

Copilot in Education: Impact on the Student Learning Experience

Alan Dennis, Antino Kim, Ge Yan
Kelley School of Business, Indiana University
November 20, 2024

Background

Generative artificial intelligence (AI) is anticipated to have a profound impact on education, with both potential benefits and challenges. Some critics argue that it may hinder students' ability to think critically and learn by encouraging over-reliance on AI to complete tasks (Selelo, 2024). Others are more optimistic, viewing it as a creative disruption that could revolutionize the centuries-old teaching paradigm and enable personalized learning experiences (Darvishi et al., 2024). Before embracing the opportunities or addressing the challenges of generative AI, it is essential to evaluate its impact on students' learning experiences. This assessment is critical for making informed decisions on how to effectively leverage generative AI in teaching and learning (Nguyen et al., 2024).

To this end, we examined students' experiences using Microsoft's generative AI tool, Microsoft 365 Copilot, in a real classroom setting where they completed actual course tasks that were graded. We are grateful to Microsoft for providing access to the necessary licenses for the tool.

Study Description

The participants were recruited from the required core course in the Master of Science in Information Systems program at the Kelley School of Business, Indiana University, in the fall of 2024. A total of 88 students (45% female) took part in the study, which was approved by the university's IRB under protocol #24611.

Participants performed two tasks (one with AI and one without AI, randomly assigned) in random order in a controlled in-person classroom setting and were monitored by exam proctors. Students were permitted to use any resources in

an open-book fashion for both tasks; the task with AI allowed the use of generative AI, while the task without AI prohibited the use of AI but still permitted other open-book resources. Each task was limited to 75 minutes, and we recorded how long each student took. We confirmed that Copilot functioned properly and that there were no technical issues that could affect the results.

Together, the two tasks constituted 10% of the course grade. The course teaching assistant graded the students' work as a part of her regular assigned duties in the course using a 15-point scale. She was not informed which submission was AI-assisted and which was not. Both tasks asked participants to assume a consultant role and propose an AI-based solution to a client problem presented in a short case. The tasks required them to clearly outline the design of the proposed solution, the associated business processes, the strategy for assessment, and the potential benefits and risks of implementation.

Upon completing both tasks, participants answered questions about their perceptions of the two dimensions. First, we assessed the perceived effectiveness of the tools. This was assessed using six items ($\alpha=.88$): The tools were useful in doing the task; The tools met my needs in doing the tasks; I am satisfied with the tools; The tools were fun to use to do the task; I would use the tools to do similar tasks in the future; I would recommend the tools to a friend. Second, we asked about the extent to which the work reflected their own knowledge and abilities (five items; $\alpha=.88$): I feel a strong sense of ownership for the work I submitted; I feel invested in the work I submitted; I used my personal judgment to do the task. I made independent decisions to do the task.

I decided what methods to use to do the task. All items were measured on 7-point Likert scales.

Results

Table 1 presents the results. AI significantly increased quality (student's grade) and reduced the time taken, both with large effect sizes (i.e., Cohen's $d > 0.8$). In practical terms, large effects are those that are "grossly perceptible," such as the "mean difference in height between 13- and 18-year-old girls" (Cohen, 1988). Participants reported increased effectiveness when using AI, also with large effect sizes (i.e., Cohen's $d > 0.8$). This aligns with our findings regarding grades and task completion time. However, the results also indicate that the use of AI led to a significantly lower perception that the work was their own. These effects were also large.

Table 1. Summary of Results

	AI	No AI		
	Mean (Std. Dev.)	Mean (Std. Dev.)	p-value	Cohen's d
Behavioral Outcomes				
Quality (Grade)	13.82 (1.03)	12.64 (1.72)	<0.001	0.83
Time (Minutes)	46.36 (23.38)	65.73 (16.10)	<0.001	0.97
Student Perceptions				
Perceived Effectiveness	6.38 (0.93)	5.49 (1.14)	<0.001	0.86
Own Work	4.92 (1.25)	5.89 (1.22)	<0.001	0.79

Interestingly, a Levene's test revealed that when AI was used, there was significantly less variance in quality ($p < .001$) with both the AI data (skewness= -1.12, kurtosis= 1.57) and non-AI data having left tails (skewness= -0.72, kurtosis= 0.13). Conversely, there was significantly greater variance in time with AI ($p < .001$), with the AI data being almost uniform in distribution (skewness= 0.05, kurtosis= -1.23) and the non-AI data being more normal but with a left tail (skewness= -0.99, kurtosis= 0.24).

Conclusion

Our results indicate that Copilot significantly enhanced students' performance. AI use increased grades by 10% and reduced the variation in grades. It also reduced the time taken by 40%, although this increased the variability in time. Students recognized these effects and reported that they perceived that AI improved their ability to perform tasks effectively. However, the use of Copilot also diminished the students' perceptions that their work was their own. In other words, the students felt they could complete the task more effectively with Copilot, but felt the final work did not fully represent their own skills and knowledge. This duality of AI improving performance while reducing the perceived authenticity of human contribution highlights the novel tension between learning efficiency and the intrinsic value of learning. These findings have important implications for integrating generative AI in education, highlighting the need to balance the tool's ability to address pain points in the learning experience and deliver timely information while providing students with opportunities to perform and demonstrate their learning.

References

- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., & Siemens, G. (2024). Impact of AI assistance on student agency. *Computers & Education*, 210, 104967.
- Nguyen, A., Kremantzis, M., Essien, A., Petrounias, I., & Hosseini, S. (2024). Enhancing Student Engagement Through Artificial Intelligence (AI): Understanding the Basics, Opportunities, and Challenges. *Journal of University Teaching and Learning Practice*, 21(06).
- Selelo, M. E. (2024). The Perpetuation of Lazy Thinking by Exploiting Writing Artificial Intelligence Tools: Perceptions from Students in the University of Limpopo. *Journal of English Education and Technology*, 4(4), 350-363.

These findings represent our initial analysis, and a full paper with a more comprehensive discussion of the implications and future research directions will follow.