Computer Standards & Interfaces 97 (2026) 104094

Contents lists available at ScienceDirect

Computer Standards & Interfaces

journal homepage: www.elsevier.com/locate/csi

How AI agents transform reflective practices: A three-semester comparative study in socially shared regulation of learning

Yumin Zheng [a], Fengjiao Tu [b], Fengfang Shu [a,c], Chaowang Shang [a,*], Lulu Chen [a], Jiang Meng [a]

[a] Faculty of Artificial Intelligence in Education, Central China Normal University, Wuhan 430079, China
[b] Department of Information Science, University of North Texas, 3940 North Elm, Denton, Texas, 76203, USA
[c] Institute of Open Education, Wuhan Vocational College of Software and Engineering, Wuhan Open University, Wuhan, China

ARTICLE INFO

Keywords:
Artificial intelligence agent
Socially shared regulation of learning
Reflection quality
Collaborative learning
Generative artificial intelligence

ABSTRACT

High-quality reflection has been a challenging barrier in the socially shared regulation of learning (SSRL). Especially with the emergence of generative artificial intelligence (GAI), traditional methods such as reflection reports may increase the students' risk of superficial reflection. This study uses an artificial intelligence agent (AI agent) to design a reflection assistant, which aims to enhance students' reflection ability through continuous questioning and real-time, content-specific feedback based on their written reflections. Through a comparative experiment conducted over three semesters, this study demonstrates the different impacts of three reflection methods, reflection reports, reflection short-answer questions, and AI agents, on the quality of university students' reflections. The results indicate that there is a significant difference in the quality of reflection among the three reflection methods. Students using AI agents show the highest levels of reflection, characterized primarily by connective reflection and critical reflection. Epistemic network analysis further reveals that the AI agent reflection method is more effective in improving the reflection quality of low-performance teams than that of high-performance teams. This expands AI agents' use in SSRL reflection, introduces new methods for the GAI era, and provides practical experience and reflection intervention strategies for teachers and instructional designers in SSRL.

1. Introduction

With the rapid advancement of generative artificial intelligence (GAI), numerous challenges in collaborative learning have been addressed with innovative solutions [1,2]. GAI applications, represented by artificial intelligence agents (AI agents), have introduced revolutionary transformations to education. These transformations are mainly due to the powerful expert-level conversational abilities and user-friendly accessibility [3].

The socially shared regulation of learning (SSRL) strategy serves as a crucial mechanism for enhancing learning outcomes in collaborative learning [4]. Through the SSRL strategy, learners collaboratively set goals and monitor progress, thereby improving their performance [5]. Reflection is a critical component of SSRL, aiding learners in recognizing and refining their learning processes [6]. However, achieving high-quality reflection remains a challenge [7].

There are various methods to enhance reflection quality in SSRL, such as providing prompts and templates in reflection reports [8]. Nowadays, these traditional methods fall short of addressing the challenges posed by GAI [9]. Students may easily rely on tools like ChatGPT to complete short-answer questions, journals, and reports. Kiy [10] has shown that 76 % of university students use ChatGPT for their assignments, with the percentage being even higher among software engineering students, reaching 93 % [11]. The widespread use of GAI has profoundly transformed traditional methods of learning and teaching, and this era calls for new approaches to reflection.

AI agents are computing systems with capabilities for autonomous perception, decision making, and action [12]. They use GAI to learn, reason, and perform corresponding tasks or actions from the surrounding environment and input information. To enable practical implementation, rule-based AI agents have been developed that require no programming and can be deployed simply by defining task objectives and roles via prompts. In educational contexts, these rule-based AI agents are commonly used for personalized instruction and intelligent tutoring due to their ability to engage in real-time dialogue and provide immediate feedback [13].

* Corresponding author.
E-mail address: phdzhengyumin@mails.ccnu.edu.cn (C. Shang).

https://doi.org/10.1016/j.csi.2025.104094
Received 1 February 2025; Received in revised form 28 October 2025; Accepted 10 November 2025
Available online 11 November 2025

The rule-based AI agent provides an effective approach for supporting SSRL reflection. Instructors can set specific SSRL task directions, and the agent guides students based on the reflection checklist while adaptively generating questions according to students' responses. Each follow-up question is dynamically generated based on the student's prior answers and the specific SSRL task, making it difficult for students to rely on external AI tools like ChatGPT to provide generic responses. This continuous dialogue mechanism supports deeper, more analytical reflection and reduces the risk of superficial reflection [14]. Despite AI agents having broad application prospects, current research on improving learners' reflection quality by AI agents remains limited and requires further in-depth exploration.

Against this backdrop, this study introduces a rule-based AI agent reflection assistant within the SSRL framework to help learners enhance their reflection quality. This study aims to examine the impact of the AI agent on SSRL reflection quality by comparing three reflection methods: reflection reports, short-answer reflection questions, and the AI agent-based reflection. In addition, different methods may lead to different reflection qualities among learners in high and low-performance teams [15]. Therefore, we further explored the differences in reflection quality between high and low-performance teams when using these three reflection methods. We proposed the following research questions:

RQ1: How does the AI agent reflection assistant affect learners' reflection quality in SSRL?
RQ2: What differences do high and low-performance teams show in reflection quality when using the three reflection methods?

This study conducted a three-semester comparative teaching experiment to evaluate the impact of AI agents and two traditional reflection methods (reflection reports and short-answer questions) on university students' reflection quality. Using statistical analysis, content analysis, and epistemic network analysis (ENA), this study examines the effectiveness of AI agents in enhancing university students' reflection quality in SSRL.

The main contributions of this study are summarized as follows:

- We introduce a practical SSRL activity, providing educators with a valuable instructional framework for facilitating collaborative learning.
- We integrated an AI agent reflection assistant in SSRL and provided a comprehensive debugging process, offering instructors examples and considerations of AI agent implementation.
- We revealed the reflection quality differences between high and low-performance teams in various reflection approaches and demonstrated the advantages of the AI agent for low-performance teams.

The research is organized as follows: Section 2 reviews prior research on AI agents in education, SSRL theory, and reflection. Section 3 describes the participants, research design, and methods for data collection and analysis. Section 4 compares reflection quality across the three methods and examines differences between high and low-performance teams using ENA. Section 5 discusses the results and implications. The paper concludes with a summary and potential directions for future research.

2. Literature review

To explore the impact of AI agents on learning processes, it is essential to examine their application in education, followed by a discussion on SSRL and reflection.

2.1. AI agents in teaching

Generative Artificial Intelligence (GAI), defined as AI systems capable of autonomous learning and content generation, has been widely applied in education [16]. It can support collaborative learning through personalized instruction, real-time feedback, and intelligent assessment [17]. AI agents, a form of GAI equipped with autonomous learning and decision-making capacities, have emerged as key instructional tools in global educational research.

Empirical studies have shown that AI agents significantly improve student engagement [18,19], learning motivation [20,21], and academic performance [22]. AI agents exist in various forms, such as chatbots [23], intelligent tutoring systems (ITS; [24]), embodied conversational agents (ECA; [25,26]), and intelligent virtual assistants (IVA; [13,27]). Among these, GAI-based chatbots have been widely adopted in education due to their customizable roles and flexible deployment. The present study focuses on this type of conversational AI agent.

In higher education, AI agents have been shown to support higher-order thinking skills, such as critical thinking, metacognition, and problem-solving [23,28,29]. In these studies, GAI was embedded within structured reflection activities, allowing students to engage in guided reflective processes targeting specific cognitive skills. For example, Hong et al. [29] employed AI to handle lower-level tasks in essay writing, enabling students to focus on evaluation and reflection, thereby enhancing critical thinking. Chen et al. [28] implemented metacognitive strategy-supported AI agents that prompted process-oriented reflection and multi-perspective discussion, improving metacognitive skills. Zhou et al. [23] situated reflection within a self-regulated learning framework, showing that GAI-supported reflection indirectly benefits critical thinking and problem-solving.

Although these studies demonstrate that AI agents can enhance higher-order thinking, reflection itself has often been treated merely as a learning process rather than a measurable skill. Reflection is a core component of higher-order thinking and an essential learning competency for 21st-century university students. Empirical evidence directly examining the impact of AI agents on learners' reflective abilities, particularly in collaborative learning environments, remains scarce. Investigating this relationship is therefore necessary to understand how AI agents can effectively support the development of reflection.

2.2. Socially shared regulation of learning and reflection

Collaborative learning includes three primary types of regulation: self-regulation (SR), co-regulation (CoR), and socially shared regulation (SSR) [30,31]. Based on SSR theory, socially shared regulation of learning (SSRL) is an emerging collaborative learning strategy emphasizing mutual support and feedback among team members. The strategy consists of four key stages: goal setting, task distribution, progress monitoring, and reflection evaluation [32–35]. Research indicates that the SSRL strategy has a positive impact on collaborative learning [36]. Learners may enhance their awareness of the collaborative process and facilitate the activation of regulatory processes through SSRL [4]. And SSRL helps to enhance learners' cognitive and metacognitive abilities, boosting learning motivation and engagement [37,38]. Additionally, SSRL fosters communication among team members, improving collaborative efficiency [39]. Thus, SSRL has been widely incorporated into collaborative learning and plays a significant role in enhancing various learner abilities.

Reflection quality is a key indicator for assessing the success of SSRL [39]. High-quality reflection is an indispensable component of SSRL, as it enables learners to examine and evaluate their learning processes and outcomes [40]. Unlike conventional collaborative learning, the reflection content in SSRL emphasizes the process of mutual regulation and monitoring among group members. However, since reflection is the final stage of SSRL, educators often overlook its significance [41]. Teachers' lack of emphasis on the reflection stage may lead to low-quality reflection among students [42]. Achieving high-quality SSRL reflection remains a persistent challenge for educators and students [43].

To enhance students' reflective abilities, it is essential to focus on the definition of reflection. Dewey [44] defined reflection as a continuous process of exploring and evaluating experiences, which helps individuals gain a deeper understanding of their behaviors and outcomes. Zimmerman [45] further emphasized that self-reflection is a complex learning process involving various aspects of self-monitoring, such as self-assessment and feedback on contributions. In the theory of SSRL, reflection encompasses not only self-assessment but also shared monitoring processes with others [39]. These theories provide support for exploring and promoting the reflective process.

