Designing and Evaluating an AI-Supported Learning Activity: How Do Students Regulate Learning with GenAI?

Yi Song[0000-0001-9037-5639], Yang Jiang[0000-0002-2195-5776], Jeremy Lee, Chunyi Ruan

ETS Research Institute, Princeton NJ 08541, USA
ysong@ets.org

Abstract. Motivated by growing interest in next-generation learning interfaces, we present the design and preliminary evaluation of an AI-enhanced learning activity that assesses and supports the development of college students' self-regulated learning (SRL) skills. Specifically, the activity integrates goal-setting support, embedded reading and note-taking tools, and a conversational agent that provides targeted scaffolding during planning, reading, and writing. In a cognitive lab study, we collected student perceptions of the activity and analyzed student-AI interaction patterns to explore how they engaged with AI support during this activity. Students perceived the activity as engaging and appropriately challenging, with self-reported learning gains and positive views of AI support. Analysis of student-AI interactions showed that students primarily leveraging AI as a writing support across drafting, composing, evaluation, and editing and more often seeking guidance and feedback than direct answers. We conclude with design implications for next-generation learning interfaces.

Keywords: GenAI, Human-AI Interaction, Interface Design, Self-Regulated Learning.

1 Introduction and Research Background

Generative Artificial intelligence (GenAI) is rapidly transforming how we learn, think, and interact with information. AI-powered learning systems generate content, provide feedback, and adapt to learners in real time. As AI tools become embedded in next-generation learning environments, learners must develop the ability to direct and manage their learning in complex, technology-rich environments [15]. Achieving this requires equipping students with self-regulated learning (SRL) skills, essential for sustained, lifelong learning [12]. Students with good SRL skills can effectively manage their learning, critically assess AI-generated information, and adapt their strategies across diverse contexts [7], whereas students who lack SRL skills may encounter difficulties with goal setting, strategy use, and motivation [6]. AI tools can augment, but not replace, human judgment**,** requiring learners to engage in critical thinking, strategic decision-making, and adaptive learning**.** To support the evolving needs of learners, research must focus on how SRL can be measured and promoted through the design and evaluation of AI-enhanced learning interfaces**.**

This study leverages GenAI not only as a learning support tool but also as a lens for assessing SRL through learners' interactions with AI. We designed an AI-enhanced learning activity that elicits observable SRL behaviors and examined both students' perceptions of the activity and their interaction patterns with the AI agent. Understanding students' perceptions is critical for interpreting how and why they engage with the AI agent. Specifically, the study addresses two research questions: (1) How do students perceive the AI-enhanced learning activity? and (2) How do students interact with AI in this activity, and how do these interactions reflect their SRL skills? By focusing on interaction-based evidence of SRL, this work contributes to designing next-generation learning environments that capture regulatory processes in situ and deepens our understanding of how students use AI to regulate their learning.

1.1 Self-Regulated Learning

SRL describes the cognitive, metacognitive, affective, motivational, and behavioral processes that learners engage in to regulate their learning [10]. Several theoretical frameworks provide insights into SRL constructs and collectively illustrate self-regulation as a dynamic process. Zimmerman's cyclic model [17] describes SRL as a three-phase process: forethought (planning and goal setting), performance (strategy use and self-monitoring), and self-reflection (evaluating outcomes and adjusting strategies). Winne and Hadwin [13] conceptualize SRL as an iterative process involving task perception, goal setting and planning, strategy implementation, and adaptation based on self-evaluation. This Four-Phase Model has a strong metacognitive perspective that recognizes self-regulated learners as active and managing their own learning through monitoring and the use of (meta)cognitive strategies. Although these SRL frameworks provide valuable foundations, they largely predate the emergence of AI, and thus do not explicitly account for the role of AI in learning processes.

