Socratic vs. Expert Agents: Conversational style effects on Programming Performance, Learning Experience, and Higher-Order Thinking

Shiqing Peng [1], Xueying Tao [2*], Heng Luo [3]
[1.2.3] Faculty of Artificial Intelligence in Education, Central China Normal University
[*] pengshiqing@mails.ccnu.edu.cn

Abstract: *This study examined how educational agent conversational styles influence seventh graders' programming learning. In a between-subjects experiment, 82 students were randomly assigned to learn with either a Socratic Agent or an Expert Agent on the AI agent platform. Results showed that: (1) Students using the Socratic Agent demonstrated greater performance improvement on coding tasks but not on objective questions, indicating that agent effects vary by task type; (2) The Socratic Agent group reported significantly higher emotional engagement and behavioral engagement, while no differences emerged on other learning experience dimensions; (3) Agent interaction style significantly influenced problem-solving ability and creativity after controlling for pre-test levels, with the Socratic Agent showing advantages; (4) Both agent styles benefited students across high and low prior-coding skills, with no significant differences between groups. The study provides empirical evidence for designing and applying educational agents in programming education.*
Keywords: Educational Agent; Conversational Style; Higher-Order Thinking; Programming Education; Learning Experience

1. Introduction

Higher-order thinking has become the core mission of programming education in the digital transformation era, serving as a key competency for junior high school students to adapt to digital environments and achieve sustainable development (Lee et al., 2024a). However, traditional programming lessons often emphasize syntax over thinking, making it difficult for teachers to meet higher-order thinking cultivation needs (Patrick & Campbell, 2025). The integration of artificial intelligence offers a new pathway where educational agents can externalize students' implicit programming thinking through real-time natural language interaction, building a bridge for personalized cultivation (Ma et al., 2024). The effectiveness of these agents depends not only on their presence but also on how they interact with students. Among design features, conversational style is particularly important because it shapes students' thinking and engagement. In this study, we focus on two representative styles: Socratic Agent and Expert Agent.

A Socratic Agent primarily supports learning through guided questioning rather than direct answers. Its core function is to lead students to explain their ideas, reflect on reasoning, and refine solutions step by step. In current educational practice, this style is widely applied in inquiry-based learning, reflective discussions, and guided problem solving. For example, a Socratic Agent may ask why a program produces a certain output, encourage comparisons of coding strategies, or prompt students to revise solutions after noticing errors. In this mode, the agent acts as a questioning partner that promotes active thinking. Previous studies suggest that this style can help learners stay cognitively active and support the development of self-regulated learning and higher-order thinking (Almelweth, 2022; Lee et al., 2024b).

In contrast, an Expert Agent mainly supports learning by providing direct explanations and immediate guidance. Its main feature is to provide clear and efficient support when students need quick guidance. In current educational practice, this style is often used in direct instruction, example-based learning, and task-oriented tutoring. For example, an Expert Agent may explain a programming concept in simple terms, provide a worked example of code, identify errors in a student’s program, or give direct suggestions for revision. In this teaching mode, the agent works more like a tutor or subject expert that helps students complete tasks efficiently. Studies have shown that this style is useful for knowledge acquisition, error correction, and learning support, especially when learners need timely help to continue a task (Farrokhnia et al., 2024).

The main difference between the two conversational styles lies in whether guidance is delivered through questions or direct explanations. This difference may influence students' programming performance, learning experience, and higher-order thinking. Most existing studies have focused on task performance or learning interest, but few have examined how Socratic and Expert conversational styles affect specific dimensions of higher-order thinking or whether task difficulty moderates these effects (Putra et al., 2023; Liu et al., 2023). Understanding these effects is critical for designing effective AI-supported programming learning environments.

To address this gap, this study examines the effects of Socratic and Expert conversational styles in junior high school programming education. A total of 82 seventh-grade students were randomly assigned to learn C++ programming with either a Socratic Agent or an Expert Agent. The research questions are as follows:

RQ1: Do the two types of educational agents have differential effects on overall programming performance and on performance across tasks of varying difficulty?

