Advances in Educational Technology and Psychology (2026)
Clausius Scientific Press, Canada

DOI: 10.23977/aetp.2026.100316
ISSN 2371-9400 Vol. 10 Num. 3

Generative AI as a Teaching Partner: A Quasi-Experimental Study on Its Role in Facilitating Classroom Dialogue and Conceptual Understanding

Zhao Yuqian[a,*], He Jingru[b], Wang Chengbo[c]

School of Education Science, Yan'an University, Yan'an, Shaanxi, 716000, China
[a]3076925834@qq.com, [b]jiabei365@qq.com, [c]wangchengbo2007@sina.com
**Corresponding author*

Keywords: Generative AI; Teaching Partner; Classroom Dialogue; Conceptual Understanding

Abstract: With the rapid advancement of AI technology, generative AI is transitioning from a tool to a collaborative partner in classroom instruction. This study employs a quasi-experimental design to investigate whether integrating generative AI as a teaching partner can effectively enhance the quality of classroom dialogue and thereby promote students' conceptual understanding. A nonequivalent control group pretest-posttest design was used, with the experimental group incorporating an AI teaching partner based on GPT-4 into instruction, engaging in multiple rounds of dialogue and concept clarification, while the control group followed traditional teaching methods. Quantitative results show that the experimental group significantly outperformed the control group in key indicators such as dialogue depth, proportion of higher-order questions, and frequency of joint reasoning ($p < 0.01$). Posttest scores on conceptual understanding were also significantly higher in the experimental group (Cohen's $d = 0.74$). Qualitative analysis further reveals that the AI's immediate feedback and heuristic questioning effectively facilitated students' shift from factual recall to explanatory and argumentative discourse. The study demonstrates that generative AI, when used as a teaching partner, holds substantial potential for enhancing classroom dialogue and deepening conceptual understanding.

1. Introduction

1.1 Research Background and Problem Statement

The application of artificial intelligence in education is undergoing a paradigm shift from content generation tools to intelligent tutoring systems, and finally to collaborative learning partners[1]. Generative AI, especially large language models (LLMs), with their powerful natural language understanding and generation capabilities, can play multiple roles in the classroom, such as question-answering facilitators, concept clarifiers, and discussion summarizers[2]. However, despite the great potential of the technology, there is a lack of systematic empirical research on whether it can effectively enhance the quality of classroom conversations and promote conceptual understanding in real classrooms.

Classroom conversations are the core mechanism of social constructivist learning theory. Vygotsky pointed out that cognitive development occurs within social interaction[3], and students gradually internalize scientific concepts through language exchanges with teachers and peers. However, in traditional classrooms, teachers' questions often focus on lower-order factual levels, and students lack sufficient opportunities for explanation, questioning, and reasoning, resulting in insufficient depth of conversation and superficial understanding of concepts.

In this context, can the intervention of generative AI break this predicament? Can it expand the cognitive boundaries of classroom interactions by providing immediate, personalized, and progressively cognitive-level dialogue support? This study poses the core question: Can generative AI, as a teaching partner, effectively promote classroom conversations and thereby enhance students' conceptual understanding?

1.2 Research Objectives and Significance

This study aims to verify the effectiveness of a generative AI-enabled human-machine collaborative dialogue teaching model through a quasi-experimental design. Its theoretical significance lies in expanding the human-machine collaborative learning theory and enriching the research perspective of dialogue teaching; its practical significance is to provide replicable AI teaching integration solutions for frontline teachers, promoting the transformation of the classroom from knowledge transmission to deep construction.

2. Background

2.1 Evolution of the Role of Generative AI in Education

Early AI education applications focused on content generation or automatic scoring, such as essay correction systems and question bank recommendation engines. With the maturity of LLMs, research began to explore the possibility of using it as an intelligent mentor, such as providing personalized explanations and generating Socratic-style questions[4][5]. In recent years, the concept of teaching partners has emerged, emphasizing that AI should be a collaborator of teachers rather than a substitute, participating in the entire process of classroom interaction[6]. Relevant empirical research is still in its infancy, and there is a lack of in-depth exploration of the causal mechanism between dialogue quality and conceptual understanding.

