Comparing Supporter, Arguer, and Comprehensive Debater: How AI Debate Agent Roles Influence Critical Thinking

Qingtang Liu[1], Lina Qin[1*], Ruiqiu Ye[1], Yunfan Zuo[1]
[1]Central China Normal University
[*]qinlina@mails.ccnu.edu.cn

Abstract: *AI debate agents show promise for cultivating critical thinking through dynamic dialogue, yet the influence of their conversational roles remains underexplored. This randomized controlled trial investigated how three distinct roles—supporter, arguer, and comprehensive debater—affect learners' critical thinking. Results showed that all three roles similarly enhanced self-perceived critical thinking but did not improve the quality of written arguments. These findings indicate that a single interaction with an AI debate agent can boost learners' confidence in their thinking skills, regardless of the agent's role. However, fostering long-term argumentation skills requires structured, iterative feedback mechanisms that address the gap between self-perception and actual ability.*

Keywords: critical thinking; pedagogical agent; debate learning

1. Introduction

The integration of artificial intelligence with education is transforming the cultivation of higher-order skills such as critical thinking. Critical thinking, with argumentation as a core component, requires learners to formulate claims, organize evidence, and evaluate viewpoints (Giri & Paily, 2020). Technology-enhanced environments have shown potential by providing argumentation scaffolds (Huang et al., 2016), yet how to stimulate deep critical thinking through real-time, dynamic conversational interactions remains underdeveloped.

Pedagogical agents based on large language models (LLMs) have pioneered adaptive, personalized debate dialogues (Winkler & Söllner, 2018). Park and Seo (2025) developed a multi-agent LLM-based debate chatbot that achieved strong alignment with human evaluators in assessing critical thinking. Concurrently, Tang et al. (2025) showed that generative AI can serve as an effective dialogic partner, supporting students' reasoning and argumentation skills. Research has shown that chatbot-assisted debates can enhance students' argumentation skills and motivation (Guo et al., 2023). However, the critical design variable of how an AI agent's systematic, real-time conversational role influences critical thinking remains under-explored.

To address this gap, this study investigates the differential effects of three distinct AI conversational roles—supporter, arguer, and comprehensive debater—on learners' critical thinking. Specifically, this study seeks to answer:

RQ1: Does interaction with an AI debate agent improve self-perceived critical thinking, and does the magnitude of improvement vary by agent role (Supporter, Arguer, Comprehensive Debater)?

RQ2: Does such interaction improve learners' written argumentation quality, and does this effect vary by agent role?

2. Methods

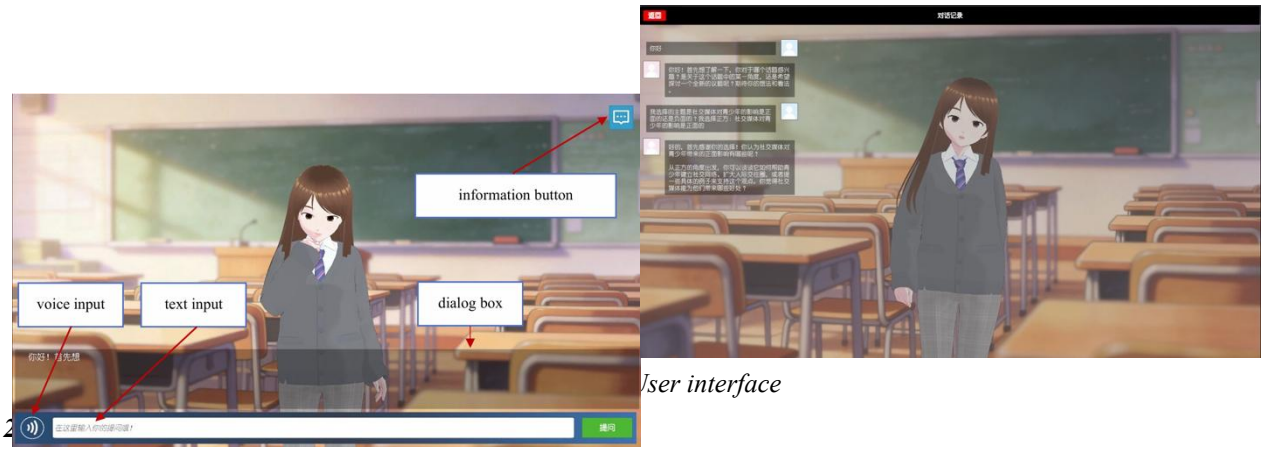
2.1. Participants

Thirty university students (7 males, 23 females, aged 18-28) were recruited. All volunteered, provided informed consent, and received an honorarium. They were randomly assigned to one of three groups (n = 10 each): Supporter, Arguer, and Comprehensive Debater.

