The Impact of Human-AI Interaction Patterns on Problem Solving, AI Literacy, and Metacognition

Wenting Sun[1] and Jiangyue Liu[2]
[1]Humboldt-Universität zu Berlin, Germany
[2]Soochow University, Jiangsu, China

suwentin@hu-berlin.de
ljy@suda.edu.cn

Abstract: Human-AI interaction, particularly in educational contexts, is a dynamic and cognitively demanding process that holds promise for enhancing goal-directed learning. Yet, there remains a scarcity of empirical studies that examine how learners' interaction with generative AI (GenAI) varies in structure and how these patterns influence distinct learning outcomes. This study investigates the relationship between human-AI interaction processes and outcomes such as AI literacy, problem-solving skills, metacognitive strategies, and task performance. We conducted an experimental study with 45 secondary school physics student teachers engaged in a GenAI-supported lesson plan assessment task. Using questionnaire responses, trace data, and prompt logs, we coded human-AI interaction behaviours based on self-regulation and cognitive processing levels. Through sequence clustering analysis, we identified two distinct interaction patterns. Both clusters showed significant improvement in task performance, but with divergent benefits. Cluster 1 exhibited diverse regulation processes characterized by exploratory, divergent prompting and low-level cognitive engagement in the early stages. This group showed significant gains in problem-solving skills through active idea generation and broad reflection. Cluster 2 demonstrated structured regulation behaviours, initiating interaction with deep-level cognitive processing and convergent prompting. These learners made more deliberate modifications and completed full self-regulated learning (SRL) cycles—planning, monitoring, and reflecting—which led to enhanced AI literacy and metacognitive strategy use. Our findings suggest that effective human-AI collaboration goes beyond prompt diversity; structured regulation behaviours serve as a key mediator between prompting and learning gains. GenAI served as both cognitive and metacognitive scaffolding, facilitating critical assessment and productive delegation. These results contribute to SRL theory in AI contexts and emphasize the importance of process-level analysis. Limitations include a small sample and limited prompt feature analysis. Future research should explore emotion-aware AI systems, multimodal interaction data, and the impact of task complexity on interaction dynamics. This study provides practical insights for educators and designers of AI-integrated learning systems. Specifically, it highlights the importance of tailoring AI scaffolds to different learner regulation styles: for exploratory learners, scaffolds can encourage broad idea generation and reflection, while for structured learners, scaffolds should support iterative planning and monitoring. These findings underline both opportunities and limitations of current GenAI use in classrooms, suggesting concrete directions for teacher practice and instructional design.

Keywords: AI literacy, Human-AI interaction, Generative AI in education, Prompt engineering, Self-regulated learning

1. Introduction

As a subset of Generative AI (GenAI), large language model (LLM)-based tools have been explored in various educational contexts. GenAI supports personalized learning materials and eases teacher workload by generating formative and summative feedback to enhance learner outcomes (Meyer et al., 2024). Some studies show learners' use of GenAI reflects their critical thinking (M. Liu et al., 2024), while others warn it may foster weak learning habits and reduce evaluative thinking (Lo et al., 2024). Understanding GenAI's role in complex problem solving requires attention to technology-, task-, learning-, and human-machine-related factors (Cetindamar et al., 2024). As learners interact with GenAI, they adapt their behaviour to better utilize its capabilities (Joksimovic et al., 2023). Therefore, detailed insights into human-AI interaction are needed to assess GenAI's educational value.

Prompting, or prompt engineering, is a dynamic process involving iterative experimentation with different strategies to affect input-output results (Zamfrescu-Pereira et al., 2023). It allows users to define and delegate tasks during problem solving (Dang et al., 2022). This dynamic becomes more critical in GenAI-supported complex tasks. Santu and Feng (2023) note that complex tasks often involve multiple steps, and LLM performance varies with prompt types. Thus, engaging learners in deep cognitive and metacognitive processes is essential when using GenAI as a support tool.

Most current studies examine AI literacy and prompt engineering in human-AI collaboration from a static lens (Knoth et al., 2024; Oppenlaender et al., 2024; Zamfrescu-Pereira et al., 2023). A sequential, temporal lens offers deeper understanding of human-AI dynamics and informs scaffolding in learning environments (Chen et al., 2025; Nguyen et al., 2024a; 2024b). To explore GenAI's role in educational problem solving, we adopt a process-oriented approach, focusing on prompting as a key factor. The research questions are: (1) What human-AI interaction patterns emerge from learners' regulatory and cognitive behaviours? (2) How do these patterns relate to AI literacy, learning performance (problem-solving ability, metacognition, task outcomes), and prompt strategies? This study aims to illuminate GenAI's complex influence on problem solving and guide its responsible integration in classrooms.

2. Theoretical Background and Hypotheses

2.1 Using GenAI for Problem Solving

In an AI-driven world, individuals must develop capabilities like self-regulated learning (SRL), sense-making, and collaboration to adapt (Markauskaite et al., 2022). Kong et al. (2024) argue that beyond conceptual AI knowledge, learners need metacognitive engagement to effectively apply AI in real-world problem solving.

