The discerning principal: What prompting–uptake behaviors characterize high performers in human–AI collaboration?

Yaxuan Xu [a], Beichen Hu [b], Bohan Dong [a], Hui Liu [a, *]

[a] College of Education, Zhejiang University, Hangzhou, China
[b] College of Liberal Arts and Sciences, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Abstract: As generative artificial intelligence (GenAI) reshapes learning, examining students' interaction behaviors is crucial for enhancing performance in human–AI collaboration. However, research has paid limited attention to students' fine-grained prompting and uptake behaviors, their sequential patterns, and the characteristics distinguishing high and low performers, particularly in developing a holistic portrayal of high performers. This study involved 59 undergraduates and conducted descriptive and random forest analyses on interaction logs from 24 high and low performers, comparing behavioral frequency, proportional distribution, and sequences. Results indicated that high performers acted as discerning principals, exhibiting five characteristics: (1) starting with a plan: they entered collaboration with a strong grasp of the task-related knowledge and concepts, integrating their own thinking into contextualized task decomposition; (2) judging with an expert eye: they internalized evaluation standards and requirements as criteria for uptake decisions and critically filtered GenAI outputs; (3) harnessing the power of "no": they viewed rejection as a quality-control strategy that drove more refined prompting; (4) reflecting in action: they perceived GenAI responses as opportunities to reconsider previous task processing and modify their task pathway when necessary; (5) orchestrating the interaction system: they developed a dual-cycle regulatory system in which multiple behavioral sequences were interconnected. These findings metaphorized high performers as discerning principals who understand what underlies the task, hold GenAI outputs to high standards, and deepen their understanding during collaboration. This study offers implications for incorporating pre-planning activities, reshaping the culture of uptake decision-making, and guiding students to focus on both task progress and the development of their understanding.

Keywords: Generative Artificial Intelligence, Prompting Behavior, Uptake Behavior, High Performer, Metaphor, Random Forest Analysis

1. Introduction

Generative artificial intelligence (GenAI) is reshaping learning paradigms and increasingly serving as an important tool for supporting students' learning activities and task completion (O'Dea, 2024; Serra & Oliveira, 2025). Although previous research has shown that GenAI can improve learning outcomes and task performance by offering interactive learning experiences and immediate, personalized feedback (Darvishi et al., 2024; ElSayary, 2024; Yang et al., 2025; Ma & Zhong, 2025), its use does not automatically guarantee high-quality learning. Instead, it may lead to excessive reliance, superficial engagement, and potentially inhibit deep learning and critical thinking (Xia et al., 2026; Kasneci et al., 2023; Lan & Chen, 2024). This suggests that students' interactions with GenAI, particularly how they formulate prompts and evaluate, filter, and take up AI-generated content, play a crucial role in shaping learning performance.

Accordingly, unpacking the "black box" of student–GenAI interaction and identifying which interaction behaviors lead to differential learning performance has become a pressing concern in educational research (Liu et al., 2024; Kim et al., 2025). Previous studies have paid particular attention to students' prompting behaviors, namely how learners formulate questions or instructions for GenAI (Liu et al., 2024; Kim et al., 2025; Güner & Er, 2025), and have revealed that high and low performers differ in their prompting strategies and interaction patterns (Liu et al., 2024; Urban et al., 2025). Some researchers have also begun to incorporate uptake behaviors into analytic frameworks of student–GenAI interaction (Shen et al., 2025). However, empirical evidence on performance-based differences in uptake behaviors remains limited, especially regarding how prompting and uptake behaviors combine and unfold sequentially.

Furthermore, just listing the frequencies or sequences of high performers' interaction behaviors with GenAI is insufficient for developing a systematic understanding of who they are as collaborators. In other words, it is not enough to know what they do; we also need to understand what kind of collaborators they are. As a cognitive tool, metaphor connects abstract concepts to familiar experiences and helps people make sense of new phenomena (Lakoff & Johnson, 1980). In the context of emerging technologies, metaphors also frame our understanding and interaction with technology (Weller, 2023), providing an integrative interpretive lens for systematically understanding the characteristics of high performers in human–AI collaboration.

Against this backdrop, the present study combines descriptive analysis and random forest analysis to explore the interaction process through which undergraduates collaborated with GenAI to complete an innovative instructional design task. Specifically, it compares high and low performers in terms of the frequency, proportional distribution, and key behavioral sequences of prompting and uptake behaviors. To provide a more systematic characterization of high performers, this study further introduces metaphor as an interpretive lens and conceptualizes high performers as a specific collaborative role. This study aims to improve student–GenAI collaboration and boost collaborative success in future educational settings.

2. Literature review

2.1. Student–GenAI interaction behaviors: Prompting and uptake

Interaction between students and GenAI can be conceptualized as a collaborative partnership rather than a one-way process of tool use (Choi et al., 2026). This interaction involves two main types of behavior: prompting, where students guide AI-generated outputs by giving various instructions and serve as the main way human–AI interaction occurs (White et al., 2023), and uptake behaviors, which are the ways students assess, evaluate, and decide whether and how to incorporate AI-generated content into their work.

The quality and relevance of AI-generated outputs depend on the structure and clarity of the user's prompts (White et al., 2023; Knoth et al., 2024). Therefore, previous studies have primarily focused on prompting behaviors and developing coding schemes that range from ask-specific categories, such as instructional design and creative problem-solving (Sun & Huang, 2025; Liu et al., 2024; Urban et al., 2025), to domain-agnostic frameworks (López-Pernas et al., 2025). Some studies have further moved beyond formal classification to focus on the content types of prompts and the cognitive features underlying them. For example, Feng (2025) analyzed interactions across four categories: socio-emotional, cognitive, metacognitive, and coordinative. Ren et al. (2026) developed a coding scheme for the cognitive network comprising cognitive construction, social interaction, metacognition, and off-task, each corresponding to specific behaviors such as idea generation and agreement. In contrast, research incorporating uptake behaviors into analytical frameworks remains relatively limited. Shen et al. (2025) included students' adoption of AI-generated ideas as a code in interaction, while Sun and Huang (2025) differentiated between direct and indirect uses of GenAI outputs in lesson planning.

Previous studies have offered insights into how students formulate prompts and their uptake patterns, but remain coarse-grained. Prompting behaviors are rarely differentiated by functions, such as task structuring or evaluation standards-setting, and uptake behaviors are often simplified to "use" or "non-use" with little focus on complex decision-making. This restricts a deeper understanding of how students strategically regulate learning through specific interaction behaviors and performance differences. Furthermore, many existing coding frameworks are task-dependent and lack generalizability across different contexts. Therefore, there is a need to develop a more fine-grained and transferable framework for student–GenAI interaction behaviors.

2.2. Student–GenAI interaction behaviors across performance levels

As understanding of student–GenAI interaction has evolved, attention has increasingly shifted to the relationship between interaction behaviors and learning performance. Evidence suggests that the quality of student–GenAI interaction is positively associated with students' final performance (Oliveira et al., 2025). Recently, researchers have compared high and low performers to identify interaction strategies indicative of better performance.

Regarding frequency, high performers interact more frequently with GenAI than low performers (Urban et al., 2025; Liu et al., 2024). In terms of specific behaviors, low performers show a stronger pattern of instructing an LLM and specifying output, primarily seeking general factual information, while high performers are more likely to provide contextual information, issue improvement commands, pose explanatory questions that include their own thoughts, and use polite language more frequently during interactions (Liu et al., 2024; Misiejuk et al., 2025).

Nevertheless, frequency alone does not fully capture strategic differences. As Celik et al. (2025) argued, interaction should be viewed as a deliberate, multi-turn process comprising iterative prompting and uptake behaviors. Researchers have begun investigating student–GenAI interaction from a behavioral-sequence perspective. Liu et al. (2024) identified that high performers exhibited a closed-loop of "generate → monitor → apply → evaluate" when accepting GenAI outputs. In programming contexts, low performers are more likely to copy AI-generated code directly, whereas high performers engage in critical application and refinement, demonstrating independent reasoning and proactive knowledge construction (Li et al., 2025). Similarly, in creative problem-solving contexts, high performers engage in mutual, iterative collaboration with GenAI as a thinking partner, whereas low performers use GenAI as a non-iterative information-gathering resource (Urban et al., 2025; Ren et al., 2026).

These findings suggest that high performers' interactions resemble a "targeted improvement partnership" characterized by more targeted AI refinements and adaptive coordination (Nguyen et al., 2024; Oliveira et al., 2025). In contrast, low performers often reflect "basic information retrieval" or "passive task delegation" (Oliveira et al., 2025). However, few studies have systematically conceptualized prompting and uptake behaviors as interconnected multi-round cognitive and decision-making processes, highlighting critical areas for further exploration.

2.3 Metaphor: A new lens for understanding human–AI collaboration

Metaphor is not merely a rhetorical device but a central feature of human thought, and its essence is to use a familiar thing or experience to understand another (Lakoff & Johnson, 1980). Metaphor functions as a cognitive framework through which people make sense of new concepts and exerts a profound influence on both cognition and practice (Lakoff & Johnson, 1980; Vallis et al., 2025).

