Scientific Research Publishing

Open Journal of Social Sciences, 2026, 14(11), 704-730
https://www.scirp.org/journal/jss
ISSN Online: 2327-5960
ISSN Print: 2327-5952

A Study on the Design and Application of Teaching Agents to Promote Deep Thinking in Junior Secondary School Students

Jie Wang[1], Xiulin Ma[2], Teng Wang[3]

[1]School of Educational Sciences, Xinjiang Normal University, Urumqi, China
[2]School of Educational Technology, Beijing Normal University, Beijing, China
[3]Students Affairs Office, Nanning Normal University, Nanning, China
Email: 15022952386@163.com

How to cite this paper: Wang, J., Ma, X. L., & Wang, T. (2026). A Study on the Design and Application of Teaching Agents to Promote Deep Thinking in Junior Secondary School Students. *Open Journal of Social Sciences, 14,* 704-730.
https://doi.org/10.4236/jss.2026.1411039

Received: October 20, 2026
Accepted: November 23, 2026
Published: November 26, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/

Open Access

Abstract

Against the backdrop of rapid advancements in artificial intelligence technology and the drive to build a leading nation in education, deep thinking, as a crucial aspect of cultivating innovative talent, represents a core consensus in the digitalization of education. This study addressed key elements and strategies in response to issues such as superficial learning, formalized inquiry, and one-dimensional assessment prevalent in junior secondary information technology classrooms. It explored the design and application of teaching agents for fostering deep thinking in information technology education. Utilizing the "Coze" platform, the corresponding teaching agents were developed, and a "Three-Dimensional Six-Stage" project-based teaching model was constructed. Employing a Design-Based Research (DBR) methodology as the core, supplemented by quasi-experimental research, the study conducted two rounds of project-based practice across two parallel groups. The results indicate that students in the experimental group achieved significantly higher post-test scores than those in the control group across all four dimensions—multi-perspective thinking, integrative thinking, critical thinking, and creative thinking ($P < .05$), while the control group exhibited a significant decline in creative thinking (pre-test M = 22.79, post-test M = 21.41; $P = .007$). The effectiveness of AI agents as an aid varied between individuals, with its efficacy closely linked to students' cognitive readiness and the quality of human-computer interaction. This model not only provides empirical guidance for the reform of information technology curriculum teaching in lower secondary schools in the digital age, but also offers valuable insights into exploring the development of students' deep thinking through human-computer collaboration.

Keywords

Teaching Agents, Deep-Thinking, Project Based Learning, Human-Computer Collaboration

1. Introduction

The ongoing integration of artificial intelligence and education has driven a profound transformation in talent development, shifting the focus from "knowledge accumulation" to "quality of thinking". Against the backdrop of digital transformation in education, innovative talent must not only be proficient in technology but also adept at collaborating with artificial intelligence to address complex, real-world problems. Consequently, the cultivation of deep thinking and related competencies is a vital component of talent development in the new century.

1.1. The Development of Deep Thinking Is an Intrinsic Requirement for Student Growth in the Intelligent Age

The *Core Competencies for Chinese Student Development* incorporates scientific spirit and practical innovation as key dimensions within the core competencies framework (Core Competencies Research Group, 2016); the *Curriculum Standards for Information Technology in Compulsory Education* (2022 *Edition*) also lists computational thinking as a core competency for the Information Technology subject, emphasizing the development of students' logical reasoning and critical thinking through algorithm design, problem decomposition, and system modelling (Ministry of Education of the People's Republic of China, 2022). At the international level, the OECD's Education 2030 framework proposes fostering students' comprehensive abilities to adapt to a complex society (Zuo et al., 2022). At the same time, UNESCO advocates strengthening the cultivation of students' higher-order abilities, such as critical and creative thinking (Holmes & Miao, 2023). It is evident that deep thinking has become both an intrinsic requirement for students' development in the intelligent age and a core consensus in the digital transformation of education. Therefore, we must attach great importance to cultivating students' deep-thinking abilities, effectively enhancing their capacity for deep thinking at the level of critical inquiry, and fostering their development into innovative talents.

1.2. Superficial Learning in Junior Secondary Information Technology Classrooms: A Serious Practical Dilemma

At present, superficial learning remains a widespread phenomenon in junior secondary Information Technology classrooms. Firstly, superficial practice is dominated by mechanical operations: teaching frequently adopts an "imitation-drill" model; when students encounter real-world information scenarios, they generally face obstacles in problem analysis, solution design, and tool selection, causing

their learning to fall into the trap of tool-skill training (Li et al., 2024). Secondly, there is a tendency for inquiry-based activities to become overly formalized: in project-based learning centered on interdisciplinary themes, "pseudo-inquiry" and "superficial inquiry" have become persistent problems in teaching. This manifests as teachers making excessive assumptions and taking a dominant role, while pupils merely follow predetermined procedures to complete verification-based activities (Yuan & Wang, 2023), which severely hinders the development of deep thinking and innovative abilities. Thirdly, the one-dimensional nature of teaching assessment and delayed feedback: traditional assessment largely relies on single tests or assignments; in programming education, there is an excessive focus on results, while the thought process is neglected, and feedback lacks personalization and timeliness. The aforementioned phenomena indicate that facilitating students' transition from superficial thinking to deep thinking has become a matter of the utmost urgency in the reform of information technology education.

1.3. The Emergence of Teaching Agents Has Provided for Fostering a New Pathway to Deep Thinking

Generative Artificial Intelligence covers a wide range of functions, but its universal nature presents a high barrier to entry for users-users must have a high level of proficiency in prompt engineering to precisely match teaching scenarios. Based on large language models, educational AI agents can provide learning support akin to that of subject experts through pre-set roles and constraints, effectively lowering the barrier to using large language models and addressing their shortcomings in educational applications (Lu et al., 2024). However, there are still relatively few Teaching Agents designed specifically for the Information Technology curriculum at the lower secondary level, with most research remaining at the theoretical stage and lacking robust practical data to support it (Liu et al., 2019; Guo et al., 2020). Consequently, implementing project-based learning through Teaching Agents in lower secondary Information Technology courses to investigate their role in the development of students' deep thinking holds significant theoretical and practical value.

1.4. Research Questions

In light of this, this study proposes three research questions: 1) What characteristics should projects aimed at developing deep thinking possess, and how should teaching activities be designed? 2) What core functions should be teaching agents possess to facilitate the development of deep thinking, and how should they be developed? 3) How can the level of deep thinking development be assessed to enable the iterative refinement and optimization of project design?

2. Research Design

2.1. Research Objectives and Subjects

This study utilizes Teaching Agents technology as its focal point, aiming to con-

struct a project-based learning model that promotes the development of deep thinking in Grade 8 students and to validate its educational effectiveness. Specifically, it comprises three sub-objectives: 1) to explore the forms and core functions of Teaching Agents that promote the development of deep thinking in Grade 8 students, thereby providing an operational development framework for Human-Agent collaborative teaching design; 2) to design an information technology project that involves in-depth interaction between students and the Teaching Agents; 3) to construct an assessment system for the development of deep thinking suitable for project-based learning in the information technology subject, enabling dynamic and multidimensional evaluation of both thought processes and outcomes.

With regard to the selection of research subjects, this study recruited pupils from two parallel Year 8 groups at a state-funded secondary school in a certain city, forming an experimental group (58 pupils) and a control group (56 pupils) based on natural groups. The selection of Year 8 pupils as the research subjects was based primarily on two considerations: firstly, Year 8 pupils are in the early stages of the formal operational stage, with their thinking gradually evolving from concrete-imaginative thinking towards abstract-logical thinking (Inhelder & Piaget, 1958), making this a critical period for fostering the development of deep thinking; the plasticity of their thinking provides favorable conditions for instructional intervention; secondly, pupils at this stage have already acquired a certain foundation in information technology and the ability to learn independently, enabling them to operate the instructional agents with relative proficiency when carrying out project activities. To ensure that the two groups of pupils were at a similar level prior to the intervention, this study conducted baseline measurements and comparisons of the two groups across dimensions such as knowledge base, computer application skills, thinking styles, academic performance, and gender ratio before the experiment. The results showed that there were no significant differences between the two groups across any of these dimensions, thereby satisfying the prerequisites for quasi-experimental research.

2.2. Research Ethics

This study has received an ethics exemption from the Ethics Committee of the College of Educational Sciences at Xinjiang Normal University. The study was conducted at a public middle school in a certain city; prior to the intervention, the research protocol was approved by the school's administration. Written informed consent forms were distributed to the parents of all participating students, and verbal or written permission was obtained from the minors participating in the study. Parents may request that their child withdraw from this study at any time without providing any reason, and such withdrawal will not result in any adverse consequences.

All Coze platform conversation records were anonymized by removing personally identifiable information such as student names and group information. Raw interaction records were stored in encrypted form, with access restricted to the

researchers involved in this study. Once data analysis is complete and the paper is published, the raw conversation records will be permanently deleted.

2.3. Research Methodology

This study adopted a research methodology system centered on the "Design-Based Research (DBR)" approach, supplemented by multiple methods, to realize a closed-loop cycle of "design-implementation-evaluation-iteration" within a real-world educational context.