2.2 Relationship between Classroom Dialogue and Conceptual Understanding

Social constructivism holds that conceptual understanding is not passive acceptance, but rather an active construction in social interaction. Mercer et al. proposed the concept of exploratory talk[7], emphasizing that students gradually deepen their understanding of scientific concepts through proposing hypotheses, providing evidence, and responding to questions in dialogue[8]. Research has found that the characteristics of high-quality dialogue include: high-order questions (such as why if... would it be like this), joint reasoning , integration of perspectives, etc., all of which are significantly positively correlated with deep conceptual understanding[9].

2.3 Previous Studies on Supporting Classroom Dialogue with Technology

Early technology tools such as online discussion boards and collaborative concept mapping software (such as CmapTools) could expand the dialogue space, but they mostly remained at the recording and presentation level, lacking guidance and generation capabilities[10]. The emergence of generative AI has made it possible to dynamically generate dialogue scaffolds: it can adjust the questioning strategy in real time based on students' answers, provide counterexamples, analogies, or follow-ups, thereby promoting cognitive conflicts and concept reconstruction[11]. However, there is currently a lack of rigorous empirical research on embedding AI systems into the actual classroom dialogue process.

3. Research Design and Methods

3.1 Research Hypotheses

H1: The quality of classroom dialogue in the experimental group (measured by indicators such as dialogue depth, proportion of high-order questions, and frequency of joint reasoning) is significantly higher than that in the control group.

H2: The post-test scores of the experimental group for concept understanding are significantly higher than those of the control group, and this effect is mediated by the quality of dialogue.

3.2 Quasi-experimental Design

An unequal control group pretest-posttest design was adopted. Two natural classes from a key middle school in a certain city (experimental group $n = 42$, control group $n = 40$) were selected. The same teacher taught the content, which was the unit on ecosystem energy flow in science. The intervention period was two weeks (a total of 6 classes). The experimental group introduced AI teaching partners during the group discussion, while the control group used the traditional discussion mode.

3.3 Participants and Context

The average age of the students was 13.4 years old. There was no significant difference in their previous science scores ($t = 0.87$, $p = 0.39$). The teaching context was a regular classroom equipped with a network environment. The AI system was deployed through the school intranet to ensure data privacy and response speed.

3.4 Intervention Measures: The Mode of AI Teaching Partners

AI Role Definition: Question Guide, generating progressive questions based on students' answers; Concept Clarifier, providing counterexamples or analogies when students have misunderstandings; Discussion Summarizer, generating concise concept summaries after group discussions and sharing them with the entire class.

Operational Process: The teacher posed the core question; Students discussed in groups of four, each group conducting multiple rounds of dialogue with the AI via tablets; The AI recorded the dialogue content in real time and intervened with questions or clarifications at key nodes; After the discussion, the AI generated a draft of the group concept map, which students could modify and improve; The teacher selected 2-3 groups for a class presentation, and the AI provided summary feedback.

Technical Tools: A customized chatbot developed based on GPT4 API, embedded with subject knowledge graph and question strategy library. The system preset three dialogue modes: progressive revelation of cognitive conflicts, analogy transfer, and others, switching automatically based on students' answers.

3.5 Measurement Tools and Data Collection

This study completed data collection using a multi-dimensional mixed-method approach centered on the core variables. Firstly, for classroom dialogue quality, the entire class was recorded and transcribed, analyzed using the Discourse Depth Level (DDL, 5-level coding) scale, and the average depth, proportion of high-order questions (Level-45), and frequency of joint reasoning (the number of times two or more students engaged in reasoning around the same argument) were calculated for each class. At the qualitative level, three typical dialogue segments from the experimental and control groups were selected for micro-interaction analysis, focusing on the specific paths through which AI triggered cognitive conflicts and promoted reasoning extension in the classroom.

Secondly, the measurement of students' concept understanding levels: On one hand, a standardized concept test paper was used for pretest and posttest, with 20 questions covering three levels: knowledge memory, application, and transfer. The internal consistency reliability $\alpha = 0.83$ was verified; On the other hand, a concept map task was used, requiring students to draw a concept map on the theme of ecosystem energy flow. The score was rated using the "complexity-association" dual-dimensional scoring method (Full score of 30 points), and the scores were independently completed by two researchers, with inter-rater reliability $\kappa = 0.91$.

Finally, two types of supplementary data were added: Randomly selected 12 students from the experimental group for semi-structured interviews to understand their perception and emotional experience of the AI collaborative partner;At the same time,the teaching teacher was required to write a reflection log after each class, recording the dynamic changes in the classroom process and details of students' classroom performance.