GenAI generates a diverse pool of potential solutions, enriching the problem space and inspiring novel ideas (Urban et al., 2024). This prompts learners to spend more time on critical thinking—selecting relevant information, refining ideas, debugging, and adjusting approaches (Yilmaz & Yilmaz, 2023; Urban et al., 2024). GenAI also supports knowledge construction and self-regulation (Lee et al., 2024), boosting learners' confidence and aiding skill development (Kong et al., 2024).

However, effective use of GenAI requires balancing cognitive strategies with technological affordances (Nguyen et al., 2024b). Despite this complexity, little research explores how learners dynamically integrate AI outputs into their problem-solving processes (Joksimovic et al., 2023).

2.2 Human-AI Interaction and AI Literacy

AI literacy—key in today's AI-rich environment—encompasses the skills to critically assess, communicate with, and use AI effectively (Long & Magerko, 2020). Perspectives vary on its components, with socio-technical competence gaining attention (Pinski & Benlian, 2023). Most AI literacy studies rely on surveys or self-reported perceptions (e.g., Pinski & Benlian, 2023; Knoth et al., 2024). While these approaches are valuable, they are limited in capturing how learners actually integrate AI outputs into complex tasks. By contrast, our study adopts a process-trace approach, analysing fine-grained behavioural sequences. This critical shift allows us to go beyond static measures of AI knowledge and instead reveal how regulation and prompting strategies dynamically shape literacy outcomes.

Studies show AI usage and evaluation are linked to behavioural engagement and higher-order thinking (Lu et al., 2024), suggesting static AI knowledge alone doesn't guarantee improved task performance. Few empirical studies connect AI literacy, human-AI interaction patterns, and learning outcomes in GenAI-supported tasks.

2.3 Prompt Engineering

Prompt engineering, more directly tied to task performance than AI literacy, involves crafting inputs that guide GenAI toward desired outcomes (Oppenlaender et al., 2024). This skill requires both linguistic knowledge and strategic use of modifiers.

Quantifying prompt quality helps visualize its relationship to other variables. For instance, Knoth et al. (2024) used component-based scores to suggest prompt engineering may be independent of general AI literacy. Lee et al. (2024) linked the variety of prompt subjects to divergent and convergent thinking. The iterative nature of prompting parallels creative processes like drafting and revising (Habib et al., 2025), supporting the democratization of creativity (Dang et al., 2022).

2.4 Human-AI Interaction and SRL

Effective prompting relies on deep cognitive engagement, similar to SRL behaviours (Dang et al., 2022; Knoth et al., 2024). Prompting often involves identifying and correcting AI errors through systematic strategy use (Zamfirescu-Pereira et al., 2023). Similarly, productive SRL learners employ structured approaches such as rereading and sequential analysis to achieve goals (Rakovic et al., 2023). To understand these strategies in GenAI-assisted tasks, analysing temporal data adds depth. Nguyen et al. (2024a) identified structured and unstructured human-AI interaction types using process mining. Lee et al. (2024) used cosine similarity to classify prompting styles.

SRL research methods offer useful parallels. Markauskaite et al. (2022) recommend unobtrusive data collection to examine human-AI interaction as a dynamic process. Some studies have applied this: e.g., Nguyen et al. (2024a, 2024b) used screen recordings and sequential analysis, while Chen et al. (2025) applied Markov models to compare help-seeking behaviours in AI-assisted groups. Urban et al. (2024) stress the need to study temporal aspects using trace data. Unlike prior studies focused on writing tasks, our research uses lesson plan assessment to analyse problem solving.

Based on prior analysis, the three hypotheses are proposed:

H1: AI literacy changes differ across human-AI interaction patterns.

H2: More structured interaction patterns are positively associated with better learning performance (problem-solving skills, metacognition, task outcomes).

H3: Different interaction patterns exhibit distinct prompt engineering features.

3. Methods

3.1 Research Context and Participants

Participants were 54 third-year student teachers enrolled in secondary school physics education at a Chinese university. Due to incomplete screen recordings, data from 9 participants were excluded, resulting in a final sample of 45 participants.

The study was conducted within an educational technology course, where students were required to design a microlesson and create a lesson plan. The GenAI-assisted lesson plan assessment task occurred after students completed their initial lesson plans. This 90-minute task involved three stages: (1) self-assessment of their own lesson plan, (2) peer assessment using AI tools, and (3) a final revision of their lesson plan before submission. Task performance was evaluated by comparing the initial and revised versions.

Participants used ERNIE Bot 3.5 V2.5.0, a large language model (LLM) developed by Baidu and similar in functionality to ChatGPT. ERNIE was chosen for its accessibility and maturity as a local AI tool. During the assessment, students had unrestricted access to both ERNIE and online search engines.