With the rise of GenAI, researchers have increasingly used metaphors to interrogate how AI is imagined in society, aiming to help people make sense of its use in teaching and learning. Vallis et al. (2025) analyzed metaphors of functions, roles, qualities, and agency to describe GenAI, revealing a diverse spectrum of perspectives. Using metaphor analysis, Jin et al. (2025) grouped GenAI metaphors into four conceptual categories: technical support (representative metaphor: high-heeled shoes), text development (compass), transformative potential (spider-man), and threat (drug). Collectively, these inquiries reveal the diverse ways in which GenAI is understood and help us think and talk about what GenAI is and what it is not.

However, previous studies have primarily focused on people's one-way perceptions of technology, emphasizing what the technology is and how people feel about it. They have provided little systematic metaphorical description of users' behavioral characteristics during interaction with GenAI, especially the distinctive behaviors exhibited by high performers. In other words, understanding what kind of technology GenAI is is not enough; we also need to understand, based on empirical evidence, what kind of collaborators high performers are.

2.4. The Present Study

Drawing on Choi et al.'s (2026) conceptualization of teacher–AI interaction, this study defined student–GenAI interaction as the dynamic exchange between students and GenAI during task completion, encompassing prompt formulation, content refinement, and uptake decisions about whether and how to incorporate AI-generated outputs into instructional design.

To address current limitations in behavioral granularity and sequential analysis, this study systematically examined fine-grained prompting and uptake behaviors. Specifically, we analyzed the frequency, proportional distribution, and sequential patterns of prompting–uptake behaviors and compared interaction pathways between high and low performers to identify key behavioral sequences predictive of performance. By introducing a metaphorical lens, we aimed to synthesize the core characteristics of high performers as collaborative partners. Accordingly, the study addressed two research questions:

RQ1: How do high and low performers differ in their frequencies and proportional distribution of prompting and uptake behaviors when engaging in student–GenAI collaborative learning tasks?

RQ2: What sequential patterns characterize the prompting–uptake behaviors of high and low performers in student–GenAI collaborative learning tasks?

3. Methodology

3.1. Participants

In this study, 59 undergraduate students at a comprehensive university in China participated in the course *Teaching Theory and Design*. The participants included 16 male and 43 female students, aged 21–23. A total of 84.75% of the participants were education majors, with the remaining 15.25% drawn from other majors. The course focused on major research domains in teaching theories and innovative instructional design. As part of the requirements, students were asked to collaboratively create a big idea–based instructional design using a GenAI tool they were already familiar with, working outside of regular class hours.

3.2. Data sources

Upon course completion, students submitted both their instructional designs and the interaction logs documenting their collaboration with the GenAI. Due to the loss of a student's interaction logs with the GenAI, that student's data was excluded from the analysis. The interaction logs submitted by the remaining 58 students were initially reviewed and cleaned. After removing invalid interactions caused by technical issues such as network interruptions, a total of 501 valid human–AI dialogue instances were retained for the analysis of students' prompting behaviors during their interactions with the GenAI. Additionally, students' big idea–based instructional designs were cross-referenced with their GenAI interaction logs to examine students' uptake of AI-generated content, specifically focusing on how AI-generated ideas and suggestions were integrated into their instructional designs.

3.3. Data analysis

We examined the student–GenAI interaction process across four stages: (1) grouping participants based on performance, (2) development and application of the coding scheme, (3) descriptive analysis, and (4) sequence analysis utilizing machine learning methods.

3.3.1. Grouping participants based on performance

Two researchers with extensive expertise in big idea–based instruction discussed the criteria for high-quality instructional design and conducted pilot scoring of students' instructional designs. After reaching consensus, they independently evaluated students' instructional designs using a 100-point scale. Consistent with the criteria used in previous studies to group high and low performance groups (e.g., Hou, 2015; Su, Li, Hu, & Rose, 2018; Zhang et al., 2022), the students with the top 20% score were identified as high performers, and the students with the bottom 20% score were identified as low performers. Therefore, 12 students were classified as high performers, and 12 students as low performers. For analytical purposes, the 24 students were labeled S1-S24. Correspondingly, each student's GenAI interaction record was coded as S1_GAI through S24_GAI, and each student's instructional design artifact was coded as S1_DESIGN through S24_DESIGN.

3.3.2. Development and application of the coding scheme

To analyze the student–GenAI interaction process, a coding scheme was developed using both deductive and inductive methods. The initial coding framework was developed by drawing on previous studies (Liu et al., 2024; López-Pernas et al., 2025; Choi et al., 2026), as well as an overview of the final instructional designs and interaction logs from all students. First, 20% of the data were independently coded by two trained coders. The coders then discussed discrepancies and refined the initial coding scheme accordingly. Following this refinement, the finalized coding scheme was applied. Interrater reliability for the final coding reached 0.887, reflecting satisfactory agreement between the two coders. Any remaining disagreements were resolved through discussion until consensus was reached for each coding decision. Table 1 and Table 2 present the final coding scheme for the prompting and uptake behaviors we proposed.

Table 1

Coding scheme for the prompting behaviors.

Dimension	Code	Code behavior	Definition	Example
Task Processing	ITP	Integrated Task Processing	Requesting GenAI to address the entire task in a single response	"Please help me think through how to design big idea–based instruction."
	DTP	Decomposed Task Processing	Breaking down the entire task into subtasks and asking GenAI to address them step by step	"Alright, let's move on to designing the third lesson."
Information Seeking	SKI	Seeking Knowledge and Information	Requesting GenAI to search for, retrieve, or synthesize relevant knowledge or information	"Is Italy's luxury goods industry considered manufacturing or a service industry?"
Task Framing and Input Specification	PC	Providing Context	Providing contextual information to GenAI	"I plan to organize the report into these sections: project title, description of the measurement scenario, measurement tools and methods, mathematical model diagrams, data records, geometric reasoning processes, measurement results and error analysis, and group reflection and improvement suggestions."

	CCT	Clarifying Concepts and Theories	Ensuring that GenAI understands task-related concepts and theories	"First, I will provide you with a paper by Professor Liu on big ideas to ensure that you correctly understand the concept of big ideas."
	CSR	Clarifying Standards and Requirements	Explicitly stating evaluation standards and requirements to GenAI	"Please note that the task must be grounded in an authentic problem context, aligned with competency-based objectives, and demonstrate an understanding of the big idea."
	PET	Providing Examples or Templates	Providing examples or templates to GenAI	"Please design the format according to the example shown in the image."
Interactive Regulation and Refinement	RES	Requesting Evaluation and Suggestions	Asking GenAI to evaluate the current state of task completion and provide feedback or suggestions	"Are there any problems with my design?"
	RIR	Requesting Iteration and Refinement	Asking GenAI to iteratively revise or improve the content	"Please further enrich the big ideas."
	FCE	Follow-up Clarification and Elaboration	Asking follow-up questions based on GenAI's responses for clarification or elaboration	"What kind of thinking is reflected in the big idea you extracted?"
User Stance and Evaluation	USD	User Statement and Decision	Expressing personal views or decisions without issuing commands or questions to GenAI	"Okay, for now we'll go with these two interdisciplinary big ideas."
	EF	Evaluative Feedback	Providing evaluative feedback on GenAI's responses	"Your design simply presents a single question for students to answer individually, and it lacks openness."

Table 2
Coding scheme for the uptake behaviors.

Code	Code behavior	Definition	Example
FU	Full Uptake	Completely and directly accepting the content generated by GenAI	The student fully takes up the big idea generated by the GenAI.
PU	Partial Uptake	Selectively using only parts of the AI-generated content	The student refines and supplements the AI-generated big idea before taking it up.
EU	Exploratory Uptake	Using AI-generated content as inspiration or a starting point to independently create a new and	The theme of the big idea generated by the AI provided directional guidance, and the student independently formulates a new big

		complete idea.	idea based on this theme.
NU	Non-Uptake	Rejecting any content generated by GenAI	The student asks the GenAI to generate a big idea but doesn't use any of the AI-generated content.

3.3.3. Descriptive analysis

To examine behavioral differences between high and low performers during interaction with GenAI, we first counted dialogue turns and the frequencies of prompting and uptake behaviors. Subsequently, given that different types of prompting behaviors functionally reflect students' strategies in task construction, regulation, and reflection, we compared proportional distributions across five prompting dimensions and four uptake behaviors. Additionally, differences in the proportional distribution within behavioral types were compared between high and low performers to identify behaviors that showed greater divergence or similarity across groups.

3.3.4. Machine learning-based sequence analysis

To identify behavioral patterns differentiating high and low performers, we employed a Random Forest (RF) classification model (Breiman, 2001). RF effectively extracts discriminative patterns from complex educational data and frequently outperforms traditional classifiers (Xu & Yin, 2021). Given our limited sample size (N = 24) and high-dimensional sequence features, RF's ensemble structure provides robustness against overfitting.

We utilized RF for exploratory analysis rather than causal inference. Specifically, Mean Decrease Impurity (MDI) scores were used to identify discriminative sequential patterns. While permutation importance was also evaluated, highly correlated features in our small sample produced unstable estimates across folds (Strobl et al., 2008). Thus, MDI was adopted as the primary feature selection metric. Given the limited sample size, SHAP explanations were similarly unstable and excluded from inference.

Common interaction analysis methods, such as Lag Sequential Analysis (LSA) and Epistemic Network Analysis (ENA), offer valuable perspectives on interaction processes. LSA is particularly well-suited to examining adjacent behavioral transitions, while ENA extends the analysis of interaction from observable behaviors to the underlying epistemic processes (Chu et al., 2025). Because the present study focuses on multi-step dependencies and performance-predictive sequence patterns, we adopted an RF-based N-gram approach, which is better suited to providing a fine-grained diagnostic of cognitive regulation (see Fig. 1 for the data processing and modeling framework).

