
2.	Searching Algorithm	Definition of searching algorithm	1.93%
		The definition of a searching algorithm and its program example	2.34%
		Explain the concept of a searching algorithm and implement the algorithm using C++	3.35%
3.	Sorting Algorithm	Definition of sorting algorithm	0.16%
		The definition of a sorting algorithm and its program example	0.95%
		Explain the concept of a sorting algorithm and implement the algorithm using C++	1.16%

3.3.2. Accuracy of ChatGPT responses with sample data.

In the process, the prompt was initiated by providing ChatGPT with sample data in the form of learning modules, so that the generated answers would focus on the provided attachments. It can be observed that each prompt in the three sub-learning topics showed improvement between the first and third prompts. An accuracy of 1% indicated a low accuracy level, while higher accuracy values signified better alignment with the sample data. This suggests an increase in similarity between the sample data and the answers generated by ChatGPT, leading

to the conclusion that, even when sample data is attached, the function of the prompt still influences the answers provided by ChatGPT (Table 2).

Table 2. Accuracy results table for answers with attaching sample data.

No.	Sub-Learning Topic	Prompt	Accuracy Level
		Prompt Initialization	
1.	Array	Definition of an array	1.32%
		Provide an explanation regarding the definition of arrays, their types, and general forms as outlined in the above module	11.36%
		Explain the definition of arrays, their types, general forms along with examples, and the implementation of the algorithm or sample programs. Answer in detail as provided in the attached module, without omissions.	15.97%
2.	Searching Algorithm	Definition of searching algorithms	3.34%
		Provide an explanation of the definition of searching algorithms, search methods, search requirements, and the formulation of search algorithms as outlined in the above module.	2.60%
		Explain the definition of searching algorithms, search methods, search techniques, and the implementation of the algorithm or sample programs as described in the module. Also provide explanations for each type of searching algorithm. Answer in detail as provided in the attached module, without omissions.	3.44%
3.	Sorting Algorithm	Definition of sorting algorithms	0.16%
		Provide an explanation of the definition of sorting algorithms, the benefits of sorted data, and sorting methods as outlined in the above module.	0.63%
		Explain the definition of sorting algorithms, the benefits of sorted data, factors affecting the effectiveness of sorting algorithms, sorting methods along with their analysis, array declaration, and the implementation of algorithms or sample programs as described in the module. Also provide explanations for each type of sorting algorithm. Answer in detail as provided in the attached module, without omissions)	2.20%

4. Conclusions

This study showed that the conventional learning method has a higher level of effectiveness compared to the use of ChatGPT, with average ratings ranging from "moderate" to "high" across various aspects such as motivation, interaction, and self-regulation. In contrast, the learning method using ChatGPT is considered less effective, with an average rating of "low" on almost all aspects. This study underscores the strengths and limitations of ChatGPT in educational contexts. While effective as a supplementary tool, ChatGPT cannot replace traditional teaching methods, especially in fostering motivation and interaction. Prompt engineering is critical to maximizing AI's potential in learning environments. Future studies should explore broader applications of AI tools to enhance educational experiences.

Acknowledgments

We would like to express our sincere thanks to the staff of the Department of Informatics Engineering, Faculty of Industrial and Informatics Technology, UHAMKA for providing the facility and assisting during the data collection and experiments.

Author Contribution

Arneitta Dwicahya Utami: Conceptualization, Methodology, Data Collection, Data Analysis, Funding Writing; Mia Kamayani Sulaeman: Conceptualization, Data Collection, Supervision; Mia Kamayani: Data Collection; Estu Siduningrum: Data Collection; Nur Chalik Azhar: Data Collection.

Competing Interest

The authors state that they have no conflicts of interest concerning the publication of this article. They also affirm that the data and the manuscript are entirely free from plagiarism.

Reference

[1] Herbold, S.; Janisz, A.H.; Heuer, U.; Kikteva, Z.; Trautsch, A. (2023). OPEN A large-scale comparison of human-written versus ChatGPT-generated essays. *Scientific Reports*, *1*, 1–11. https://doi.org/10.1038/s41598-023-45644-9.