3.2 Data Collection and Analysis Procedures

3.2.1 Data collection

Questionnaires: AI Literacy (AIL) scale from Pinski and Benlian (2023) was used, excluding the "overall" item due to concerns over reliability (Knoth et al., 2024). Problem Solving Skills (PSS) scale was adapted from Heppner & Petersen (1982), covering problem-solving confidence, approach-avoidance style, and personal control. Metacognitive Strategies scale was adapted from Pintrich et al. (1991), measuring rehearsal, elaboration, organization, critical thinking, and metacognitive self-regulation. All items were translated into Chinese (participants' mother tongue) and rated on a 5-point Likert scale. Cronbach's alpha indicated good internal consistency ($\alpha = 0.896$).

Human-AI Interaction Trace Data: All lab computers used Windows 10 and a custom-built screen recorder ("PSR.exe") that captured user actions and screens as .mht files. Python scripts were developed to extract structured logs from these files, including timestamped logs of mouse movements, clicks, keystrokes, action windows, and action content.

Prompt Engineering Assessment: Prompt datasets were manually extracted from screen captures. Two raters independently coded the prompts, reaching 90% inter-rater agreement. Features (details in Table 1) were evaluated using: Component-based scoring (Eager & Brunton, 2023): e.g., number of unique verbs, focus terms, contexts, constraints, etc. Divergent-Convergent Thinking metrics (Lee et al., 2024): e.g., number of prompt turns, prompt word count, prompt diversity.

Lesson Plan Evaluation: Two raters evaluated lesson plans (before and after AI use) using the rubric from König et al. (2021), which includes six criteria: content transformation, task creation, adaptation to learners, clarity of objectives, contextualization, and instructional phasing (max score: 33).

Table 1: Prompt components

Prompt features	Indicators
Verb, focus, context, focus and constraints, alignment, constraints and limitations	n_unique_verb; n_unique_focus; n_unique_context; n_unique_focus and constraints; n_unique_ alignment; n_unique_ constraints and limitations
Divergent-convergent thinking	n_prompt_turn; n_word_prompt; n_unique_prompt; n_word_prompt_under_5; n_subject_prompt

3.2.2 Analytical methods

To address RQ1, raw trace data were first mapped into SRL processes, followed by sequential clustering and process mining to detect patterns of human-AI interaction aligned with SRL theory.

Human-AI Interaction Coding Scheme. A hybrid coding approach was used. First, theory-driven method was used. Initial code scheme was created based on SRL process coding from Raković et al. (2023), and human-AI action categories from Liu et al. (2024) and Nguyen et al. (2024a). Then data-driven was employed. It means the created code scheme was refined to suit the observed behaviour in this dataset (see Table 2). Actions were coded from raw trace logs and aggregated into SRL process labels at 5-second intervals, based on the frequency of high-level actions.

Table 2: The SRL process and action library for Human-AI interaction.

SRL processes (12)	**Action label (44)**
Planning	First-reading task lists/evaluation rubrics
Reading	Reading AI outputs; Re-Reading AI outputs; Reading lesson plan
Navigation	File-searching
Monitoring	Re-read task lists; Check screen-records setting; Re-read evaluation rubrics
Orientation	Download & Upload-file; Refresh- Learning management system Rename lesson plan; Login AI or Learning management system; WORD revision setting
High-Interaction with AI	Prompt-follow up; Search-engine for the same task; Interaction-other AI tools for the same task
Low-Interaction with AI	Prompt-typed; Prompt-pasted from other materials; Prompt-file uploaded; Prompt-polish words; Prompt-new; Prompt-old; Prompt-cancel; Explore AI interface
Low-AI outputs Elaboration	Copy Paste-AI outputs; Copy Paste All-AI outputs; Delete Pasted-AI outputs; Replace using AI outputs
High-AI outputs Elaboration	Add explanation based on AI outputs; Delete unrelated pasted AI outputs; Paraphrase based on AI outputs
High-modify	Add-Screenshot; Paste Not-AI outputs; Delete Not-AI outputs; Reorganize sections; Add-text; Text position modification
Low-modify	Modification Distance; Serial number modification; Modification Format; Undo; Save; Try to change
Courseware	Modify self-Courseware

Human-AI Interaction Sequence Clustering and Process Mining. To identify distinct human-AI interaction patterns, we used **sequence clustering**, a method that groups learners based on similarities in their interaction processes. In practice, we combined Optimal Matching (OM) to detect subsequence similarities and Ward's Clustering (WC) to group learners into clusters (Murtagh & Legendre, 2014). OM was implemented with the TraMineR package in R (Gabadinho et al., 2011), and WC was performed using the cluster package (Maechler et al., 2015). This approach enabled us to distinguish two typical patterns of regulation and prompting while keeping the analysis grounded in sequential SRL-based actions. To further interpret the micro-level behavioural sequences within each identified cluster, we applied process mining techniques, a method increasingly used in human-AI interaction research (Nguyen et al., 2024a; Chen et al., 2025). Specifically, we employed the Fuzzy Miner approach using Disco software (by Fluxicon) to visualize the flow and frequency of interactions, offering insights into the temporal structure and cognitive regulation patterns within each interaction type.

Analysis for RQ2

To explore the relationships between human-AI interaction patterns and learning performance outcomes, several quantitative analyses were conducted using SPSS. For baseline differences, a Mann-Whitney U test was first performed to compare the pre-test scores across identified interaction clusters. This ensured that differences in post-test outcomes were not confounded by prior knowledge or individual learner characteristics.

