Exploring the Impact of ChatGPT on Student Interactions in Computer-Supported Collaborative Learning

Han Kyul Kim[1], Shriniwas Nayak[1], Aleyeh Roknaldin[1], Xiaoci Zhang[1], Marlon Twyman[1], Stephen Lu[1]

[1]University of Southern California

{hankyulk, spnayak, roknaldi, xiaocizh, marlontw, sclu}@aaai.org

Abstract

The growing popularity of generative AI, particularly ChatGPT, has sparked both enthusiasm and caution among practitioners and researchers in education. To effectively harness the full potential of ChatGPT in educational contexts, it is crucial to analyze its impact and suitability for different educational purposes. This paper takes an initial step in exploring the applicability of ChatGPT in a computer-supported collaborative learning (CSCL) environment. Using statistical analysis, we validate the shifts in student interactions during an asynchronous group brainstorming session by introducing ChatGPT as an instantaneous question-answering agent.

Introduction

A generative AI such as ChatGPT brings a paradigm shift in education settings with its potential usage in providing personalized learning or virtual assistance to instructors (Ansah 2023; Lo 2023; Qadir 2023). However, concerns related to hallucination (Ji et al. 2023; Alkaissi and McFarlane 2023) and academic integrity (McMurtrie 2023; Rudolph, Tan, and Tan 2023) have been raised, casting doubt on the broader applicability of generative AI in education.

The issues of factual accuracy and fair assessment from generative AI are crucial when the primary objective of learning activities is to acquire accurate factual information. Yet, it's essential to recognize that such an objective, in line with established education frameworks like Bloom's taxonomy (Bloom et al. 1956; Kropp, Stoker, and Bashaw 1966), represents only a fraction of the learning objectives necessary for effective educational experiences. When learning activities are designed to empower students to analyze, evaluate, or create new ideas, as categorized in Bloom's taxonomy, fostering students' creativity becomes a core issue (Egan et al. 2017; Al-Samarraie and Hurmuzan 2018). For example, in brainstorming activities, the quantity or novelty of the ideas become crucial metrics to track (Fu et al. 2015; Hong and Chiu 2016).

In this paper, we present a unique application of ChatGPT within the context of asynchronous computer-supported collaborative learning (CSCL), in which the focal learning objective is to enhance students' creative thinking. Our contribution not only expands the application of generative AI in educational settings but also provides initial statistical analyses illustrating the changes in student interaction patterns resulting from integrating ChatGPT into a group brainstorming activity.

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Experiment setting & methods

CSCL context

Our application and analysis draw from student chat data gathered in a graduate-level product engineering course during the spring of 2023. Over the final two weeks of the semester, 12 cohorts, each comprising 4 to 5 students, were randomly generated. Each week, cohort members participated in asynchronous online group discussions for three days to generate a list of creative product design ideas aligned with specific customer requirements. At the end of each week, their ideas were presented to the entire class and reviewed by both their classmates and instructors.

Among 12 cohorts, 6 cohorts, labeled as `Type A`, were randomly selected to incorporate ChatGPT as an additional question-answering agent. Students in these cohorts were provided with the ability to tag ChatGPT in their messages. Subsequently, our internal system interfaced with ChatGPT's OpenAI API to retrieve and post its responses to the tagged messages on the group's chat. Conversely, the remaining 6 cohorts, referred to as `Type B`, were not provided with any option to directly interact with ChatGPT during their discussions. The specific number of students in each cohort and comprehensive statistics on the number of messages in each cohort are provided in Tables 5 $\sim$ 7 in the Supplementary materials.

Evaluation methods

To evaluate the impact of ChatGPT on student interactions during group brainstorming activities, we employed the Collaborative Learning Conversation Skills Taxonomy (CLCST) (Soller 2001). It classifies the interaction-related skills commonly used by students in CSCL environments. Two experts familiar with CLCST manually identified messages in the discussions that exhibited the students' active learning, as defined in Soller (2001) as follows:

- Request: ask for help/advice in solving the problem or in understanding a team-mate's comment.

- Inform: direct or advance the conversation by providing information or advice.
- Motivate: provide positive feedback and reinforcement.