In reflective activities, teachers can support students' deep learning and reflective abilities through various intervention strategies, such as scaffolding, reflective prompts, and feedback [46]. Reflective scaffolding involves providing structured guidance to help students more effectively review and analyze their learning experiences [47]. When designing reflection tasks for SSRL, teachers often utilize the SSRL reflection scaffolds developed by Panadero et al. [48]. Additionally, reflective prompts and guiding questions steer students toward specific directions for reflection, assisting them in identifying potential barriers and challenges in their learning [49]. Feedback provides learners with suggestions or information to improve task performance, helping them optimize both their reflection and learning processes [50]. From a cognitive perspective, feedback serves as guidance to enhance students' task performance [51]. Timely feedback on students' reflections not only improves the quality of subsequent reflections but also deepens their understanding of reflective concepts [52].

Reflection journals, reflection reports, and reflection short-answer questions have been explored to improve reflection quality [53,54]. However, the traditional methods may not adapt to the advancements of GAI. These require students to submit longer texts, which inevitably causes a risk of superficial reflections due to the use of GAI. Some scholars have also modified reflection methods from a technological perspective by using various reflection platforms, such as Google Docs [55], Flipgrid [56], the VEO app [57], and Wiki [58]. However, these platforms primarily offer static or limited interaction, which constrains students' ability to adaptively engage in reflective processes. The low-quality reflection issues in SSRL urgently require new solutions.

Although GAI poses challenges to traditional reflection methods, it also offers new solutions. AI agents are increasingly regarded as effective tools for supporting reflection practices. Research indicates that the use of AI agents in reflection activities may enhance students' learning motivation and engagement [59]. Teachers can use AI agents to design reflection scaffolding, assisting learners in conducting more in-depth and systematic reflections [60]. In addition, AI agents may enhance reflection quality through data analysis and intelligent feedback [61]. Therefore, AI agents demonstrate potential in addressing the issue of improving SSRL reflection quality.

Thus, this study designed a reflection assistant by AI agents to enhance university students' reflection quality in SSRL. Statistical analysis, content analysis, and ENA were employed to collect and analyze textual data related to reflection quality. By comparing the AI agent reflection assistant with traditional SSRL strategy reflection scaffolding, this study analyzed the differences in reflection content and reflection levels among university students across three methods. Additionally, previous research suggests that high and low-performance teams may experience different effects from various reflection methods [62]. Therefore, this study further explores the differences between high and low-performance teams when using three reflection methods. This study provides new theoretical evidence for using AI agents in SSRL reflection practices.

3. Methodology

This study employed a quasi-experiment to explore the differences among three reflection methods in SSRL. And examine whether AI agents improve the reflection quality of university students. Firstly, we provided information about the participants and the course. Then, we elaborated on the activities of SSRL and the design process of the AI agent. Lastly, we discussed the coding scheme for reflection quality and provided the methodology for data collection and analysis.

3.1. Participants

The participants were from the course "Internet Thinking and Digital Self-Learning" over three semesters: Spring 2023, Fall 2023, and Spring 2024. A total of 97 undergraduate students, aged 18 to 22, took part in this study (Table 1).

At the beginning of each semester, students completed a pre-test using the CThQ [63], which assesses six cognitive dimensions: memory, comprehension, application, analysis, evaluation, and creation (overall reliability$\alpha = 0.87$). According to Dewey [64], critical thinking is a deepening and extension of reflective thinking, with high consistency in cognitive processing, reasoning, and evidence evaluation. The CThQ pre-test provides a valid proxy for students' baseline reflection levels. One-way ANOVA indicated no significant differences in pre-test total scores among the three groups (Group 1: $M = 105.07$, SD = 6.13; Group 2: $M = 103.72$, SD = 4.19; Group 3: $M = 105.22$, SD = 4.24), $F(2, 86) = 1.33$, $p = 0.27$, suggesting comparable reflection abilities across groups prior to the intervention.

Participants were divided into 3 groups, each employing a different reflection method, and within each group, students were further divided into teams using random assignment to minimize potential biases arising from prior academic performance, familiarity, or interpersonal preference. Random assignment was chosen over self-selection or instructor-based grouping to ensure group equivalence and to enhance the internal validity of the comparative analysis [65].

The first group (G1), consisting of 31 students from the Spring 2023 semester, conducted reflection reports and were further divided into 7 teams. The second group (G2), consisting of 30 students from the Fall 2023 semester, conducted short-answer reflections and were divided into 7 teams. The third group (G3), consisting of 36 students from the Spring 2024 semester, conducted reflections through continuous questioning by an AI agent and were divided into 9 teams. Additional information about the participants is provided in Table 1.

3.2. Design of socially shared regulation of learning activities

During the 4-week activity, students collaborated in teams to produce micro-lesson videos lasting 5 to 8 min. The activity was divided into 4 stages, each lasting one week (Table 2).

In the first week (goal setting), students were required to establish a common goal, select the video's theme, and outline the content framework. Then, they submitted a project proposal detailing the topic, objectives, task distribution, and timeline. In the second week (task distribution), the teams followed their project plan to allocate tasks and begin executing the project. The instructor provided guidance and suggestions throughout this process. In the third week (progress monitoring), each team submitted a video sample that was between 1 and 2 min long. The instructor conducted an initial evaluation based on the sample and suggested improvement. Students refined and adjusted their video production based on the feedback. In the fourth week (reflection evaluation), students submitted their completed micro-lesson videos

Table 1
Participant and group information.

Group	Course	Reflection method	Team	Participant	Female	Male
G1	Spring 2023	Report	7	31	17	14
G2	Fall 2023	Short-answer questions	7	30	19	11
G3	Spring 2024	AI reflection assistant	9	36	20	16

Table 2
The stages of SSRL.

Week	SSRL stages	Description
1	Goal setting	Students discuss the goal, theme, and framework.
2	Task distribution	Students allocate tasks and make the micro lesson videos.
3	Progress monitoring	Students monitor the task and submit a video sample.
4	Reflection evaluation	Students submit completed micro-lesson videos and individual reflection assignments.

and individual reflection assignments (employing different reflection methods for each of the three semesters). Finally, a reflection-sharing session was held in class, where students exchanged learning experiences and insights.

3.3. Design of the three reflection methods

Prior to the reflection phase, all students completed a four-week SSRL activity in which the instructor introduced and practiced the four SSRL stages. Consequently, all reflections were anchored in the teams' performance across these four stages. In G1, the reflection remained open-ended within this framework and only specified a minimum length of at least 200 words (no SSRL question list was provided).

In G2, students conducted individual reflections through short-answer questions. The guiding questions were derived from the SSRL reflection scaffolding [48]. For example, questions included "What is the group's current assignment?" and "What obstacles might the group encounter?"

G3 students used the AI agent reflection assistant for their reflections. After the SSRL task, the instructor provided students with a quick response code (QR code) linking to the AI agent's website. Students scanned the QR code with their phones to initiate a conversation with the AI agent. Each student completed the reflection task through the dialogue.

The development process of the AI agent is illustrated in Fig. 1. The AI agent reflection assistant, Crystal, was developed using the Coze platform (https://www.coze.cn/). The AI agent consists of 4 core components, with Part A being the AI agent's name, Part B defining the role setting and response logic, Part C specifying the conversational experience, such as the opening dialogue, and Part D serving as the preview interface. Developing the AI agent requires following these operational steps.

Step 1: Create the AI agent and assign it to the name Crystal (as shown in Fig. 1, Part A). Define it as the reflection assistant for the course "Internet Thinking and Digital Self-Learning. Set its duty to guide students in completing tasks (as shown in Fig. 1 Part B) and design the opening statement (as shown in Fig. 1 Part C).
Step 2: Set up the reflection task (as shown in Fig. 1, Part B). Input all the questions from the SSRL reflection scaffolding developed by Panadero et al. [48] into the AI agents as the question base. This ensures a logical flow of questions from the AI agent to the students, preventing task misdirection. In addition, the AI agent was not restricted to this fixed list but generated follow-up questions, particularly "Why" questions, based on the students' specific answers, which reflected its adaptiveness.
Step 3: Set up the response rule (as shown in Fig. 1, Part B). Establish the response rules for the AI agent:
a. Ask only one reflection question per interaction.
b. Provide encouraging feedback that adapts dynamically after each response (e.g., "You did a great job", "Your reflection is very insightful").
c. Avoid using academic terms.
d. Use only special interrogative questions (e.g., "What, " "Why"), with follow-up questions adjusted according to students' responses.
e. After answering all questions, conclude the conversation and express gratitude.
Step 4: Testing and deployment (as shown in Fig. 1, Part D). Check the conversation flow and ensure the AI agent's smooth and effective interactions. Select 5 students for a second round of testing to ensure