2.3.1. Design-Based Research (DBR)

This study integrates design and research, carefully devising teaching activities and educational intervention programs to address specific educational issues within real-world educational contexts, while continuously iterating and refining these based on practical data. In this study, DBR runs throughout the entire process of project design, teaching implementation, and effectiveness evaluation (Design-Based Research Collective, 2003). The two rounds of project implementation correspond to two rounds of DBR iteration; in each round, issues are promptly identified based on feedback from the previous round, and the functions of the intelligent agents and the teaching model are continuously optimized.

2.3.2. Quasi-Experimental Research Methodology

The experimental group utilized the teaching agents designed in this study for learning support, while the control group continued with traditional teaching methods; following the experiment, the effectiveness of the agent's project was verified by comparing learning outcomes. As the researchers were unable to randomly assign students to different groups, this study employed natural groups as the unit of grouping, constituting a quasi-experimental design (Jia, 2008). To control for extraneous variables, the two groups were kept consistent in terms of teaching staff, curriculum content, timetable arrangements, and project themes; the sole difference was the introduction of the Teaching Agents to support the experimental group.

Both rounds were implemented in the two intact classes described above by the same information technology teacher, with the same content, timetable, and project themes. In the experimental group, the two projects were carried out in accordance with this model, and the students interacted with the agent throughout the lessons. In the control group, the same two projects were completed through the school's traditional approach—teacher explanation and demonstration followed by the students' imitative practice—without the teaching agent and without the six-stage scaffolding; in the first round, this condition is referred to as traditional teaching, and in the second round as traditional project-based learning. It should be noted, therefore, that the "Three-Dimensional Six-Stage" model was applied only in the experimental group; in the control group it was replaced by the traditional teaching sequence. The comparison consequently tests the combination of the teaching agent and the six-stage model against traditional instruc-

tion rather than the teaching agent alone. Instructional exposure was equivalent across the two conditions: the information technology course was timetabled as one lesson per week, and each project was implemented over three consecutive weeks (Project 1, from 13 to 29 October 2025; Project 2, from 10 to 28 November 2025).

2.3.3. Questionnaire Survey Method

Data was collected through the use of standardized questionnaires, and the collected data was processed and analyzed using SPSS software. For this study, a Deep-Thinking Assessment Scale was developed and administered to pupils in the two groups at three time points—the pre-test, the first post-test (after the first round), and the second post-test (after the second round)—to track changes in the various dimensions of deep thinking.

2.3.4. Classroom Observation Method

Classroom observations were used to record pupils' learning states, behavioral performance, and levels of participation, thereby assisting teachers in understanding the learning situation and adjusting teaching strategies in a timely manner. The researcher conducted non-participatory observations in the classroom during each round of project-based learning, focusing on the frequency and depth of interaction between pupils and the teaching agents, their questioning behavior, and collaborative inquiry activities, and compiled classroom observation records.

2.3.5. Interview Method

Initial interviews with subject teachers were conducted to understand the current state of teaching; following each experimental round, interviews with both teachers and pupils were carried out to assess the effectiveness of the AI project and gather suggestions for improvement. The interviews adopted a semi-structured format, centering on questions such as: Did the AI help you think? In which aspects did you encounter difficulties? And how would you like the AI to be improved? The transcribed interview content was used as qualitative evidence to supplement the quantitative results.

2.4. Research Process

In line with the DBR (Design-Based Research) approach, this study divides the research process into six stages: research preparation, project design, prototype development, teaching implementation, effectiveness analysis, and summary and reflection. The conceptual diagram for this study is shown in Figure 1.

3. For Cultivating Deep Thinking: Exploration of and Design of a Model

The design of learning projects centered on intelligent agents must adhere to the requirements of specific learner groups, learning objectives, and teaching contexts. It is essential to clarify what the intelligent agents can do and what they need

to do in order to complete their functional design, thereby constructing a project-based teaching model supported by the intelligent agents.

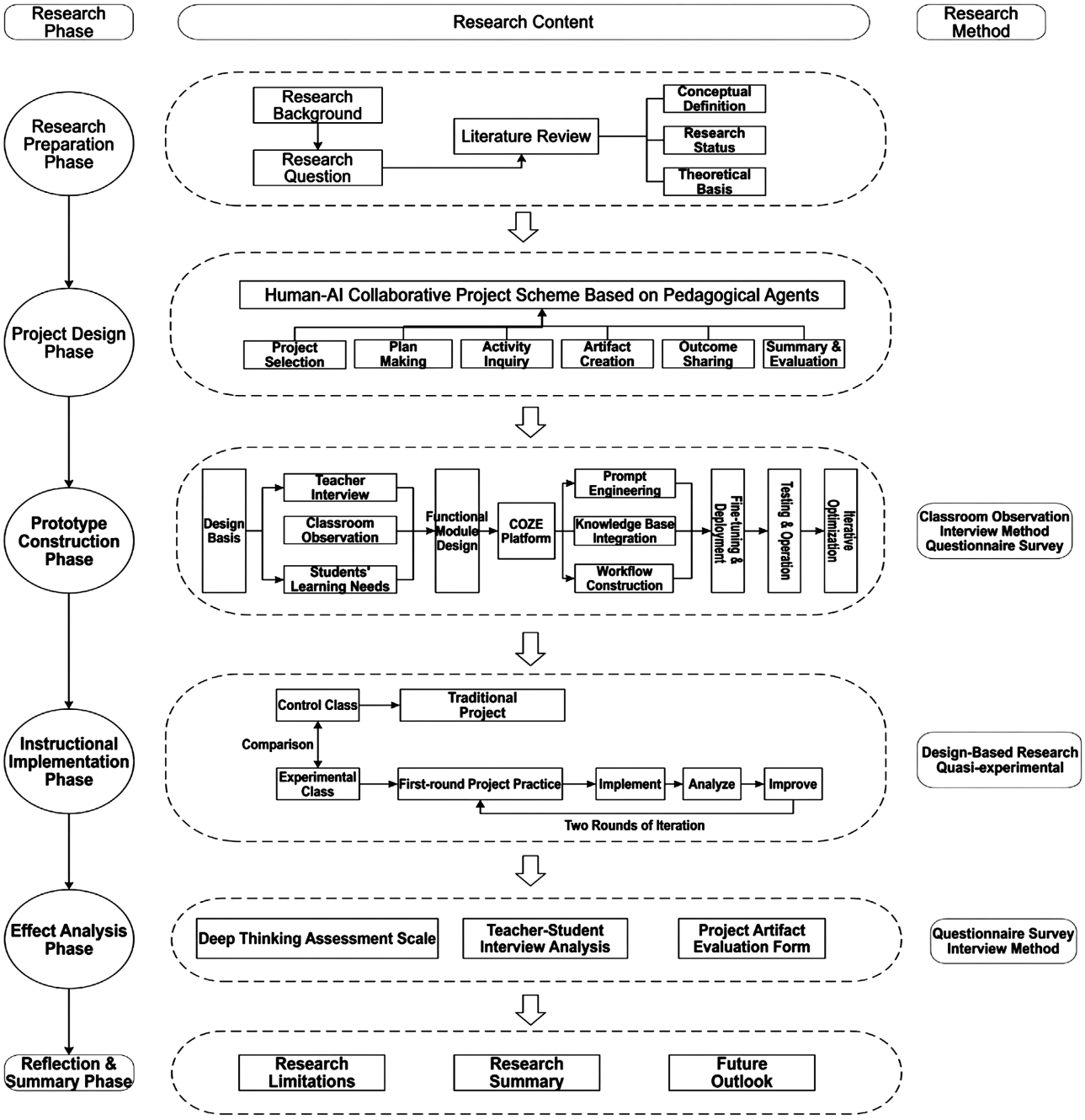


Figure 1. Research concept map.

3.1. Explanation and Structural Definition of Deep Thinking

In the field of research on the cultivation of deep thinking, existing academic perspectives can primarily be summarized into two representative schools of thought. The first is the cognitive process school, represented by Bloom. This school, starting from the internal structure of cognitive processes, systematically divides them

into levels ranging from shallow to deep—such as remembering, understanding, applying, analyzing, evaluating, and creating—so as to achieve a deep mastery of knowledge. Whilst this framework provides a classic analytical dimension for identifying and describing the external manifestations of deep thinking, its focus lies in the hierarchical classification of thinking outcomes; however, it does not adequately explain the specific cognitive methods and pathways required to achieve the objectives at each level of deep thinking. Secondly, there is the deep learning approach represented by project-based learning. Guo Hua points out that deep learning is a process in which, under the guidance of a teacher, students engage in active inquiry and problem-solving around challenging themes to achieve the development of deep thinking (Guo, 2016). Lou Li argues that deep learning can stimulate students' motivation to learn and their capacity for deep thinking through various forms, such as reflective learning, project-based learning, and task-driven learning (Lou, 2025). Whilst these learning formats create a learning context (vehicle) for achieving deep learning, they do not provide a substantive explanation of the specific ways in which deep learning occurs for students within this context.