To ensure annotation consistency, we computed Cohen's kappa (Cohen 1960) between the two annotators, resulting in a consistency score of 93.2%. This high value indicates almost perfect agreement between the two annotators (McHugh 2012).

For all our statistical analyses, we conducted repeated ANOVA, incorporating the inclusion of ChatGPT in a cohort and the week of the brainstorming session as two independent covariates. To account for the variation in the number of messages and students in each cohort, we use the proportions of active learning messages within each cohort as a dependent variable. Upon obtaining significant results from the F-test, we employed the Tukey method (Tukey 1949) for subsequent multiple pairwise comparisons, allowing us to mitigate group-wise variation.

Results

When comparing the level of active learning in overall interactions within `Type A` and `Type B` cohorts, we did not observe any statistical significance (Table 8 in the Supplementary materials). While the impact of ChatGPT was not evident across all student interactions, this result does not preclude the possibility of statistical significance within different types of student interactions. As ChatGPT was exclusively included in `Type A`, we can categorize the student interactions within `Type A` into two distinct types: student-student interactions and student-ChatGPT interactions.

Student-student interaction

As students from cohorts in `Type A` were required to tag ChatGPT in their messages to induce its response, all of their messages without the tag were regarded as intended for other students. For `Type B`, where ChatGPT was not available as a question-answering agent, all of their interactions were categorized as student-student interactions.

Table 1: ANOVA results on the ratio of active learning within student-student interactions, indicating statistically significant group-wise variation resulting from introducing ChatGPT

Source of Variation	SS	df	F
ChatGPT	0.307	1	**4.909***
Week	0.031	1	0.501
ChatGPT $\times$ Week	0.080	1	1.285
Residual	1.253	20	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Based on the statistical analysis from Tables 1 and 2, cohorts in `Type B` exhibited a 22.6 percentage point higher ratio of active learning in student-student interactions. This finding implies that even though there was no major change in the overall student interaction, the internal dynamics of student interactions, particularly the level of active learning, were significantly altered by incorporating ChatGPT.

Table 2: Impact of ChatGPT on the ratio of active learning in student-student interaction based on Tukey's method, indicating a higher ratio of active learning in `Type B` cohorts

Comparison	Difference	Adj. p-value
With ChatGPT vs. Without ChatGPT	0.226	**0.037**

Student-ChatGPT interaction

When analyzing student-ChatGPT interactions, we exclusively focus on cohorts in `Type A`, comparing the levels of active learning in messages directed towards ChatGPT and those addressed to other students within the same cohort.

Table 3: ANOVA results on the ratio of active learning between student-ChatGPT and student-student interactions, indicating statistically significant group-wise variation between the types of interactions

Source of Variation	SS	df	F
To ChatGPT	1.722	1	**24.765*****
Week	0.000	1	0.000
ChatGPT $\times$ Week	0.008	1	0.11
Residual	1.391	20	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Impact of ChatGPT on the ratio of active learning in student-ChatGPT interaction based on Tukey's method, indicating a higher level of active learning between student interactions directed to ChatGPT

Comparison	Difference	Adj. p-value
To ChatGPT vs. To students	-0.5357	**0.000**

The statistical results presented in Tables 3 and 4 indicate a significant statistical difference in the level of active learning between student interactions directed towards ChatGPT and those addressed to other students. On average, student-ChatGPT messages exhibited a 53.57 percentage points higher level of active learning compared to student-student interactions. This result suggests that the active learning process within cohorts was becoming more bilateral between students and ChatGPT, even in the presence of other students in the cohort.

Conclusion

Our paper presents preliminary findings from an interesting application of ChatGPT in group brainstorming activities in the context of CSCL. Our statistical analysis quantitatively confirms that ChatGPT diminishes the level of active learning interactions between students, a crucial element for successful collaborative learning. This decrease is primarily attributed to a relatively higher proportion of active learning interactions occurring predominantly between individual students and ChatGPT.

While our initial exploration sheds light on this phenomenon, further analysis is imperative to unravel its underlying causes fully. Moreover, a deeper dive into ChatGPT's influence on the tangible outcomes of group brainstorming sessions remains an essential avenue for future investigation. It's worth noting that our experimental design did not explicitly account for the potential novelty effect of introducing ChatGPT, a factor we aim to address in our future work. Nevertheless, it is essential to continue exploring the potential of generative AI in assisting students in creative thinking-related tasks. It provides a promising educational setting to leverage generative AI's capability without being significantly limited by its hallucination issues.