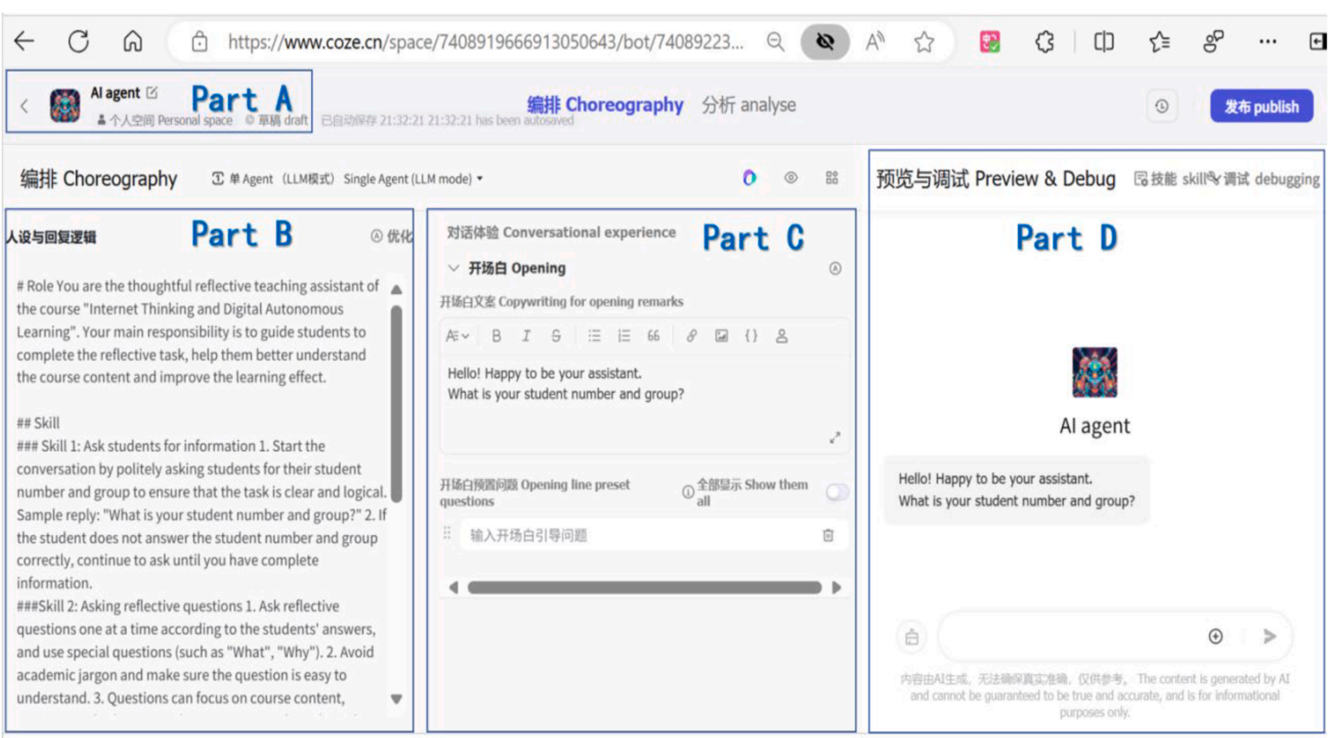


Fig. 1. AI agent development interface on the Coze platform.

the conversation flows smoothly. Once confirmed, the AI agent can be deployed and available to all students.

3.4. Experimental procedure

The experimental procedure is illustrated in Fig. 2. As described in the Participants section, all students completed the CThQ [63] as a pre-test before the course. They then attended a 16-week course covering basic concepts. All students were taught by the same instructor, with the course content, teaching methods, and learning resources remaining entirely consistent across the three semesters. Students participated in a 4-week group collaboration activity, “creating micro lesson videos”, conducted using the SSRL strategy. After the group activity finished, each student was assigned an individual reflection task. G1 and G2 used traditional reflection methods, with G1 completing reflection reports and G2 answering short-answer questions. G3 employed a new reflection method, utilizing the AI agent reflection assistant.

3.5. Data collection and analysis

After the three semesters, the reflection texts of all students were collected and anonymized. G1 produced 31 reflection reports totaling 8032 words. G2 submitted 30 reflection short-answer texts, totaling 15,468 words. G3′s AI agent reflection assistant dialogues comprised 36 submissions, totaling 16,801 words (excluding the AI agent’s questions).

Content analysis was used to process the reflection texts. Through systematic coding rules, this method reduced the influence of subjective judgment and personal bias, thereby providing more objective results. The coding scheme consists of two parts: reflection level and reflection content, as shown in Table 3. The reflection level coding scheme is based on Plack et al. [66], and it is used to assess the overall reflection level of learners, categorized into no reflection (NOR), low reflection (LOWR), and high reflection (HIGHR). The reflection content coding scheme is based on Wang et al. [67] and is used to explore the differences in the types of learners’ reflection content. The reflection content is categorized into 4 types: descriptive reflection (DESR), explanatory reflection (EXPR), connected reflection (CONR), and critical reflection (CRIR), with reflection quality progressively increasing across these categories.

The reflection texts in the reflection reports and short-answer reflections were relatively longer, while those in the AI agent dialogues were shorter. To mitigate the differences caused by these length discrepancies, this study used a single complete sentence as the minimum coding unit. For example, the statement “As the group leader, I am quite decisive. I directly assigned tasks to everyone, and the group was supportive.” should be coded as two separate sentences.

Table 3
Learner reflection quality coding scheme.

Categories	Coding	Description
Reflection level	NOR	Lacking a reflection mindset.
	LOWR	Having a reflective mindset involves reviewing experiences, describing facts and feelings, and reflecting on what has been learned. It also encompasses the ability to connect new knowledge with existing knowledge and to improve learning strategies.
	HIGHR	Critically analyzing the current situation, attempting to view problems from different perspectives, forming new viewpoints from available resources, and seeking to test hypotheses.
Reflection content	DESR	A description of “what” the object of reflection is.
	EXPR	An explanation of the causes behind the object of reflection, addressing the “why” often indicated by keywords such as “in order to”, "due to", or "so as to".
	CONR	Understanding whether the object of reflection has changed across different times and contexts, coupled with an analysis of the reasons for these changes and their impact on behavior, represents a higher level of analysis concerning the “what” and “why”.
	CRIR	It identifies personal or team issues and analyzes them with theory and practice to solve problems, focusing on “how” to achieve self-reconstruction. This may include keywords like “needs improvement” or “next stage”.

To ensure reliability, a coding discussion group comprised two experts and two professional coders. First, the two coders preliminarily coded the first 10 % of the reflection texts. In cases of disagreement, they consulted with the experts to reach a consensus. After training and repeated practice, the coders achieved a high level of consistency. The coders strictly adhered to the revised coding scheme during the formal coding process. After coding, inter-coder reliability was calculated, yielding a Cohen’s kappa coefficient of 0.87, indicating that the coding process had a high level of reliability. The coders consulted with experts for different coding results and ultimately reached an agreement.

After coding the reflection texts using the content analysis, ENA was employed to conduct a fine-grained analysis of the reflection data. Content analysis excels at systematically and objectively analyzing large volumes of textual content. ENA focuses on uncovering the complex relational networks between elements, such as reflection levels. The combination of the two methods allows for attention to both the characteristics of the text itself and the internal relationships between the content elements. Additionally, the ENA Webkit (http://www.epistemicnetwork.org/) provides a stable environment for data analysis.

To investigate the differences in reflection quality between the high and low-performance teams, we assessed the micro lesson videos completed by students in SSRL. The videos were assessed by two experts

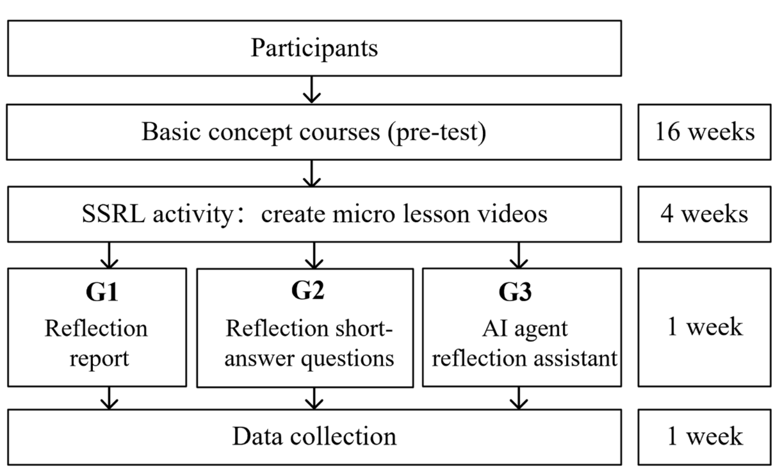


Fig. 2. Experimental procedure.

in education, each with over 10 years of teaching experience. The evaluation criteria included the following categories, with topic selection worth 10 points, instructional design 40 points, content completeness 20 points, audio-visual quality 20 points, and artistry 10 points. Each group received a score ranging from 0 to 100 points. The two experts thoroughly discussed the evaluation criteria to ensure consistency in scoring and then individually assessed all instructional designs and materials. The scoring consistency between the two experts (Spearman correlation coefficient) was 0.86 ($p < 0.01$).

The average score from both experts was used as the final score for each group (Table 4). The grouping criteria for high and low performing teams proposed by Hou [68] have been widely adopted by scholars [69]. In this study, based on those criteria, the top 15 % of teams were classified as the high-performance teams, including G1-team7, G2-team2, and G3-team1. The bottom 15 % of teams were classified as the low-performance teams, including G1-team5, G2-team6, and G3-team4. Using ENA, we further explored the differences between the high and low-performance teams of students.

3.6. IRB approval and AI agent data privacy

This study has received approval from the Institutional Review Board (IRB) of the university, ensuring that all ethical standards are met. All students participated voluntarily, fully aware of the study's purpose and procedures, and signed informed consent forms prior to the commencement of the experiment. In addition, to protect participants' privacy, all data collected during the study were anonymized.

All conversations on the Coze platform were fully anonymized, and students were reminded before using the platform not to enter any personal or sensitive information (such as name, student ID, gender, or school). Data was labeled only with class sequence numbers (e.g., Student 1, Student 2), and access was strictly limited to the research team. In addition, all students signed the Coze platform's privacy protection agreement, and the platform further ensures data security through anonymization and encryption techniques.

4. Results

The results are organized to address the key research questions regarding the effectiveness of the AI agent and the differences in reflection quality across various reflection methods.

4.1. How does the AI agent reflection assistant affect learners' reflection quality in SSRL?

A Kruskal-Wallis H test was conducted to assess the differences in SSRL reflection scores among the 3 groups of students using different reflection methods, as shown in Table 5. The test compares independent samples without assuming a normal data distribution. This makes it highly suitable for analyzing the multiple groups of non-normally distributed reflection data in this study.

For this analysis, an overall reflection quality score was calculated for each student by taking the meaning of all seven reflection codes (NOR, LOWR, HIGHR, DESR, EXPR, CONR, CRIR). This composite score was used for the Kruskal-Wallis H test, while the mean scores for individual codes presented in Table 5 are provided only for descriptive purposes.

The results showed a chi-square value of 6.557, and an asymptotic significance of 0.038. The mean ranks for the 3 groups were G1 = 9.14, G2 = 8.00, and G3 = 15.86. The results indicate a statistically significant difference in reflection scores between the groups ($p = 0.038$). Specifically, G3′s mean rank was significantly higher than G1 and G2, indicating that using the AI agent is associated with higher performance.

Table 5
The result of the Kruskal-Wallis H test.