2.2. The AI Debate Agent System and Role Design

We developed an AI debate agent system built on the Unity engine, powered by the Qwen2.5-3B LLM and integrated with Microsoft Azure Speech Services. It features an embodied character with synchronized speech, lip movements, and body language for enhanced social presence and immersion. All debate dialogues are systematically recorded (Fig.1).

The three distinct agent roles were designed using the CRISPE prompt framework (Liu et al., 2025): Supporter bolsters the participant's statements with examples or theoretical perspectives; Arguer challenges the participant's statements through rebuttals or contradictory evidence; Comprehensive Debater dynamically switches between supporting and arguing roles based on dialogue context.



2 *User interface*

This study employed a randomized pretest-posttest intergroup design. The agent role was the independent variable, with argumentative writing scores and self-assessed critical thinking as dependent variables. The experiment was conducted in a single 50-minute session (Fig.2). The procedure consists of four phases: (1) Learning Phase: Participants studied materials on argument structure and the Toulmin model (El Majidi, Janssen, & de Graaff, 2021) to ensure a common understanding of debate frameworks, (2) Pre-test Phase: Participants completed a demographic questionnaire and a 5-item Critical Thinking Disposition Scale. They then wrote a 300-word argumentative essay on a pre-assigned topic, (3) Dialogue Phase: Participants engaged in a debate with their assigned AI agent. They chose a topic from five predefined social issues, all balanced for complexity. The conversation continued until the participant signaled its end. The system recorded all interactions, (4) Post-test Phase: Participants completed a 12-item Critical Thinking Self-Assessment Scale and wrote a second argumentative essay on a different, counterbalanced topic.

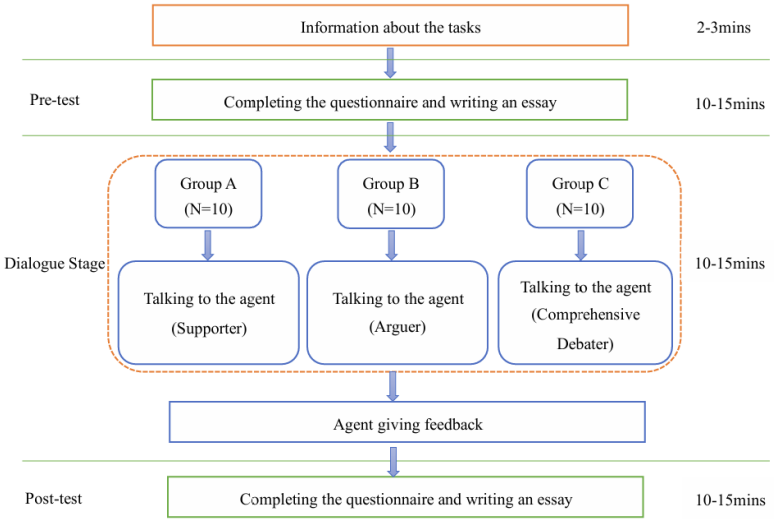


Fig.2. The experimental procedure

2.4. Measuring Tools

Three instruments were used: (1) Argumentative Essay Writing Task adapted from GRE "Analyze an Issue" tasks for pre- and post-tests with counterbalanced topics; (2) Critical Thinking Disposition Scale (pre-test, 5 items) measuring

baseline disposition; (3) Critical Thinking Self-Assessment Scale (post-test, 12 items) measuring self-perceived skills. All self-report scales employed a 5-point Likert scale.

Essay quality was assessed by two blinded raters using a 5-point scale across seven dimensions (organization, adequacy, elaboration, etc.) adapted from El Majidi et al. (2021). Inter-rater reliability was good (ICC = 0.77 and 0.78 for pre- and post-test).

2.5. Data analysis

Due to the small sample size (n = 10 per group), nonparametric tests were used in SPSS 27.0. Kruskal-Wallis H tests examined between-group differences at baseline and post-test. Wilcoxon signed-rank tests compared within-group changes from pre- to post-test. The significance level was set at $p < .05$.

3. Results

Baseline homogeneity was confirmed: no significant differences among groups in critical thinking disposition (χ^2 (2) = 0.25, p = .881) or argumentative writing scores (χ^2 (2) = 0.65, p = .722).