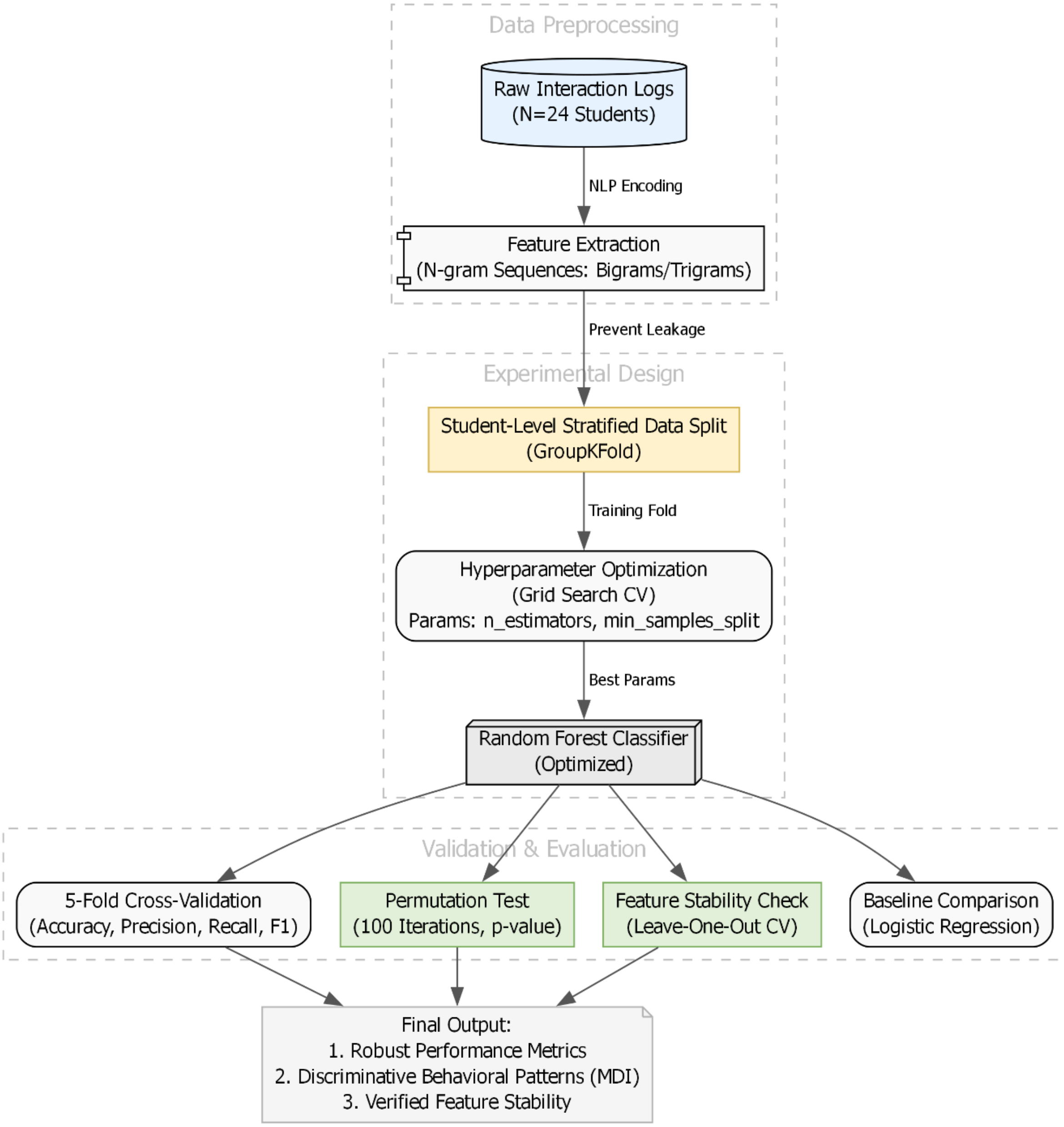

Fig. 1. Data processing and modeling framework. (*Note:* This flowchart illustrates the comprehensive analytical pipeline: transforming raw human–AI interaction logs into N-gram behavioral sequence features, utilizing a Random Forest model for feature selection, and rigorously validating the results through stratified cross-validation, permutation testing, and Leave-One-Out Cross-Validation (LOOCV) for feature stability.)

Interaction logs were converted into behavioral sequences. We extracted bigrams and trigrams using N-gram modeling to capture fundamental interaction units (e.g., "Action → Feedback → Adjustment"). Longer sequences (e.g., 4-grams) produced sparse features and reduced stability. Hyperparameters were optimized via grid search (optimal: n_estimators = 50, min_samples_split = 4, max_depth = None) using scikit-learn in Python (Pedregosa et al., 2011).

To prevent data leakage, we applied student-level stratified cross-validation (GroupKFold), ensuring that all sequences from a single student remained exclusively in either the training or the testing fold. Model performance was evaluated using accuracy, precision, recall, and F1-score. A permutation test (100 iterations) assessed whether performance significantly exceeded chance levels (Good, 2000). Finally, a Leave-One-Out Cross-Validation (LOOCV) check confirmed feature stability, with core sequences demonstrating 70% to 100% Top-10 inclusion rates.

4. Results

4.1 Differences in the frequency and proportional distribution of prompting and uptake behaviors between high and low performers

Frequency analyses were conducted on dialogue turns and different types of interaction behaviors. Regarding dialogue turns, high performers engaged in 8 to 143 turns, while low performers ranged from 1 to 28 turns. High performers performed 549 prompting behaviors (Mean = 45.75) and 391 uptake behaviors (Mean = 32.58). Low performers demonstrated 151 prompting behaviors (Mean = 12.58) and 110 uptake behaviors (Mean = 9.17), revealing substantial differences in interaction frequency.

A stacked bar chart was employed to illustrate the frequencies and proportional distributions of various interaction behavior types across groups (see Fig. 2), revealing clear differences in the composition of prompting and uptake behaviors between high and low performers. Regarding prompting behaviors, 32.06% of prompts generated by high performers were categorized as task framing and input specification, followed by interaction regulation and refinement (29.51%) and task processing (20.26%). In contrast, the most prevalent prompting behavior among low performers was task processing (36.42%), followed by task framing and input specification (27.78%) and interaction regulation and refinement (22.84%). Regarding uptake behaviors, NU was most frequently observed among high performers, whereas PU was more frequent among low performers. Exploratory EU was the least used uptake behavior in both groups.

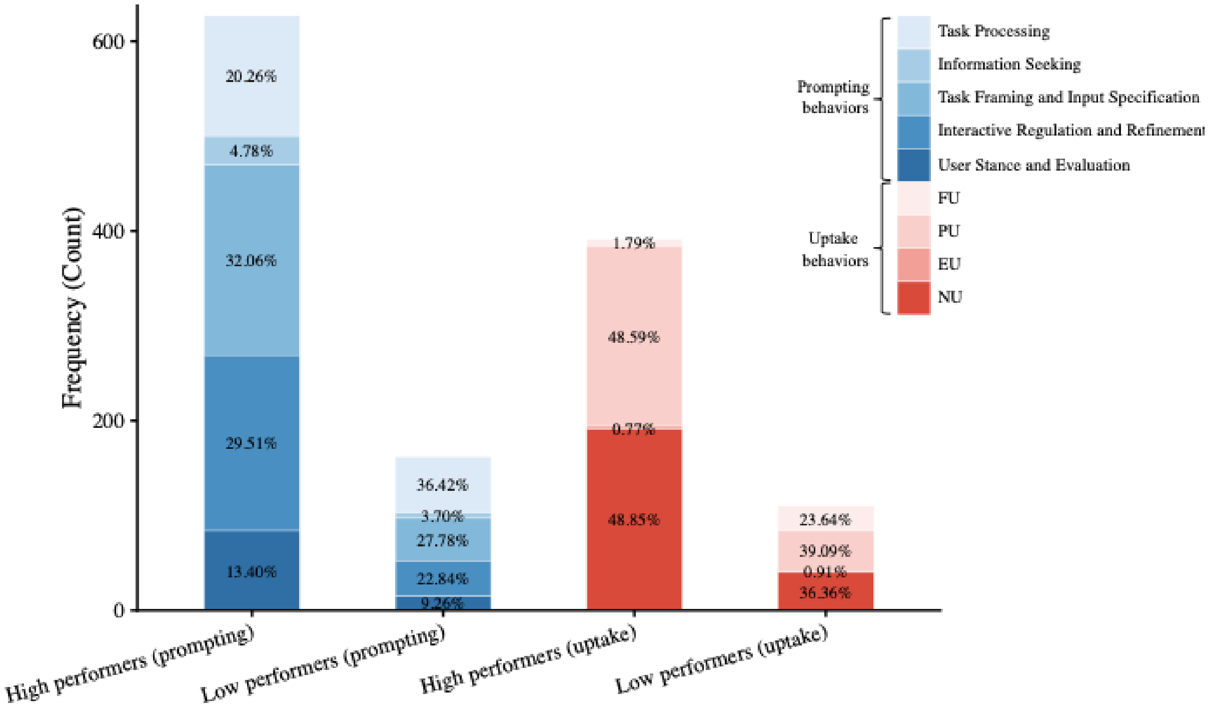

Fig. 2. Frequency and proportional distribution of prompting and uptake behaviors among high and low performers

To further examine proportional differences in the use of corresponding behaviors between high and low performers, a dumbbell plot was created (see Fig. 3). Statistical analyses showed both differences and similarities in prompting and uptake behavior types across performance groups. Regarding prompting behaviors, the largest proportional difference was observed in ITP, with low performers exhibiting a 10.89% higher proportion than high performers. The second-largest difference appeared in RES, where high performers had an 8.79% higher proportion than low performers. The third largest difference was observed in PC, with high performers exceeding low performers by 6.79%. Additionally, several prompting behaviors exhibited comparable proportions across the groups. The proportion of CCT was nearly identical (Δ% = −0.01%). FCE appeared at similar rates (Δ% = 0.21%), as did PET (Δ% = −1.03%). Regarding uptake behaviors, the most notable proportional differences were observed in FU and NU. Low performers exhibited a 21.85% higher proportion of FU compared to high performers, while their proportion of NU was 12.49% lower than that of high performers.