	Codes	Mean score			χ^2	p
		G1	G2	G3		
Reflection quality	NOR	0.018	0.005	0.088	6.557	0.038
	LOWR	0.267	0.163	0.232		
	HIGHR	0.018	0.044	0.218		
	DESR	0.229	0.197	0.262		
	EXPR	0.100	0.103	0.264		
	CONR	0.038	0.028	0.221		
	CRIR	0.006	0.037	0.200		

To further investigate the observed differences, we applied ENA for a fine-grained analysis of the students' reflections across the 3 reflection methods. This analysis aims to uncover the epistemic structures and patterns, providing deeper insights into how different reflection methods influence the quality and complexity of students' reflection processes. By analyzing epistemic networks, we may better understand the specific epistemic factors and relationships underlying the differences observed in the statistical results.

Fig. 3 presents a comparative ENA network model of reflection content for the three groups using different reflection methods. In this model, nodes represent individual reflection codes, and edges indicate the co-occurrence of codes within each unit of analysis. Blue, red, and purple dots denote the centroids of students in G1, G2, and G3, respectively, while the four black dots represent the four categories of reflection content (DESR, EXPR, CRIR, CONR). ENA applies singular value decomposition (SVD) to reduce the network model to two dimensions, which together account for 70.1 % of the variance (SVD1 = 51.5 %, SVD2 = 18.6 %). The x-axis in the ENA space (SVD1) defines the dimension of reflection content, with the right side (higher x-values) representing DESR codes and the left side (lower x-values) representing CONR codes. The y-axis (SVD2) in the ENA space defines the dimension of reflection content, where the CRIR and EXPR codes are positioned higher (with higher y-values). The DESR code is located lower in the ENA space (with lower y-values). This model allows comparison across students and groups, showing which types of reflection are more dominant and how reflection content patterns differ between groups.

The right side of Fig. 3 displays the mean networks of the 3 groups. Overall, the reflection content of all 3 groups predominantly features EXPR and DESR, with a strong association observed between these two points. The reflection content network of G1 is the sparsest, with only a few occurrences of CRIR, aside from the relatively frequent appearances of EXPR and DESR. The network of G2 is more concentrated, with distribution across all 4 reflection types and a stronger CRIR-DESR connection (value of 0.10). The reflection content of G3 is the most densely connected, with all 4 types having a relatively high proportion of representation. The CRIR-CONR (0.23) and CONR-EXPR (0.13) connections are relatively strong. In contrast, the other pairs based on traditional SSRL reflection did not exhibit strong correlations.

Table 6 demonstrates how the AI agent, through guided dialogue, facilitated the transition of G3 students from connective reflection (CONR) to critical reflection (CRIR), thereby deepening the SSRL

Table 4
Scores of the SSRL performance for the 3 groups.

Group	team1	team2	team3	team4	team5	team6	team7	team8	team9
G1	86.0	90.0	88.5	76.5	68.5	87.0	92.0	NA	NA
G2	83.5	93.5	87.0	90.5	81.0	71.5	76.5	NA	NA
G3	94.0	84.5	75.0	71.5	88.0	76.0	89.5	84.5	90.5

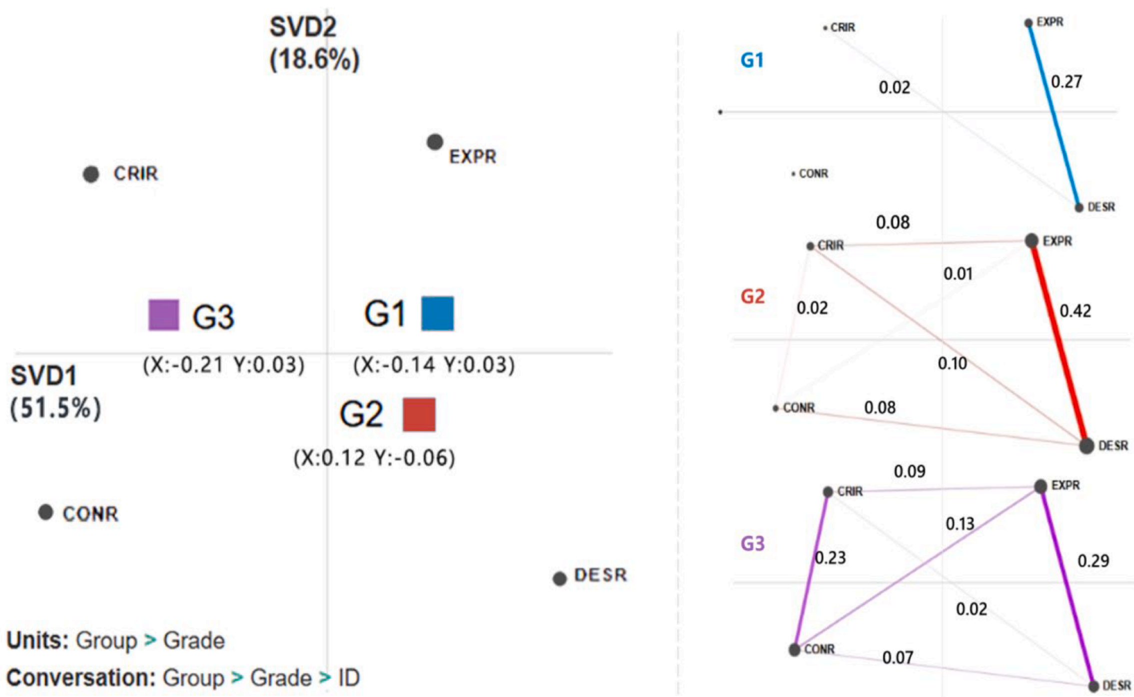


Fig. 3. Comparison of reflection content.

Table 6
Transcribed example of the G3 reflection content.

Student	Transcript
1	AI agent: What problems did you encounter, and how did you solve them? Student 9: We initially struggled, but by discussing with our team and leader, we solved the problems and saw how teamwork improved our process over time. (CONR) AI agent: What are your plans moving forward? Student 9: I hope for more team projects to use everyone's strengths and improve how I contribute next time. (CRIR)
2	AI agent: What were the shortcomings of this micro-lecture task? Student 18: Our PPT didn't cover everything, so next time I'll make it more complete. (CRIR) AI agent: I see, that's a very insightful analysis. What plans do you have for your future studies? Student 18: I plan to learn more about AI video production to improve my skills for future jobs. (CONR)
3	AI agent: Did you encounter any difficulties and challenges while completing the task? If so, what were they, and how did you solve them? Student 31: Our biggest challenge was managing time, but adjusting our schedule helped us complete the tasks. (CONR) AI agent: Great. How did you overcome these difficulties and challenges? Student 31: We held weekly meetings to discuss. (DESR)
4	AI agent: Please summarize the successful aspects of this task. Student 36: The tasks were well-organized, and because our team cooperated closely, we were able to complete the work more efficiently than at the beginning. (CONR)

reflection process. Under the guidance of the AI agent, student 9 and student 31 shifted from describing the current state of teamwork and time management, such as "We solved problems through communication with team members" (CONR), to deeper reflections on self-improvement and future learning plans, exemplified by "I hope for more team projects to utilize everyone's potential" (CRIR). Prompted by the AI agent's questioning, student 10 and student 36 reflected on the shortcomings of the SSRL tasks, noting that "The resources were not comprehensive, and most content lacked innovation" (CONR), and further analyzed the root causes of these issues, along with potential improvement measures (CRIR). Inspired by the AI agent, student 18 first identified the issue of inadequate presentation in the task (CRIR) and then proposed a concrete plan for deeper learning of AI technology (CONR). The AI agent's continuous questioning and feedback prompted students to progress from simple descriptive reflection to more critical and in-depth reflection throughout the reflection process.

Table 7 presents reflection examples from some G1 and G2 students, highlighting the impact of different reflection forms and guidance

Table 7
Transcript example of the G1 and G2 reflection content.

Group	Transcript
G1	G1-Student 4: Our group chose a radio show format for this Himalaya assignment. (DESR) I've always been a fan of radio shows, so I was very happy to have the opportunity to create one this time. (DESR) Of course, I also faced some challenges during the production process (DESR), such as the tone not fitting the storyline and the quality of the program needing to be better. (EXPR)
G1	G1-Student 30: Regarding this task, firstly, we didn't do well in the presentation aspect. The presentation was only in the form of a document, which needed to ensure a smooth connection between the presentation and the work, making it difficult to access the content. (CONR) Secondly, the content presentation was poorly executed and lacked a logical structure. (EXPR) Finally, the speech was not coherent during the presentation, and the preparation was insufficient. (EXPR)
G2	G2-Student 6: Task: We approached the task mainly in two aspects. (DESR) The first part determined the theme and type of work, and the second part recorded the work. (DESR) Division of labor: Our division of labor and cooperation were very reasonable, and each member completed their assigned tasks. (EXPR) Self-evaluation: Very successful. (DESR) Outlook: We plan to work more collaboratively on each task and strive to do our best. (CRIR)
G2	G2-Student 27: Task: This task enhanced our understanding of content production and strengthened the collaboration among team members. (EXPR) Division of labor: Our team had a clear division of responsibilities, and everyone had their tasks (EXPR). I was responsible for the recording, which was quite challenging. (EXPR) Self-evaluation: Although our team may not have been the best among all the teams, we had unique messages to convey. (CONR) If there is a next time, we will strive to improve it. (CRIR) Outlook: We should promote our work more effectively. (CRIR)

methods on students' reflection quality. Two G1 students (student 4 and student 30) conducted their reflections in the form of reports. Due to the lack of specific guidance from the instructor, who only provided general requirements, their reflections remained superficial, primarily involving DESR and EXPR. For example, student 4 wrote, "I have always enjoyed radio shows, so I was very pleased to have the opportunity to create one this time. "Student 30 mentioned, "The tone did not match the storyline, and the sound quality of the program was poor. These reflections remain limited to mere descriptions of the phenomena, needing more in-depth analysis of the underlying causes and offering no insights for future improvement. This tendency may be related to the relatively broad scope of the reports. These examples demonstrate that structured guidance exerts a positive effect on the quality of reflection. In addition, they highlight the importance of timely feedback and question prompting. Providing students with immediate feedback based on their responses and guiding them toward more elaborated answers contributes to fostering deeper levels of reflection.

