THE IMPACT OF CHATGPT ON STUDENTS' PERFORMANCE IN SOLVING PROBABILITY TASKS

Vehid Kurtić, MA, dr. Naida Bikić, van. prof., Edina Kadrić Durmiš
Univerzitet u Zenici, Filozofski fakultet Zenica, BiH

ABSTRACT
The integration of artificial intelligence (AI) in education has gained increasing attention, particularly in enhancing students' ability to solve tasks in mathematics. This study investigates the effect of ChatGPT on university students' performance in solving probability tasks. A case study methodology was used, involving five students who participated in a three-week experimental program. Participants completed a pre-test without AI assistance and a post-test with ChatGPT support. The primary objective was to evaluate whether ChatGPT significantly improves the accuracy of task completion across different levels of complexity. Results show a substantial improvement in students' performance after AI intervention. The mean score rose from 7.80 (39.00%) in the pre-test to 14.10 (70.50%) in the post-test. Improvement was most notable in advanced-level tasks, where the success rate increased from 17.50% to 47.50%. Statistical analysis using repeated measures ANOVA confirmed a significant main effect of test phase ($F = 9.529$, $p = 0.037$), indicating that students achieved higher scores with ChatGPT.
These findings show that AI-driven tools like ChatGPT are effective in helping students solve tasks in probability theory. The consistent improvement across task complexities highlights ChatGPT's potential as a valuable educational aid, promoting deeper engagement with mathematical concepts.

Keywords: ChatGPT, artificial intelligence, probability, mathematics education, AI-assisted learning, solving tasks

1. INTRODUCTION
Probability is a foundational concept in mathematics education and a critical tool for interpreting statistical information, assessing risk, and making informed decisions. Yet, it remains one of the most conceptually difficult topics for students, who often struggle with its abstract nature and counterintuitive results. These difficulties are compounded by limited instructional time in curricula and an emphasis on formulaic procedures over conceptual understanding, resulting in shallow comprehension and persistent misconceptions [6]. Innovative solutions are needed to help learners engage more meaningfully with probability, particularly those that offer tailored support and real-time feedback. In this context, artificial intelligence (AI) tools like ChatGPT have emerged as potential learning aids. As a virtual tutor, it offers step-by-step guidance, clarification of concepts such as conditional probability and distributions, and instant feedback. These features hold promise for making abstract topics more approachable and supporting individual learning pathways. However, questions remain about its pedagogical impact, especially whether such tools can meaningfully improve students' actual performance or simply assist in answer generation [7,11].
While early studies have reported positive student attitudes toward ChatGPT's usability, speed, and convenience [3,4], empirical evidence on its effect on quantitative learning outcomes, particularly in mathematics, is still emerging. Kurtić et al. [6] found that students using ChatGPT

in probability tasks showed significant gains in performance, but they also continued to perceive the material as difficult, suggesting that confidence and comprehension may not always align. This points to the need for further research focused on measurable academic outcomes, especially in topics like probability that require conceptual depth and procedural fluency.

This study aims to examine the impact of ChatGPT on students' performance in university-level probability tasks, with additional reflection on students' learning experience. Specifically, it addresses the following questions:

1. *How does ChatGPT influence students' performance in solving probability tasks?*
2. *How do students perceive their learning experience when using ChatGPT for probability tasks?*

By analyzing students' performance before and after AI-assisted instruction, this study provides evidence on the practical effectiveness of ChatGPT in probability education. To contextualize these questions, the following section reviews relevant literature on the educational applications, cognitive impact, and ethical considerations surrounding ChatGPT in academic settings.