Regarding performance and strategy changes (pre vs. post), a comparison was conducted to analyse within-group changes in AI Literacy (AIL), Problem-Solving Skills (PSS), Metacognitive Strategies (MS), and Lesson Plan Quality. Paired-sample t-tests were used for each variable within each cluster, and Cohen's d was calculated to estimate effect sizes for each change.

Prompt Engineering Features: To explore whether different human-AI interaction patterns were associated with differences in prompting behaviour, a Mann-Whitney U test was applied to compare prompt engineering features across clusters.

Correlations Between Prompting and Interaction Patterns: To examine the relationships between prompt engineering features and interaction behaviours, a Spearman correlation analysis was conducted.

4. Results

In total, 31,015 SRL process were coded in human-AI interaction and 276 prompts were extracted.

4.1 RQ1: What Human-AI Interaction Patterns can be Identified in the Context of Lesson Plan Assessment Task?

As shown in Figure 1 (left), 45 sequential datasets were clustered into two groups: Cluster 1 (n = 18) and Cluster 2 (n = 27). The agglomerative coefficient of 0.76 indicated strong cluster cohesion. Figure 1 (right) displays the SRL process distributions, with Cluster 1 showing longer revision durations, suggesting greater engagement in cognitive and metacognitive activities.

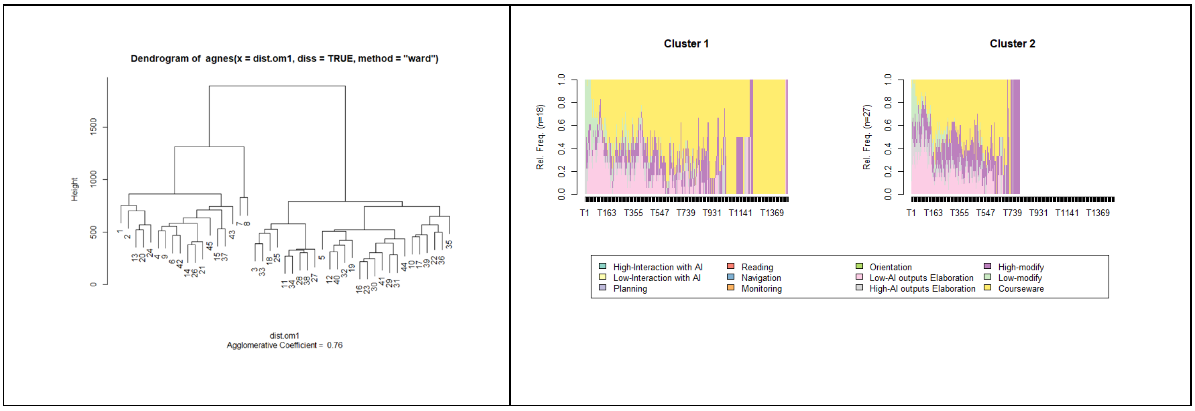

Figure 1: Ward's clustering results (left) and Sequence distribution (right) for 45 human-AI interaction process sequences

Process mining revealed distinct human-AI interaction patterns between the two clusters. The human-AI interaction behaviours of cluster1 and 2 can be found in Figure 2 and 3 respectively. Arrows indicate the main activity sequences. The darker the colour of a box, the higher the frequency of that behaviour. As shown in Figure 2, Cluster 1 began the task in three ways: Navigation (file search), Low-modify (lesson plan editing), and Planning (reading task lists/rubrics). Two main paths emerged: one progressed from Navigation→Monitoring →Orientation→Low→High AI-output elaboration→High-modify, and the other from Planning→Low→High AI interaction→Reading. This indicates students either focused on formatting issues or copied AI outputs directly, with interpretation often occurring before prompting or reading, rather than after. Cluster 1 exhibited diverse but less reflective regulation, as monitoring and orientation behaviours declined in later phases.

By contrast, Cluster 2 followed more structured paths (Figure 3) (seen in the next page). They prioritized deep processing—starting from Planning → Low AI interaction, followed by a double loop: High AI interaction↔Reading↔High-modify, suggesting systematic evaluation and integration of AI outputs into content. This was followed by Low-modify, monitoring, and re-planning, indicating iterative SRL behaviour: planning, performing, reflecting, and replanning.