3.6 Data Analysis Methods

This study employed a mixed research paradigm combining quantitative and qualitative methods for data analysis. The two methods complemented and cross-validated each other. The specific operational path is as follows: At the quantitative analysis level, ANCOVA is used to control the pre-test scores and compare the post-test differences between groups; independent sample t-tests are conducted for the dialogue indicators; the PROCESS macro is employed to test the mediating effect of dialogue quality.

At the qualitative analysis level, thematic analysis is performed on the interviews and dialogue segments, extracting core categories such as the progressive nature of student cognitive conflicts in AI questioning, etc., and conducting triangulation verification with the quantitative results.

4. Research Results

4.1 Analysis of Classroom Dialogue Quality

Quantitative results (Table 1) show that the experimental group had significantly higher dialogue depth ($M = 3.72$, $SD = 0.45$) compared to the control group ($M = 2.89$, $SD = 0.51$), with a t-value of 7.83 and a p-value less than 0.001; the proportion of high-level questions increased by 42%; the frequency of joint reasoning per class reached 8.3 times on average, while the control group had only 3.1 times.

In the qualitative analysis, a typical dialogue from the experimental group is as follows:

Student A: Energy is transferred from plants to herbivores, and then to carnivores.

AI: What role do decomposers play in this process? Where does the energy they obtain come from?

Student B: Decomposers eat dead organisms... So they also obtain energy.

AI: Very good. Then, if decomposers suddenly disappear in a certain ecosystem, how would the energy flow change?

Student C: Maybe the energy would get stuck in the dead organisms and not be recycled...

AI: Does this mean that the energy flow would stop? Or would it continue in other ways?

This segment shows that the AI, through progressive questioning, guides the students to shift from a linear transmission perspective to a circular perspective, achieving a conceptual reconfiguration.

Table 1: Comparison of Quantitative Indicators for Dialogue Quality (M ± SD).

Indicators	Experimental group (n=42)	Control group (n=40)	t	Cohen's d	p
Dialogue depth score (1-5)	3.72±0.45	2.89±0.51	7.83	1.09	<.001
Proportion of high-order questions (%)	48.6±7.2	34.2±6.5	9.41	1.32	<.001
Frequency of joint reasoning per class	8.3±1.6	3.1±1.2	16.20	2.02	<.001

4.2 Analysis Results of Conceptual Understanding Level

The ANCOVA results (Table 2) show that after controlling for the pre-test scores, the post-test scores of the experimental group were significantly higher than those of the control group ($F(1, 79) = 24.63$, $p < 0.001$, $\eta^2 = 0.24$). In the concept map scoring, the experimental group showed a particularly significant improvement in the dimension of correlation. More students were able to correctly establish the bidirectional connections between the energy decomposers cycles.

Table 2: Pre-test and post-test scores for concept understanding and concept mapping (ANCOVA controls pre-test)

Variables	Post-test adjusted mean (SE) of the experimental group	Post-test adjusted mean (SE) of the control group	F(1,79)	η^2	p
Concept test score (0-100)	78.4(1.5)	68.7(1.6)	24.63	.24	<.001
Total score of concept map (0-30)	22.8(0.8)	18.3(0.8)	15.71	.17	<.001
Among them: Dimension of Relevance	9.6(0.4)	7.2(0.4)	19.40	.20	<.001

4.3 Supplementary Findings

During the student interviews, 85% of the students in the experimental group stated that the AI questions made them think more deeply, but some students also mentioned that sometimes the AI questions were too difficult and they would get stuck. The teacher reflection log pointed out that the intervention of AI shifted the teacher's role from a knowledge provider to a dialogue orchestrator, requiring more precise timing of the intervention.

5. Discussion

5.1 Interpretation of Main Findings

This study confirmed that generative AI, as a teaching partner, can effectively enhance the quality of classroom dialogue by providing cognitive scaffolds and extending dialogues, thereby promoting conceptual understanding. The immediacy and personalization features of AI broke through the bottleneck in traditional classrooms where teachers could not respond to all groups simultaneously, enabling each student to receive support within the zone of proximal development.

5.2 Theoretical Implications

The research expanded the human-machine collaborative cognitive model: AI is not only an information provider but also a co-constructor of the dialogue. Through generating cognitive conflicts and analogical prompts, it activates students' metacognitive monitoring and promotes conceptual reconstruction. Additionally, the research also enriched the dialogue learning theory, revealing how technology is embedded in social interaction and changes the discourse power structure (students shift from passive responses to active questioning).