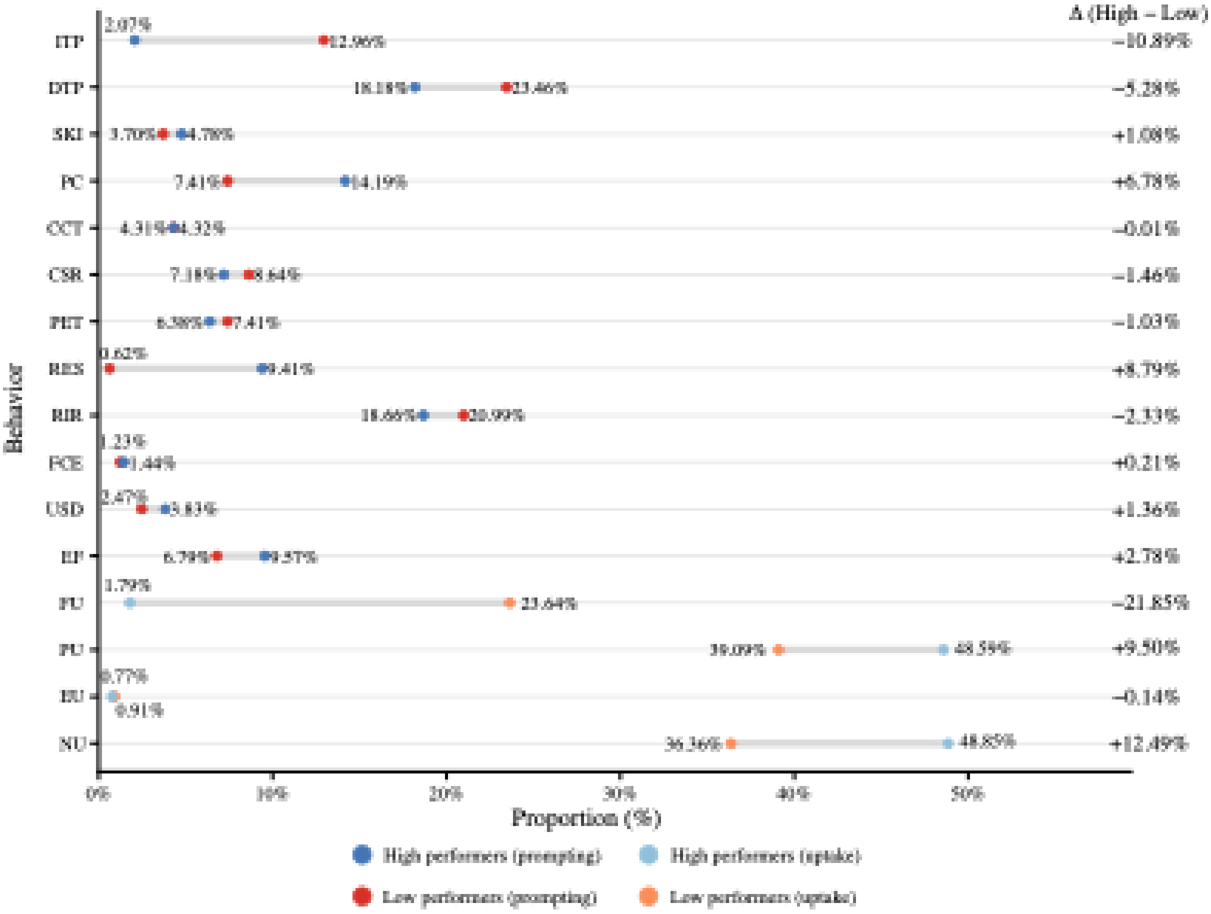

Fig. 3. Proportional differences in prompting and uptake behaviors among high and low performers

4.2. Behavioral sequences associated with high performance in human-AI collaborative learning

4.2.1 Classification Model Performance and Stability Analysis

To validate the effectiveness of behavioral sequences in differentiating learning performance, we evaluated the optimized Random Forest (RF) classification model against a Logistic Regression baseline.

As shown in Table 3, the optimized RF model consistently outperformed the baseline across all metrics. The RF model achieved an average accuracy of 0.8300, representing an 8.0 percentage-point improvement over the baseline (0.7500). Notably, the RF model yielded a high precision of 0.9500 (vs. 0.8667), indicating that the identified behavioral patterns may serve as reliable indicators of effective collaborative strategies. Furthermore, the higher F1-score (0.7981 vs. 0.7200) suggests the sequence modeling approach captures behavioral distinctions missed by the baseline. By incorporating sequential order, the model distinguishes interaction contexts, such as determining whether FU directly follows CSR (low performers) or whether PU is preceded by PC and DTP (high performers).

Table 3

Model performance comparison (5-fold cross-validation)

Model	Mean Accuracy	Standard Deviation (σ)	Precision	Recall	F1-score
Random Forest(Optimized)	0.8300	0.1744	0.9500	0.7333	0.7981
Logistic Regression(Baseline)	0.7500	0.1549	0.8667	0.6667	0.7200

Through hyperparameter optimization, performance variance was effectively controlled. The slightly higher standard deviation of the RF model (σ = 0.1744) compared to the baseline (σ= 0.1549) likely reflects the sensitivity of ensemble methods to training-set composition in small-sample cross-validation. A subsequent permutation test confirmed the model's performance significantly exceeded chance levels ($p < 0.01$).

4.2.2 Identification of key predictive behavioral sequences

To identify sequential patterns differentiating high and low performers, we analyzed the MDI feature importance scores produced by the RF model. Both groups frequently exhibited basic sequences such as RIR → PU and DTP → CSR, suggesting a baseline level of interaction competence with the AI. However, the feature importance of these sequences was notably low, falling outside the top 10 predictors. Relying solely on these common behaviors is therefore insufficient to distinguish performance levels.

Consequently, we identified discriminative behavioral sequences and grouped them into three categories: (a) task planning, (b) uptake decision, and (c) post-uptake regulation. Collectively, these dimensions form a behavioral regulation cycle in human-AI collaboration. Fig. 4 illustrates the frequencies and predictive importance of these key sequences across the two groups.

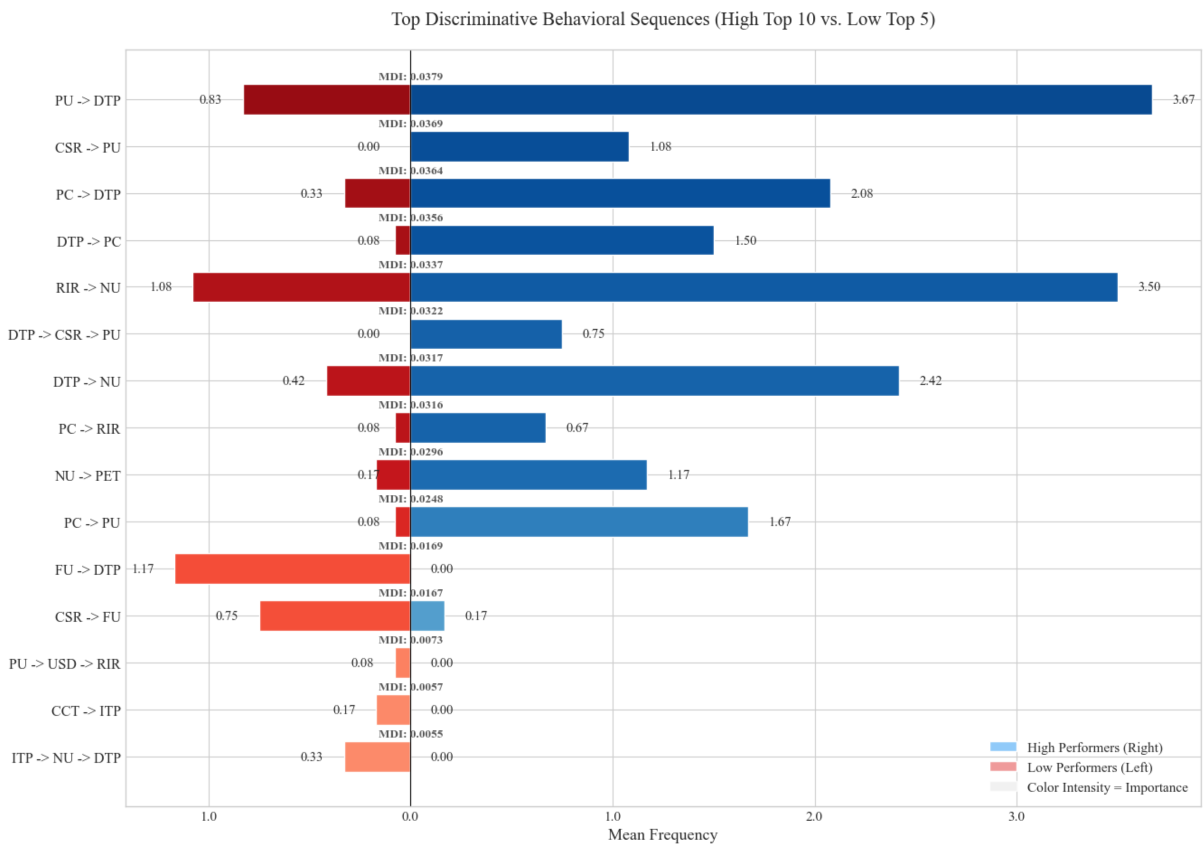

Fig. 4. Average frequencies and predictive importance of key behavioral sequences between high and low performers. (*Note*: Blue bars extending to the right denote the high-performing group, whereas red bars extending to the left denote the low-performing group. Color intensity reflects feature importance (MDI values) derived from the Random Forest model.)