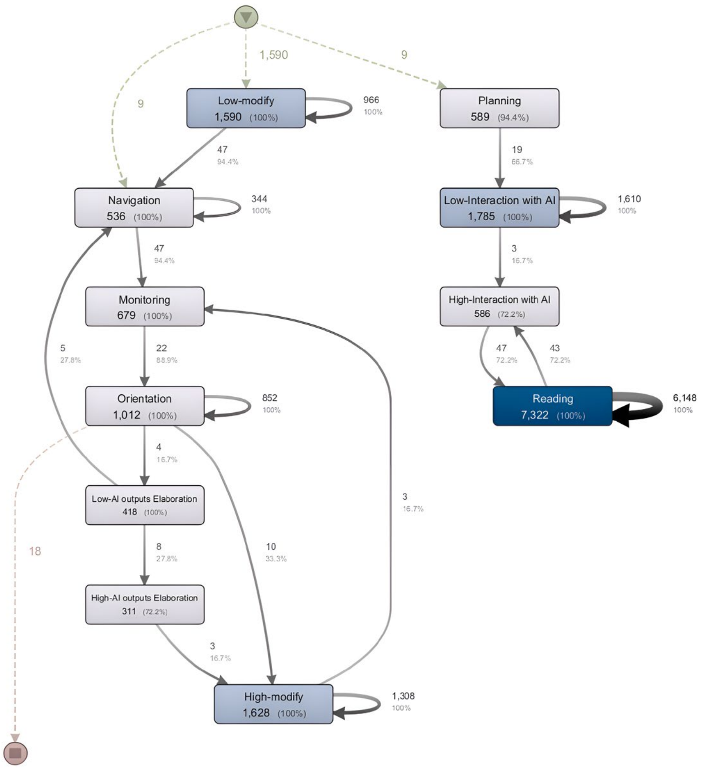

Figure 2: Cluster 1 human-AI interaction process maps-diverse regulation process

4.2 RQ2: How are Human-AI Interaction Patterns relating to Learning Performance?

To assess the impact of human-AI interaction patterns on learning performance, a Mann-Whitney U test was first conducted to compare pre-test differences between clusters. As shown in Table 3, no significant differences were found, indicating similar baseline characteristics. No significant post-test differences were found either, suggesting that a single test may be insufficient to capture GenAI's learning impact.

Table 3: Descriptive analysis of the PSS, MS, AIL, and lesson plan

Cluster 1 (N=18)			**Cluster 2 (N=27)**			**Mann-Whitney U test**	
Variables	**Mean**	**SD**	**Variables**	**Mean**	**SD**	**Z**	**Sig. (2-tailed)**
Pre-AIL	3.24	0.44	Pre-AIL	3.19	0.59	-0.51	0.61
Post-AIL	3.30	0.46	Post-AIL	3.44	0.63	-1.08	0.28
Pre-PSS	3.28	0.42	Pre-PSS	3.39	0.35	-1.19	0.23
Post-PSS	3.48	0.41	Post-PSS	3.44	0.35	-0.19	0.85
Pre-MS	3.55	0.41	Pre-MS	3.51	0.62	-0.91	0.37
Post-MS	3.57	0.35	Post-MS	3.69	0.54	-0.98	0.33
Pre-lesson plan	26.39	3.84	Pre-lesson plan	26.11	4.08	-0.26	0.80
Post-lesson plan	28.89	2.54	Post-lesson plan	28.19	4.74	-0.15	0.88

Note. Pre=Pre-test, Post=Post-test, AIL=AI literacy, PSS=Problem solving skills, MS=Metacognitive strategies.

For H1 and H2, paired sample t-tests were used to examine performance changes within clusters. As shown in Table 4, AI literacy did not significantly improve in Cluster 1 ($t = -0.66$, $p > 0.05$), but showed a significant gain in Cluster 2 ($t = -3.54$, $p < 0.05$, $d = 0.42$), suggesting that structured regulation enhanced AI literacy.

Table 4: Paired sample t-test results for H1

Clusters	**Pairs**	**Mean**	**SD**	**SE**	**t**	**df**	**Cohen's d**
Cluster 1	Pre-AIL - Post-AIL	-0.06	0.41	0.1	-0.66	17	0.14
Cluster 2	Pre-AIL - Post-AIL	-0.25	0.37	0.07	-3.54	26	0.42**

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

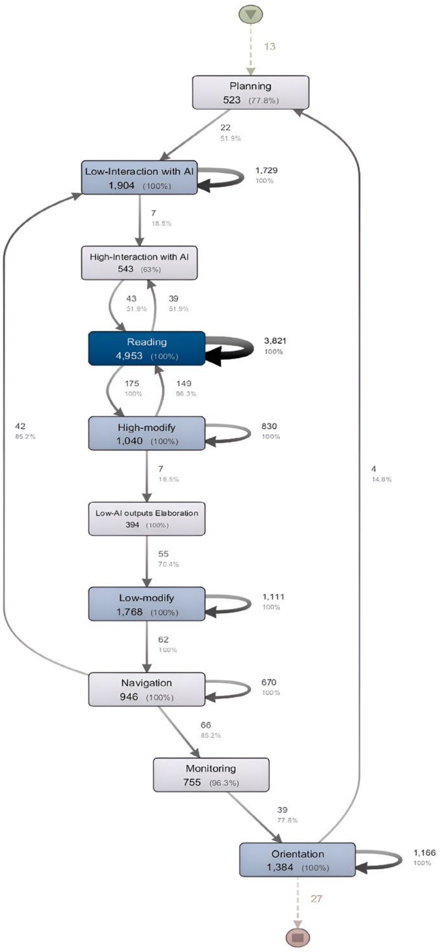

Figure 3: Cluster 2 human-AI interaction process maps-structured regulation process

Table 5 shows both clusters improved significantly in task performance (lesson plans), with a stronger effect in Cluster 1 (d = 0.77 vs. 0.47). Cluster 1 also showed significant gains in problem-solving skills (PSS) ($t = -2.23$, $p < 0.05$, $d = 0.48$), while Cluster 2 improved in metacognitive strategies (MS) ($t = -0.85$, $p < 0.01$, $d = 0.32$). These results imply diverse regulation supports PSS, while structured regulation enhances MS.