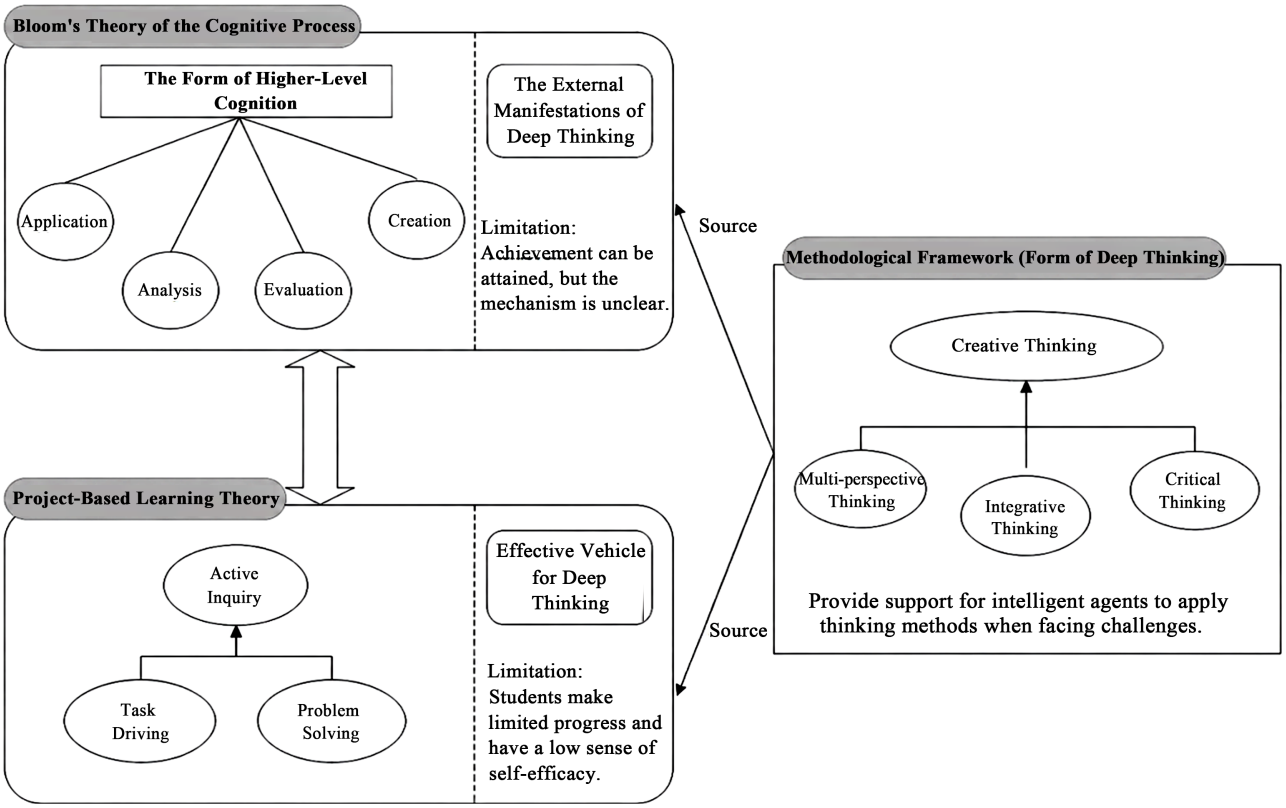


Figure 2. Methodological structure of deep thinking.

In response to this methodological shortcoming, this study deconstructs deep thinking from a methodological perspective into four interrelated core dimensions: multi-perspective thinking, integrative thinking, critical thinking, and cre-

ative thinking (Dai & Yang, 2023). These four dimensions are not isolated from one another but collectively constitute the overall structure of deep thinking. Multi-perspective thinking provides the raw material for integrative and critical thinking; integrative thinking links disparate pieces of information to form a systematic understanding; critical thinking scrutinizes and examines that understanding; and creative thinking transforms this deepened understanding into innovative outcomes, as shown in **Figure 2**.

3.2. Agents for Deep-Thinking and Their Functional Design

3.2.1. The Logical Starting Point for the Design of Agent Functions

The determination of functional modules for the intelligent agents was based on a survey of teaching starting points. This study gathered evidence of needs from three sources: teacher interviews, classroom observations, and student questionnaires. Teacher interviews revealed that the school offers only one information technology lesson per week; with a heavy teaching workload, teachers largely follow a set routine to deliver basic instruction, paying insufficient attention to deep thinking. Classrooms are dominated by traditional lectures combined with practical exercises, and superficial learning is widespread. Most teachers have already used GenAI in the classroom, but are concerned that its universality may lead students to search for irrelevant content or directly copy generated answers, thereby undermining independent thinking (Zhang, 2024). Consequently, they hope that the AI agents can assist students in problem-solving through guided questioning, restrict the scope of questions and answers, and provide real-time assessment of classroom performance. Classroom observations revealed that student engagement was poor, with a lack of enthusiasm in answering questions and a low task completion rate; overall, students exhibited a lackadaisical and haphazard approach to learning. Student questionnaires revealed that awareness of general-purpose large language models (such as Doubao and DeepSeek) was high, with 80 per cent of students having heard of them or using them occasionally, while only 7.2 per cent had never heard of them; however, awareness of Teaching Agents was low, with 66.6 per cent of students having never heard of them and only 6.7 per cent having used them. Nevertheless, students showed a high level of acceptance towards the use of Teaching Agents in the classroom, with the majority stating they were "very willing" or "quite willing". In terms of functional expectations, "answering questions and resolving doubts," "companionship," and "learning methods" ranked in the top three. reflecting students' dual needs for both practicality and emotional support. The above evidence collectively establishes the design principle that the AI Agents should "guide rather than answer on the student's behalf," and directly informs the selection of functional modules.

3.2.2. Core Functions That Teaching Agents Should Possess

Addressing the four dimensions of multi-perspective thinking, integrative thinking, critical thinking, and creative thinking, the teaching AI integrates five categories of core functional modules: multi-solution generation, interdisciplinary

perspectives (Yuan et al., 2025), intelligent guided Q&A, knowledge visualization, and classroom thinking reports. These five core functional modules form an intersecting grid with the four dimensions of Deep-Thinking, providing comprehensive support for the realization of these dimensions.

Based on the functional characteristics of current mainstream Teaching Agents and the service capabilities of their development platforms, this study identifies the key functions of the AI Agents as the configuration of persona and response logic, workflow construction, knowledge base integration, and the optimization of the interactive experience. Regarding persona and response logic configuration, the AI Agents' subject-matter expert identity and linguistic style are defined in accordance with the cognitive level of Grade 8 students, ensuring that its responses are both professional and tailored to the pupils' comprehension abilities; in terms of workflow construction, processes such as problem diagnosis, step-by-step guidance, and solution comparison were designed (Wang et al., 2025) to ensure that the AI intervenes at critical junctures by providing scaffolding rather than direct answers (Li & Wang, 2024); regarding knowledge base integration, key subject knowledge points and interdisciplinary materials were incorporated to support the generation of multiple solutions and interdisciplinary perspectives; in terms of interactive experience design, emphasis was placed on response speed and the tone of guidance to prevent technical alienation from undermining the quality of human-computer interaction. In response to feedback from the first round of implementation indicating that critical thinking had not improved significantly, targeted improvements were made to the AI agents' interaction design in the second round. These included the addition of scaffolding-style support, such as "code fill-in-the-blanks" and "debugging incorrect code," avoiding the direct provision of answers and instead offering students a framework for thinking, thereby stimulating their critical questioning, as shown in **Figure 3**.

3.3. Design of the "Three-Dimensional Six-Stage" Project-Based Teaching Model

Within the framework of project-based learning, this study has developed the "Three-Dimensional Six-Stage" project-based teaching model (Liu & Zhong, 2002). The "three-dimensional" refers to the three core elements of project-based learning—real-world contexts, intelligent scaffolding, and cognitive development: real-world contexts provide the problem-based framework for deep thinking (He, 1997). Intelligent scaffolding is provided by the teaching agents through dynamic support, while cognitive development serves as the overarching objective throughout the project. The "Six-stage" refers to the six phases of project implementation, namely: context introduction, problem definition, solution design, collaborative inquiry, presentation of outcomes, and evaluation and reflection. The six stages are interlinked, with the Teaching Agent playing distinct roles at each stage: during the context introduction stage, it provides background information from multiple perspectives; during the problem definition stage, it assists in defining and breaking down the problem; during the solution design stage, it generates multiple

solutions for comparison; during the collaborative inquiry stage, it supports in-depth inquiry through guided questioning; during the outcome presentation stage, it facilitates visual integration; and during the evaluation and reflection stage, it generates a classroom thinking report to support formative assessment. This model embeds the Teaching Agents into every stage of project-based learning, ensuring that technical support is deeply integrated with cognitive development objectives (Sang & Wang, 2025).