First, task planning refers to how students structure tasks prior to uptake. High performers demonstrated a tendency toward proactive contextualized task decomposition. The reciprocal interaction between PC → DTP (Rank 3, MDI = 0.0364, Mean: High 2.08 vs. Low 0.33) and DTP → PC (Rank 4) represents a central pattern. Furthermore, the significance of PC → RIR (Rank 8) indicates that contextual prompts often trigger subsequent interaction cycles. This suggests that context in the high-performing group functions as a central organizing element rather than mere background information. Conversely, low performers more frequently exhibited the CCT → ITP sequence (MDI = 0.0057), indicating a tendency to issue broad, holistic task requests lacking fine-grained scaffolding.

Second, the uptake decision captures cognitive gating, the evaluative decision process through which students accept, partially take up, or reject AI-generated outputs. CSR → PU emerged as a key predictor of high performance (Rank 2, MDI = 0.0369, Mean: High 1.08 vs. Low 0.00). This indicates that high performers tend to partially take up GenAI outputs after screening them against explicit standards. This criteria-driven evaluation is also evident in DTP → CSR → PU (Rank 6). High performers' quality control equally extends to rejection behaviors; DTP → NU (Rank 7) reveals strict screening of sub-task outputs. In contrast, low performers were characterized by CSR → FU (MDI = 0.0167) and ITP → FU (MDI = 0.0050), indicating a tendency toward wholesale acceptance of AI outputs regardless of prior constraints.

Third, post-uptake regulation concerns how students advance subsequent tasks after completing an uptake decision. PU → DTP is the strongest predictor of high performance (Rank 1, MDI = 0.0379). High performers demonstrated workflow continuity by utilizing partial uptake to guide the next planning phase. Subsequently, RIR → NU (Rank 5) reveals that high performers decisively reject subpar AI generation during iterations. Finally, NU → PET (Rank 9) uncovers a scaffolded rejection strategy: upon rejecting AI content, high performers provide specific examples or templates for refinement. Conversely, low performers primarily exhibited FU → DTP (MDI = 0.0169). This chronological order reveals a disjointed workflow in which students initiate the next sub-task only after fully accepting the preceding AI output, rather than using intermediate evaluation or refinement steps.

4.2.3 Structural analysis of interaction patterns*: From linear execution to systemic orchestration*

While Section 4.2.2 identified specific predictive behavioral sequences, Fig. 5 illustrates how these sequences interconnect to form distinct interaction systems. This structural visualization indicates that high performance stems from the organized synergy of multiple behaviors rather than isolated actions.

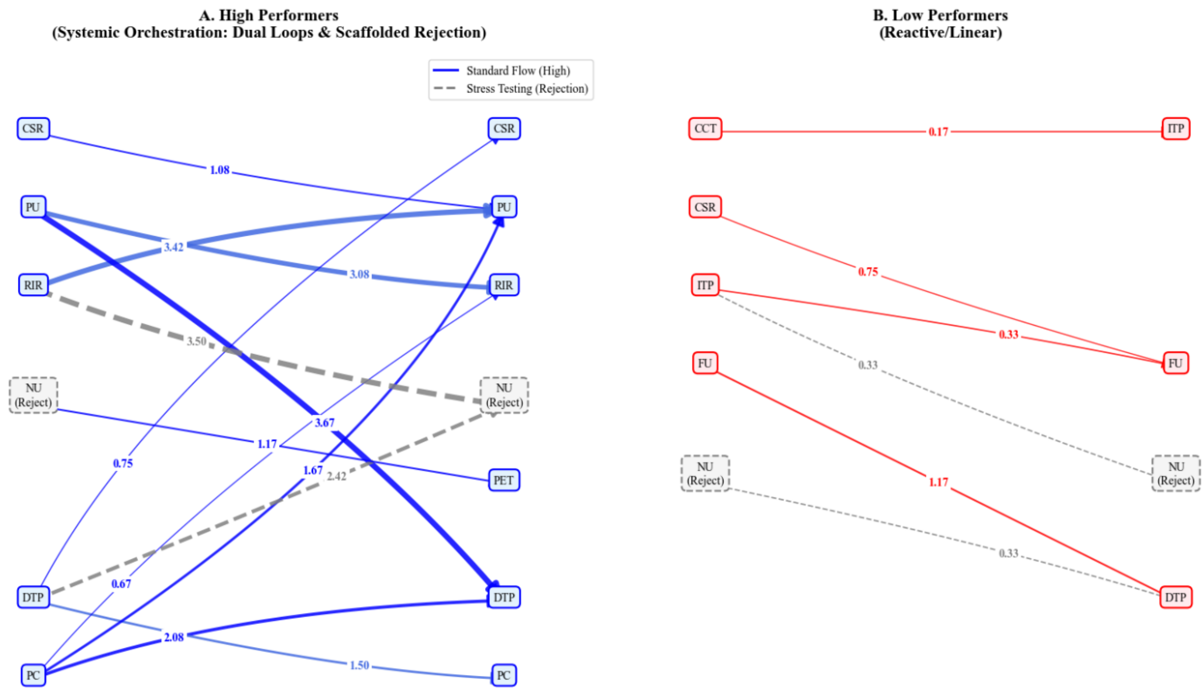

Fig. 5. Comparison of prompt-uptake behavioral transition patterns between high and low performers. (*Note*: Panel A illustrates the strategic, iterative loop structure of high performers, whereas Panel B presents the linear, terminal uptake path characteristic of low performers. Line thickness indicates the frequency of behavioral transitions.)

As shown in Fig. 5 (Panel A), high performers present a highly interconnected interaction system. In the lower half, the bidirectional flow (crossing lines) between PC and DTP reflects a dynamic reciprocal pattern, forming a "contextualized task decomposition loop." In the upper half, a closed circuit emerges between RIR and PU. Instead of following a linear path, high performers use iterative outputs as inputs for subsequent optimization guided by high standards. Concurrently, the prominent path to NU (grey dashed line) shows how they safeguard quality by rejecting substandard content.

Crucially, the PU → DTP transition (the thickest path) bridges these two loops, forming a unified system in which uptake decisions guide subsequent task planning. This indicates that uptake serves as a transition to the next planning phase rather than a terminal endpoint. Additionally, paths from the PC pointing upward to the RIR and PU show how contextual information permeates the interaction. Meanwhile, the extension from NU to PET illustrates how rejection is repurposed into proactive instructional guidance. Together, these configurations reflect a tendency toward iterative refinement and self-correction.

In contrast, the interaction patterns of low performers (Fig. 5, Panel B) exhibit linear and terminal characteristics. Their dominant paths (e.g., CSR → FU, ITP → FU) flow directly to FU without feedback loops. Notably, the FU → DTP transition appears as a unidirectional extension rather than an internal loop. This points to a chronological misalignment, where students initiate subsequent tasks only after fully accepting previous outputs. Such "checklist-style execution" suggests that low performers treat the AI primarily as a unidirectional information provider rather than a collaborative partner.

5. Discussion and implications

5.1. Discussion

This study reveals that disparities in student–GenAI collaboration extended beyond the frequency and proportional distribution of behaviors to how these behaviors were connected and coordinated. Compared to low performers, high performers acted as discerning principals: they possessed a strong grasp of task-related knowledge and concepts, held GenAI outputs to high standards, and continuously refined task pathways while deepening their own understanding. These characteristics are demonstrated in the following five key behavioral characteristics.