Table 5: Paired sample t-test results for H2

Clusters	**Pairs**	**Mean**	**SD**	**SE**	**t**	**df**	**Cohen's d**
Cluster 1	Lesson plan-A – Lesson plan-B	-2.5	4.08	0.96	-2.6	17	0.77*
	Pre-PSS -Post-PSS	-0.20	0.38	0.09	-2.23	17	0.48*
	Pre-MS-Post-MS	-0.01	0.36	0.09	-0.66	17	0.04
Cluster 2	Lesson plan-A – Lesson plan-B	-2.07	2.85	0.55	-3.78	26	0.47***
	Pre-PSS -Post-PSS	-0.05	0.33	0.06	-0.85	26	0.15
	Pre-MS-Post-MS	-0.19	0.34	0.06	-0.85	26	0.32**

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

For H3, a Mann-Whitney U test was used to compare prompting features between clusters. As shown in Table 6, Cluster 1 had significantly higher values in n_word_prompt ($Z = -2.75$, $p < 0.01$), n_unique_verb ($Z = -2.1$, $p < 0.05$), and n_unique_focus ($Z = -2.63$, $p < 0.01$), with effect sizes of -0.49, -0.37, and -0.46, respectively. This suggests Cluster 1 used longer prompts with more varied verbs and focus expressions. Based on Lee et al. (2024), Cluster 1 demonstrated a divergent thinking approach in prompt engineering, while Cluster 2 used a more convergent strategy.

Table 6: Comparison of prompt features by Mann-Whitney U (H3)

Prompt features	Mean (Cluster 1)	Mean (Cluster 2)	Z	Effect Size	Sig. (2-tailed)
n_prompt_turn	7.28	5.3	-1.61	-0.28	0.107
n_subject_prompt	6	3.93	-1.91	-0.34	0.057
n_unique_prompt	6.94	4.89	-1.67	-0.29	0.095
n_word_prompt_under_5	0.67	0.56	-0.26	-0.04	0.799
n_word_prompt	173.22	95.59	-2.75	-0.49	0.006**
n_unique_verb	4.78	3.04	-2.1	-0.37	0.036*
n_unique_focus	4.72	2.67	-2.63	-0.46	0.009**
n_unique_context	1.11	1	-0.49	-0.08	0.624
n_unique_focus and condition	1.72	1.26	-0.65	-0.11	0.516
n_unique_alignment	0.78	0.33	-1.76	-0.27	0.078
n_unique_constraints and limitations	0.67	0.3	-1.91	-0.26	0.056

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

5. Discussions and Implications

RQ1: Our findings confirm prior research on human-AI interaction patterns (Nguyen et al., 2024a). Cluster 1 showed an exploratory, iterative approach (diverse regulation, divergent prompting), while Cluster 2 followed a goal-oriented method (structured regulation, convergent prompting). High-AI interaction and high-modify behaviours in Cluster 2 involved critical engagement—reading, prompting follow-ups, evaluating, and integrating outputs into lesson plans. While both clusters engaged in higher-order thinking, Cluster 2 demonstrated tighter sequences connecting AI interaction to meaningful modifications. Unlike Cluster 1, Cluster 2 also showed more complete SRL cycles (planning, monitoring, reflection) and iterative behaviour, possibly due to slightly higher pre-test problem-solving scores (PSS = 3.38 > 3.28). This supports the idea that stronger problem-solving skills foster reflective human-AI interactions (Chiu et al., 2024).

RQ2: Consistent with Knoth et al. (2024), our findings indicate that AI literacy alone is insufficient for effective prompting. However, Cluster 2's structured regulation and convergent prompting significantly improved AI literacy, echoing findings from Chen et al. (2025). GenAI acted as a cognitive and metacognitive scaffold, helping learners understand the human-AI collaboration from a socio-technical lens.

GenAI boosted task performance across clusters, consistent with prior research (Yilmaz & Yilmaz, 2023; Urban et al., 2024). However, the type of human-AI interaction influenced how performance improved. Cluster 1's divergent prompting encouraged idea generation and reflection, which supports PSS development through active exploration. In contrast, Cluster 2's structured approach enabled fast, meaningful iterations by offloading routine tasks to GenAI, fostering metacognitive strategies and efficient cognitive work (Markauskaite et al., 2022).

While Knoth et al. (2024) found direct links between prompting and AI literacy, our results suggest that regulation mediates this link. Cluster 1's use of more unique verbs and focus terms didn't lead to better AI literacy, possibly due to lack of reflection. This implies that divergent prompting without deep regulation may hinder conceptual clarity.