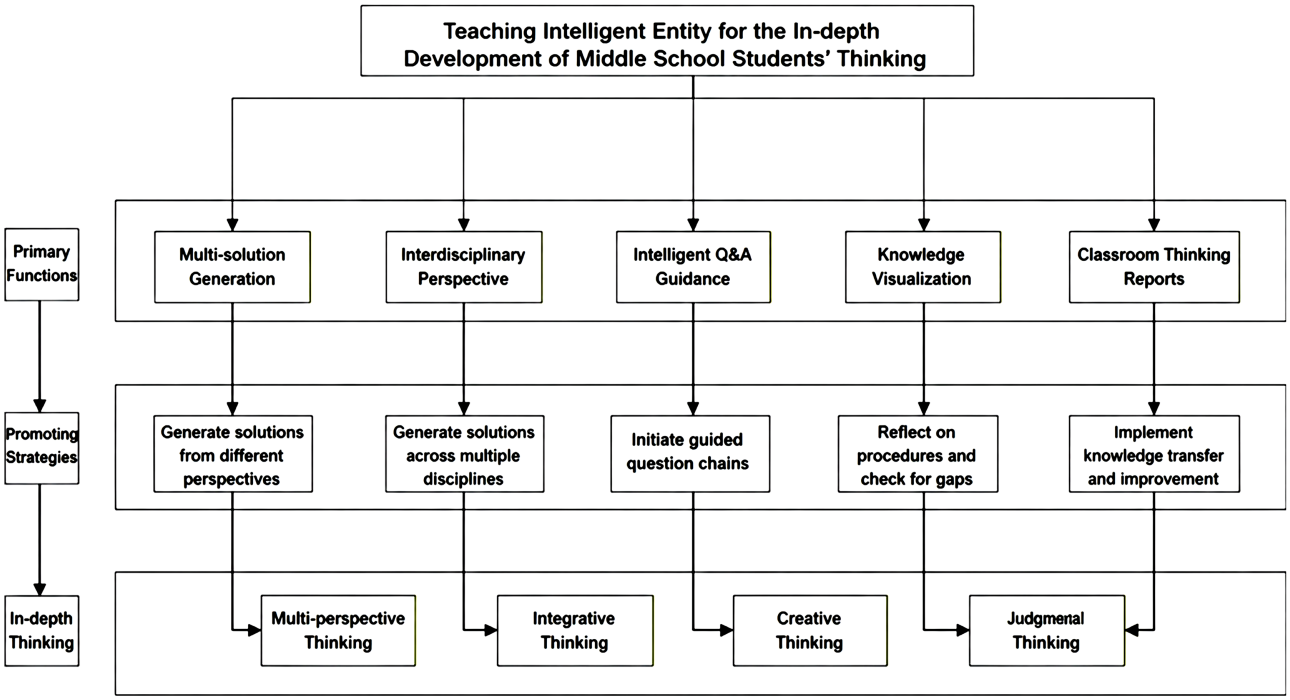


Figure 3. Functional framework diagram of the teaching agents for the development of deep thinking in Grade 8 students.

3.4. Design of Research Tools

The conduct of this study relies on two key types of tools: firstly, agent tools, which form the environmental foundation for the research; and secondly, assessment tools for deep thinking levels, which are crucial for validating the study's effectiveness.

3.4.1. Intelligent Agents Creation Platform: Coze

"Coze" is launched by *ByteDance*, an AI bot development platform. Through its visual interface and modular components, it enables users with no programming background to rapidly build AI agents. It supports flexible switching between multiple models (such as *Doubao* and *DeepSeek-R*1). It is adapted to various scenarios, including conversational interaction and process automation, while being particularly well-suited to lightweight development needs in sectors such as education and office automation. Since its launch, it has become an important player in the domestic AI development and application field thanks to its low entry barrier, adaptability to multiple scenarios, and ecosystem integration capabilities.

This study selected *Coze* as the platform for creating Teaching Agents, primarily based on the following three considerations: Firstly, it supports low-code or no-code creation of roles and the addition of constraints, enabling teachers to design teaching roles and interaction models suited to the cognitive level and thinking patterns of Grade 8 students. Secondly, the visual workflow editing and knowledge base customization features enable teachers to intuitively and rapidly construct thought-provoking chains of questions and provide targeted subject-specific knowledge. Thirdly, the real-time feedback and dialogue logging functions for teachers provide strong support for pupils' thought processes and delivering personalized interventions, making the process of cultivating deep thinking visible, guidable, and assessable.

To make the intervention reproducible, the configuration of the agent is specified here. The agent was created in single-agent (LLM) mode, so that a student conversed with one agent rather than switching among multiple agents, which reduced the cognitive load and shortened the response latency. After comparing the models available on Coze, the text model Doubao 1.5 Pro 32k was adopted as the global model for its performance in knowledge, code, and reasoning, whereas the multimodal model Doubao 1.6 Ultra-Fast was used in the AI_answer workflow for its rapid reasoning and its multi-step problem-solving capability; the preset Balanced mode was selected for generation diversity, so as to balance the randomness and the accuracy of the output. The core prompts were written around six components—role, objective, skills, workflow, output format, and constraints—which together defined the agent's identity, working goal, capabilities, procedure, response format, and the scope of content it should not address. The agent was released on Coze for use by the students after all of these settings had been finalized.

3.4.2. Deep Thinking Assessment Tools

Drawing on the components of deep thinking proposed by Ye Xiu (Ye, 2023), Shao Zebin (Shao, 2018), and Bloom (Anderson, 2009) and others, and drawing on the research findings of the author's team, the author posits that the modes of deep thinking primarily include multi-perspective thinking, critical thinking, integrative thinking, and creative thinking. These four cognitive abilities can serve as first-level dimensions for the cultivation and development of students' deep thinking; these four primary dimensions represent a progression from lower-order to higher-order thinking, and they are also the foundation for the teaching design and assessment in this study.

To ensure the assessment has a practical structure, this study further subdivides the first-level dimensions into 12 second-level dimensions (Zhang et al., 2014). Under multi-perspective thinking, there are divergent thinking and the ability to switch perspectives; integrative thinking, drawing on the requirement to "establish cross-disciplinary thematic learning activities" as set out in the *Compulsory Education Information Technology Curriculum Standards* (2022 *Edition*), comprises associative judgment, big-idea thinking, inductive analysis, and evaluation

and integration. Critical thinking is based on Facione's California Critical Thinking Disposition Test, which comprises the sub-dimensions of element identification and judgment, truth-seeking, and cognitive maturity; creative thinking is based on Williams' Creative Thinking Scale (Facione, 1990; Williams, 2003), comprising the desire to explore knowledge, imagination, and a spirit of adventure and challenge. During the compilation process, adjustments were made to the second-level dimensions through consolidation and selection: the overlapping elements of "curiosity" and "thirst for knowledge" in critical and creative thinking were merged into "desire to explore knowledge"; "adventurousness" and "challenge-seeking" were merged into "spirit of adventure and challenge"; "open-mindedness" was incorporated into "ability to shift perspectives"; "analytical ability" and "systematization ability" were incorporated into "evaluation and integration"; and the "self-confidence" dimension, which was not required by the curriculum standards, was removed. The scale comprises 26 items and employs a five-point Likert scale; items 6, 13, 19, and 26 are reverse-scored items. The items are drawn from four sources: self-developed items, the curriculum standards, the *Facione* scale, and the Williams scale. To test the stability and reliability of the scale, a pilot study was conducted with 166 Year 8 pupils from three randomly selected groups at a partner school. After excluding six invalid questionnaires, 160 valid questionnaires were recovered, yielding a response rate of 96.38 percent. Analysis using SPSS 24.0 revealed that the overall Cronbach's α coefficient for the scale was .899, with α coefficients for each dimension exceeding .7. The specific results are presented in Table 1. Factor analysis indicated that all variables clustered into four dimensions, showing a high degree of consistency with the anticipated dimensions.

Table 1. Reliability tests for survey data.

Dimension	Number of Items	Cronbach's α Coefficient
Multiperspective Thinking	6	.724
Integrative Thinking	7	.762
Critical Thinking	6	.725
Creative Thinking	7	.748

To verify the validity of the questionnaire used in this study, the author conducted validity analyses on the survey data. The specific results are presented in Table 2.

The KMO test results show that the KMO value is .848, which exceeds .8, indicating that the questionnaire is suitable for factor analysis. Additionally, the Sig value of the Bartlett's sphericity test is .000, with a significance level less than .05, indicating that there are significant correlations among the variables. The survey data exhibit a high degree of structural coherence and are suitable for factor analysis.

Table 2. KMO and Bartlett's tests for validity.

KMO Coefficient of Measurement Adequacy	.848	
Bartlett's Sphericity Test	Approximate Chi-Square	1989.649
	Degrees of Freedom	325
	Significance	.000

Factor analysis revealed that all variables clustered into four dimensions, which closely align with the expected survey dimensions, indicating that this scale effectively covers the research questions. Therefore, it can be confirmed that the questionnaire has good structural validity and is suitable for the research needs.