RQ2: Do the two types of educational agents differentially affect the various dimensions of students' learning experience?

RQ3: Do the two types of educational agents differentially affect the various dimensions of students' higher-order thinking?

2. Research Methods

2.1. Research Participants

The study involved seventh-grade students from a middle school in China. A total of 82 students (42 males and 40 females), aged 12 to 13 years, participated in the study. Based on pre-test programming scores, students were divided into two classes of 41 each to ensure homogeneity between groups before the experiment.

2.2. Research Design and Procedure

2.2.1. Agent Development and Optimization

Two programming teaching agents with distinct conversational styles were developed on the Coze platform. Both agents shared a unified C++ knowledge base covering grammar rules and key programming concepts. The Socratic Agent employed a graded guidance approach, using progressive questioning to help students identify errors and construct solutions without directly providing answers. The Expert Agent focused on efficient problem-solving, delivering error localization, correction solutions, and knowledge explanations directly. Both agents underwent multiple testing rounds, and they were finalized once average teacher ratings reached ≥4.0. The key difference between the two styles lies in whether guidance is delivered through questions or direct explanations. Typical interaction examples for the Socratic Agent and Expert Agent are shown in Figure 1 and Figure 2, respectively.

Figure 1

Example interaction with the Socratic Agent

Figure 2

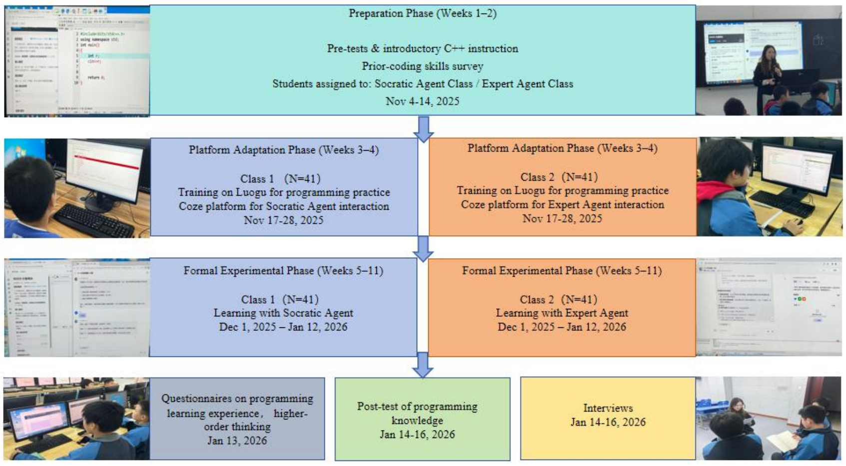
Example interaction with the Expert Agent

2.2.2. Experimental Procedure

The experiment was conducted from November 4, 2025, to January 16, 2026, in three phases. The same teacher delivered all instruction, and both classes followed identical schedules and content. The preparation phase (Weeks 1–2) included pre-tests and introductory C++ instruction. During the platform adaptation phase (Weeks 3–4), students practiced on Luogu and interacted with their assigned agent on Coze. The formal experimental phase (Weeks 5–11) involved learning programming with the assigned agent. Questionnaires were collected at completion, followed by interviews from January 14–16, 2026. Figure 3 outlines the overall procedure.