2. LITERATURE REVIEW

The integration of artificial intelligence (AI) tools, particularly large language models such as ChatGPT, has rapidly become a focal point of educational research, as institutions seek to understand their pedagogical implications. This interest is especially strong in higher education, where ChatGPT is being explored for its potential to support student learning, motivation, and engagement in complex subjects, including mathematics and probability. Ajlouni et al. [1] examined ChatGPT through the lens of constructivist learning theory, which emphasizes the active construction of knowledge through experience and interaction. Their findings suggest that ChatGPT can enhance student engagement and critical thinking, although theoretical gaps remain in fully aligning AI use with constructivist principles. Similarly, Cavazos et al. [4] moved beyond ethical debates to empirically assess how students perceive and use ChatGPT. Their factor analysis revealed multiple dimensions of AI use—information gathering, response generation, and self-improvement reflecting the complex cognitive roles ChatGPT can play in learning. Expanding this scope, Lo et al. [7] synthesized 72 studies in a systematic review that frames engagement in three dimensions: behavioral, emotional, and cognitive. While ChatGPT was linked to increased behavioral and emotional engagement manifested in enjoyment and academic involvement the cognitive dimension showed mixed results. This aligns with concerns about reduced critical thinking and overreliance on AI, a point echoed by Romero et al. [9], who applied the UTAUT2 model to analyze students' acceptance of ChatGPT in Spain. They found that motivation and prior experience were critical to the adoption of AI tools, reinforcing the need for user-centered educational design. In Latin American contexts, Guadalupe et al. [5] studied Peruvian students and highlighted both the accessibility and motivational potential of ChatGPT in digital learning, while also cautioning against issues of creativity loss and academic dishonesty. These concerns were further emphasized by Čuturilo and Plavec [8], who noted that institutional policies often lag behind student usage patterns, leading to ethical ambiguities in AI-supported academic work. Urban et al. [11] investigated the cognitive impacts of ChatGPT through the HHAIR framework, focusing on metacognition and self-efficacy in problem-solving contexts. They found that while ChatGPT can boost performance and confidence, it may also distort students' self-assessment of their capabilities. This is particularly relevant in tasks requiring deep conceptual understanding, such as those in probability, which is the focus of the present study. Other studies have examined ChatGPT's influence in specific educational domains. Ali et al. [2] explored its motivational role in English language learning, finding that self-determination and perceived autonomy played a major role in student engagement. Likewise, Thuy [10] reported that Vietnamese students valued

ChatGPT's language support, but also identified concerns about source credibility and citation, reflecting broader challenges in AI-mediated learning. From a sociotechnical standpoint, Baek et al. [3] emphasized that student use of ChatGPT is shaped by individual and institutional factors, including socio-economic status and academic discipline. Together, these studies provide a multifaceted foundation for exploring the impact of ChatGPT on students' learning experiences. They underscore ChatGPT's dual nature as both an effective support tool and a source of pedagogical concern especially in relation to self-confidence, conceptual understanding, and academic performance. This study contributes to the literature by examining these dimensions through a focused, task-specific lens, assessing ChatGPT's effectiveness in improving student outcomes in a core area of mathematics.

3. METHODOLOGY

This study employs a qualitative case study approach to examine the impact of using ChatGPT on students' success in solving probability tasks. The research was conducted in January 2024 with a sample of five mathematics and computer science students from the University of Zenica, who voluntarily participated and possessed basic prior knowledge in the field. During the three-week experimental program, students worked on exercises from the course *Probability and Statistics*, solving tasks in two phases: first without ChatGPT support (pre-test), and then with its assistance (post-test). The tests consisted of tasks organized into three difficulty levels: basic (3 tasks, 2 points each), intermediate (3 tasks, 3 points each), and advanced (2 tasks, 4 points each), covering combinatorics, conditional probability, probability distributions, and more complex mathematical techniques. The independent variable was the presence of ChatGPT support (used or not used), while the dependent variable was accuracy in solving tasks, measured by the total score and performance across difficulty levels. The main hypothesis of the study was:

- *H_1 – the use of ChatGPT significantly improves the accuracy of solving probability tasks among students.*

4. RESULTS AND FINDINGS

The tests were conducted in two phases: before and after the use of ChatGPT, with a maximum score of 20 points. The goal was to track the students' progress in solving probability tasks by assessing their abilities and understanding of tasks of varying difficulty levels.

As shown in *Table 1*, the students scored a total of 39 points out of 100 on the first test, while on the second test they scored 70.5 points, representing an approximately 80% increase. This improvement indicates a significant enhancement in solving probability tasks after using ChatGPT, which enabled a better understanding of key concepts, improved skills in solving probability tasks, and greater confidence in applying probability, as presented in [6].

At the basic level, the average success rate was 68.33%, as seen in *Table 1*, indicating a good grasp of fundamental probability concepts. However, as the tasks became more difficult, success rates decreased: at the advanced level, the success rate was only 17.5%, while at the intermediate level (which includes tasks involving conditional probabilities and distribution functions) the success rate was 38.33%. The average score on the first test was 7.80 (39%), as displayed in *Table 1*, suggesting that students had a solid understanding of basic concepts but required additional practice to master more complex ideas.

On the post-test, the average score was 14.1 (70.5%), as shown in *Table 1*, representing significant progress. Success in the intermediate category increased to 73.33%, while in the advanced category (which involves the application of integrals and geometric probability) the success rate rose to 47.5%. These results indicate a positive impact of ChatGPT on students' ability to understand and apply complex probability principles.