1.2 Potential and Challenges of AI Tools for SRL

GenAI has the potential to enhance SRL by providing immediate, adaptive, and context-specific support, assisting learners in setting goals, generating ideas, synthesizing information, and refining their understanding through interactive feedback [5,7]. Unlike static resources, GenAI can personalize responses based on user input, providing tailored explanations. Given that evaluation and reflection are crucial components of SRL [14], AI tools can serve as scaffolds, prompting learners to articulate their reasoning, identify knowledge gaps, and navigate complex tasks. Recent technology advances increasingly integrate learning analytics, GenAI, and adaptive scaffolding to both measure and support learners' regulatory processes. For example, Li et al. [8] introduced FLoRA, a hybrid human-AI system that captures learning trace data and models learners' SRL behaviors to deliver real-time, personalized scaffolds across planning, monitoring, and reflection phases. Cheng et al. [3] showed that students' question-asking behaviors with GenAI serve as observable indicators of SRL, with more targeted and knowledge-seeking questions linked to improved writing performance. These tools reflect a shift toward AI systems that not only infer SRL from process data but also shape it through context-aware scaffolding, positioning AI as both a measurement instrument and a collaborative partner in regulating learning.

While AI provides a rich source of external regulation, the extent to which learners actively engage in metacognitive monitoring and self-evaluation when using AI remains an open question. One major concern is that students may over-rely on AI-generated content without critically evaluating its accuracy [16]. Research suggests that students often struggle with metacognitive monitoring, leading to misplaced confidence in AI-generated responses, particularly in complex tasks that require higher-order thinking and nuanced judgment. For example, despite improving immediate task performance, ChatGPT may reconfigure SRL processes in ways that promote cognitive offloading, shallow engagement, and reduced metacognitive control, raising concerns about long-term learning [4]. Unlike human instructors who encourage critical thinking and exploration, AI may lead students to skip the evaluation phase, undermining their ability to internalize feedback and improve future performance [2]. Additionally, AI assistance can shape help-seeking behaviors, fostering a dependency on external validation. As AI becomes more integrated into education, it is critical to understand how to balance its benefits with strategies that reinforce SRL, such as reflection and evaluation.

2 AI-Enhanced Learning Activity

Using an evidence-centered design approach [9], we developed an AI-enhanced learning activity to elicit and measure SRL skills (see Fig. 1.). The activity engages students with a real-world issue, requiring critical analysis of multiple sources to develop an informed position in an essay. The activity mirrors the demands of integrated reading and writing tasks commonly found in higher education, where students read to learn by extracting, synthesizing, and communicating complex information effectively [1].

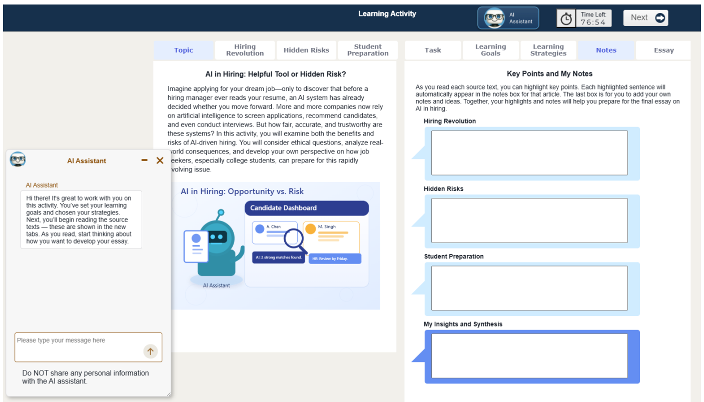

Fig. 1. A screenshot of the AI-enhanced learning activity.

The activity begins with a scenario, introducing the issue of AI use in hiring. Students then select and rank their learning goals and strategies. This goal-setting phase encourages students to articulate their intentions and take ownership of the learning process. Next, students read multiple source texts to gain background knowledge and different perspectives about this issue. As they read, students must actively construct knowledge, navigate conflicting information, and make sense of complex ideas. Highlighting and note-taking tools are embedded in the task to support metacognitive regulation as students identify, organize, and synthesize key information. In the culminating task, students write an essay demonstrating their understanding of the issue, which requires them not only to comprehend the content but also to synthesize ideas across sources and communicate ideas clearly and persuasively. Writing on controversial issues draws on a range of cognitive and metacognitive skills, requiring students to consider multiple perspectives and construct well-supported arguments, yet research shows that these skills are often underdeveloped [11]. The activity is designed to be completed within 90 minutes, and a timer is provided on the top bar of the interface.