Figure 3

Experimental design and study procedure



2.3. Data Collection

Data were collected through pre-test and post-test programming tests and online questionnaires. The students' learning performance is measured through programming tests which comprised objective questions (multiple-choice and fill-in-the-blank) on basic grammar and control structures, as well as coding tasks assessing practical coding skills. The questionnaire included three sections. The first section of online questionnaires collected basic student information. The second section, the Learning Experience Scale (Davis, 1989; Gunuc & Kuzu, 2015), measured learning interest, self-efficacy, emotional engagement, behavioral engagement, cognitive engagement, perceived usefulness, and perceived ease of use (4 items each). The third section, the Higher-Order Thinking Scale (Hwang et al., 2018), assessed problem-solving ability (4 items), critical thinking (4 items), and creativity (3 items). All items used a 5-point Likert scale and were adapted from established studies. Reliability and validity were confirmed with Cronbach's alpha for Higher-Order Thinking at 0.962 and for Learning Experience at 0.846. KMO values were 0.874 and 0.752, with significant Bartlett's tests ($p < 0.001$), indicating good construct validity.

2.4. Data Analysis

A series of statistical analyses were conducted to examine the effects of agent conversational style on students' programming performance, learning experience, and higher-order thinking. The independent variables were agent interaction style and task type. The dependent variables were programming performance, learning experience, and higher-order thinking. Statistical analyses were performed using SPSS 26.0, with effect sizes (partial η^2) reported. Independent samples t-tests examined group differences in performance improvement. To further examine the effects of agent interaction style on each dependent variable, three sets of analyses were conducted: ANCOVA for different task types controlling for pre-test scores, Mann-Whitney U tests for learning experience dimensions, and one-way ANCOVA for higher-order thinking dimensions controlling for pre-test differences.

3. Results

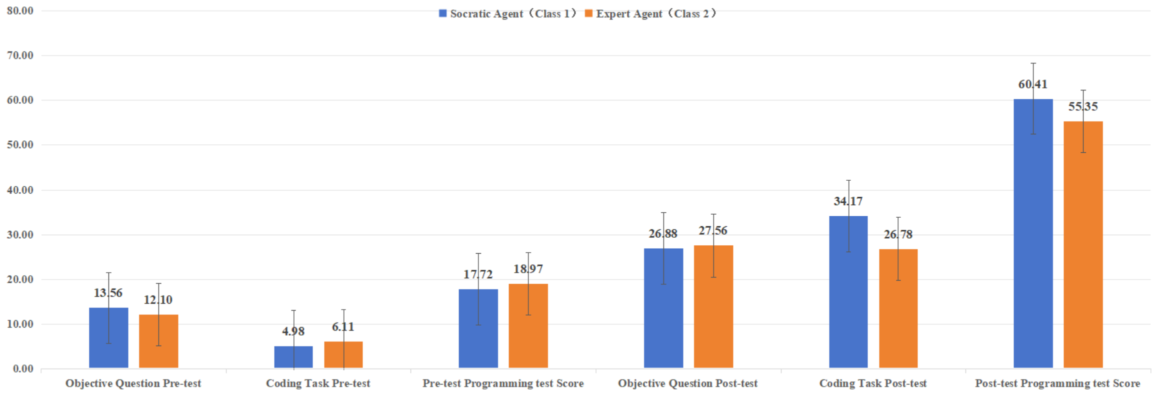
3.1. Results Analysis

3.1.1. Impact of Different Agent Conversational Styles on Students' Programming Performance

To answer RQ1, independent samples t-tests were conducted to compare performance improvement between the two groups. As shown in Figure 4, students in the Socratic Agent group showed slightly greater performance improvement than those in the Expert Agent group. However, the t-test results indicated that this difference was not statistically significant ($p > .05$), and no significant differences in gain scores were found between the two groups.

Figure 4

Comparison of Pre-test and Post-test Scores Across Task Types and Agent Groups



Further ANCOVAs examined the effects of agent conversational style on different task types, controlling for pre-test scores. As shown in Table 1, after controlling for pre-test scores, the agent style significantly affected coding task scores but not objective question scores. Although both groups improved from pre-test to post-test, the Socratic Agent group showed substantially greater gains on coding tasks (from 4.98 to 34.17) compared with the Expert Agent group (from 6.11 to 26.78). In contrast, both groups showed similar improvements on objective questions, with the Socratic Agent group increasing from 13.56 to 26.88 and the Expert Agent group from 12.10 to 27.56 (see Figure 1).

