The Impact of Interacting with Generative Artificial Intelligence on College Students' Creativity Performance

Yue Tong Wu[1]*, Minyu Bin[1], Weiting Liao[1], Yi Yuan[1], Xinyu Liu[1], YunYing Yang[1]

[1]South China Normal University, China

*2024020983@m.scnu.edu.cn

Abstract:*This study leverages the PISA 2022 Innovative Thinking Assessment to investigate the impact of generative AI(GAI) on college students, encompassing three dimensions: Generating diverse ideas (GI), Evaluating and Improving existing ideas (EI), and Generating Creative ideas (GC). This study reveals the strengths of GAI in supporting assessment and improvement skills, and also points out its shortcomings in promoting divergent thinking and creative expression, which provides some references for the research on GAI assisting students to enhance their creativity.*

Keywords: GAI, Creativity, PISA 2022, Evaluating and Improvement Ideas, Higher Education

1.Introduction

GAI offers new opportunities for the creative development of university students by autonomously generating texts, images and solutions (Habib et al., 2023). The strong capabilities of text generation, translation and dialogue make it possible for GAI to support assessment practices (Nikolic et al., 2023), gradually making it a widely used new tool in education (Crompton & Burke, 2023). Zhu et al. (2025) meta-analysis found that long-term interaction with GAI is effective in improving students' performance in school, and that GAI is more effective in higher education (Chaudhry et al., 2023). According to the dual pathway to creativity model, the generation of creativity relies on the collaborative effects of divergent thinking (generating multiple possibilities) and convergent thinking (selecting the best solution) (Runco & Acar, 2012). In this framework, GAI can supports divergent thinkines that the intervention of technological tools can influence an individual's creativity performance at different levels along the continuum from everyday creativity (mini-c) to exceptional creativity (big-c) (Kaufman & Beghetto, 2009). By rapidly generating a large number of ideas, it facilitates the optimisation process of convergent thinking by providing structured suggestions. Understanding how GAI interaction with postgraduate students affects students' creative performance is important for fostering students' creativity. The PISA 2022 innovative thinking test framework provides a standardised tool for assessing creativity in college students, and its three core dimensions are highly consistent with the creative generation process. By introducing generative artificial intelligence into the process of college students completing the PISA 2022 test on innovative thinking, this study will explore the following questions.

1.Does the intervention of GAI have an impact on the creativity performance of university students in the PISA test?

2.Are there differences in this impact across the three dimensions of generating diverse ideas, evaluating and improving existing ideas, and generating creative ideas?

2.Methods

This study selected 16 students majoring in modern educational technology from a university in South China (14 females and 2 males, all familiar with AI tools). The participants were required to complete 10 questions from the four categories of PISA 2022 (written expression, visual expression, social and scientific problem-solving) within the specified time(OECD, 2023). The participants first answered the questions independently, then collaborated with the GAI(Tencent Yuanbao) to form the final answers, and uploaded the process data. Of the ten questions, three were designed to test whether students could generate different ideas, three were designed to test whether they could evaluate and improve ideas, and four were designed to test whether they could generate creative ideas. After the test, we formulated a scoring framework for each unit based on the official PISA 2022 scoring criteria, and conducted a scoring statistics on the answers

submitted by the students. After excluding the data of the two students who did not submit complete responses, the responses of the remaining 14 students (2 males and 12 females) were statistically analysed and the total score statistics are shown in Figure 1. Based on the score results, We selected typical students for case studies in order to further explain the reasons for the differences.

3.Results

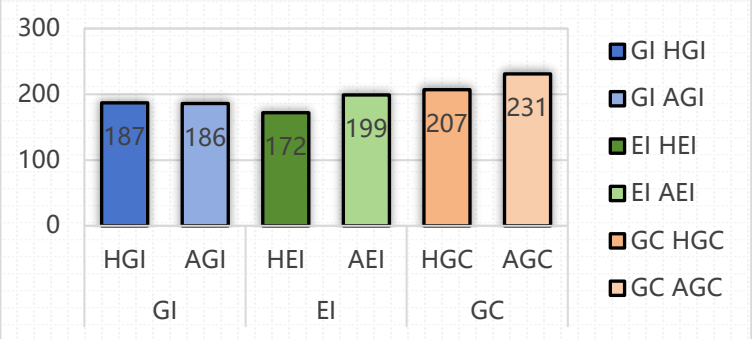

Figure 1. The scores of students' innovative thinking in each dimension

Note.