First, high performers typically had a solid grasp of task-relevant knowledge and concepts before collaborating with GenAI, enabling them to create clear pre-plans. This positioned them as principals who "start with a plan" by offering well-organized, clearly defined contexts. Descriptive analysis showed that high performers' prompting behaviors focused more on task framing and input specification (32.06%), especially providing context more frequently ($\Delta\%$ = 6.79%), consistent with the findings of Misiejuk et al. (2025). Random forest analysis identified PC → DTP and DTP → PC as key sequences for high performance, indicating that high performers embedded personal thinking within contexts when decomposing tasks. For example, S10 provided GenAI with a predetermined assessment task theme, "campus news reporting practice", and asked for help in identifying assessment dimensions. This process helps narrow the generative space, enhancing the accuracy and quality of GenAI outputs (Nam et al., 2024; Abdelghani et al., 2023). In contrast, although low performers exhibited a higher proportion of DTP ($\Delta\%$ = −5.28%), their task decomposition lacked stable coupling with PC. This suggests that high performance is defined not just by decomposition but by whether it is grounded in personal and anticipatory thinking. The PC → RIR sequence among high performers further supports this interpretation. Additionally, contextualized prompting frequently led high performers to partially use GenAI responses (PC → PU). The high predictive value and frequent occurrence of PC in key behavioral sequences among high performers indicate that GenAI's effectiveness as a complementary cognitive artifact depends on active human engagement (Krakowski, 2025). Conversely, the CCT → ITP sequence identified in low performers suggests that while they demonstrate some awareness of verifying shared understanding between AI and humans to facilitate effective communication (Geroimenko, 2025), they failed to integrate sub-tasks with anticipatory thinking. These findings align with distributed cognition theory (Hutchins, 1995), which posits that cognition is distributed across dynamic interactions among individuals, tools, and the surrounding environment (Salomon, 1997; Hollan et al., 2000), and conceptualizes AI as an extension of human cognition (Guo et al., 2025). In this study, low performers tended to engage in limited personal deliberation and transfer substantial cognitive load to GenAI, thereby forming an AI-dominant distributed cognition pattern. Conversely, high performers demonstrated more optimized cognitive load allocation, whose anticipatory thinking enabled reciprocal development between personal thinking and GenAI support. Such human-originated cognitive input is essential for the effective functioning of distributed cognition systems (Salomon, 1997).

Second, high performers were not merely principals who articulated requirements; they evaluated GenAI outputs with an expert eye by internalizing external standards and requirements as criteria for uptake decisions. Random forest analysis revealed that DTP → CSR → PU and CSR → PU were key predictive sequences of high performance, while CSR → FU characterized low performers. These findings indicate that CSR does not automatically guarantee high performance; instead, the key lies in whether students internalize these standards and requirements as personal evaluation criteria. Specifically, after clarifying standards for GenAI, high performers screened and re-evaluated generated content against internalized criteria, demonstrating a critical, selective uptake strategy. This finding aligns with that of Li et al. (2025). In contrast, low performers exhibited a more linear uptake pathway, tending toward FU even after specifying standards and requirements. Another key sequence identified among low performers, ITP → FU, confirms their heavy reliance on GenAI and weaker tendency for independent filtering. These behavioral patterns indicate a pragmatic orientation toward maximizing efficiency or outcome, which may also be characterized as a form of "smart loafing" (Amoozadeh et al., 2024; Stieglitz et al., 2022). Thus, differences in uptake following CSR highlight whether standards and requirements are activated as regulatory functions in decision-making, and can be interpreted by metacognitive monitoring (Urban et al., 2025). Learners with stronger metacognitive skills are better equipped to evaluate outputs against predefined standards, which enables them to suppress automatic uptake tendencies induced by the processing fluency and trigger more deliberate, analytical processing and integration (Tankelevitch et al., 2024).

Third, high performers viewed NU not as a sign of collaborative failure, but as an essential act of judgment and quality control grounded in high standards. They harnessed the power of "no" and followed that rejection with more refined prompting. Descriptive analysis revealed that the gap in NU between high and low performers was the second largest ($\Delta\% = 12.49\%$). Random forest analysis identified RIR → NU and DTP → NU as key predictive sequences for high performance. After assessing the quality and structure of AI-generated content, learners might actively end the current collaborative process and redesign their collaborative pathways. This pattern could be attributed to higher levels of critical thinking among high performers, who are better at analyzing, evaluating, and synthesizing information to make reasoned decisions (Gerlich, 2025). Moreover, it highlights a clear human judgment decision in human–AI collaboration, underscoring the importance of maintaining a dynamic balance between human judgment and GenAI capabilities (Zhang et al., 2025). Additionally, how learners continue the collaboration after NU warrants closer examination. The sequence NU → PET emerged as a key predictor of high performance, suggesting that high performers tended to use examples or templates to provide clearer guidance for subsequent turns. In contrast, the sequence ITP → NU → DTP emerged as a key predictor of low performance, suggesting that low performers often rejected GenAI outputs after integrated task processing but failed to provide concrete constraints or reference points during subsequent task decomposition. From the perspective of self-regulated learning theory (Zimmerman, 2000), learners in this study engaged in causal attribution after rejecting AI-generated content and subsequently adjusted their strategies (Panadero & Alonso-Tapia, 2014). High performers often attributed unsatisfactory outputs to a lack of clarity and specificity in their prompts, so they provided templates or examples to eliminate ambiguity around the request (Kulkarni & Tupsakhare, 2024). In contrast, low performers were more likely to attribute issues to insufficient task decomposition but failed to incorporate context-specific personal thinking to improve outputs.

Fourth, high performers viewed uptake not as the end of a task-processing cycle, but as an opportunity to re-examine the task and deepen their understanding, exhibiting a process of “reflecting in action”. Random forest analysis showed that PU→DTP predicted high performance, while FU→DTP predicted low performance. A closer look at the interaction logs revealed that while both groups moved on to the next sub-task after an uptake decision, some high performers revisited previously completed parts following partial uptake in response to GenAI’s output. For example, after extracting the big ideas, S2 asked GenAI to design the instructional sequence based on them. Following partial uptake of the proposed sequence, the student reflected on the previously extracted big ideas and asked the GenAI to refine them from both disciplinary and interdisciplinary perspectives. This behavior isn’t just about simple revision; it represents a recalibration of direction based on the learner’s current understanding of the task, making human–AI collaboration a dynamic, iterative, and revisable process. Additionally, task decomposition following partial uptake can help students better identify the limits of their current understanding and deepen their understanding of the task through interaction. This also supports the idea that high performers are more likely to regard GenAI as a cognitive partner or a tutor, which helps promote reflection, revision, and improvement of their ideas (Hwang et al., 2020).

Finally, high performers’ success stemmed from a dual-cycle regulatory system in which multiple behavioral sequences were highly connected and hierarchically coupled. Like discerning principals, they orchestrated the system through a complex, interconnected, and dynamically synchronized way of thinking. This system comprises two interconnected cycles: a contextualized decomposition cycle, where the bidirectional flow between PC and DTP combines learners’ personal anticipatory thinking with task decomposition to ensure both accuracy and structural coherence in task processing; and an iterative refinement cycle, centered on RIR and PU behaviors to improve quality through standards-based review. Importantly, the PU → DTP transition connects these two cycles, turning uptake decisions from endpoint actions into catalysts for a new round of task planning. This finding explains why certain common behavioral sequences occur in both high and low performers’ interactions, yet do not lead to high performance among low performers. Specifically, high performers can integrate these effective behavioral sequences into the dual-cycle system, activating their regulatory potential. In contrast, the same sequences among low performers tend to exist in isolation, exhibiting linear and terminal characteristics. Consequently, even when high-quality interaction behaviors occur, their contribution to performance remains limited. Additionally, the behavioral transition patterns of high performers demonstrate a more advanced scaffolding mechanism. This mechanism reflects the widespread integration of learners’ own thinking into contexts, as well as a shift from non-uptake to more accurate prompt refinement, turning potential “collaborative failure” into opportunities for strategic growth.

5.2. Implications

The present study suggests that high-performing human–AI collaboration depends less on the precision of single prompts than on whether learners enter collaboration with a strong grasp of task-related knowledge and concepts. High performers utilized their understanding to design, evaluate, and adjust their task-processing pathways. Accordingly, educational support should move beyond a narrow focus on “how to ask better questions”. Instead, it should prioritize cultivating learners’ capacities as discerning principals in collaboration, enabling them to set direction, evaluate outputs, dynamically adjust collaborative processes, and deepen their understanding of the task throughout the interaction.

First, the findings demonstrate that high-quality human–AI collaboration relies more on learners’ grasp of task-related knowledge and concepts, and on their prior thinking beyond prompting techniques. Compared to high performers who often integrated their contextual insights, low performers often turned to GenAI without personal thinking and pre-plans, shifting higher-order cognitive work to the GenAI. This disparity reflects how learners assign cognitive responsibility in human–AI collaboration. Therefore, support should extend beyond training students in prompting techniques to help them clarify their cognitive roles. Students should take responsibility for understanding the task and setting its direction before interaction begins. Educators can incorporate guiding questions such as “What concepts or theories are relevant to it?” and “What initial ideas have I already formed?” to help students assess their readiness before engaging with GenAI. They can then use that understanding to define the boundaries and structure of the interaction and craft their prompts accordingly. Such support helps students planning human–GenAI interaction from a starting point of prior human thinking, which may encourage complementary effects and help prevent GenAI from becoming a competing substitute that fosters overreliance and skill degradation (Barrett et al., 2012; Chen & Chan, 2024).