Table 1.

Summary of ANCOVA for the Effect of Agent Style on Objective Question and Coding Task Scores

Dependent Variable	F	P	Partial η^2	R^2
Objective Question Score	0.820	0.368	0.010	0.122
Coding Task Score	9.403	0.003**	0.106	0.285

a. ***p < 0.001; **p < 0.01; *p < 0.05

3.1.2. Effects on Students' Learning Experience

To answer RQ2, Mann–Whitney U tests were conducted to compare learning experience dimensions between the two groups. As shown in Table 2, the Socratic Agent group scored significantly higher on emotional engagement ($p = 0.041$) and behavioral engagement ($p = 0.032$). No significant differences were observed in other dimensions, suggesting that agent conversational style primarily influences emotional and behavioral engagement rather than students' subjective perceptions.

Table 2.

Mann-Whitney U Test Results for Learning Experience Dimensions by Agent Style

Learning Experience Dimension	Agent Style	M	SD	Z	P	Effect Size (r)
Learning Interest	Socratic Agent	4.043	1.116	-1.182	0.237	0.130
	Expert Agent	3.890	0.880			
Self-Efficacy	Socratic Agent	3.302	0.704	-0.181	0.856	0.020
	Expert Agent	3.353	0.583			
Perceived Usefulness	Socratic Agent	4.220	0.844	-1.022	0.307	0.110
	Expert Agent	3.988	0.963			
Perceived Ease of Use	Socratic Agent	4.183	0.754	-1.157	0.247	0.130
	Expert Agent	3.963	0.809			
Cognitive Engagement	Socratic Agent	4.098	0.836	-1.481	0.138	0.160
	Expert Agent	3.756	0.954			
Emotional Engagement	Socratic Agent	4.238	0.707	-2.048	0.041*	0.230

Learning Experience Dimension	Agent Style	M	SD	Z	P	Effect Size (r)
	Expert Agent	3.799	0.932			
Behavioral Engagement	Socratic Agent	4.055	0.770	-2.144	0.032*	0.240
	Expert Agent	3.585	0.921			

a. ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

3.1.3. Effects on Students' Higher-Order Thinking

To answer RQ3, higher-order thinking dimensions were compared between the two agent styles. As shown in Table 3, the Socratic Agent achieved higher adjusted means on problem-solving ability and creativity compared with the Expert Agent, with significant differences for problem-solving ($p = 0.031$) and creativity ($p = 0.049$). For critical thinking, the Socratic Agent scored slightly higher, but the difference was not statistically significant ($p = 0.166$).

Table 3.

Summary of ANCOVA for the Effect of Agent Style on Higher-Order Thinking Dimensions

Higher-Order Thinking Dimension	Agent Style	M	SD	F	P	Partial η^2
Problem-Solving Ability	Socratic Agent	4.140	0.835	4.803	0.031*	0.057
	Expert Agent	3.760	0.941			
Critical Thinking	Socratic Agent	4.010	0.756	1.959	0.166	0.024
	Expert Agent	3.750	0.871			
Creativity	Socratic Agent	4.102	0.837	3.988	0.049*	0.048
	Expert Agent	3.768	0.893			

3.1.4. Differences Across Subgroups with Different Prior-coding Skills

To examine whether the effects of agent conversational style differed by prior coding skills, participants were divided into high and low initial ability subgroups based on the pre-test score median. Mann–Whitney U tests conducted within each subgroup revealed no significant differences between the Socratic and Expert Agent groups ($p > 0.05$), indicating that both agent styles had comparable effects regardless of prior-coding skills. These findings suggest that neither style widens the achievement gap and that both can be effectively applied in mixed-ability classrooms.