GI: Dimension of generating diverse ideas

HGI: Score of students completing independently

AGI: Score of students completing in collaboration with GAI

EI: Dimension of evaluating and improving existing ideas

HEI: Score of students completing independently

AEI: Score of students completing in collaboration with GAI

GC: Dimension of generating creative ideas

HGC: Score of students completing independently

AGC: Score of students completing in collaboration with GAI

3.1.Analysis of the Generating Diverse Ideas (GI)

The Wilcoxon signed-rank test indicated no significant difference between AGI and HGI (Z=-0.447, p=0.655, two-tailed), with 85.7% of the pairs (12/14) having identical scores. The negligible effect size (r=-0.12) further supports the lack of a meaningful difference. The intervention of GAI did not affect students' performance in the dimension of generating diverse ideas.

3.2.Analysis of the Evaluation and Improvement of Ideas (EI)

The score of the HEI group (M=12.29, SD=1.33) was lower than that of the AEI group (M=14.21, SD=2.16). The paired sample t-test results showed that the difference between the two groups was significant (p=0.015, p<0.05). The effect size analysis indicated that the difference was of moderate effect (Cohen's d=-0.75, 95% CI [-1.33, -0.14]), suggesting that the score of the HEI group was significantly lower than that of the AEI group.

Table 1.The results of the paired sample T-test for the EI dimension

	n	M	SD	p	t	Cohen's d
HEI	14	12.29	1.326	0.015*	-2.791	-0.746
AEI	14	14.21	2.155			

Note: *p<0.05

3.3.Analysis of the Generating Creative Ideas (GC)

The paired-sample t-test results indicated that the difference between the scores of the HGC group (M=14.79, SD=3.81) and the AGC group (M=16.50, SD=2.88) did not reach statistical significance, p=0.135 (p>0.05). The effect size analysis showed that the effect size was small (Cohen's d=-0.43, 95%CI [-0.97, 0.13]), and the score difference between HGC and AGC was not significant.

Table 2.The results of the paired sample T-test for the GC dimension

	n	M	SD	p	t	Cohen's d
HGC	14	14.79	3.806	0.135	-1.593	-0.426
AGC	14	16.50	2.876			

Note: *p<0.05

3.4.Analysis the question texts of typical students

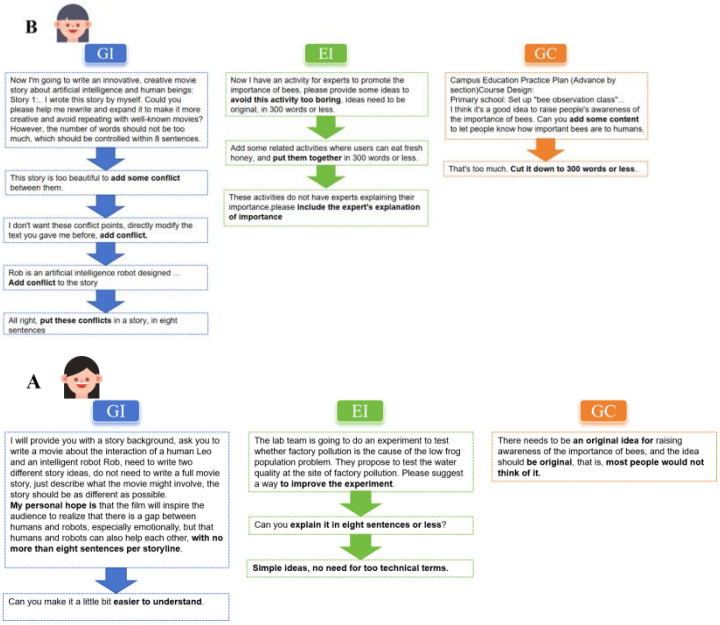

Figure 2. Example text of Student A and B's questioning of GAI in the GI, EI, and GC dimensions

The case analysis shows that students A and B only improved in the EI dimension, and A's GC score decreased after collaboration. This is attributed to the fact that they adopted a problem-oriented strategy in this dimension, conducting deep cognitive interaction and scheme optimization through the immediate feedback of GAI. In contrast, the scores in the GI dimension were limited by inefficient communication (B) or the simple reliance on AI-generated content (A). In the GC dimension, both students showed shallow interaction, such as directly adopting the AI output or only making superficial modifications, resulting in stagnant or even declining scores (A). Thus, the significant increase in the EI dimension can be attributed to the fact that students were more inclined to engage in substantive problem-solving strategy discussions with AI rather than merely performing low-level task processing, Their questions all involve deep-level cognitive interactions such as ideas, understanding, and thinking.