Second, high performers exhibited a clear understanding-driven uptake pattern, frequently engaging in revisiting, rejecting, and replanning activities. This underscores the importance of a strong grasp of task-related knowledge and concepts, enabling students to make rational uptake decisions based on internalized criteria. Such understanding also allows them to reject GenAI outputs more decisively and exercise greater control over both completed work and subsequent task processing. Therefore, human–GenAI collaborative learning requires more than developing their own understanding of the task; it necessitates reshaping the “culture of uptake decision-making” by guiding students to shift their idea of success from a result-oriented focus on “getting answers quickly” to a process-oriented calibration of both task handling and understanding. Specifically, support can be provided in three ways. First, educators should emphasize criteria-based uptake by incorporating collaborative processes into assessments, using process documentation or decision logs to encourage reflection on the basis for uptake (Wang et al., 2026), embedding adaptive metacognitive scaffoldings into interactions (Liu et al., 2026), and fostering students’ ability to critically evaluate AI-generated content (Xu et al., 2025). Second, the meaning of rejection should be redefined; rejection is no longer a failure but a strategic opportunity for quality improvement. It transforms GenAI’s “imperfect responses” into chances for strategic optimization. Educators can support this shift by presenting examples of high-quality rejection and helping students reconceptualize rejection. Third, uptake should be viewed not as the endpoint of collaboration but as a trigger for ongoing cognitive regulation. Teachers can motivate learners to regularly reflect on uptake decisions and plan subsequent steps accordingly, ensuring that collaborative pathways remain revisable and guided by human judgment.

Finally, the study shows that high-performing human–AI collaboration relies neither on a single high-quality prompt nor a linear "prompt–generate–complete" path. Instead, it hinges on a system where multiple interactional sequences interweave and cycle. This behavioral pattern reflects high performers' complex and interconnected thinking and supports the simultaneous deepening of task processing and understanding. This does not diminish the value of prompting frameworks and prompt engineering for unlocking GenAI's potential, but emphasizes that learners need to act as discerning principals. They should not only articulate task demands and make supervisory decisions based on their understanding, but also dynamically adjust both the collaborative process and the learning pace. Therefore, educators should guide students to monitor both task progress and the development of their understanding. Such metacognitive monitoring ensures that learners maintain ownership of their understanding throughout collaboration and that each interaction fosters its development. Specifically, metacognitively oriented plugins or tools should be developed to facilitate students' self-regulated learning across these dimensions. Simultaneously, educators should provide continuous support, assessing students' performance and needs to offer necessary assistance (Xu et al., 2025).

6. Conclusion, limitation, and future directions

As GenAI rapidly transforms students' learning patterns, examining human–AI interaction in detail becomes increasingly important for improving and applying GenAI -supported learning modes (Tu & Hwang, 2024). However, fine-grained analyses of students' prompting–uptake behaviors remain limited. To address this gap, we developed a coding scheme comprising 12 prompting behaviors and 4 uptake behaviors to analyze interaction logs. Using descriptive and random forest analyses, the study identified behavioral differences between high and low performers in their interactions with GenAI.

Drawing on a metaphorical lens, this study conceptualizes high performers as discerning principals. Grounded in a strong grasp of task-related knowledge and concepts, they evaluated GenAI outputs against high standards, regulated the collaborative process, and deepened their understanding of the task during human–AI interaction. The main findings are as follows: (1) prior to collaboration, high performers possessed a solid understanding of the task-related knowledge and concepts and established pre-plans, embedding personal insights into task decomposition, which differentiates their distributed cognition patterns from those of low performers; (2) high performers internalized evaluation standards and requirements to guide critical uptake, acting as principals with high standards; by contrast, low performers tended toward linear and full uptake, a disparity potentially stemming from differences in metacognitive monitoring ability; (3) high performers utilized non-uptake for quality control and refined prompts accordingly, underscoring the importance of critical thinking and adaptive attribution; (4) following uptake, some high performers leveraged GenAI responses as opportunities to rethink prior task handling, revisit and adjust their collaborative pathways, embodying a "reflecting in action" approach that highlights uptake's transformative role in human–AI collaboration; (5) high performers' success stemmed from a dual-cycle regulatory system in which multiple behavioral sequences were highly interconnected and hierarchically coupled, reflecting the complex and interwoven nature of their thinking. Taken together, these findings offer new perspectives for optimizing students' performance in collaborating with

GenAI. Future instructional guidance should prioritize the "think first, then work" principle, incorporating structured pre-planning to help students clarify their understanding and the boundaries of cognitive responsibility between themselves and GenAI, reshape the culture of uptake decision-making while providing multiple forms of practical support, and help students maintain simultaneous attention to task progress and the development of their own understanding. Ultimately, this study not only extends the empirical boundaries of student–GenAI interaction research but also provides implications for improving student–GenAI interaction and informing the future development of GenAI tools.

Nevertheless, this study has several limitations. First, the small, unbalanced sample from an education course at a comprehensive university in China may limit the generalizability of the findings. Future studies should adopt diverse sampling to improve external validity. Second, the study focused primarily on the task context of instructional design. Future research could therefore examine student–GenAI interaction patterns and strategies across a wider range of learning tasks and problem-solving contexts. Third, the analysis of prompting and uptake behaviors relied mainly on students' interaction records with GenAI and their instructional design products. Future studies could incorporate more multimodal data and adopt a wider range of data collection methods, such as think-aloud protocols. In addition, it should be acknowledged that even when learners engage in similar interaction behaviors, the quality of those behaviors may still differ. A more in-depth analysis of behavioral quality, therefore, warrants further investigation. Improvements in these areas would enable a deeper understanding of student–GenAI interaction patterns and provide educators with richer insights into how to guide human–AI collaborative learning and enhance its effectiveness.

CRediT authorship contribution statement

Yaxuan Xu: Writing – review & editing, Writing – original draft, Methodology, Data curation, Formal analysis, Visualization, Conceptualization. Beichen Hu: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Visualization, Conceptualization. Hui Liu: Conceptualization, Supervision. Bohan Dong: Formal analysis.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Abdelghani, R., Sauzéon, H., & Oudeyer, P. Y. (2023). Generative AI in the classroom: Can students remain active learners?. *arXiv:2310.03192*. https://doi.org/10.48550/arXiv.2310.03192

Amoozadeh, M., Daniels, D., Nam, D., Kumar, A., Chen, S., Hilton, M., Srinivasa Ragavan, S., & Alipour, M. A. (2024). Trust in generative AI among students: An exploratory study. In *Proceedings of the 55th ACM technical symposium on computer science education v. 1* (pp. 67–73). https://doi.org/10.1145/3626252.3630842

Barrett, M., Oborn, E., Orlikowski, W. J., & Yates, J. (2012). Reconfiguring boundary relations: Robotic innovations in pharmacy work. *Organization Science*, *23*(5), 1448–1466. https://doi.org/10.1287/orsc.1100.0639

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Celik, I., Kontkanen, S., Laru, J., & Dalyanci, A. A. (2025). Co-constructing adaptive lesson plans with GAI: Students' intelligent-TPACK and prompt engineering strategies. *Computers & Education*, *241*, Article 105485. https://doi.org/10.1016/j.compedu.2025.105485

Chen, Z., & Chan, J. (2024). Large language model in creative work: The role of collaboration modality and user expertise. *Management Science*, *70*(12), 9101–9117. https://doi.org/10.1287/mnsc.2023.03014

Choi, S., Jeong, S., Park, S., Kim, Y., & Han, I. (2026). Analyzing teacher–AI interaction patterns across teacher experience and AI proficiency in student-centered lesson design. *Teaching and Teacher Education*, *169*, Article 105266. https://doi.org/10.1016/j.tate.2025.105266

Chu, X., Wang, M., Spector, J. M., Chen, N. S., Chai, C. S., Hwang, G. J., & Zhai, X. (2025). Enhancing the flipped classroom model with generative AI and Metaverse technologies: insights from lag sequential and epistemic network analysis. *Educational technology research and development*, *73*(3), 1539-1567. https://doi.org/10.1007/s11423-025-10457-2

Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., & Siemens, G. (2024). Impact of AI assistance on student agency. *Computers & Education*, *210*, Article 104967. https://doi.org/10.1016/j.compedu.2023.104967

ElSayary, A. (2024). An investigation of teachers' perceptions of using ChatGPT as a supporting tool for teaching and learning in the digital era. *Journal of Computer Assisted Learning*, *40*(3), 931–945. https://doi.org/10.1111/jcal.12926

Feng, S. (2025). Group interaction patterns in generative AI-supported collaborative problem solving: Network analysis of the interactions among students and a GAI chatbot. *British Journal of Educational Technology*, 56, 2125–2145. https://doi.org/10.1111/bjet.13611

Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, *15*(1), 6. https://doi.org/10.3390/soc15010006

Geroimenko, V. (2026). Feedback and iterative refinement: Improving AI responses through interaction cycles. In *Beyond and after prompt engineering: The future of AI communication*. Cham: Springer. https://doi.org/10.1007/978-3-032-04569-0_11.

Good, P. (2000). *Permutation tests: A practical guide to resampling methods for testing hypotheses*. Springer Science & Business Media.