The Deep-Thinking Assessment Scale developed in this study comprises four dimensions, each containing several items scored on a five-point Likert scale. Reliability analysis showed that the Cronbach's α coefficients for each dimension were .724, .762, .725, and .748, respectively, all exceeding .7, indicating good scale reliability; validity tests showed a KMO value of .848, which exceeds .8, and a Sig value of .000 in Bartlett's sphericity test, indicating suitability for factor analysis. Furthermore, the fact that all variables clustered into the four dimensions was consistent with expectations, demonstrating good structural validity. It should be noted that this is a self-report instrument: it captures the deep-thinking tendencies that students report across the four dimensions, that is, their self-reported deep-thinking levels, rather than their demonstrated performance on objective tasks; the results reported below should therefore be read as self-reported tendencies. The assessment was administered at three time points—pre-test, first post-test, and second post-test—to ensure that the dynamic process of Deep-Thinking development could be captured.

4. Two Rounds of Teaching Practice: Practical Implementation and Analysis of Outcomes

Under the "Three-Dimensional Six-Stage" project-based teaching model, this study conducted two rounds of project-based practice centered on the themes "Discovering the Beauty of Winter: AI-Assisted Text Creation" and "Making Ingenious Use of Multi-Branch Structures to Build an Intelligent Assessment System". The first round focused on language-based text creation, while the second round centered on programming-based problem-solving, thereby testing the effectiveness of Teaching Agents in supporting deep thinking across different subject areas (Zhai, 2020) and achieving the objective of "promoting model optimization through design iteration" (Guo, 2016).

4.1. First Round of Project Practice

4.1.1. Project Design and Implementation

The first-round project, "Discovering the Beauty of Winter: AI-Assisted Text Creation", took "The Beauty of Winter" as its theme. This represented a reorganiza-

tion of Lesson 2, "Understanding Machine Learning", from Unit 1 (Can Machines Predict?) and Lesson 1, "AI-Generated Content", from Unit 3 (Can Machines Create?) in the second volume of the Tsinghua Information Technology textbook. By utilizing the Teaching Agents' capabilities—such as multi-solution generation and an interdisciplinary perspective—students were guided to engage in text creation with AI assistance. The project was conducted according to the "Three-Dimensional Six-Stage" model: during the contextual introduction stage, the teacher used textual and visual materials depicting winter scenes to create an authentic context, evoking students' life experiences and emotional responses; in the problem-definition stage, students interacted with the teaching AI to clarify the core task of "how to express the beauty of winter through writing" and broke this down into sub-questions such as perspectives for observation, writing techniques, and emotional expression; in the proposal-design stage, the teaching AI generated multiple proposals for the same creative theme, allowing students to compare, select, and combine these from different perspectives to form preliminary creative ideas. During the collaborative inquiry stage, pupils work in groups to produce texts and provide peer feedback, while the teaching AI offers support on writing techniques through guided questioning; during the presentation stage, each group showcases their work and explains their creative approach; during the evaluation and reflection stage, Teaching Agents generates a classroom thinking report, and the teacher uses this report to guide pupils in reviewing their thought processes. The entire project spans several weeks, during which pupils maintain frequent human-computer interaction with the Teaching Agents throughout.

4.1.2. Results of the First Round of Implementation

Following the first round of implementation, a paired-sample t-test was conducted on the experimental group's levels of deep thinking. The results showed that students in the experimental group demonstrated varying degrees of improvement across all dimensions of deep thinking, with an overall average increase of 4.88 points (from 79.10 to 83.98). Specifically, the *P*-values for the three dimensions of multi-perspective thinking ($t = -5.134$, $P = .000$), integrative thinking ($t = -3.083$, $P = .003$), and creative thinking ($t = -2.330$, $P = .023$) were all less than .05, indicating statistically significant differences; however, the improvement in critical thinking was not significant ($P = .230$). The specific test results are shown in Table 3.

The reasons for this phenomenon may be as follows: as this was the students' first time participating in project-based learning assisted by educational AI agents, and their first encounter with such agents, they experienced a sense of novelty and a certain degree of technological awe. Consequently, their mindset was primarily focused on learning and adaptation, with relatively little critical questioning. Classroom observations corroborated this assessment: in the early stages, students largely viewed the AI as an "authoritative answerer" and a source of knowledge upon which they could directly rely; consequently, they rarely questioned its outputs. Interaction between students and the AI was predominantly characterized

by receptive question-and-answer exchanges, while in-depth questioning, comparative inquiry, and dialectical critical thinking were relatively lacking. Nevertheless, the average score for critical thinking showed a slight improvement, indicating that project-based learning utilizing educational AI is effective in cultivating deep thinking among Grade 8 students.

Table 3. Pre- and post-test analysis of the depth of thinking among students in the first round of the experimental group.

Dimension	Time point	Number of Cases	Mean	Standard Deviation	t	*P*
Multi-perspective thinking	Pre-test	58	17.52	3.147	−5.134	.000
	Post-test	58	19.64	2.647		
Integrative Thinking	Pre-test	58	20.76	3.445	−3.083	.003
	Post-test	58	22.12	3.212		
Critical Thinking	Pre-test	58	19.05	2.453	−1.214	.230
	Post-test	58	19.52	2.817		
Creative Thinking	Pre-test	58	21.78	3.038	−2.330	.023
	Post-test	58	22.71	3.314		
Deep Thinking (Overall)	Pre-test	58	79.10	10.001	−4.051	.000
	Post-test	58	83.98	10.093		

4.1.3. First Round of Feedback and Improvements

Addressing the issue of the lack of significant improvement in critical thinking, this study further identified the causes through interviews with teachers and pupils: on the one hand, pupils lacked the awareness and habit of critically examining the AI's outputs; on the other hand, the AI's interaction design was geared towards "providing solutions" rather than "encouraging questioning," and thus failed to effectively stimulate critical thinking. Consequently, the second round of implementation involved targeted improvements to the AI's interactive design. These included the addition of scaffolding support such as "code fill-in-the-blanks" and "debugging erroneous code," as well as the design of guided questioning sessions such as "Do you agree with the AI's proposal?" and "Can you identify any issues in the AI's response?" This shifted the AI's role from "providing answers" to "guiding critical analysis," avoiding the direct provision of answers while offering students a framework for their thinking.

4.2. Second Round of Project Practice

4.2.1. Project Design and Implementation

The second-round project, "Making Effective Use of Multi-Branch Structures to Build an Intelligent Assessment System," centered on the theme of an "intelligent assessment system." Utilizing the improved Teaching Agents, it guided students to develop logically rigorous, systematic thinking through multi-branch struc-

tured programming. The project was likewise conducted according to the "Three-Dimensional Six-Stage" model: during the contextual introduction phase, real-world scenarios such as the scoring of campus activities were used to introduce the concept of "multi-branch decision-making"; in the problem-definition stage, students used the Teaching Agents to break down the task of "building an intelligent assessment system" into sub-problems such as conditional judgment, branch design, and boundary handling; in the solution-design stage, the Teaching Agents generated multiple branching structure proposals, and students compared their logical rigor before selecting the optimal solution; in the collaborative inquiry stage, students completed the programming implementation and debugging, while the Teaching Agents provided scaffolding through methods such as "code fill-in-the-blanks" and "debugging erroneous code," guiding students to proactively identify and rectify issues; during the outcome presentation stage, Each group demonstrates the evaluation system's performance and explains its logical design; during the evaluation and reflection stage, the Teaching Agent generates a classroom thinking report to support teachers in conducting formative assessment and reflection. Compared to the first round, the second round places greater emphasis on students' questioning and debugging behaviors during their interactions with the Teaching Agents, shifting human-computer interaction from "accepting answers" to "critical debugging."

4.2.2. Outcomes of the Second Round of Practice

Because the two groups were not equivalent at baseline, the second-round outcomes were examined through a baseline-referenced, longitudinal comparison of the existing scores rather than an unadjusted test of significance. When each group's change from the pre-test to the post-test after two rounds was considered, the experimental group rose from 79.10 (SD = 10.001) to 86.50 (SD = 10.617), a gain of 7.40 points (Cohen's d = .72), whereas the control group fell from 81.07 (SD = 12.946) to 77.80 (SD = 10.567), a loss of 3.27 points (Cohen's d = −.28); the between-group difference in the post-test mean was 8.70 points (95% CI [4.77, 12.63]). It should be acknowledged, however, that because only one intact class was assigned to each condition, statistical adjustment cannot remove class-level confounders, and these comparisons therefore remain exploratory. Compared with the first-round post-test, the overall score of the experimental group rose from 83.98 to 86.50, an increase of 2.52, while the overall score of the control group changed from 77.63 to 77.80, an increase of only .17. Looking at the individual sub-dimensions, the experimental group showed improvements in multi-perspective thinking (+.27), integrative thinking (+.97), critical thinking (+.60), and creative thinking (+.67), all of which exceeded those of the control group; conversely, the control group experienced a decline in multi-perspective and critical thinking. Notably, the experimental group's critical thinking scores shifted from a non-significant improvement in the first round to a positive improvement in the second round, indicating that the refined interactive design of the AI Agents was effective in stimulating critical questioning.