Leveraging GenAI and prompt engineering, we developed a conversational AI agent, performing the role of a college instructor. AI support reflects core components of SRL, such as goal setting, strategy use, monitoring, and adaptation, as emphasized across multiple theoretical frameworks [13]. To provide targeted and relevant support, we first identified three stages of the learning activity: planning, reading, and writing. Then we set up specific prompting for the AI support aligned with each stage. For example, in the planning stage, the AI support focuses on selecting important goals and choosing effective strategies; in the reading stage, the AI support prioritizes reading strategies, such as highlighting key points, synthesizing information across sources, and evaluating ideas; in the writing stage, the AI agent helps students strengthen their essays through guidance and feedback. Additionally, the AI agent encourages students' critical thinking by avoiding giving away direct answers. Instead, it uses open-ended prompts, probing questions, and reflective feedback to guide students in evaluating information, considering alternatives, and articulating their own reasoning. In prompting, we provided detailed information about the activity, including all the texts, flow of the activity, and task directions. The AI agent also has access to student task responses.

An iterative process was applied to evaluate and refine the AI-generated responses, grounded in learning theories and SRL models. The refinement process involved multiple cycles of prompt engineering, in which candidate prompts were revised and tested against a set of hypothetical inputs that represent diverse interactions from learners.

3 Cognitive Lab Study Method

3.1 Data and Procedure

We conducted a cognitive lab study to examine the usability of the AI-enhanced learning activity and collect preliminary evidence of students' SRL skills through analyzing their interactions with the AI agent. Ten students (6 female, 3 male, 1 other) from three universities participated. This student sample included seven seniors, one freshman, one junior, and one graduate across four majors (computer science, education, social science, business). All the students reported their overall GPA as 3 points or higher.

Prior to the study, students completed an online survey about SRL skills and AI experience. Each student completed an online study session with a researcher. Each session started with an overview of the research. When the student worked on the activity, they shared their screen, enabling the researcher to observe their behavior and progress. After the student completed the activity, the researcher conducted an interview to collect student feedback and perception of the activity and AI agent. The student was asked to rate the design aspects (e.g., task difficulty, topic interest, activity engagement, AI helpfulness) on a 5-point Likert-type scale, and explained why they think so. The student also self-rated their knowledge about the AI in hiring issue before and after the activity on a 7-point scale. Each study session lasted up to two hours.

3.2 Annotation of Students' Interaction with AI

Informed by existing SRL frameworks [13,17], we developed an annotation framework to capture SRL behaviors indicated in the chats sent by students when interacting with AI. The framework consisted of two levels of coding. First, we identified the SRL dimension: task definition, goal setting and planning, enacting reading strategies, enacting writing strategies, monitoring, and regulating motivation/affect. Second, we identified specific SRL skills within a dimension (Table 1). Further, we identified the request type: direct answer, guidance, or feedback, indicating student agency in learning.

Table 1. SRL Annotation Framework

Dimension	Skill
Task definition	Activity requirements
	AI capability
	Reading task requirements
	Writing task requirements
Goal setting and planning	Goal setting
	Strategy selection
Enacting reading strategies	Reading comprehension
	Key points
	Summary of articles
	Notes of articles
	Comparison across articles
	Synthesis of ideas
Enacting writing strategies	Drafting
	Composition
	Evaluation
	Editing
Monitoring	Task progress monitoring
	AI response monitoring
	Time monitoring
Regulating motivation/affect	Positive regulation
	Negative regulation

There were 214 total chat turns, counting both student and AI messages. Our annotation focused on 107 student-sent chat messages to the AI agent. Two researchers independently annotated all the messages, achieving good reliabilities (98% agreement and $k = .96$ for the SRL dimension; 84% agreement and $k = .81$ for the SRL skill; and 81% agreement and $k = .71$ for the request type). The researchers discussed all the cases with different codes and reached an agreement on the final codes.

3.3 Task Responses

The platform captured and saved students' task responses, including their selected goals and strategies, highlighted texts from readings, notes about the readings, and essays. A

6-point rubric was developed to evaluate the overall quality of essays, in terms of position, arguments, consideration of different perspectives, and reflection on how students should prepare for AI-driven hiring. Two researchers independently annotated the 10 essays, achieving a 70% exact agreement and 100% adjacent agreement (score difference is no more than 1 point), with a weighted quadratic kappa of .92. The researchers discussed the three essays with different scores and reached an agreement on the final scores.