Table 1. pre-test and post-test results

	Basic (6 pts)		*Intermediate (6 pts)*		*Advanced (8 pts)*		*Total (20 pts)*	
	pre-test	*post-test*	*pre-test*	*post-test*	*pre-test*	*post-test*	*pre-test*	*post-test*
Student 1	4.5	5.5	1.0	4.5	2.0	3.0	7.5	13.0
Student 2	6.0	6.0	6.0	6.0	4.0	6.0	16.0	18.0
Student 3	2.0	6.0	1.0	5.0	0.0	5.0	3.0	16.0
Student 4	3.0	6.0	0.5	4.0	0.0	2.0	3.5	12.0
Student 5	5.0	6.0	3.0	2.5	1.0	3.0	9.0	11.5
Average	4.1	5.9	2.3	4.4	1.4	3.8	7.8	14.1

The results obtained suggest that the use of ChatGPT had an impact on students' performance in the post-test, confirming the main hypothesis (H_1) that ChatGPT significantly improves the accuracy of solving probability tasks among students. To confirm this claim, repeated measures analysis was applied in the study. This method allows for the evaluation of changes in the results of the same students before and after the intervention, which is crucial for understanding the effectiveness of the program. One of the key prerequisites for conducting repeated measures ANOVA is the assumption of sphericity, which was tested using Mauchly's Test. The results indicated that this assumption for Test factor sphericity was not violated. Therefore, the Greenhouse-Geisser and Huynh-Feldt corrections were applied to ensure the validity of the results. *Figure 1* best illustrates the students' progress in solving the post-test with the support of ChatGPT.

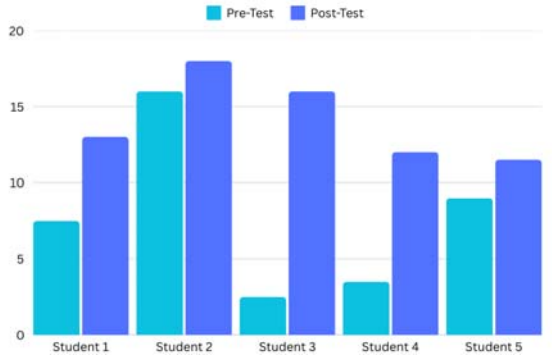

Figure 1. Total results on pre-test and post-test by student

Below, we provide an individual interpretation of each student's improvement, including the effect observed in terms of students' confidence and understanding of probability [6].

- Student 1 improved test score by 74%, and both confidence and conceptual understanding showed noticeable growth.
- Student 2 achieved a 12.5% increase in score. Despite the relatively modest performance gain, this student reported that ChatGPT was highly beneficial to their learning experience.
- Student 3 demonstrated the most substantial progress, with a remarkable 540% increase in score. ChatGPT had a transformative impact on confidence and comprehension.
- Student 4 enhanced the performance by 243%. ChatGPT significantly contributed to improved understanding and problem-solving ability.
- Student 5 recorded a 27.8% increase in score. This student highlighted ChatGPT's value in helping clarify key concepts in probability.

Furthermore, the results presented in *Table 2* reveal a statistically significant main effect of the test ($F = 9.529$, $p = 0.037$), indicating that students performed substantially better on the post-test compared to the pre-test. This finding confirms the effectiveness of the experimental program and provides empirical support for hypothesis H_1, which posits that students achieve significantly better results on the post-test than on the pre-test.

Table 2. Tests of Within-Subjects Effects

Source	*Type III Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*
Test	33.075	1	33.075	9.529	0.037

This finding aligns with previous research [6], which emphasized that the integration of AI tools, such as ChatGPT, leads to an improvement in students' performance, especially in solving probability tasks. As Kurtić et al. [6] also observed, students were more confident with basic tasks but faced greater challenges with complex tasks, even when using AI tools.

5. DISCUSSION

The results of this study demonstrate that the use of ChatGPT had a positive impact on students' performance in solving probability tasks. Improvements were recorded across all levels of task complexity, indicating that ChatGPT effectively supports student learning, not only for basic tasks but also for more challenging exercises. The lack of significant interaction between test type and task difficulty suggests that the observed progress was consistent regardless of cognitive demand.

Although this study did not directly assess the influence of ChatGPT on students' self-confidence or motivation, the findings can be interpreted in light of existing literature. According to Kurtić et al. [6], students' prior experience with AI tools significantly influences their behavioral intention and usage, with perceived usefulness and enjoyment being the most relevant motivators. The improved performance in the post-test may reflect the positive attitudes students held toward ChatGPT and their belief in its usefulness as a learning aid. Existing research also highlights potential challenges. Urban et al. [11], through the HHAIR theory, point to the misalignment between students' self-perception and actual performance, underscoring the need for metacognitive training when integrating AI tools in education. While our study did not address ethical aspects, relevant literature (see [2] and [8]) emphasizes the growing concern around the misuse of ChatGPT for paraphrasing or generating assignment content, often in contradiction with institutional guidelines. In conclusion, while the findings of this study confirm ChatGPT's value in supporting students' learning outcomes in probability tasks, they also open up questions about its broader impact on the learning process.