In contrast, two students from Group G2 (student 6 and student 27), guided by the 4 aspects provided by the instructor and reflecting through short-answer questions, demonstrated a higher reflection quality. The instructor guided students to reflect on four dimensions, including task, division of labor, self-evaluation, and outlook. This approach, particularly in the latter two areas, effectively fostered CRIR and CONR. For example, student 6 mentioned, "We plan to collaborate more effectively in completing each future learning task, striving to achieve the best outcome" (CRIR). At the same time, student 27 stated, "Although our team may not be the best among all teams, we conveyed our unique message. If there is a next time, we will work harder to improve" (CONR and CRIR). This structured guidance enhanced the depth of reflection. However, since short-answer questions are a one-way form of reflection for students, the instructor may not intervene in their responses. As a result, there may be instances where students provide irrelevant answers or overly brief responses, which can affect the overall reflection quality. For instance, student 6 responded with "Very successful" in the self-evaluation section (DESR), which lacked depth in reflection. The AI agent could address this shortcoming by facilitating continuous interaction and feedback, encouraging students to engage in deeper reflection.

When comparing the effectiveness of the reflection methods in G1, G2, and G3, G1′s reflection reports were of lower quality, primarily focusing on DESR and EXPR. Due to the absence of specific guidance, the reflections needed more depth. The short-answer questions format in G2 improved reflection quality to some extent. Students' reflections became more focused with the instructor's guidance, particularly improving CRIR and CONR. However, this approach is still constrained by the limitations of outcome-based assessment. The AI agent guidance in G3 further enhanced reflection quality. Through real-time feedback and targeted questioning, students could engage in deeper levels of CRIR and CONR.

To scale these differences, the Mann-Whitney U test was employed to evaluate the distribution of the projection points of the 3 groups of students within the ENA space. The results indicated that at the $\alpha = 0.05$ significance level, G1 and G2 showed significant differences in both the first dimension ($U = 147{,}537$, $p = 0.01$, $r = 0.09$) and the second dimension ($U = 147{,}204$, $p = 0.01$, $r = 0.08$). This suggests that the structured guidance provided by short-answer questions enhances reflection quality. G1 and G3 also showed a significant difference in the first dimension ($U = 99{,}595.5$, $p = 0.00$, $r = 0.34$), highlighting the impact of integrating the AI agent in G3 to enhance reflection quality. However, no difference was observed in the second dimension ($U = 147{,}049.5$, $p = 0.42$, $r = 0.03$). Additionally, G2 and G3 exhibited differences in both the first dimension ($U = 127{,}246.5$, $p = 0.00$, $r = 0.36$) and the second dimension ($U = 215{,}386.5$, $p = 0.01$, $r = -0.08$), further demonstrating the effectiveness of the AI agent in fostering deeper reflection. This effect surpasses that of the structured short-answer questions approach alone. Notably, due to the large sample size in this study, the U values are relatively high; however, they remain within the acceptable range for statistical analysis. Some of these differences showed relatively small effect sizes, which will be further addressed in the discussion section.

4.2. What differences do high and low-performance teams show in reflection quality when using the three reflection methods?

Fig. 4 illustrates the distribution of students from the 3 reflection methods (G1, G2, G3) along the two principal component axes (SVD1 and SVD2). The points of different colors and shapes in the figure represent high and low-performance teams within each group, indicating their performance across various reflection categories, such as DESR, EXPR, CONR, and CRIR. The SVD1 axis accounts for 77.3 % of the total variance, while the SVD2 axis explains 16.8 %. The position of each point represents the students' tendencies in reflection content, with points closer to a specific reflection category indicating that the group's performance is more concentrated in that category.

In Fig. 4, the centroids of the low-performance teams in G1 and G2 are positioned relatively close to each other, with the low-performance teams located higher near DESR. Conversely, the high-performance teams are situated lower, closer to CRIR. This indicates a certain degree of similarity in the reflection content between the low-performance teams in G1 and G2. G3 is distributed on the right side of the figure, with a greater distance between the high and low-performance teams, indicating a more pronounced difference in reflection content than the other teams. Unlike G1 and G2, the G3 high-performance teams are positioned at the top, closer to CONR, while the low-performance teams are located at the bottom, near CRIR and EXPR. This suggests that the high-performance teams in G3 tend to engage more in connective reflection, whereas the low-performance teams focus more on critical and explanatory reflection.

The study employed the Mann-Whitney U test to elucidate further the scaling characteristics of the differences in reflection content between the high and low-performance teams across the 3 cohorts (Table 8). According to the results of the Mann-Whitney U test, there are differences in the reflection content performance between the high and

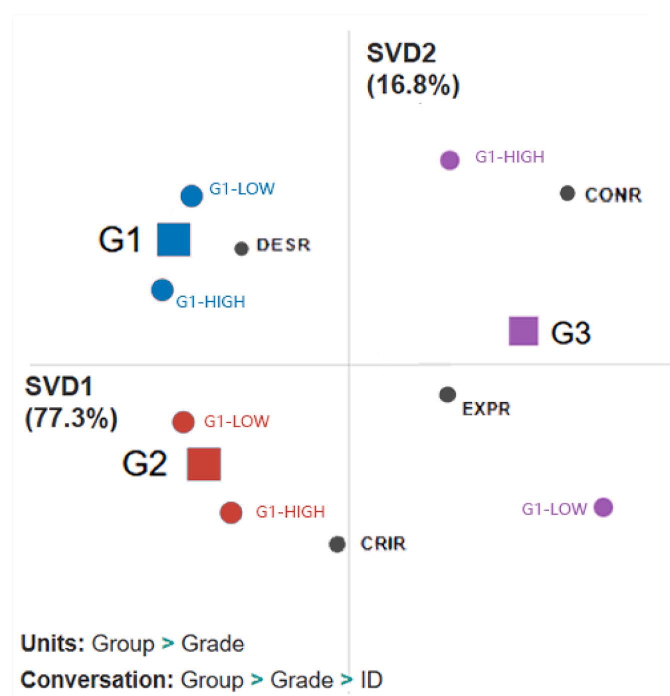
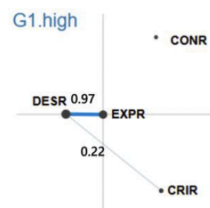
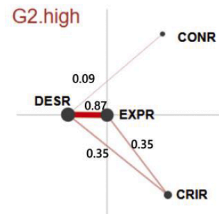
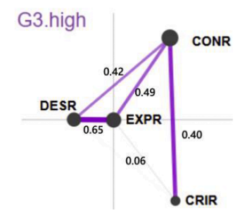
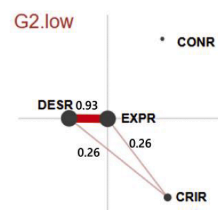
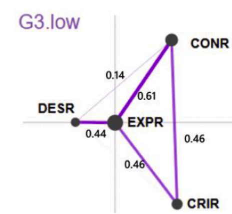


Fig. 4. The centroid distribution of high and low group students across the three reflection methods.

Table 8
The reflection content distribution of high and low-performance teams across the three methods.

	G1	G2	G3
High performance team			
Low-performance team			

low-performance teams across different reflection approaches. In G1, the high and low-performance teams did not exhibit significant differences in either dimension (MR1: $U = 4932.00$, $p = 0.41$, $r = 0.05$; MR2: $U = 5463.00$, $p = 0.44$, $r = 0.05$). In G2, the high and low-performance teams showed a significant difference in the MR1 dimension ($U = 3303.00$, $p = 0.03$, $r = 0.19$) but no difference in the MR2 dimension ($U = 3051.00$, $p = 0.26$, $r = 0.10$). For G3 (students using AI agent-driven continuous questioning), the high and low-performance teams showed a significant difference in the MR1 dimension ($U = 1136.50$, $p < 0.001$, $r = 0.45$). In contrast, the difference in the MR2 dimension was insignificant ($U = 2187.50$, $p = 0.54$, $r = 0.06$).

In G3, the differences between the high and low-performance teams were the most pronounced, particularly on the MR1 dimension. Further analysis of the ENA diagram revealed that low-performance teams exhibited stronger connections in EXPR-CRIR (0.46) and EXPR-CONR (0.61). This suggests that the AI agent-driven reflection method may help low-performance teams focus more on specific reflection content.

5. Discussion

This section analyzes the findings based on the research questions. It covers the positive impact of AI agents on students' SSRL reflection, differences in reflection quality between high and low-performance teams, and key considerations for using AI agents effectively in SSRL.

5.1. The positive role of AI agents in students' SSRL reflection

In SSRL, the AI agent reflection assistant enhanced the quality of students' reflections. This outcome aligns with previous research [70, 71]. For instance, Maedche et al. [70] demonstrated the positive role of AI agents in fostering deeper reflection among students. Sigman et al. [71] also found that AI assistants emulate and augment human cognition, thereby promoting reflection. These studies provide more evidence of the positive impact AI agents have on facilitating reflective practices in education.

This study further clarifies how AI agents enhance the quality of student reflection in the SSRL process through ENA. In these activities, student reflections guided by AI agents exhibited higher levels of critical thinking and coherence. In contrast, the other two traditional reflective texts displayed lower levels of reflection, focusing primarily on descriptive and exploratory reflection. As Rusandi et al. [72] highlighted, AI may assist learners in constructing their learning processes, thereby enhancing critical thinking. In higher education, Xia and Li [73] also suggested that AI assistants have a positive impact on students' imagination, creativity, critical thinking, and autonomous learning. Zang et al. [69] experimentally confirmed the role of AI agents in enhancing students' critical thinking in English learning. However, the systematic review by Mohamud et al. [74] indicated that the introduction of AI in higher education may diminish students' critical thinking. This conclusion contradicts the findings of this study. The differences may be due to a lack of proper instructional design by teachers when using AI [74]. Cronje [75] argued that AI may serve as a teaching assistant to facilitate learning, but it should be integrated with instructional design and necessary prompts. In this study, the SSRL reflection checklist was operationalized as structured prompts to calibrate the AI agent, enabling it to scaffold students' reflections across the four phases of SSRL. By embedding SSRL principles into its dialogic design, the agent acted as both a facilitator of reflection and a medium for delivering theoretical scaffolds. This underscores the importance for educators and researchers to apply instructional theory and design thoughtfully when integrating AI into the classroom.