In terms of the quality of the work, the experimental group's average score remained higher than that of the control group, and a distinct outlier with a high score was observed, indicating that, with the assistance of the AI Agents, individual students were able to demonstrate exceptional creativity and technical proficiency; conversely, the control group exhibited three outliers with low scores, suggesting that under the traditional teaching model, a significant proportion of students may not have fully understood the tasks, or may have lacked effective support and motivation to learn. Teacher interviews further corroborated this difference: students in the experimental group demonstrated a stronger desire to explore and greater problem-solving abilities, and were able to engage in in-depth thinking guided by the AI Agents; students in the control group encountered more difficulties when writing code, and fewer were able to run their programs successfully.

4.2.3. Feedback from Teacher and Student Interviews

Based on interviews with teachers and pupils, pupils in the experimental group explicitly affirmed the improvements in the agents' response speed and guidance logic. In particular, designs such as "code fill-in-the-blanks" and "debugging erroneous code" not only avoided providing direct answers but also offered pupils a framework for thinking, facilitating a shift from "fear of making mistakes" to "proactive debugging." Their thinking patterns evolved from the divergent imagination of the first round to a more logical and systematic approach. Several students stated that during the "debugging of erroneous code" process, they were required to actively consider "where the error lies, why it occurred, and how to correct it"; this process helped to develop their logical reasoning and critical thinking skills. Teachers also reported that the improved AI Agents were better able to guide students toward deeper thinking, resulting in a more proactive and thorough problem-solving process. In contrast, students in the control group generally found programming lessons tedious and difficult to understand; their learning approach was characterized by passivity and dependence. Although they were able to mechanically memorize syntax rules, they were unable to transfer their knowledge, which indirectly highlights the shortcomings of traditional project-based learning—lacking intelligent scaffolding in stimulating deep thinking.

4.3. Analysis of Research Findings and Outcomes

4.3.1. Teaching Agents Effectively Enhance Students' Levels of Deep Thinking

Independent-samples t-tests were first performed to compare the pre-intervention levels of each deep thinking dimension between the experimental group and the control group, so as to examine the baseline equivalence of the two groups (see Table 4). At the pre-test, there were no statistically significant between-group differences in multi-perspective thinking, integrative thinking, critical thinking, creative thinking and the overall score of deep thinking ($P > .05$). It indicated that students in the two groups were comparable at baseline before the intervention.

Table 4. Pre-test analysis of deep thinking levels in the experimental and control groups.

Dimension	Group	Number of Cases	Mean	Standard Deviation	t	*P*
Multi-perspective thinking	Experimental group	58	17.52	3.147	−1.312	.192
	Control group	56	18.34	3.538		
Integrative Thinking	Experimental group	58	20.76	3.445	−.649	.518
	Control group	56	21.23	4.315		
Critical Thinking	Experimental group	58	19.05	2.453	.572	.569
	Control group	56	18.71	3.740		
Creative Thinking	Experimental group	58	21.78	3.038	−1.429	.156
	Control group	56	22.79	4.405		
Deep Thinking (Overall)	Experimental group	58	79.10	10.001	−.910	.365
	Control group	56	81.07	12.946		

Following two rounds of practical sessions, normality tests were conducted on the pre- and post-test data for deep thinking in the two groups; both sets of data were found to follow a normal distribution, and independent samples t-tests were subsequently performed. Because both datasets satisfied the normality assumption, the parametric independent-samples t-test was considered appropriate and was therefore adopted for the between-group comparison. The results showed that the experimental group significantly outperformed the control group in multi-perspective thinking (t = 3.854, $P = .000$), integrative thinking (t = 4.748, $P = .000$), critical thinking (t = 3.001, $P = .003$), creative thinking (t = 2.990, $P = .003$), and overall performance (t = 4.382, $P = .000$). The experimental group's overall score of 86.50 was significantly higher than that of the control group (77.80), demonstrating a strong advantage as expected, as shown in Table 5.

The significant post-test between-group difference may stem not only from improvements in the experimental group but also from potential score changes within the control group. To unpack the within-group changes in the control group and interpret the source of post-test between-group differences, paired-samples t-tests were implemented for the control group. Descriptive statistics, mean changes and effect sizes of pre-test and post-test are summarized in Table 6. The paired test confirmed a significant decline in creative thinking for the control group: pre-test M = 22.79, SD = 4.405; post-test M = 21.41, SD = 3.340, with a mean difference of −1.38 and Cohen's d = −.35, and this decrease was statistically significant ($P = .007$). Other dimensions also showed slight score reductions without statistical significance: multi-perspective thinking (pre-test M = 18.34, SD = 3.538; post-test M = 17.70, SD = 3.202, Cohen'd = −.19), integrative thinking (pre-test M = 21.23, SD = 4.315; post-test M = 20.29, SD = 3.323, Cohen's d = −.24), critical thinking (pre-test M = 18.71, SD = 3.740; post-test M = 18.41, SD = 3.190,

Cohen's d = −.09), and overall deep thinking (pre-test M = 81.07, SD = 12.946; post-test M = 77.80, SD = 10.567, Cohen's d = −.28).

Table 5. Analysis of deep thinking levels between the experimental and control groups after two rounds of practical sessions.

Dimension	Group	Number of Cases	Mean	Standard Deviation	t	*P*
Multi-perspective thinking	Experimental group	58	19.91	2.940	3.854	.000
	Control group	56	17.70	3.202		
Integrative Thinking	Experimental group	58	23.09	2.969	4.748	.000
	Control group	56	20.29	3.323		
Critical Thinking	Experimental group	58	20.12	2.890	3.001	.003
	Control group	56	18.41	3.190		
Creative Thinking	Experimental group	58	23.38	3.675	2.990	.003
	Control group	56	21.41	3.340		
Deep Thinking (Overall)	Experimental group	58	86.50	10.617	4.382	.000
	Control group	56	77.80	10.567		

Table 6. Pre-test and post-test descriptive statistics for the control group.

Dimension	Group	Test Time	n	Mean (M)	SD	Mean Change (Post - Pre)	Cohen's d
Multi-perspective Thinking	Control Group	Pre-test	56	18.34	3.538	—	—
	Control Group	Post-test	56	17.70	3.202	−.64	−.19
Integrative Thinking	Control Group	Pre-test	56	21.23	4.315	—	—
	Control Group	Post-test	56	20.29	3.323	−.94	−.24
Critical Thinking	Control Group	Pre-test	56	18.71	3.740	—	—
	Control Group	Post-test	56	18.41	3.190	−.30	−.09
Creative Thinking	Control Group	Pre-test	56	22.79	4.405	—	—
	Control Group	Post-test	56	21.41	3.340	−1.38	−.35
Deep Thinking (Overall)	Control Group	Pre-test	56	81.07	12.946	—	—
	Control Group	Post-test	56	77.80	10.567	−3.27	−.28

Figure 4 visually compares pre-test and post-test mean scores of deep thinking dimensions for the control group, where error bars denote standard deviations. The figure clearly shows no spontaneous improvement of deep thinking indicators under traditional instruction; creative thinking declined significantly while other dimensions only dropped marginally. This finding suggests that traditional teaching cannot sustain the stable development of students' higher-order thinking

such as creative thinking, and further supports the positive effect of the intervention on deep thinking from the perspective of within-group changes in the control group.

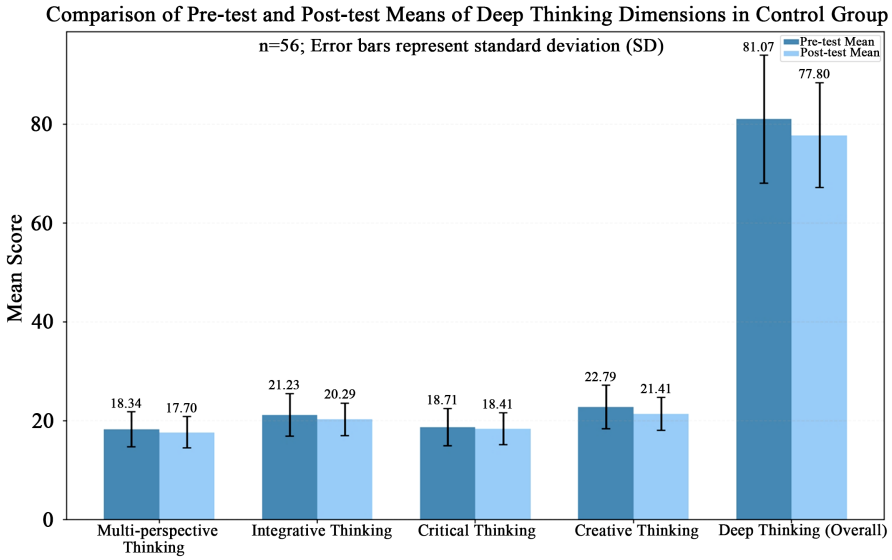


Figure 4. Comparison of pre-test and post-test means of deep thinking dimensions in the control group.