[2] Zhou, M.; Duan, N.; Liu, S.; Shum, H.Y. (2020). Progress in Neural NLP: Modeling, Learning, and Reasoning. *Engineering*, *6*(3), 275–290. https://doi.org/10.1016/j.eng.2019.12.014.

[3] Xue, Y.; Chen, H.; Bai, G.R.; Tairas, R. (2024). Does ChatGPT help with introductory programming? An experiment of students using ChatGPT in CS1. *ICSE 2024 - Software Engineering Education and Training*, *1*, 1–11. https://doi.org/10.1145/3639474.3640076.

[4] Joshi, I.; et al. (2022). ChatGPT in the classroom: An analysis of its strengths and weaknesses for solving undergraduate computer science questions. *Proceedings of the ACM*, *1*, 625–631. https://doi.org/10.1145/3626252.3630803.

[5] Jackaria, P.M.; Hajan, B.H.; Mastul, A.R.H. (2024). A comparative analysis of the rating of college students' essays by ChatGPT versus human raters. *International Journal of Learning, Teaching, and Educational Research*, *23*(2), 478–492. https://doi.org/10.26803/ijlter.23.2.23.

[6] Septiyani, S.; Sutabri, T. (2024). Analisis perbandingan antara cara konvensional dan Chat GPT dalam keberhasilan mahasiswa mengerjakan tugas menggunakan metode pengamatan partisipan. *Indonesian Journal of Multidisciplinary*, *2*, 391–397.

[7] Hamzah, F.; Mujib, A.; Firmansyah. (2022). Efektivitas pembelajaran blended learning menggunakan Schoology pada pelajaran matematika. *DELTA: Jurnal Ilmiah Pendidikan Matematika*, *10*(1), 95–104.

[8] Ekawardhana, N.E. (2020). Efektivitas pembelajaran dengan menggunakan media video conference. *Seminar Nasional Ilmu Terapan*, *1*, 1–7. [Online]. Available: https://ojs.widyakartika.ac.id/index.php/sniter/article/view/218.

[9] Eom, S.B.; Ashill, N. (2016). The determinants of students' perceived learning outcomes and satisfaction in university online education: An update. *Decision Sciences Journal of Innovative Education*, *14*, 185–215. https://doi.org/10.1109/CONMEDIA46929.2019.8981813.

[10] Saleh, A.M.; Abuaddous, H.Y.; Alansari, I.S.; Enaizan, O. (2022). The evaluation of user experience of learning management systems using UEQ. *International Journal of Emerging Technologies in Learning*, *17*(7), 145–162. https://doi.org/10.3991/ijet.v17i07.29525.

[11] Kochanek, M.; et al. (2024). Improving training dataset balance with ChatGPT prompt engineering. *Electronics*, *13*(12), 1–20. https://doi.org/10.3390/electronics13122255.

[12] Masus, S.B.; Fadhilaturrahmi, F. (2020). Peningkatan keterampilan proses sains IPA dengan menggunakan metode eksperimen di sekolah dasar. *Jurnal Pendidikan dan Konseling*, *2*(2), 161–167. https://doi.org/10.31004/jpdk.v2i1.1129.

[13] Chen, C.A.; Hsieh, C.W.; Chen, D.Y. (2021). Can training enhance public employees' public service motivation? A pretest–posttest design. *Review of Public Personnel Administration*, *41*(1), 194–215. https://doi.org/10.1177/0734371X19872244.

[14] Pratama, A.; Faroqi, A.; Mandyartha, E.P. (2022). Evaluation of user experience in integrated learning information systems using User Experience Questionnaire (UEQ). *Journal of Information Systems and Informatics*, *4*(4), 1019–1029. https://doi.org/10.51519/journalisi.v4i4.394.

[15] A prompt pattern catalog to enhance prompt engineering with ChatGPT. (accessed on 1 November 2025 Year) Available online: http://arxiv.org/abs/2302.11382.

[16] Nurakhim, B.; Madiistriyatno, H.; Wijayanto, A. (2023). The influence of leadership on employee performance in Class IIA Central Institution, Karawang. *Return: Study of Management, Economics, and Business*, *3*(4), 400–408. https://doi.org/10.57096/return.v3i04.93.