4. Discussion

4.1. Matching Agent Style with Task Type

The Socratic Agent outperformed the Expert Agent on coding tasks but not on objective questions, highlighting that the effect of agent conversational style depends on task type. For well-structured objective questions, direct answers from Expert Agents reduce cognitive load, whereas ill-structured coding tasks benefit from guided scaffolding that helps students construct solutions independently (Akçapınar & Sidan, 2024; Ma et al., 2024; Dan et al., 2024). In this study, students using the Socratic Agent showed substantially greater gains on coding tasks, while both groups improved similarly on objective questions. These findings indicate that neither agent style is inherently superior. It is more important to match the agent to the task. In practice, instructors could use Expert Agents for foundational concepts such as syntax and data types, and Socratic Agents for problem-solving, debugging, and algorithm design, combining the efficiency of direct guidance with the depth of guided reasoning.

4.2. Socratic Agent Improves Learning Experience and Higher-Order Thinking

The Socratic Agent enhanced students' emotional engagement, behavioral engagement, problem-solving ability, and creativity. Emotional engagement increased because dialogue simulates teacher-student interaction (Lin et al., 2025), and behavioral engagement improved as inquiry-based interaction requires active thinking without providing direct answers (Al Mamun & Lawrie, 2023). Problem-solving ability benefited from iterative questioning cycles (Du et al., 2025), and creativity was stimulated through divergent thinking prompts (Lin et al., 2026). Critical thinking showed a non-significant trend, likely due to the short intervention and students' developmental stage (Hwang et al., 2018). These results suggest that Socratic Agents can be effectively used to foster deeper engagement and higher-order thinking in programming courses.

4.3. Equity of Agent Effects Across Prior-Coding Skills

Analysis revealed no significant differences in the effects of the two agent styles between high and low ability students. This finding indicates that both approaches are equitable across learner populations, addressing concerns about technology widening achievement gaps (Huang et al., 2024). Future intelligent tutoring systems could build on this equity by incorporating adaptive mechanisms that adjust interaction styles based on real-time student performance (Lee et al., 2024a). For example, systems might employ more directive guidance when students struggle with basic concepts, then shift to inquiry-based approaches when they demonstrate readiness for deeper challenges. Such dynamic adaptation could further optimize learning outcomes while preserving the fairness demonstrated in the current study.

4.4. Theoretical and Practical Implications

This study offers theoretical and practical implications for programming education. Theoretically, it extends prior work by showing that agent interaction effects depend on task type, validates Socratic pedagogy in human-agent learning contexts at the junior secondary level, and demonstrates how Socratic Agents can support educational equity across prior coding skills. Practically, instructors could use Expert Agents to teach foundational concepts (e.g., syntax, data types) and Socratic Agents with graded questioning to foster problem-solving during debugging or algorithm design. Matching agent style to task demands and learner needs, rather than adopting a single style, may better promote engagement and higher-order thinking. Such integration could help shift programming education toward a greater focus on cultivating higher-order thinking skills.

5. Limitation and Future Direction

This study has several limitations due to its design and conditions. The short experimental period may not fully capture students' higher-order thinking development. The single-school sample limits generalizability. Future research should use multimodal data (e.g., eye-tracking, EEG, epistemic network analysis) to more precisely examine cognitive processes, and adopt longitudinal designs to explore long-term effects, while further optimizing the Socratic Agent's questioning and adaptability for broader application in programming education.

Acknowledgments

This paper presents the research findings of the 2024 Hunan Provincial Educational Information Technology Research Project entitled “Strategies for Improving Computer Programming Ability among Rural Primary and Secondary School Students” (Project No.: HNETR24055), funded by the Department of Education of Hunan Province.

References

Akçapınar, G., & Sidan, E. (2024). AI chatbots in programming education: guiding success or encouraging plagiarism. *Discover Artificial Intelligence*, 4(1), 87.

Almelweth, H. (2022). The effectiveness of a proposed strategy for teaching geography through artificial intelligence applications in developing secondary school students’ higher-order thinking skills and achievement. *Pegem Journal of Education and Instruction*, 12(3), 169-176.