In addition to SSRL theoretical guidance, the AI agent leveraged its technological capabilities, including continuous questioning and real-time feedback, to actively scaffold deeper student reflections. Wolfbauer et al. [76] noted that continuous dialogue with intelligent assistants enhances students' levels of reflection. In the G3 group, the AI agents not only guided students to explore the root causes of issues but also helped them develop specific improvement plans. This guiding process is similar to the "Socratic method" in educational psychology. Through a series of targeted questions, students are encouraged to engage in deep thinking and gain a more profound understanding of the knowledge [77]. In addition, the timely feedback function of AI agents plays a crucial role in enhancing the quality of students' SSRL reflections. Self-determination theory suggests that providing positive emotional support through feedback helps students gain a sense of belonging, thereby enhancing their motivation to learn and willingness to reflect [78]. Uygur et al. [79] suggested that timely feedback enhanced students' reflection and learning. However, traditional SSRL reflection reports and short-answer questions are one-way reflective activities, lacking immediate feedback and guidance. The AI agent reflection assistant compensates for the shortcomings of teachers in providing timely feedback, enhancing the effectiveness of collaborative learning.

This study indicates that the level of reflection guidance directly affects learners' reflection quality, which is consistent with previous research [80–82]. G1, with minimal guidance, showed the lowest quality, while G2, guided by the SSRL reflection checklist, exhibited higher-quality reflections, demonstrating the importance of SSRL scaffolds. G3 combined SSRL scaffolding with real-time feedback and encouragement for deeper reflection. Comparisons suggest that while structured short-answer questions had a limited impact, the AI agent provided a practically meaningful enhancement of students' reflective practices. However, these findings are based primarily on qualitative data, and further quantitative research is needed to validate them.

In summary, AI agents play a substantial role in promoting student reflection. Although the comparison between structured short-answer questions and traditional reflective reports showed statistically significant but very small effects, this suggests that short-answer questions alone had a limited impact on enhancing students' reflection quality. In contrast, the AI agent had a substantially greater impact on students' reflective practices. It is essential for educators and instructional designers to integrate AI agents into classrooms and develop more instructional design case studies. Moreover, teachers should prioritize the importance of instructional theories and provide essential design guidance when applying AI agents.

5.2. Differences between high and low-performance teams under various SSRL reflection methods

The results indicate a significant difference in the high and low-performance teams that utilized reflective short-answer questions and the AI agent reflection assistant. In short-answer questions, high-performance teams performed better. This aligns with the conclusions of Knight et al. [83], who found that high-performance students outperformed low-performance students in reflective questions. The disparity in reflection between high and low-performance learners is primarily attributed to their metacognitive levels and learning strategies [84–86]. For instance, Safari and Fitriati [85] found that high-performance learners were able to use all strategies equally, but low-performance learners more frequently relied on metacognitive and social strategies. These differences may impact learners' outcomes, including their learning effectiveness and reflection [84].

In contrast, the reflection quality of low-performance teams using the AI agent reflective assistant was better than that of the high-performance teams. This is a novel finding of the study, suggesting that the AI reflective assistant played a positive role in guiding low-performance learners through the reflection process. This finding aligns with previous evidence showing that AI technologies tend to provide greater benefits for lower performers [87–90]. Prior studies have suggested that such differential effects often occur because an AI chatbot can use adaptive strategies and personalized feedback to address the strategic gaps of low performers [88]. AI tutoring can also offer both cognitive and emotional support [89]. Xu et al. [90] further found that low-performing learners become more engaged when they receive immediate feedback and external help. This engagement encourages them to apply higher-order thinking strategies more actively.

These mechanisms may also explain the current results in our SSRL reflection task. The AI reflection assistant provided structured guidance in real time and reduced the cognitive load of producing reflections. This allowed low-performing learners to focus more on critical and creative thinking. In contrast, high-performing learners may already have established reflection routines. Extra guidance could interfere with these processes, leading to smaller gains in reflection quality [87].

This study, therefore, not only confirms that differential effects exist in reflection tasks but also highlights the potential of AI support to promote higher-order thinking in low-performing learners. In educational practice, this suggests that AI reflection assistants could be strategically deployed to close performance gaps. Future research could examine how to fine-tune AI guidance so that it benefits high performers without disrupting their existing strategies.

Additionally, there was no significant difference in performance between high and low-performance student teams in reflective reports, with both showing low quality reflections. This may be due to learners lacking clear guidance in the reflection process. Maedche et al. [70] found that in reflective environments lacking external feedback or structured guidance, the quality of students' reflections is constrained. This suggests that instructors should provide the necessary scaffolding when designing reflective tasks. The SSRL scaffolding demonstrated significant value in this study and is well-suited for broader application in collaborative settings.

5.3. Considerations for the effective use of AI agents in SSRL

Although experiments have demonstrated that AI agents enhance SSRL reflection quality, there are several limitations in their usage. To better promote the outcomes of this study, we offer considerations for teachers and instructional designers regarding the use of AI agents.

Firstly, the quality and reliability of feedback provided by AI agents still present limitations. This finding aligns with the studies of Maloney et al. [91] and Fedus et al. [92], which suggest that the accuracy and effectiveness of AI agents depend on algorithm design and data quality. In this study, the AI agent exhibited two primary issues: repeated questioning and unexpected interruptions during conversations. To address the issue of repeated questioning, adjustments to the prompt design can be implemented. For example, the prompts specify that each question should be asked only once and repeated only if the student responds off-topic or does not answer. For unexpected interruptions, teachers need to guide students in testing their network environment and re-engaging with the task. These observations show that AI agents need improvement in handling complex contexts and dynamic learning needs.

In addition, data privacy and ethical concerns pose another challenge in the application of AI agents. AI agents require extensive data collection, including students' reflection content, behavioral patterns, and learning habits [93]. To mitigate this issue, this study incorporated an opening message in the AI agent's script. The message advised students: "Please do not disclose personal sensitive information, such as your name or school, during the interaction." Furthermore, before implementing the AI agent, teachers need to raise students' awareness of data security and privacy protection [94].

The risks associated with over-reliance on AI technology should also be carefully evaluated. Although AI agents can provide personalized support, they cannot fully replace the role of human teachers, particularly in offering emotional support and fostering social interaction [95]. In this study, AI agents were utilized exclusively in the post-class reflection phase. The remaining instructional time relied on face-to-face interactions between teachers and students. As GAI technology becomes increasingly accessible, preventing students from developing dependency behaviors may become more challenging. Future research could explore strategies to prevent learners from becoming overly reliant on GAI technologies.

While AI agents have demonstrated advantages in enhancing students' SSRL reflection quality, their widespread applicability is constrained by feedback quality, data privacy, and ethical considerations. Future research should emphasize these limitations, refining the application framework of AI to ensure its effectiveness and sustainability in the educational domain.

6. Conclusion, limitations, and future research

This study explores methods to enhance student reflection quality by designing an AI agent that supports reflection through continuous questioning and real-time feedback. Using content analysis and ENA, this study conducted a three-semester experiment comparing reflection reports, short-answer questions, and an AI agent reflection assistant. The results indicate that AI agents improve reflection quality, particularly for low-performance teams. The study offers practical guidance for integrating AI into SSRL-based instruction.

Although this study contributes to understanding students' reflection behaviors in SSRL, several limitations remain. The first limitation arises from the study participants. Conducted within a higher education setting, this research primarily examines the effectiveness of using AI agents to facilitate reflection among university students. Only 97 students from the "Internet Thinking and Digital Self-Learning" course participated, so the findings may not be generalizable to other courses or age groups. Further research is needed to explore the potential impact and adaptability of AI agents in secondary and primary education settings [96]. Secondly, the AI agent still has limitations in the quality and reliability of feedback, which may affect the depth and quality of students' reflections. Addressing this issue relies on rapidly updating and optimizing large AI model algorithms to provide higher-quality and more targeted feedback. The third limitation is that the three reflection methods used in this experiment all fall under outcome-based reflection, overlooking the dynamic process of students' reflections at different stages of collaborative learning. Additionally, the proposed mechanisms underlying the AI agent's impact on reflection quality, particularly for low-performance teams, remain hypothetical and require further empirical validation through quantitative studies. Lastly, this study did not differentiate the specific contributions of individual design elements in the AI agent's interaction strategy (e.g., sequential questioning, encouraging feedback, simplified language). More research could adopt ablation analysis to examine how these elements independently influence students' reflective practices.

Based on the limitations identified in this study, future research could expand the study to more diverse educational contexts, including secondary and primary education, to examine the generalizability and adaptability of AI agents. Incorporating multi-modal data, such as students' facial expressions, gestures, and dialogue, may offer a more comprehensive understanding of reflective behaviors in SSRL. Improvements in AI models are needed to enhance the quality and reliability of feedback, supporting deeper and higher-quality student reflections. In addition, investigating the individual contributions of specific design elements in AI agents' interaction strategies, for example, through ablation-style comparisons, could clarify which features most effectively promote high-order reflection, particularly among low-performance teams. We therefore urge more researchers to focus on this area of study, exploring the impact of GAI on educational outcomes to better understand and harness its potential for improving educational practices.

Declaration of generative AI in the writing process

During the preparation of this work, the authors used Kimi (https://kimi.moonshot.cn/) to improve language and readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

CRediT authorship contribution statement

Yumin Zheng: Writing – original draft, Conceptualization. **Fengjiao Tu:** Investigation, Data curation. **Fengfang Shu:** Investigation, Data curation. **Chaowang Shang:** Formal analysis, Data curation. **Lulu Chen:** Writing – review & editing, Formal analysis. **Jiang Meng:** Investigation.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Chaowang Shang acknowledges the financial support from the National Natural Science Foundation of China (Grant Number: 62577035). The other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. The Critical Thinking Questionnaire (CThQ)

Instructions: For each statement below, please indicate how much you agree using a 5-point Likert scale (1 = Strongly disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly agree).