Supplementary materials

Table 5: Number of students in each cohort

	Type A	Type B
Cohort #1	5	4
Cohort #2	4	5
Cohort #3	5	4
Cohort #4	5	5
Cohort #5	5	5
Cohort #6	5	5
Total	29	28

Table 6: Summary of messages in Type A

Type A			
Cohort number	**Total number of student messages**	**Avg number of words per message**	**Total number of ChatGPT messages**
Cohort #1	35	20.371	13
Cohort #2	83	13.41	28
Cohort #3	92	31.011	10
Cohort #4	115	23.191	34
Cohort #4	38	26.789	18
Cohort #5	46	24.0	28

Table 7: Summary of messages in Type B

Type B		
Cohort number	**Total number of student messages**	**Avg number of words per message**
Cohort #1	25	17.56
Cohort #2	85	14.729
Cohort #3	30	11.767
Cohort #4	61	15.18
Cohort #4	21	29.691
Cohort #5	71	19.239

Table 8: ANOVA results on the level of active learning within all student interactions

Source of Variation	SS	df	F
ChatGPT	0.012	1	0.202
Week	0.080	1	1.289
ChatGPT $\times$ Week	0.032	1	0.515
Residual	1.237	20	

References

Al-Samarraie, H.; and Hurmuzan, S. 2018. A review of brainstorming techniques in higher education. *Thinking Skills and creativity*, 27: 78–91.

Alkaissi, H.; and McFarlane, S. I. 2023. Artificial hallucinations in ChatGPT: implications in scientific writing. *Cureus*, 15(2).

Ansah, D. B.-A. L. O. 2023. Education in the era of Generative Artificial Intelligence (AI): Understanding the potential benefits of CHATGPT in promoting teaching and learning. *SSRN Electronic Journal*.

Bloom, B. S.; Engelhart, M. D.; Furst, E.; Hill, W. H.; and Krathwohl, D. R. 1956. Handbook I: cognitive domain. *New York: David McKay*.

Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1): 37–46.

Egan, A.; Maguire, R.; Christophers, L.; and Rooney, B. 2017. Developing creativity in higher education for 21st century learners: A protocol for a scoping review. *International Journal of educational research*, 82: 21–27.

Fu, K.; Murphy, J.; Yang, M.; Otto, K.; Jensen, D.; and Wood, K. 2015. Design-by-analogy: experimental evaluation of a functional analogy search methodology for concept generation improvement. *Research in Engineering Design*, 26: 77–95.

Hong, H.-Y.; and Chiu, C.-H. 2016. Understanding how students perceive the role of ideas for their knowledge work in a knowledge-building environment. *Australasian Journal of Educational Technology*, 32(1).

Ji, Z.; Lee, N.; Frieske, R.; Yu, T.; Su, D.; Xu, Y.; Ishii, E.; Bang, Y. J.; Madotto, A.; and Fung, P. 2023. Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12): 1–38.

Kropp, R. P.; Stoker, H. W.; and Bashaw, W. 1966. The validation of the taxonomy of educational objectives. *The Journal of Experimental Education*, 34(3): 69–76.

Lo, C. K. 2023. What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Education Sciences*, 13(4).

McHugh, M. L. 2012. Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3): 276–282.

McMurtrie, B. 2023. Teaching: Will ChatGPT change the way you teach. *The Chronicle of Higher Education*.

Qadir, J. 2023. Engineering Education in the Era of ChatGPT: Promise and Pitfalls of Generative AI for Education.

In *2023 IEEE Global Engineering Education Conference (EDUCON)*, 1–9.

Rudolph, J.; Tan, S.; and Tan, S. 2023. ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, 6(1).

Soller, A. 2001. Supporting social interaction in an intelligent collaborative learning system. *International journal of artificial intelligence in education*, 12(1): 40–62.

Tukey, J. W. 1949. Comparing individual means in the analysis of variance. *Biometrics*, 99–114.