Theoretically, SRL frameworks help explain human-AI learning. SRL variables—technology-, task-, learning-, and human-machine-related—intertwine in GenAI-supported problem solving (Cetindamar et al., 2024). More research is needed to validate SRL in AI contexts.

To assess AI literacy and prompting in goal-directed learning, multimodal approaches are needed—beyond surveys, including prompt features and task delegation (Pinski et al., 2023). Analysing human-AI interaction processes offers micro-level insights into AI literacy. Profiling users' prompting behavior can support training and scaffolded learning.

Implications for Practice. For teachers, our findings suggest that GenAI can be positioned differently depending on learners' interaction styles. For exploratory students (Cluster 1), teachers may design tasks that prompt idea generation and wide-ranging exploration, with AI scaffolds encouraging reflection before task submission. For structured students (Cluster 2), scaffolds can emphasise iterative cycles of planning, monitoring, and revision, guiding learners to critically evaluate and selectively integrate AI outputs. For instructional designers, these insights point to the need for adaptive scaffolds that respond to learners' self-regulation tendencies rather than offering uniform prompts.

6. Conclusions

Our study identified two human-AI interaction patterns through sequence clustering and analyzed their process and outcome differences. Both clusters showed significant task performance improvement. The diverse regulation cluster used more unique verbs and focus on prompting, enhancing problem-solving skills, while the structured regulation cluster showed gains in AI literacy and metacognitive strategies. These findings highlight how human-AI interaction patterns influence different learning outcomes and the need to develop varied capabilities for AI-assisted problem solving.

Several limitations exist. The small sample size limits generalizability. Future research with larger, more diverse populations and multiple languages would yield more robust, real-world insights. Our prompt engineering analysis was limited to two feature types; Including help-seeking behaviour could deepen understanding of SRL in AI-assisted tasks (Chen et al., 2025).

Future studies should incorporate multimodal or multichannel analyses, such as detecting emotions during interaction, which current trace data miss. Leveraging LLMs' emotion detection could improve human-AI interaction. Beyond AI literacy and prompting, research should explore impacts on higher-order thinking (e.g., critical thinking, creativity) and how task complexity and delegation affect interaction and performance (Pinski et al., 2023).

Ethics Declaration: Ethical clearance was not required for this study.

AI Declaration: AI tools (specifically ChatGPT by OpenAI) were used during the writing process to assist with grammar refinement and language polishing. All research design, data analysis, and interpretation were conducted solely by the authors.

References

Cetindamar, D., Kitto, K., Wu, M., Zhang, Y., Abedin, B. and Knight, S. (2024) "Explicating AI literacy of employees at digital workplaces", IEEE Transactions on Engineering Management, Vol 71, pp 810–823. https://doi.org/10.1109/TEM.2021.3138503

Chen, A., Xiang, M., Zhou, J., Jia, J., Shang, J., Li, X. and Fan, Y. (2025) "Unpacking help-seeking process through multimodal learning analytics: A comparative study of ChatGPT vs Human expert", Computers & Education, Vol 226, 105198. https://doi.org/10.1016/j.compedu.2024.105198

Dang, H., Mecke, L., Lehmann, F., Goller, S. and Buschek, D. (2022) "How to prompt? Opportunities and challenges of zero- and few-shot learning for human-AI interaction in creative applications of generative models", arXiv preprint, https://arxiv.org/pdf/2209.01390

Eager, B. and Brunton, R. (2023) "Prompting higher education towards AI-augmented teaching and learning practice", Journal of University Teaching and Learning Practice, Vol 20, No. 5. https://doi.org/10.53761/1.20.5.02

Gabadinho, A., Ritschard, G., Müller, N. S. and Studer, M. (2011) "Analyzing and visualizing state sequences in R with TraMineR", Journal of Statistical Software, Vol 40, No. 4, pp 1–37. https://doi.org/10.18637/jss.v040.i04

Heppner, P. and Petersen, C. H. (1982) "The development and implications of a personal problem-solving inventory", Journal of Counseling Psychology, Vol 29, No. 1, pp 66–75. https://doi.org/10.1037/0022-0167.29.1.66

Joksimovic, S., Ifenthaler, D., Marrone, R., De Laat, M. and Siemens, G. (2023) "Opportunities of artificial intelligence for supporting complex problem-solving: Findings from a scoping review", Computers and Education: Artificial Intelligence, Vol 4, 100138. https://doi.org/10.1016/j.caeai.2023.100138

Kong, S. C., Cheung, M. Y. W. and Tsang, O. (2024) "Developing an artificial intelligence literacy framework: Evaluation of a literacy course for senior secondary students using a project-based learning approach", Computers and Education: Artificial Intelligence, Vol 6, 100214. https://doi.org/10.1016/j.caeai.2024.100214

König, J., Krepf, M., Bremerich-Vos, A. and Buchholtz, C. (2021) "Meeting cognitive demands of lesson planning: Introducing the CODE-PLAN Model to describe and analyze teachers' planning competence", The Teacher Educator, Vol 56, No. 4, pp 466–487. https://doi.org/10.1080/08878730.2021.1938324