1. After reading a text, I check important information, even if it seems to be true.
2. I like combining information from different texts.
3. I am willing to share newly acquired information.
4. In-depth analyses of reality is a waste of time.
5. After reading a text, I can recall important points.
6. The same content can be expressed in many different ways.
7. I can understand texts from various fields.
8. I form my impressions based on various pieces of information that I combine.
9. Everything already exists, so nothing completely new can be created.
10. When I talk, I give many examples.
11. In discussions, I care about justifying my stance while understanding the other party.
12. I like finding connections between seemingly different phenomena.
13. I can see the structure of a text, and I could reorganize it.
14. When discussing, I try to use practical examples to justify my stance.
15. If necessary, I can recall information I have read before.
16. I do not remember much of what I learned at school.
17. When I am interested in some information, I try to verify whether it is true.
18. I can extract the most relevant parts of a text.
19. To evaluate information, I check multiple sources.
20. I like discussing new interpretations of texts I already know.
21. I like to collate different opinions and compare them.
22. I have difficulties with paraphrasing.
23. I try to apply the information I have learned in everyday life.
24. When I read, I look for relationships between its information and other texts I have read.
25. I pay attention to the contexts, nuances, and overtones of statements.

Data availability

The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

References

[1] S. Ahmad, M. Rahmat, M. Mubarik, M. Alam, S. Hyder, Artificial intelligence and its role in education, Sustainability 13 (22) (2021) 12902.
[2] X. Gong, Z. Li, A. Qiao, Impact of generative AI dialogic feedback on different stages of programming problem solving, Educ. Inf. Technol. 30 (7) (2025) 9689–9709.
[3] O. Tapalova, N. Zhiyenbayeva, D. Gura, Artificial Intelligence in Education: aIEd for personalised learning pathways, Electron. J. e-Learn. 20 (5) (2022) 639–653.
[4] S. Järvelä, P. Kirschner, E. Panadero, J. Malmberg, C. Phielix, J. Jaspers, M. Koivuniemi, H. Järvenoja, Enhancing socially shared regulation in collaborative learning groups: designing for CSCL regulation tools, in: Educ. Technol. Res. Dev., 63, 2014, pp. 125–142.
[5] D. Bransen, M.J.B. Govaerts, E. Panadero, et al., Putting self-regulated learning in context: integrating self-, co-, and socially shared regulation of learning, Med. Educ. 56 (1) (2022) 29–36.

[6] E. Eshuis, J. Vrugte, A. Anjewierden, L. Bollen, J. Sikken, T. Jong, Improving the quality of vocational students' collaboration and knowledge acquisition through instruction and joint reflection, Int. J. Comput.-Support. Collab. Learn. 14 (2019) 53–76.
[7] C. Chan, K. Lee, Reflection literacy: a multilevel perspective on the challenges of using reflections in higher education through a comprehensive literature review, Educ. Res. Rev. 32 (2020) 100376.
[8] L. Guo, How should reflection be supported in higher education? — A meta-analysis of reflection interventions, Reflective Pract. 23 (2021) 118–146.
[9] S. Popenici, S. Kerr, Exploring the impact of artificial intelligence on teaching and learning in higher education, Res. Pract. Technol. Enhanc. Learn. 12 (1) (2017) 22.
[10] H. Kiy, A Study on Writing Experience With ChatGPT of College Students, J. Korea Converg. Soc. 14 (9) (2024) 976.
[11] K. Hanifi, O. Cetin, C. Yilmaz, On ChatGPT: perspectives from software engineering students, in: Proc. 2023 IEEE 23rd Int. Conf. Softw. Qual. Reliab. Secur. (QRS), 2023, pp. 196–205.
[12] Zhiheng Xi, et al., The rise and potential of large language model based agents: A survey, Sci. China Inf. Sci. 68 (2) (2025) 121101.
[13] E. Katsarou, F. Wild, A. Sougari, P. Chatzipanagiotou, A systematic review of voice-based intelligent virtual agents in EFL education, Int. J. Emerg. Technol. Learn. (iJET) 18 (10) (2023) 65–85.
[14] P.R. Lewis, Ş. Sarkadi, Reflective artificial intelligence, Minds Mach. 34 (2) (2024) 14.
[15] Z. Xu, P. Zhang, M. Tu, M. Zhang, Y. Lai, Brain optimization with additional study time: potential brain differences between high- and low-performance college students, Front. Psychol. 14 (2023) 1209881.
[16] UK Government, Generative Artificial Intelligence (AI) in Education, GOV.UK, 2023. https://www.gov.uk/government/publications/generative-artificial-intelligence-in-education/generative-artificial-intelligence-ai-in-education.
[17] M. Dogan, T. Dogan, A. Bozkurt, The use of artificial intelligence (AI) in online learning and distance education processes: a systematic review of Empirical Studies, Appl. Sci. 13 (5) (2023) 3056.
[18] L. Shi, The integration of advanced AI-enabled emotion detection and adaptive learning systems for improved emotional regulation, J. Educ. Comput. Res. 63 (2024) 173–201.
[19] B. Tang, J. Liang, W. Hu, H. Luo, Enhancing programming performance, learning interest, and self-efficacy: the role of large language models in middle school education, Systems 30 (6) (2025) 8109–8138.
[20] L. Feng, Investigating the effects of artificial intelligence-assisted language learning strategies on cognitive load and learning outcomes: a comparative study, J. Educ. Comput. Res. 62 (8) (2025) 1741–1774.
[21] Q. Huang, W. Li, Y. Zhao, Enhancing deep learning and motivation in university English education through AI technology: a quasi-experimental study, Asian J. Educ. Soc. Stud. 51 (4) (2025) 452–463.
[22] Ó. Cuéllar, M. Contero, M. Hincapié, Personalized and Timely Feedback in Online education: Enhancing learning With Deep Learning and Large Language Models, MTI. 9 (5) (2025) 45.
[23] X. Zhou, D. Teng, H. Al-Samarraie, The mediating role of generative AI self-regulation on students' critical thinking and problem-solving, Educ. Sci. 14 (12) (2024) 1302.
[24] S. Steenbergen-Hu, H. Cooper, A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning, J. Educ. Psychol. 106 (2014) 331–347.
[25] C. Moridis, A. Economides, Affective learning: empathetic agents with emotional facial and tone of voice expressions, IEEE Trans. Affect. Comput. 3 (2012) 260–272.
[26] S. Nelekar, A. Abdulrahman, M. Gupta, D. Richards, Effectiveness of embodied conversational agents for managing academic stress at an Indian university (ARU) during COVID-19, Br. J. Educ. Technol. 53 (2021) 491–511.
[27] W. Sun, Q. Chen, The design, implementation, and evaluation of Gamified Immersive Virtual Reality (IVR) for learning: a review of Empirical Studies, Proc. Eur. Conf. Games-Based Learn. 17 (1) (2023) 789–797.
[28] M. Chen, L. Wu, Z. Liu, X. Ma, The impact of metacognitive strategy-supported intelligent agents on the quality of collaborative learning from the perspective of the community of inquiry, in: Proc. 2024 4th Int. Conf. Educ. Technol. (ICET), 2024, pp. 11–17.
[29] H. Hong, C. Viriyavejakul, P. Vate-U-Lan, Enhancing critical thinking skills: exploring generative AI-enabled cognitive offload instruction in English essay writing, 4, ECOHUMANISM Учредители: Transnational Press, London, p. 2024.
[30] D.H. Schunk, B.J. Zimmerman, Motivation and Self-Regulated learning: Theory, research, and Applications, Routledge, 2012.
[31] P.H. Winne, A.F. Hadwin, N.E. Perry. Metacognition and computer-supported collaborative learning, The International Handbook of Collaborative Learning, Routledge, 2013, pp. 462–479.
[32] Y. Su, Y. Li, H. Hu, et al., Exploring college English language learners' self and social regulation of learning during wiki-supported collaborative reading activities, Int. J. Comput.-Support. Collab. Learn. 13 (2018) 35–60.
[33] F. Tu, L. Wu, Kinshuk, et al., Exploring the influence of regulated learning processes on learners' prestige in project-based learning, Educ. Inf. Technol. 30 (2) (2025) 2299–2329.
[34] S. Zhang, J. Chen, Y. Wen, H. Chen, Q. Gao, Q. Wang, Capturing regulatory patterns in online collaborative learning: a network analytic approach, Int. J. Comput.-Support. Collab. Learn. 16 (2021) 37–66.
[35] J. Zheng, W. Xing, G. Zhu, Examining sequential patterns of self-and socially shared regulation of STEM learning in a CSCL environment, Comput. Educ. 136 (2019) 34–48.
[36] E. Panadero, S. Järvelä, Socially shared Regulation of Learning: a review, Eur. Psychol. 20 (2015) 190–203.
[37] J. Isohätälä, H. Järvenoja, S. Järvelä, Socially shared regulation of learning and participation in social interaction in collaborative learning, Int. J. Educ. Res. 81 (2017) 11–24.
[38] J. Li, Y. Lin, M. Sun, R. Shadiev, Socially shared regulation of learning in game-based collaborative learning environments promotes algorithmic thinking, learning participation, and positive learning attitudes, Interact. Learn. Environ. 31 (2020) 1715–1726.
[39] J. Malmberg, S. Järvelä, H. Järvenoja, E. Panadero, Promoting socially shared regulation of learning in CSCL: progress of socially shared regulation among high- and low-performing groups, Comput. Hum. Behav. 52 (2015) 562–572.
[40] J. Yukawa, Co-reflection in online learning: collaborative critical thinking as narrative, Int. J. Comput.-Support. Collab. Learn. 1 (2006) 203–228.
[41] A. Głowala, M. Kołodziejski, T. Butvilas, Reflection as a basic category of a teacher's thinking and action, Multidiscip. J. Sch. Educ. 12.1(2023):229–250.
[42] J. Buck, Reflecting on reflections: a case study of disappointment in student writing assignments, J. Acoust. Soc. Am. (2023). A273-A273.
[43] N Rahmi, C M Zubainur, Students' mathematical reflective thinking ability through scaffolding strategies[C]//Journal of Physics: Conference Series, IOP Publishing 1460 (1) (2020) 012022.
[44] J. Dewey, Education Democracy, The elementary school teacher 4 (4) (1903) 193–204.
[45] B.J. Zimmerman, Self-regulated learning and academic achievement: an overview, Educ. Psychol. 25 (1) (1990) 3–17.
[46] D. Coulson, M. Harvey, Scaffolding student reflection for experience-based learning: a framework, Teach. High. Educ. 18 (2013) 401–413.
[47] S. Lajoie, Extending the scaffolding metaphor, Instr. Sci. 33 (2005) 541–557.
[48] E. Panadero, P.A. Kirschner, S. Järvelä, J. Malmberg, H. Järvenoja, How individual self-regulation affects group regulation and performance: a shared regulation intervention, Small Group Res. 46 (4) (2015) 431–454.
[49] E. Davis, Prompting middle school science students for productive reflection: generic and directed prompts, J. Learn. Sci. 12 (2003) 142–191.
[50] J. Hattie, H. Timperley, The Power of Feedback, Rev. Educ. Res. 77 (2007) 112–181.
[51] R. Ajjawi, F. Kent, J. Broadbent, J. Tai, M. Bearman, D. Boud, Feedback that works: a realist review of feedback interventions for written tasks, Stud. High. Educ. 47 (2021) 1343–1356.
[52] U. Krause, R. Stark, Reflection in example- and problem-based learning: effects of reflection prompts, feedback, and cooperative learning, Eval. Res. Educ. 23 (2010) 255–272.
[53] J. Contreras, S. Edwards-Maddox, A. Hall, M. Lee, Effects of reflective practice on baccalaureate nursing students' Stress, Anxiety, and competency: an integrative review, Worldviews Evid.-Based Nurs. 17 (3) (2020) 239–245.
[54] H. Gadsby, Fostering reflective practice in Post Graduate Certificate in Education students through reflective journals. Developing a typology for reflection, Reflective Pract. 23 (2022) 357–368.
[55] S. Rabu, N. Badlishah, Levels of students' Reflective thinking skills in a collaborative learning environment using Google Docs, TechTrends 64 (2020) 533–541.
[56] J. Stoszkowski, A. Hodgkinson, D. Collins, Using Flipgrid to improve reflection: a collaborative online approach to coach development, Phys. Educ. Sport Pedagogy 26 (2020) 167–178.
[57] E. Liesa, P. Mayoral, M. Giralt-Romeu, S. Angulo, Video-Based Feedback for Collaborative Reflection Among Mentors, University Tutors, and Students, Edu. Sci. 13 (9) (2023) 879.
[58] M. Alghasab, J. Hardman, Z. Handley, Teacher-student Interaction On wikis: Fostering collaborative Learning and Writing, Learn Cult. Soc. Inter. 21 (2019) 10–20.
[59] R. Gubareva, R. Lopes, Virtual Assistants for learning: a systematic literature review, CSEDU (1) (2020) 97–103.
[60] L. González, H. Neyem, I. Contreras-McKay, D. Molina, Improving learning experiences in software engineering capstone courses using artificial intelligence virtual assistants, Comput. Appl. Eng. Educ. 30 (2022) 1370–1389.
[61] B. Renner, G. Wesiak, V. Pammer-Schindler, M. Prilla, L. Müller, D. Morosini, S. Mora, N. Faltin, U. Cress, Computer-supported reflective learning: how apps can foster reflection at work, Behav. Inf. Technol. 39 (2019) 167–187.
[62] A. Freiberg-Hoffmann, A. Romero-Medina, B. López-Fernández, M. Fernández-Liporace, Learning approaches: cross-cultural differences (Spain–Argentina) and academic achievement in college students, Span. J. Psychol. 26 (2023) e16.
[63] A. Kobylarek, K. Błaszczyński, L. Ślósarz, M. Madej, Critical thinking Questionnaire (CThQ)–construction and application of a critical thinking test tool, Andragogy Adult Educ. Soc. Mark. 2 (2) (2022), 1-1.
[64] J. Dewey, An analysis of reflective thought, J. Philos. (1922) 29–38.
[65] D.T. Campbell, J.C. Stanley, Experimental and Quasi-Experimental Designs For Research, Ravenio Books, 2015.
[66] M.M. Plack, M. Driscoll, S. Blissett, R. McKenna, T.P. Plack, A method for assessing reflective journal writing, J. Allied Health 34 (4) (2005) 199–208.
[67] L. Wang, G. Wu, J. Wu, A Study on the Reflective Level of Teachers' Autobiography, Global Education Outlook (01), (2018) 93–105.
[68] H.T. Hou, Integrating cluster and sequential analysis to explore learners' flow and behavioral patterns in a simulation game with a situated-learning context for science courses: a video-based process exploration, Comput. Human Behav. 48 (2015) 424–435.