4 Preliminary Findings

4.1 Task Performance

In the learning activity, students were asked to identify their learning goals and strategies before proceeding to the reading and writing tasks. The top three goals identified by students are: (1) I want to build a solid understanding of how AI is used in hiring (80%); (2) I want to understand how AI in hiring might affect my own career opportunities (60%); and (3) I want to form and defend my own position on whether AI should be used in hiring (50%). The top three strategies selected by students are: (1) Highlight key ideas and take notes while reading (70%); (2) Plan the essay structure before starting to write (70%); and (3) Review my essay draft to see if my ideas are clear and well supported (60%). Student essay scores ranged from 2 to 6 points, with a mean of 4.2 (SD = 1.48). There was a small, non-significant correlation between student essay scores and the number of student-AI interactions (r = .31, p = .389), possibly due to the small sample size.

4.2 Student-AI Interaction

Students' use of AI varied. Among the 10 participants, one did not interact with AI at all, four students only had one or two interactions, while the rest sent AI from 10 to 28 messages. The most frequent SRL dimension of student messages was enacting writing strategies (69.2%), followed by enacting reading strategies (16.8%). This finding was related to the nature of the integrated reading and writing activity. It also suggested that students used AI primarily during the writing phase. The monitoring dimension was 7.5%, and the task definition dimension was 5.6%. Other dimensions were rare.

Students used AI mostly to conduct essay evaluation (26.2%), followed by essay editing (16.8%) and essay composition (15%). Both essay drafting and summary of source texts were 11.2%. No other categories were larger than 5%. This result clearly showed that students turned to AI at multiple writing stages. In terms of the request type, asking AI to provide guidance was the largest (38.1%), followed by asking for feedback (34.3%), and lastly asking AI to provide a direct answer (27.6%).

4.3 Student Perception

In the post-activity interview, we collected student feedback on the design of this AI-enhanced learning activity. Overall, students perceived the reading task as easy (M = 1.6, SD = .70) and the writing task as normal (M = 2.9, SD = .77) in terms of difficulty.

They found the topic interesting (*M* = 4.5, *SD* = .71) and activity engaging (*M* = 4.2, *SD* = .92). Seven students perceived the AI responses as helpful or very helpful, while three students reported not using AI or using it minimally and therefore could not meaningfully evaluate it (*M* = 4.2, *SD* = 1.07). Students' self-rated average scores of their knowledge about the topic increased after the completion of the activity (Pre-activity: *M* = 3.4, *SD* = 1.26; Post-activity: *M* = 5.1, *SD* = 1.20).

Students shared their perspectives on the activity design and offered several recommendations. Overall, the students viewed the task layout and flow positively, describing the interface as easy to navigate and the organization of reading, note-taking and writing as generally clear. At the same time, they pointed out a few design issues that could be addressed. First, one student commented that notes are displayed in the order highlighted rather than in the order in which they appear in the articles, and multiple students thought the note boxes were small. Second, multiple students suggested adding scaffolding for the essay task (e.g., breaking down into writing phases, and prompting for integrating evidence). Third, students felt that the AI support was inconsistent at times, with responses that were not always well aligned with their questions or writing needs.

5 Discussion

Overall, the findings of our cognitive lab study suggest that students responded positively to the AI-enhanced learning activity, perceiving it as engaging and appropriately challenging. Students' self-reported knowledge gains, alongside favorable perceptions of task interest and AI helpfulness, indicate that the activity may have supported students' learning about the issue. Analysis of student-AI interactions revealed substantial variation in how students engaged with the AI, ranging from minimal to intensive use. Students most frequently used AI to enact writing strategies across drafting, composing, editing, and evaluation phases, indicating that they primarily viewed AI as a writing support. Requests for guidance and feedback were more common than requests for direct answers, a pattern consistent with the AI agent's design to avoid doing the task for students and with the study context in which students were aware they were being observed. The observed correlation between the number of student-AI interactions and their essay scores should be interpreted as preliminary due to the small sample size. For next steps, we will examine how students' SRL behaviors, as reflected in their AI interactions and writing processes, relate to the quality of their task performance.