4.Discussion

In conjunction with the Case Study, the significant improvement in the EI dimension may be related to the fact that students entered the cognitive level of the discussion earlier during the interaction with the GAI. According to Suriano et al. (2024), students who possess a high level of trust and positive attitudes when interacting with AI tools such as ChatGPT are more inclined to delve deeper into cognitive dimensions, such as critical analyses and improvement of existing ideas. In addition, the non-significance of the GI and GC dimensions may be related to the limitations of the GAI tool and the nature of the task. On the one hand, generating diverse ideas (GI) requires students to have high divergent thinking skills, whereas the output of GAI tools is often based on schematic generation of existing data, which may limit students' original thinking (An et al., 2025). Meanwhile, the tool we used in the experiment, Tencent Yuanbao, does not directly help students to draw pictures or modify posters, and it is not as powerful as text in processing pictures, which may further explain the lack of significant improvement in the GC dimension. On the other hand, this study used PISA2022 test questions, which are relatively simple for college students, and the evaluation settings of the questions only assessed the number of diverse ideas in the GI dimension. With or without the collaboration of GAI, students were able to complete the requirements of the question well in this dimension.

This study also has some limitations, including a small sample size, the lack of comprehensive behavioral coding, and the possible ceiling effect due to the PISA tasks designed for young students. The experimental procedure might have increased the cognitive burden on the participants. In future research, we will use a comprehensive GAI testing platform to reduce the task burden on the participants and employ behavioral sequence analysis to explore the internal mechanism by which GAI affects creative performance.

Reference

An, S., Zhang, S., Guo, T., Lu, S., Zhang, W., & Cai, Z. (2025). Impacts of generative AI on student teachers' task performance and collaborative knowledge construction process in mind mapping-based collaborative environment. *Computers & Education*, *227*, 105227. https://doi.org/10.1016/j.compedu.2024.105227

Chaudhry, I. S., Sarwary, A. M., Ghaleb, & Chabchoub, H. (2023). Time to Revisit Existing Student's Performance Evaluation Approach in Higher Education Sector in a New Era of ChatGPT — A Case Study. *Cogent Education*, *10*(1). https://doi.org/10.1080/2331186x.2023.2210461

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, *20*(22), 1–22. https://doi.org/10.1186/s41239-023-00392-8

Habib, S., Vogel, T., Thorne, E., & Xiao, A. (2023). How Does Generative Artificial Intelligence Impact Student Creativity? *Journal of Creativity*, *34*(1). https://doi.org/10.1016/j.yjoc.2023.100072

Kaufman, J. C., & Beghetto, R. A. (2009). Beyond Big and Little: the Four C Model of Creativity. *Review of General Psychology*, *13*(1), 1–12. https://doi.org/10.1037/a0013688

Nikolic, S., Daniel, S., Haque, R., Belkina, M., Ghulam Mubashar Hassan, Grundy, S., Lyden, S., Neal, P. R., & Sandison, C. (2023). ChatGPT versus engineering education assessment: a multidisciplinary and multi-institutional

benchmarking and analysis of this generative artificial intelligence tool to investigate assessment integrity. *European Journal of Engineering Education*, *48*(4), 1–56. https://doi.org/10.1080/03043797.2023.2213169

OECD (2023), PISA 2022 Assessment and Analytical Framework, PISA, OECD Publishing, Paris, https://doi.org/10.1787/dfe0bf9c-en.

Runco, M. A., & Acar, S. (2012). Divergent Thinking as an Indicator of Creative Potential. *Creativity Research Journal*, *24*(1), 66–75. https://doi.org/10.1080/10400419.2012.652929

Suriano, R., Plebe, A., Acciai, A., & Fabio, R. A. (2024). Student interaction with ChatGPT can promote complex critical thinking skills. *Learning and Instruction*, *95*(0959-4752), 102011–102011. https://doi.org/10.1016/j.learninstruc.2024.102011

Zhu, Y., Liu, Q., & Zhao, L. (2025). Exploring the impact of generative artificial intelligence on students' learning outcomes: a meta-analysis. *Education and Information Technologies*, *30*, 16211–16239. https://doi.org/10.1007/s10639-025-13420-z