3.1. The Impact of Different AI Agent Roles on Self-Perceived Critical Thinking (RQ1)

First, Kruskal-Wallis H tests revealed no statistically significant differences among the three agent roles on any post-test outcomes. This includes self-perceived critical thinking and argumentative writing scores (all $p > .05$). Complete results are presented in Table 1.

Second, Wilcoxon signed-rank tests on the full sample indicated that interaction with an AI debate agent led to a significant overall enhancement in learners' self-perceived critical thinking (Z = -4.79, $p < .001$; see Table 2).

Table 1. Enhancement in Self-Perceived Critical Thinking and the Effect of Agent Roles

Dimension	M ± SD	χ^2 (df = 2)	*p*
Self-Perceived Critical Thinking	48.33 ± 5.89	3.95	0.139
Argumentative Writing	23.50 ± 2.90	1.09	0.579

3.2. The Impact of Different AI Agent Roles on Argumentative Writing Quality (RQ2)

In contrast to the positive subjective outcomes, the interaction did not lead to a significant improvement in objective argumentative writing quality at the overall level (p = .925; see Table 2). Furthermore, consistent with RQ1 results, no differential effect by agent role was found for writing quality (p = .579).

Table 2. Wilcoxon Signed-Rank Test Results for Key Variables (Within-Subject Comparisons, Full Sample)

Dimension	Pre-test (M ± SD)	Post-test (M ± SD)	Z	*p*
Self-Perceived Critical Thinking	18.83 ± 2.77	48.33 ± 5.89	-4.79	<.001
Argumentative Writing	23.48 ± 2.77	23.50 ± 2.90	-0.09	0.925

4. Discussion and Conclusion

This study offers two key insights. First, a single interaction with any AI debate agent significantly enhanced learners' self-perceived critical thinking, with no significant differences among the three roles. This highlights AI-driven dialogue's potential to boost learner confidence and engagement in debate contexts.

Second, despite gains in self-perception, no improvement was observed in objective argumentation quality. This divergence underscores that complex argumentation skills require prolonged, iterative practice (Kellogg & Raulerson, 2007). It also reveals a critical gap between confidence and actual proficiency, suggesting self-report measures alone are insufficient for evaluating skill development.

From a theoretical perspective, these findings suggest that the pedagogical role of an AI agent—whether supportive, challenging, or adaptive—may not produce differential effects in short-term interactions, but such effects could emerge

over extended engagement. From a practical standpoint, for single-session activities aimed at enhancing engagement, any AI debate role proves effective; however, fostering long-term argumentation skill development requires multi-iteration interventions with structured writing feedback.

This study has limitations, including its single-intervention design, small sample, and reliance on self-reported data. Future research should: (1) adopt longitudinal designs to explore cumulative effects of different roles, (2) analyze dialogue processes combined with physiological data to understand cognitive and emotional influences, and (3) examine matching effects between learner individual differences and AI roles.

Acknowledgements

This research was supported by the New Engineering New Form Textbook Construction Project of Hubei Provincial Department of Education (No. XGK02019) and the China University Industry-University-Research Innovation Fund (No. 2024XL009).

References

El Majidi, A., Janssen, D., & de Graaff, R. (2021). The effects of in-class debates on argumentation skills in second language education. System, 101, 102576.

Giri, V., & Paily, M. U. (2020). Effect of scientific argumentation on the development of critical thinking. Science & Education, 29, 673–690.

Guo, K., Zhong, Y., Li, D., & Chu, S. K. W. (2023). Effects of chatbot-assisted in-class debates on students' argumentation skills and task motivation. Computers & Education, 203, Article 104862.

Huang, C. J., Chang, S. C., Chen, H. M., Tseng, J. H., & Chien, S. Y. (2016). A group intelligence-based asynchronous argumentation learning-assistance platform. Interactive Learning Environments, 24(7), 1408–1427.

Kellogg, R. T., & Raulerson, B. A. (2007). Improving the writing skills of college students. Psychonomic Bulletin & Review, 14, 237–242.

Liu, Z. S., Quan, L., Wang, Y. G., et al. (2025). Impact of prompt frameworks on user search behavior in conversational search. Information Studies: Theory & Application, 48(11), 28–40. (in Chinese)

Park, B., & Seo, K. (2025). Assessing critical thinking through a multi-agent LLM-based debate chatbot. In Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25), Article 521. New York, NY: ACM.

Tang, K. S., Cooper, G., & Kendeou, P. (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, 34(2), 215–232.

Winkler, R., & Söllner, M. (2018). Unleashing the potential of chatbots in education: A state-of-the-art analysis. In Academy of Management Proceedings (Vol. 2018, No. 1, p. 15903). Briarcliff Manor, NY: Academy of Management.