6. CONCLUSION

This study demonstrates that integrating ChatGPT into probability education can meaningfully enhance students' performance of solving tasks. By offering step-by-step guidance, real-time feedback, and accessible explanations, ChatGPT supports greater accuracy and efficiency in task completion—findings consistent with previous research [6]. While students also expressed more confidence during AI-supported tasks, the complexity of probability remained evident, indicating that AI can assist but not replace deep cognitive engagement.

ChatGPT also contributed to a shift in students' perception of probability, making abstract concepts feel more approachable. s[11], increased confidence may not always reflect true understanding, underscoring the need for thoughtful instructional design and metacognitive awareness. Ethical considerations are equally important. While ChatGPT promotes creative problem-solving [4], concerns remain regarding plagiarism, overreliance, and uncritical use [8].

Institutions must therefore provide clear guidelines and promote AI literacy to ensure responsible use. Based on these findings, the following recommendations are proposed:

- **Educators** should integrate ChatGPT within blended learning models and receive training to guide its ethical and pedagogical use.
- **Policymakers** should establish enforceable standards that ensure data privacy, equity, and academic integrity.
- **AI Developers** should enhance accuracy, transparency, and alignment with curricular goals, working closely with educators.

In sum, ChatGPT offers valuable support in probability instruction, but its effectiveness depends on responsible integration.

7. REFERENCES

[1] A. O. Ajlouni, F. A. Wahba, and Abdallah, "Students' Attitudes Towards Using ChatGPT as a Learning Tool: The Case of the University of Jordan," *Int. J. Interact. Mob. Technol.*, vol. 17, no. 18, pp. 99–117, 2023, doi: 10.3991/ijim.v17i18.41753.

[2] J. K. M. Ali, M. A. A. Shamsan, T. A. Hezam, and A. A. Q. Mohammed, "Impact of ChatGPT on Learning Motivation: Teachers and Students' Voices," *J. English Stud. Arabia Felix*, vol. 2, no. 1, pp. 41–49, Mar. 2023, doi: 10.56540/jesaf.v2i1.51.

[3] C. Baek, T. Tate, and M. Warschauer, " 'ChatGPT Seems Too Good to be True': College Students' Use and Perceptions of Generative AI," *Comput. Educ. Artif. Intell.*, vol. 7, p. 100294, 2024, doi: 10.1016/j.caeai.2024.100294.

[4] J. T. Cavazos, K. A. Hauck, H. M. Baskin, and C. M. Bain, "ChatGPT Goes to College: Exploring Student Perspectives on Artificial Intelligence in the Classroom," *Teach. Psychol.*, 2024, doi: 10.1177/00986283241268829.

[5] A. Guadalupe et al., "Effect of Chat GPT on the Digitized Learning Process of University Students," *J. Namibian Stud.*, vol. 33, pp. 1–15, 2023, doi: 10.59670/jns.v33i.411.

[6] V. Kurtić, N. Bikić, and E. Kadrić Durmiš, "Enhancing Students' Confidence and Understanding in Probability Through ChatGPT: An Analysis of AI's Impact on Learning Experiences," *SaZnanje*, vol. 4, no. 4, pp. 942–957, 2024. [Online]. Available: https://www.ceeol.com/search/article-detail?id=1270723

[7] C. K. Lo, K. F. Hew, and M. S. Jong, "The Influence of ChatGPT on Student Engagement: A Systematic Review and Future Research Agenda," *Comput. Educ.*, vol. 219, p. 105100, 2024, doi: 10.1016/j.compedu.2024.105100.

[8] N. Čuturilo and J. Plavec, "Writing with AI: University Students' Use of ChatGPT," *J. Lang. Educ.*, vol. 9, no. 4, 2023. [Online]. Available: https://cyberleninka.ru/article/n/writing-with-ai-university-students-use-of-chatgpt

[9] M. Romero, M. S. Ramírez-Montoya, M. B. Fernández, and F. Lara Lara, "Use of ChatGPT at University as a Tool for Complex Thinking: Students' Perceived Usefulness," *J. New Approaches Educ. Res.*, vol. 12, no. 2, p. 323, 2023, doi: 10.7821/naer.2023.7.1458.

[10] T. Thuy, "The Perception by University Students of the Use of ChatGPT in Education," *Int. J. Emerg. Technol. Learn.*, vol. 18, no. 17, pp. 4–19, 2023, doi: 10.3991/ijet.v18i17.39019.

[11] M. Urban et al., "ChatGPT Improves Creative Problem-Solving Performance in University Students: An Experimental Study," *Comput. Educ.*, vol. 215, p. 105031, 2024, doi: 10.1016/j.compedu.2024.105031.