Al Mamun, M. A., & Lawrie, G. (2023). Student-content interactions: Exploring behavioural engagement with self-regulated inquiry-based online learning modules. *Smart learning environments*, 10(1), 1.

Dan, Y., Lei, Z., Gu, Y., Li, Y., Yin, J., Lin, J., ... & Qiu, X. (2024). Educhat: A large language model-based conversational agent for intelligent education. In *China Conference on Knowledge Graph and Semantic Computing* (pp. 297-308). Singapore: Springer Nature Singapore.

Davis F D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 1989: 319-340.

Du, X., Du, M., Zhou, Z., & Bai, Y. (2025). Facilitator or hindrance? The impact of AI on university students' higher-order thinking skills in complex problem solving. *International Journal of Educational Technology in Higher Education*, 22(1), 39.

Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in education and teaching international*, 61(3), 460-474.

Gunuc, S., & Kuzu, A. (2015). Student engagement scale: development, reliability and validity. *Assessment & Evaluation in Higher Education*, 40(4), 587-610.

Hwang, G. J., Lai, C. L., Liang, J. C., Chu, H. C., & Tsai, C. C. (2018). A long-term experiment to investigate the relationships between high school students’ perceptions of mobile learning and peer interaction and higher-order thinking tendencies. *Educational Technology Research and Development*, 66(1), 75-93.

Huang, K. L., Liu, Y. C., & Dong, M. Q. (2024). Incorporating AIGC into design ideation: A study on self-efficacy and learning experience acceptance under higher-order thinking. *Thinking Skills and Creativity*, 52, 101508.

Ilgun Dibek, M., Sahin Kursad, M., & Erdogan, T. (2025). Influence of artificial intelligence tools on higher order thinking skills: a meta-analysis. *Interactive Learning Environments*, 33(3), 2216–2238.

Lee, H. Y., Chen, P. H., Wang, W. S., Huang, Y. M., & Wu, T. T. (2024a). Empowering ChatGPT with guidance mechanism in blended learning: Effect of self-regulated learning, higher-order thinking skills, and knowledge construction. *International Journal of Educational Technology in Higher Education*, 21(1), 16.

Lee, D., Arnold, M., Srivastava, A., Plastow, K., Strelan, P., Ploeckl, F., ... & Palmer, E. (2024b). The impact of generative AI on higher education learning and teaching: A study of educators’ perspectives. *Computers and Education: Artificial Intelligence*, 6, 100221. https://doi.org/10.1016/j.caeai.2024.100221.

Lin, H. C. K., Tseng, C. H., & Chen, N. S. (2025). Enhancing programming education. *Educational Technology & Society*, 28(2), 279-294.

Lin, C. S. (2026). Co-Creating With AI: How Communication Students Learn, Think, and Rely on Intelligent Systems. *Journal of Educational Computing Research*, 64(1), 155-179.

Liu, C. C., Liu, S. J., Hwang, G. J., Tu, Y. F., Wang, Y., & Wang, N. (2023). Engaging EFL students' critical thinking tendency and in-depth reflection in technology-based writing contexts: A peer assessment-incorporated automatic evaluation approach. *Education and Information Technologies*, 28(10), 13027-13052.

Ma, Q., Shen, H., Koedinger, K., & Wu, S. T. (2024). How to teach programming in the ai era? using llms as a teachable agent for debugging. *In International Conference on Artificial Intelligence in Education* (pp. 265-279). Cham: Springer Nature Switzerland.

Patrick, P. M., Yip, S. Y., & Campbell, C. (2025). Artificial Intelligence and Higher-Order Thinking: A Systematic Review of Educator and Student Experiences and Perspectives in Higher Education. *Higher Education Quarterly*, 79(4), e70069.

Putra, F. W., Rangka, I. B., Aminah, S., & Aditama, M. H. (2023). ChatGPT in the higher education environment: perspectives from the theory of high order thinking skills. *Journal of Public Health,* 45(4), e840-e841.