Lee, H. Y., Chen, P. H., Wang, W. S., Huang, Y. M. and Wu, T. T. (2024) "Empowering ChatGPT with guidance mechanism in blended learning: Effect of self-regulated learning, higher-order thinking skills, and knowledge construction", International Journal of Educational Technology in Higher Education, Vol 21, No. 1, 16. https://doi.org/10.1186/s41239-024-00447-4

Liu, M., Zhang, L. J. and Biebricher, C. (2024) "Investigating students' cognitive processes in generative AI-assisted digital multimodal composing and traditional writing", Computers & Education, Vol 211, 104977. https://doi.org/10.1016/j.compedu.2023.104977

Lo, C. K., Hew, K. F. and Jong, M. S. Y. (2024) "The influence of ChatGPT on student engagement: A systematic review and future research agenda", Computers & Education, 105100. https://doi.org/10.1016/j.compedu.2024.105100

Long, D. and Magerko, B. (2020) "What is AI literacy? Competencies and design considerations", in R. Bernhaupt, F. Mueller, D. Verweij, J. Andres, J. McGrenere, A. Cockburn, et al. (Eds.), CHI 2020 Proceedings, pp 1–16. ACM. https://doi.org/10.1145/3313831.3376727

Lu, K., Zhu, J., Pang, F. and Shadiev, R. (2024) "Understanding the relationship between college students' artificial intelligence literacy and higher order thinking skills using the 3P model: The mediating roles of behavioral engagement and peer interaction", Educational Technology Research and Development, pp 1–24. https://doi.org/10.1007/s11423-024-10434-1

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. and Hornik, K. (2015) "Cluster: Cluster analysis basics and extensions", R Package, Version 2.0.3.

Markauskaite, L., Marrone, R., Poquet, O., Knight, S., Martinez-Maldonado, R., Howard, S. and Siemens, G. (2022) "Rethinking the entwinement between artificial intelligence and human learning: What capabilities do learners need for a world with AI?", Computers and Education: Artificial Intelligence, Vol 3, 100056. https://doi.org/10.1016/j.caeai.2022.100056

Meyer, J., Jansen, T., Schiller, R., Liebenow, L. W., Steinbach, M., Horbach, A. and Fleckenstein, J. (2024) "Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions", Computers and Education: Artificial Intelligence, Vol 6, 100199. https://doi.org/10.1016/j.caeai.2023.100199

Nguyen, A., Hong, Y., Dang, B. and Huang, X. (2024a) "Human-AI collaboration patterns in AI-assisted academic writing", Studies in Higher Education, pp 1–18. https://doi.org/10.1080/03075079.2024.2323593

Nguyen, A., Ilesanmi, F., Dang, B., Vuorenmaa, E. and Järvelä, S. (2024b) "Hybrid Intelligence in Academic Writing: Examining Self-Regulated Learning Patterns in an AI-Assisted Writing Task", in HHAI 2024: Hybrid Human AI Systems for the Social Good, pp 241–254. IOS Press. https://doi.org/10.3233/FAIA240198

Oppenlaender, J., Linder, R. and Silvennoinen, J. (2024) "Prompting AI art: An investigation into the creative skill of prompt engineering", International Journal of Human–Computer Interaction, pp 1–23. https://doi.org/10.1080/10447318.2024.2431761

Pinski, M., Adam, M. and Benlian, A. (2023, April) "AI knowledge: Improving AI delegation through human enablement", in Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, pp 1–17. https://doi.org/10.1145/3544548.3580794

Pinski, M. and Benlian, A. (2023) "AI literacy – towards measuring human competency in artificial intelligence", Hawaii International Conference on System Sciences 2023 (HICSS-56). https://aisel.aisnet.org/hicss-56/cl/ai_and_future_work/3

Pintrich, P. R., Smith, D. A. F., Garcia, T. and McKeachie, W. J. (1991) "A manual for the use of the motivated strategies for learning questionnaire (MSLQ)", Ann Arbor, MI: The University of Michigan.

Raković, M., Iqbal, S., Li, T., Fan, Y., Singh, S., Surendrannair, S. and Gašević, D. (2023) "Harnessing the potential of trace data and linguistic analysis to predict learner performance in a multi-text writing task", Journal of Computer Assisted Learning, Vol 39, No. 3, pp 703–718. https://doi.org/10.1111/jcal.12769

Santu, S. K. K. and Feng, D. (2023) "Teler: A general taxonomy of LLM prompts for benchmarking complex tasks", arXiv preprint, arXiv:2305.11430. https://doi.org/10.48550/arXiv.2305.11430

Urban, M., Děchtěrenko, F., Lukavský, J., Hrabalová, V., Svacha, F., Brom, C. and Urban, K. (2024) "ChatGPT improves creative problem-solving performance in university students: An experimental study", Computers & Education, Vol 215, 105031. https://doi.org/10.1016/j.compedu.2024.105031

Yilmaz, R. and Yilmaz, F. G. K. (2023) "The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation", Computers and Education: Artificial Intelligence, Vol 4, 100147. https://doi.org/10.1016/j.caeai.2023.100147