5.3 Practical Suggestions

For teachers: Design core question chains to guide AI to generate progressive questions around key concepts; Train students on how to interact with AI, such as learning to ask questions, doubt, and integrate AI viewpoints; Teachers themselves need to transform their roles, from answer givers to dialogue designers and reflection promoters.

For developers: Enhance AI's subject-specific teaching knowledge (PCK), embed common misunderstandings and counterexamples of specific concepts in the library; Develop dialogue dashboards to visually represent the depth of each group's dialogue in real time, assisting teachers in decision-making; Add an emotional perception module to avoid AI questions being too difficult, causing students to feel frustrated.

5.4 Research Limitations and Future Directions

This study has the following limitations: First, the sample size is small, with only two classes selected for the controlled experiment; second, the applicable subject dimension is single, focusing only on the science class scenario for teaching intervention; third, the intervention duration is limited, with the overall teaching practice period being only two weeks; fourth, the used AI system is a task-customized version, and its universality in different teaching scenarios needs to be verified.

Based on these limitations, this study needs to further research in the following aspects in the future: Verify the adaptability of the AI dialogue model in different humanities and social science disciplines; Track the long-term longitudinal changes in students' higher-order thinking abilities and learning emotional attitudes under a longer intervention period; Compare the differential impacts of various AI interaction modes on classroom interaction experiences; Establish a long-term ecological mechanism for the collaboration of AI, teachers, and students.

6. Conclusion

This study, through a rigorous quasi-experimental design, systematically verified the significant efficacy of generative AI as a teaching partner in improving classroom dialogue quality and conceptual understanding. The research found that AI not only generates high-quality questions and feedback but also plays the role of a cognitive activator in social interaction, promoting students' transition from passive acceptance to deep construction. This achievement provides key empirical support for the theoretical construction and practical promotion of the human-machine collaborative teaching model, and also indicates the transformative potential of generative AI in innovating classroom interaction and moving towards deep learning. In the future, we call on more researchers and educators to join hands to explore the unlimited possibilities of AI empowerment in education in a more responsible and creative way.

References

[1] Roll, I., & Wylie, R. (2016). Evolution and Revolution in Artificial Intelligence in Education. International Journal of Artificial Intelligence in Education, 26(2), 582–599.

[2] Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274.

[3] Liang Aimin. Research on the Construction and Application of Social Constructivist Learning Theory from Vygotsky's Perspective of Psychological Development [J]. Shandong Foreign Language Teaching, 2011(3).

[4] Jayaweera, C., & Dorr, B. J.(2026). BLADE: Better Language Answers through Dialogue and Explanations. arXiv:2604.03236.

[5] Li Haifeng, Wang Wei. Investigation of Human-Machine Debate Method: An Exploration of a Debate-Based Intelligent Conversation Robot Support Model for Cultivating Students' Higher-order Thinking Skills [J]. Research in Educational Telecommunications, 2024.

[6] Tan, S., & Wu, Y.(2025). Generative AI as a Dialogic Partner: Enhancing Multiple Perspectives, Reasoning, and Argumentation in Science Education with Customized Chatbots. Journal of Science Education and Technology.

[7] Filatova, A., Didenko, A., & Ravedovskaya, U.(2025). From Lecturer to Chatbot: Evaluating a Hybrid Teaching Model with GPT-4 as Co-facilitator. Journal of Teaching English for Specific and Academic Purposes, 13(4), 529–538.

[8] Zhong Qiquan. "A Preliminary Discussion on Classroom Discourse Analysis" [J]. Global Education Outlook, 2013, 42(11): 10-20.

[9] El Fathi, T., Saad, A., Larhzil, H., Lamri, D., & Al Ibrahmi, E. M.(2025). Integrating generative AI into STEM education: enhancing conceptual understanding, addressing misconceptions, and assessing student acceptance. Disciplinary and Interdisciplinary Science Education Research, 7, 6.

[10] Zhang Xuefu. Research on Knowledge Model Organization and Collaboration Mechanism Based on Concept Maps - Taking CmapTools as an Example [J]. Modern Library and Information Technology, 2005, 21(11): 34-38.

[11] Yuan, S., Liu, M., Huang, X., & Zhai, X. (2024). Generative AI as a Socratic tutor: Scaffolding student reasoning through adaptive dialogue in science classrooms. Computers & Education: Artificial Intelligence, 6, 100212. https://doi.org/10.1016/j.caeai.2024.100212