A paired-samples t-test was performed on the pre- and post-test data for the experimental group, with the results shown in **Table 7**. The analysis of the data revealed that, following two rounds of practical sessions, the students made marked progress in multi-perspective thinking, integrative thinking, critical thinking, creative thinking, and the overall dimension. The *P*-values for all dimensions were less than .05, indicating significant changes in all five areas. The paired-sample t-test for the control group showed that the overall mean decreased from 81.07 to 77.80 ($P = .090$), with a significant decline in creative thinking ($P = .007$). This indicates that, in a traditional project-based learning environment lacking systematic thinking scaffolding and intelligent support, the development of students' deep thinking faces a clear bottleneck, thereby indirectly confirming the necessity and effectiveness of introducing Teaching Agents and structured teaching models.

4.3.2. Individual Differences in the Effectiveness of Agent Use

A scatter plot was created showing the initial levels of deep thinking among students in the experimental group and their levels after two rounds of practice, dividing them into four quadrants: the "Potential Breakthrough group", the "Strength Growth group", the "Stable Performance group", and the "Individual Attention group". The use of the AI Agents had a positive impact on the majority of students, with the most significant effect observed in the "Potential Breakthrough group", indicating that the scaffolding provided by the AI Agents effectively activated their developmental potential; students in the "Strength Growth

group", meanwhile, demonstrated sustained and efficient growth from a high starting point. For students in the "Individual Attention group," further analysis of backend logs and individual interviews is required to investigate whether they failed to benefit due to insufficient usage frequency, cognitive barriers, or motivational issues, so that targeted interventions can be implemented. This indicates that the effectiveness of the AI assistant varies depending on individual adaptability and is closely related to students' cognitive readiness and the quality of human-computer interaction. As shown in Figure 5.

Table 7. Analysis of the experimental group's deep thinking levels following two rounds of practice.

Dimension	Time point	Number of Cases	Mean	Standard Deviation	Standard error of the mean	*P*
Multiperspective thinking	Pre-test	58	17.52	3.147	.413	.000
	Post-test	58	19.91	2.940	.386	
Integrative Thinking	Pre-test	58	20.76	3.445	.452	.000
	Post-test	58	23.09	2.969	.390	
Critical Thinking	Pre-test	58	19.05	2.453	.322	.005
	Post-test	58	20.12	2.890	.380	
Creative Thinking	Pre-test	58	21.78	3.038	.399	.000
	Post-test	58	23.38	3.675	.482	
Deep Thinking (Overall)	Pre-test	58	79.10	10.001	1.313	.000
	Post-test	58	86.50	10.617	1.394	

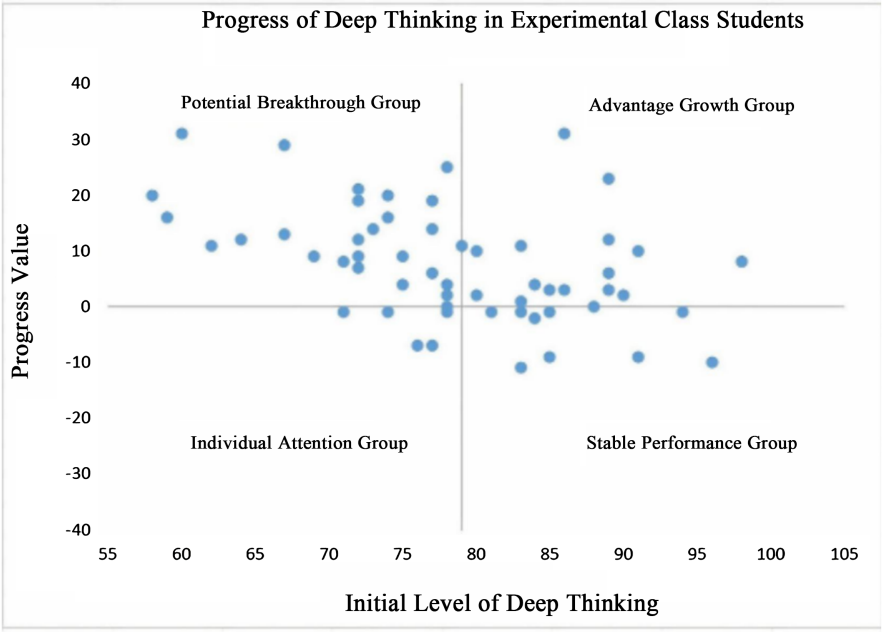


Figure 5. Progress in deep thinking within the experimental group.

4.3.3. The Promotion Effects Vary Structurally across Different Dimensions

There are structural imbalances in the promotional effects of the Teaching AI Agents across different dimensions of deep thinking. The cultivation of critical thinking proved particularly challenging; the first round of improvements did not reach a significant level, and it was only after the second round of intervention that significant, albeit relatively limited, progress was achieved. This phenomenon is closely related to the interaction design paradigm of intelligent agent systems: single agents are often constrained by their fixed response patterns and have limited capacity to stimulate deep questioning and dialectical thinking; whereas consciously guiding students, through instructional design, to critically examine the agents' outputs can demystify technological authority, thereby more effectively stimulating students' critical thinking (Hu et al., 2026).

It is worth noting that multi-perspective thinking and integrative thinking demonstrated sustained and significant improvement across both rounds of practice. This indicates that AI agents possess greater advantages in procedural and structured thinking training. Such as assisting with knowledge integration and providing diverse informational perspectives, and are more readily empowered by technological means. This uneven pattern of promotion suggests that the role of AI agents in fostering deep thinking should be clearly defined: they can serve as core enabling tools in dimensions of thinking that are easily structured, whereas for unstructured thinking, such as critical and creative thinking, which involves value judgments and breakthrough associations. It is necessary to rely on multi-agent systems to construct a supportive interactive ecosystem (Wu et al., 2024), while strengthening teachers' humanistic guidance and peer-to-peer collaborative debate, thereby jointly forming a complementary and symbiotic development system. Existing meta-analyses have also indicated that the promotional effects of educational agents on non-cognitive abilities, such as learning motivation and engagement, are generally modest and do not reach statistical significance, suggesting that they still have limitations in stimulating and sustaining students' intrinsic motivation (Wu et al., 2024).

5. Conclusion and Outlook

5.1. Research Conclusion

This study constructed a "Three-Dimensional Six-Stage" project-based teaching model integrating teaching AI agents and empirically tested its role in promoting the development of students' deep thinking. The study reached the following four key conclusions.

Firstly, high-quality Teaching Agents can effectively enhance students' levels of deep thinking; students in the experimental group performed significantly better than those in the control group across all dimensions, and the effectiveness of the Teaching Agents was also corroborated by feedback from teacher and student interviews;

Secondly, there are individual differences in the effectiveness of the AI agents' use; this is closely related to students' cognitive readiness and the quality of human-computer interaction, necessitating differentiated support for students at different cognitive levels;

Thirdly, there are structural differences in the promotional effects across the various dimensions; critical thinking must be stimulated by guiding students to critically examine the outputs of the Teaching Agents;

Fourthly, as one of the three key elements of the project, the Teaching Agents—which align with the philosophy of deep thinking—can also effectively enhance the quality of project work. Results from two rounds of practical work demonstrate that, with the assistance of intelligent agents, the overall quality, completeness, and innovativeness of students' work have seen significant and consistent improvements: through structured cognitive scaffolding, the intelligent agents have raised the level of completion and standardization of the work, ensuring an overall rise in the quality baseline; furthermore, by stimulating deep thinking activities, they have enhanced the complexity and innovativeness of outstanding work. This implies that intelligent agents are not merely tools for task completion, but also catalysts for cognitive engagement.

5.2. Reflections and Outlook

This study still has certain limitations: firstly, the sample was drawn from only one school, and the sample size was limited; secondly, the research period was relatively short; thirdly, the assessment of deep thinking relied on scales, which failed to capture the dynamic processes of thought accurately.

This study's intervention was delivered at the class level, whereas the measures were collected at the individual level. Because only one intact class was assigned to each condition and no data on the students' daily performance were collected, the analysis did not take into account the dependence of students within the same classroom, and the intervention effect cannot be separated from class-level effects. The results of this study should therefore be regarded as exploratory associations, and the generalizability of its conclusions should be treated with caution.

In summary, future research should expand the scope of the sample and extend the study period to explore the role of multi-agent systems in supporting students' higher-order thinking. It should further optimize the personalized support provided by the agents and the design of human-computer interaction, thereby providing a more robust theoretical and practical foundation for the collaborative development of students' deep thinking through human-computer collaboration.

Author Contributions

Conceptualization, J.W. and X.M.; methodology, J.W.; software, J.W.; validation, J.W., X.M. and T.W.; formal analysis, J.W.; investigation, J.W. and T.W.; resources, J.W.; data curation, J.W.; writing—original draft preparation, J.W.; writing—review and editing, X.M. and T.W.; visualization, J.W.; supervision, X.M.;

project administration, X.M. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Anderson, L. W. (2009). *Bloom's Taxonomy of Educational Objectives.* Foreign Language Teaching and Research Press.