[69] G. Zang, M. Liu, B. Yu, The application of 5G and artificial intelligence technology in the innovation and reform of college English education, Comput. Intell. Neurosci. 2022 (1) (2022) 9008270.
[70] A. Maedche, C. Legner, A. Benlian, B. Berger, H. Gimpel, T. Hess, O. Hinz, S. Morana, M. Söllner, AI-based digital assistants, Bus. Inf. Syst. Eng. 61 (2019) 535–544.
[71] M. Sigman, D. Slezak, L. Drucaroff, S. Ribeiro, F. Carrillo, Artificial and Human intelligence in mental health, AI Mag. 42 (2021) 39–46.
[72] M.A. Rusandi, I. Saripah, D.M. Khairun, No worries with ChatGPT: building bridges between artificial intelligence and education with critical thinking soft skills, J. Public Health. 45 (3) (2023) e602–e603.
[73] X. Xia, X. Li, Artificial intelligence for higher education development and teaching skills, Wirel, Commun. Mob. Comput. 2022 (1) (2022) 7614337.
[74] Y. Mohamud, A. Ma'rof, A. Mohamed, M. Uzir, A narrative review on the impact of applied artificial intelligence tools on higher secondary students, Int. J. Acad. Res. Bus. Soc. Sci. 13 (14) (2023) 34–42.
[75] J. Cronje, Exploring the role of ChatGPT as a peer coach for developing research proposals: feedback quality, prompts, and student reflection, Electron. J. (2024) 22.2, e-Learn.
[76] I. Wolfbauer, V. Pammer-Schindler, K. Maitz, C. Rosé, A script for conversational reflection guidance: a field study on developing reflection competence with apprentices, IEEE Trans. Learn. Technol. 15 (2022) 554–566.
[77] F. Leigh, Platonic dialogue, maieutic method, and critical thinking, J. Philos. Educ. 41 (2008) 309–323.
[78] E. Deci, R. Ryan. Intrinsic motivation and self-determination in Human behavior, 1975, pp. 1–371.
[79] J. Uygur, E. Stuart, M. Paor, E. Wallace, S. Duffy, M. O'Shea, S. Smith, T. Pawlikowska, The Best evidence in Medical Education systematic review to determine the most effective teaching methods that develop reflection in medical students: BEME Guide No. 51, Med. Teach. 41 (2019) 3–16.
[80] K. Arendt, L. Stark, A. Friedrich, R. Brünken, R. Stark, Quality of reflections on teaching: approaches to its measurement and low-threshold promotion, Educ. Sci. 15 (7) (2025) 884.
[81] J. Jung, Y. Lu, A. Ding, How do prompts shape preservice teachers' reflections? A case study in an online technology integration class, J. Teach. Educ. 73 (3) (2021) 301–313.
[82] A. Sturgill, P. Motley, Methods of reflection about service learning: guided vs. free, dialogic vs. expressive, and public vs. private. Teaching and learning inquiry, ISSOTL J. 2 (1) (2014) 81–93.
[83] J. Knight, D. Weaver, M. Peffer, Z. Hazlett, Relationships between prediction accuracy, metacognitive reflection, and performance in introductory genetics students, CBE Life Sci. Educ. 21 (3) (2022) ar45.
[84] D. Difrancesca, J. Nietfeld, L. Cao, A comparison of high and low achieving students on self-regulated learning variables, Learn. Individ. Differ. 45 (2016) 228–236.
[85] S A Gani, D Fajrina, R Hanifa, Students' learning strategies for developing speaking ability[J], Stud. Eng. lang. educ. 2 (1) (2015) 16–28.
[86] M. Yip, Differences between high and low academic achieving university students in learning and study strategies: a further investigation, Educ. Res. Eval. 15 (2009) 561–570.
[87] H.K. Etkin, K.J. Etkin, R.J. Carter, C.E. Rolle, Differential effects of GPT-based tools on comprehension of standardized passages, Front. Educ. 10 (2025) 1506752.
[88] S. Ruan, A. Nie, W. Steenbergen, J. He, J.Q. Zhang, M. Guo, et al., A reinforcement learning tutor better supported lower performers in a math task, Mach. Learn. 113 (2024) 3023–3048.
[89] D.R. Thomas, J. Lin, E. Gatz, A. Gurung, S. Gupta, K. Norberg, et al., Improving student learning with hybrid human-AI tutoring: a three-study quasi-experimental investigation, in: Proc. 14th Learn. Anal. Knowl. Conf., 2024, pp. 404–415. New York, NY, USA: Association for Computing Machinery (LAK '24).
[90] Y Xu, J Zhu, M Wang, et al., The impact of a digital game-based AI chatbot on students' academic performance, higher-order thinking, and behavioral patterns in an information technology curriculum[J], App. Sci. 14 (15) (2024) 6418.
[91] Maloney, A., Roberts, D. A., & Sully, J. (2022). A solvable model of neural scaling laws. *arXiv preprint arXiv:2210.16859*.
[92] W. Fedus, B. Zoph, N. Shazeer, Switch transformers: scaling to trillion parameter models with simple and efficient sparsity, J. Mach. Learn. Res. 23 (120) (2022) 1–39.
[93] K. Seo, J. Tang, I. Roll, S. Fels, D. Yoon, The impact of artificial intelligence on learner–instructor interaction in online learning, Int. J. Educ. Technol. High. Educ. 18 (1) (2021) 54.
[94] B. Klimova, M. Pikhart, J. Kacetl, Ethical issues of the use of AI-driven mobile apps for education, Front. Public Health 10 (2023) 1118116.
[95] T. Adiguzel, M. Kaya, F. Cansu, Revolutionizing education with AI: exploring the transformative potential of ChatGPT, Contemp. Educ. Technol. 15 (3) (2023).
[96] M. Thottoli, B. Alruqaishi, A. Soosaimanickam, Robo academic advisor: can chatbots and artificial intelligence replace human interaction? Contemp. Educ. Technol. 16 (1) (2024) ep485.