Taken together, these findings highlight the multifaceted roles AI can play in student writing activities and point to opportunities for future design work that more explicitly scaffolds strategic and reflective uses of AI. In particular, the activity could be improved by refining the note-taking interface, adding scaffolds for the essay task, and improving the relevancy of AI responses. Our research has multiple implications on designing next-generation learning interfaces: (1) Interfaces should make task phases explicit and deliver phase-specific AI scaffolds to better align support with learners' evolving needs; (2) Interfaces should adopt guidance and feedback strategies rather than giving direct answers, so that learners are encouraged to practice their metacognitive skills; and (3) Interfaces should embed reflection checkpoints, which may help learners monitor understanding, reflect on goals, and evaluate performance.

References

1. Association of American Colleges and Universities: *The LEAP vision for learning: Outcomes, practices, impact, and employers' view*. Author, Washington, DC (2011)
2. Atchley, P., Pannell, H., Wofford, K., Hopkins, M., Atchley, R.A.: Human and AI collaboration in the higher education environment: Opportunities and concerns. *Cognitive Research: Principles and Implications* 9, Article 20 (2024)
3. Cheng, Y., Fan, Y., Li, X., Chen, G., Gašević, D., Swiecki, Z.: Asking generative artificial intelligence the right questions improves writing performance. *Computers and Education: Artificial Intelligence* 8, 100374 (2025)
4. Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., Gašević, D.: Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology* 56, 489–530 (2025)
5. Jin, S.H., Im, K., Yoo, M. et al.: Supporting students' self-regulated learning in online learning using artificial intelligence applications. *International Journal of Educational Technology in Higher Education* 20, 37 (2023)
6. Karaoglan Yilmaz, F.G., Yilmaz, R.: Learning analytics as a metacognitive tool to influence learner transactional distance and motivation in online learning environments. *Innovations in Education and Teaching International* 58(5), 575–585 (2021)
7. Lai, J.W.: Adapting self-regulated learning in an age of generative artificial intelligence chatbots. *Future Internet* 16, 218 (2024)
8. Li, X., Li, T., Yan, L., Li, Y., Zhao, L., Raković, M., Molenaar, I., Gašević, D., Fan, Y.: FLoRA: An advanced AI-powered engine to facilitate hybrid human-AI regulated learning. *Computers & Education* 243, 105527 (2025)
9. Mislevy, R.J., Steinberg, L.S., Almond, R.G.: On the structure of educational assessm105527 (*surement: Interdisciplinary Research and Perspectives* 1(1), 3–67 (2003)
10. Panadero, E.: A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology* (2017)
11. Raković, M., Winne, P.H.: SR-WMS: A typology of self-regulation in writing from multiple sources. In: Wang, Y., Joksimović, S., San Pedro, M.O.Z., Way, J.D., Whitmer, J. (eds.) *Social and Emotional Learning and Complex Skills Assessment*. Springer, Cham (2022)
12. Roth, A., Ogrin, S., Schmitz, B.: Assessing self-regulated learning in higher education: A systematic literature review of self-report instruments. *Educational Assessment, Evaluation and Accountability* 28, 225–250 (2016)
13. Winne, P.H., Hadwin, A.F.: Studying as self-regulated learning. In: Hacker, D.J., Dunlosky, J., Graesser, A. (eds.) *Metacognition in Educational Theory and Practice*, pp. 277–304. Lawrence Erlbaum, Hillsdale, NJ (1998)
14. Winne, P.H., Perry, N.E.: Measuring self-regulated learning. In: Boekaerts, M., Pintrich, P.R., Zeidner, M. (eds.) *Handbook of Self-Regulation*, pp. 531–566. Academic Press, San Diego (2000)
15. World Economic Forum: *The future of jobs report 2020*. World Economic Forum (2020)
16. Zamfirescu-Pereira, J., Wong, R.Y., Hartmann, B., Yang, Q.: Why Johnny can't prompt: How non-AI experts try (and fail) to design LLM prompts. In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–21 (2023)
17. Zimmerman, B.J.: Attaining self-regulation: A social cognitive perspective. In: Boekaerts, M., Zeidner, M., Pintrich, P.R. (eds.) *Handbook of Self-Regulation*, pp. 13–39. Academic Press, San Diego, CA (2000)