Core Competencies Research Group (2016). Core Competencies for Chinese Student Development. *Chinese Journal of Education, No. 10,* 1-3. (In Chinese) https://www.zhangqiaokeyan.com/academic-journal-cn_journal-chinese-society-education_thesis/020129520269.html

Dai, X., & Yang, X. (2023). Explorations into Cultivating Higher-Order Thinking in Senior Secondary Physics: A Case Study of the "Periodic Motion" Unit. *Physics Teaching, 45,* 2-6. (In Chinese) https://wljx.ecnu.edu.cn/assets/userfiles/PDF/141651_20230316195048.pdf

Design-Based Research Collective (2003). Design-Based Research: An Emerging Paradigm for Educational Inquiry. *Educational Researcher, 32,* 5-8. https://doi.org/10.3102/0013189X032001005

Facione, P. A. (1990). *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction (ERIC Document No. ED 315423).* ERIC.

Guo, H. (2016). Deep Learning and Its Significance. *Curriculum, Teaching Material and Method, 36,* 25-32. (In Chinese) https://doi.org/10.19877/j.cnki.kcjcjf.2016.11.005

Guo, J., Rong, Q., & Hao, J. (2020). A Review of Research on the Application of Artificial Intelligence in Teaching Abroad. *E-Education Research, 41,* 91-98, 107. (In Chinese) https://doi.org/10.13811/j.cnki.eer.2020.02.013

He, K. (1997). Constructivist Teaching Models, Teaching Methods and Instructional Design. *Journal of Beijing Normal University (Social Sciences), No. 5,* 74-81. (In Chinese) https://cit.bnu.edu.cn/xydt/53642.html

Holmes, W., & Miao, F. (2023). *Guidance for Generative AI in Education and Research.* UNESCO Publishing.

Hu, J., Yan, T., & Zhang, T. (2026). Overcoming Cognitive Outsourcing: Design of a Human-Computer Collaborative Debate System Aimed at Critical Thinking. *E-Education Research, 47,* 54-63. (In Chinese) https://doi.org/10.13811/j.cnki.eer.2026.03.007

Inhelder, B., & Piaget, J. (1958). *The Growth of Logical Thinking from Childhood to Adolescence: An Essay on the Construction of Formal Operational Structures.* Basic Books. https://doi.org/10.1037/10034-000

Jia, X. (2008). How Primary and Secondary School Teachers Conduct Topic Research (IV): Educational Experimental Research Method in Educational Research. *Theory and Practice of Education, No. 11,* 44-46. (In Chinese) https://kns.cnki.net/kcms2/article/abstract?v=YsZ6lxCJneQ3lNOSYhaVJeUhu1PFNzVvprsaNTzzqNEDEYt9mrHlglN21WbQkHE6Q0MA2y3dQSVXSMK6oe_RIGVV9w0cJgWtjQ0Z0LW2QvTn0_ob0OFtkx3ujAqByttL3F8CKLTCUtjtJiATJbZVIRmXF6xcx-h4d8cId77r3pkhFWHmuh4UNSmIevvJUsx8fAN-EVBz8Hw=&uniplatform=NZKPT&language=CHS

Li, F., Xi, S., & Xiong, Z. (2024). Information Technology Curriculum Standards Oriented to Digital Literacy and Skills and Their Teaching Implementation. *China Distance Education, 44,* 52-59, 96. (In Chinese) https://doi.org/10.13541/j.cnki.chinade.2024.11.003

Li, H., & Wang, W. (2024). The Human-Machine Debate Inquiry Method: An Exploration of a Model for Cultivating Students' Higher-Order Thinking Skills Supported by a Debate-Oriented Intelligent Conversational Robot. *E-Education Research, 45,* 106-112, 128. (In Chinese) https://doi.org/10.13811/j.cnki.eer.2024.03.015

Liu, J., & Zhong, Z. (2002). A Study on the Project-Based Learning (PBL) Model. *Foreign Education Research, No. 11,* 18-22. (In Chinese) https://www.cnki.com.cn/Article/CJFDTOTAL-WGJY200211004.htm

Liu, Q., Ba, S., & Luo, L. (2019). A Review of the Mechanism of Educational Agents'Effects on Cognitive Learning. *Journal of Distance Education, 37,* 35-44. (In Chinese) https://doi.org/10.15881/j.cnki.cn33-1304/g4.2019.05.005

Lou, L. (2025). *Theory and Practice of Unit-Based Teaching in Senior Secondary English Aimed at Deep Learning.* Master's Thesis, Zhejiang Normal University.

Lu, Y., Yu, J., & Chen, P. (2024). Construction and Application of Teaching Agents Based on Large Models. *China Educational Technology, No. 7,* 99-108. (In Chinese) https://kns.cnki.net/kcms2/article/abstract?v=YsZ6lxCJneTZ4C9kVIEMNQzLNMH78g5UfD2suoe_oSyOMOlEDS4ui6QQDCViP98Txgr2bk1KMnw8-8YK07P0M0Z9zpVzVNqyGjDlrr85L5O3K7WsATqq0_97xvJqz38szjtfnWcTzG0y22d2OjXltcPYgyeRB0G9SP2FdWRnsLO_NF77UbSierHmS8Gfu7zu03EyfrrgpckHj6wDG4EHeg==&uniplatform=NZKPT&language=CHS

Ministry of Education of the People's Republic of China (2022). *Curriculum Standards for Information Technology in Compulsory Education (2022 Edition).* Beijing Normal University Press.

Sang, G., & Wang, J. (2025). Digital Technology Empowering Project-Based Learning: Real-World Scenarios, Logical Frameworks, and Ecosystems. *E-Education Research, 46,* 22-29. (In Chinese) https://doi.org/10.13811/j.cnki.eer.2025.09.002

Shao, Z. (2018). On Deep Thinking in Educational Research. *Research in Higher Education of China, 39,* 48-54. (In Chinese) https://fx.gfkd.chaoxing.com/detail_38502727e7500f2648270fa1daa487e70ab3213fc33c70ba1921b0a3ea255101fc1cf1fbb4666ae6535f4c85ff926e2941f56312f7b213b37b9430f6894f0614148c02d174aa5ff8d4b6a9ee27091aaa

Wang, D., Chen, Z., Shao, W. et al. (2025). From "Negative Energy" to "Empowerment": Design of Thought-Chain Prompts Based on LLMs and Construction of AI Agents for Teaching and Research—Taking Intelligent Analysis of Classroom Teaching as an Example. *China Educational Technology, No. 3,* 111-117. (In Chinese) https://doi.org/10.3969/j.issn.1006-9860.2025.03.015

Williams, F. E. (2003). The Williams Creativity Propensity Scale. *China New Era, No. 22,* 89-90. (In Chinese)

Wu, Y., Jiang, Y., Chen, Y., & Zhang, W. (2024). Multi-Agent Systems Supported by Large Language Models: Technical Pathways, Educational Applications, and Future Prospects. *Open Education Research, 30,* 63-75. (In Chinese) https://doi.org/10.13966/j.cnki.kfjyyj.2024.05.007

Ye, X. (2023). *Deep Thinking: A Methodology for Leapfrog Growth That Cuts through Complexity to Reach the Essence.* Posts & Telecom Press.

Yuan, L., & Wang, Y. (2023). Dilemmas and Countermeasures of Interdisciplinary Teach-

ing in Primary and Secondary Schools under the Background of Digital Education. *E-Education Research, 44,* 87-94. (In Chinese) https://doi.org/10.13811/j.cnki.eer.2023.12.012

Yuan, L., Xu, J., & Liang, S. (2025). An Agent-Empowered Human-Machine Collaborative Teaching Support Model for Interdisciplinary Themes. *E-Education Research, 46,* 87-94. (In Chinese) https://doi.org/10.13811/j.cnki.eer.2025.03.012

Zhai, H. (2020). PBL-Based Project Design for Junior Secondary School Information Technology. *Theory and Practice of Education, 40,* 62-64. (In Chinese) https://kns.cnki.net/kcms2/article/abstract?v=YsZ6lxCJneT03juvuoUBN07Y_k0WVDcpjXGxJSnmSLhf6AD238Fc5OlObryav50nbdX9PtOY9Y4LQGQvE9UcE4zPVFIBCtF1_XrlrrOEhDU1dT3MktUFa4tZ4O6UI8RL2ITZB8v_L33aJSbfMrRI0LUhghmTpogS9GC-ZqShxcXMEhzrFe4rX_7ahl_7Rmhz6xY8WOhu4wtvSPFLTuCeQ==&uniplatform=NZKPT&language=CHS

Zhang, H., Wu, X., & Wang, J. (2014). Objectives of Deep Learning and the Construction of an Evaluation System. *China Educational Technology, No. 7,* 51-55. (In Chinese) https://doi.org/10.3969/j.issn.1006-9860.2014.07.008

Zhang, Y. (2024). Creating a Generative Classroom: The Development and Application of Teaching Agents. *China Education News.*

Zuo, H., Zhu, M., & Cao, Y. (2022). Dimensions, Cultivation Strategies, and Implications of Students' Composite Competencies from the Perspective of "OECD 2030". *Theory and Practice of Education, 42,* 9-14. (In Chinese) https://jyll.cbpt.cnki.net/portal/journal/portal/client/paper/61239359fe47a6e7fea9be271ffd0225