Güner, H., Er, E. (2025). AI in the classroom: Exploring students' interaction with ChatGPT in programming learning. *Education and Information Technologies*, *30*, 12681–12707 https://doi.org/10.1007/s10639-025-13337-7

Guo, W., Liang, Z., Wang, C., Li, X., Hu, H., Chen, S., Yu, Q., & Zhao, Q. (2025). Student-AI collaborative creative problem-solving: The role of human agency. *Computers & Education*, *239*, Article 105433. https://doi.org/10.1016/j.compedu.2025.105433

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *7*(2), 174–196. https://doi.org/10.1145/353485.353487

Hou, H. T. (2015). 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. *Computers in Human Behavior*, *48*, 424–435. https://doi.org/10.1016/j.chb.2015.02.010

Hwang, G. J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *1*, Article 100001. https://doi.org/10.1016/j.caeai.2020.100001

Jin, F., Sun, L., Pan, Y., & Lin, C. H. (2025). High heels, compass, spider-man, or drug? Metaphor analysis of generative artificial intelligence in academic writing. *Computers & Education*, *228*, Article 105248. https://doi.org/10.1016/j.compedu.2025.105248

Kasneci E., Sessler K., Küchemann S., Bannert M., Dementieva D., Fischer F., Gasser U., Groh G., Günnemann S., Hüllermeier E., Krusche S., Kutyniok G., Michaeli T., Nerdel C., Pfeffer J., Poquet O., Sailer M., Schmidt A., Seidel T., Kasneci G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kim, J., Yu, S., Lee, S.-S., & Detrick, R. (2025). Students' prompt patterns and its effects in AI-assisted academic writing: Focusing on students' level of AI literacy. *Journal of Research on Technology in Education*, 1–18. https://doi.org/10.1080/15391523.2025.2456043

Knoth, N., Tolzin, A., Janson, A., & Leimeister, J. M. (2024). AI literacy and its implications for prompt engineering strategies. *Computers and Education: Artificial Intelligence*, *6*, Article 100225. https://doi.org/10.1016/j.caeai.2024.100225

Krakowski, S. (2025). Human-AI agency in the age of generative AI. *Information and Organization*, *35*(1), Article 100560. https://doi.org/10.1016/j.infoandorg.2025.100560

Kulkarni, N. D., & Tupsakhare, P. (2024). Crafting effective prompts: Enhancing ai performance through structured input design. *Journal of Recent Trends in Computer Science and Engineering (JRTCSE)*, *12*(5), 1–10. https://doi.org/10.70589/JRTCSE.2024.5.1

Lakoff, G., & Johnson, M. (1980). Metaphors we live by. University of Chicago Press.

Lan Y. J., Chen N. S. (2024). Teachers' agency in the era agency of LLM and generative AI: Designing pedagogical AI agents. *Educational Technology & Society*, *27*(1), 1–17. https://doi.org/10.30191/ETS.202401_27(1).PP01

Li, S., Liu, J., & Dong, Q. (2025). Generative artificial intelligence-supported programming education: Effects on learning performance, self-efficacy and processes. *Australasian Journal of Educational Technology*, *41*(3), 1–25. https://doi.org/10.14742/ajet.9932

Liu, J., Li, S., & Dong, Q. (2024). Collaboration with generative artificial intelligence: An exploratory study based on learning analytics. *Journal of Educational Computing Research*, *62*(5), 1014–1046. https://doi.org/10.1177/07356331241242441

Liu, J., Zhang, Y., Li, W., Wang, Q., Niu, P., & Zhang, X. (2025). Adaptive vs. planned metacognitive scaffolding for computational thinking: Evidence from generative AI-supported programming in elementary education. *Computers & Education*, *241*, Article 105473. https://doi.org/10.1016/j.compedu.2025.105473

López-Pernas, S., Misiejuk, K., Kaliisa, R., & Saqr, M. (2025). Capturing the process of students' AI interactions when creating and learning complex network structures. *IEEE Transactions on Learning Technologies*, *18*, 556–568. https://doi.org/10.1109/TLT.2025.3568599

Ma, N., & Zhong, Z. (2025). A meta-analysis of the impact of generative artificial intelligence on learning outcomes. *Journal of Computer Assisted Learning*, *41*(5), Article e70117, https://doi.org/10.1111/jcal.70117

Misiejuk, K., López-Pernas, S., Kaliisa, R., Saqr, M. (2025). Learning together: Modeling the process of student-AI interactions when generating learning resources. In *Proceedings of TEEM 2024* (pp. 448–458). *Lecture Notes in Educational Technology*. https://doi.org/10.1007/978-981-96-5658-5_45

Nam, D., Macvean, A., Hellendoorn, V., Vasilescu, B., & Myers, B. (2024). Using an LLM to help with code understanding. In *Proceedings of the ieee/acm 46th international conference on software engineering* (pp. 1-13). https://doi.org/10.1145/3597503.3639187
Nguyen, A., Hong, Y., Dang, B., & Huang, X. (2024). Human-AI collaboration patterns in AI-assisted academic writing. *Studies in Higher Education*, *49*(5), 847–864. https://doi.org/10.1080/03075079.2024.2323593
O'Dea, X. (2024). Generative AI: Is it a paradigm shift for higher education? *Studies in Higher Education*, *49*(5), 811–816. https://doi.org/10.1080/03075079.2024.2332944
Oliveira, M., Zednik, C., Bombaerts, G., Sadowski, B., & Conijn, R. (2025). Assessing students' DRIVE: A framework to evaluate learning through interactions with generative AI. *Computers and Education: Artificial Intelligence*, *9*, Article 100497. https://doi.org/10.1016/j.caeai.2025.100497
Panadero, E., & Alonso-Tapia, J. (2014). How do students self-regulate? Review of Zimmerman's cyclical model of self-regulated learning. *Anales de psicologia*, *30*(2), 450–462. https://doi.org/10.6018/analesps
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas , J. Alexandre, P., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.
Ren, W., Li, J., Pi, Z., Guo, J., & Li, X. (2026). How do high-performers and low-performers differently engage in collaborative creative problem solving with a conversational GAI chatbot?. *Thinking Skills and Creativity*, *60*, Article 102134. https://doi.org/10.1016/j.tsc.2026.102134
Salomon, G. (1997). *Distributed cognitions: Psychological and educational considerations.* Cambridge University Press.
Serra, P., & Oliveira, Â. (2025). AI-powered prompt engineering for education 4.0: Transforming digital resources into engaging learning experiences. *Education Sciences*, *15*(12), Article 1640. https://doi.org/10.3390/educsci15121640
Shen, W., Lin, X. F., Liu, J., Liang, X., Chen, R., Lai, X., & Zheng, X. (2025). The process of undergraduates' collaboration with a generative artificial intelligence chatbot: Insights from conversation content and epistemic network analysis. *Journal of Computer Assisted Learning*, *41*(6), Article e70140. https://doi.org/10.1111/jcal.70140
Stieglitz, S., Mirbabaie, M., Möllmann, N. R., & Rzyski, J. (2022). Collaborating with virtual assistants in organizations: Analyzing social loafing tendencies and responsibility attribution. *Information Systems Frontiers*, *24*(3), 745–770. https://doi.org/10.1007/s10796-021-10201-0
Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, *9*(1), Article 307. https://doi.org/10.1186/1471-2105-9-307
Su, Y., Li, Y., Hu, H., & Rosé, C. P. (2018). Exploring college English language learners' self and social regulation of learning during wiki-supported collaborative reading activities. *International Journal of Computer-Supported Collaborative Learning*, *13*, 13–35. https://doi.org/10.1007/s11412-018-9269-y
Sun, Y., & Huang, Y. (2025). Exploring the Role of Generative AI in Collaborative Lesson Planning for Pre-Service Teachers. *Journal of Computer Assisted Learning*, *41*(4), Article e70098. https://doi.org/10.1111/jcal.70098
Tankelevitch, L., Kewenig, V., Simkute, A., Scott, A. E., Sarkar, A., Sellen, A., & Rintel, S. (2024). The metacognitive demands and opportunities of generative AI. In *Proceedings of the 2024 CHI conference on human factors in computing systems*. https://doi.org/10.1145/3613904.3642902

Tu, Y. F., & Hwang, G. J. (2024). University students' conceptions of ChatGPT-supported learning: a drawing and epistemic network analysis. *Interactive Learning Environments*, *32*(10), 6790–6814. https://doi.org/10.1080/10494820.2023.2286370

Urban, M., Lukavský, J., Brom, C., Hein, V., Svacha, F., Děchtěrenko, F., & Urban, K. (2025). Prompting for creative problem-solving: A process-mining study. *Learning and Instruction*, *99*, Article 102156. https://doi.org/10.1016/j.learninstruc.2025.102156

Vallis, C., Wilson, S., & Casey, A. (2025). Fear and awe: Making sense of generative AI through metaphor. *Journal of Interactive Media in Education*, *2025*(1), 1–15. https://doi.org/10.5334/jime.972

Wang, S., Xu, L., Liu, J., & Zhai, Y. (2026). Addressing the challenges of AI-generated assignment submissions in education: Insights and strategies. J*ournal of Data Science*, *24*(1), 254–260. https://doi.org/10.6339/25-JDS1208

Weller, M. (2023). *Metaphors of Ed Tech (1st ed.)*. Athabasca University Press.

White J., Fu Q., Hays S., Sandborn M., Olea C., Gilbert H., Elnashar A., Spencer-Smith J., Schmidt D. C. (2023). A prompt pattern catalog to enhance prompt engineering with ChatGPT. *arXiv: 2302.11382*. https://doi.org/10.48550/arXiv.2302.11382

Xia, L., An, X., Li, X., & Dong, Y. (2026). Perceptions of generative artificial intelligence (AI), behavioral intention, and use experience as predictors of university students' learning agency in generative AI-supported contexts. *Journal of Educational Computing Research*, *64*(1), 92–125. https://doi.org/10.1177/07356331251382853

Xu, Q., & Yin, J. (2021). Application of random forest algorithm in physical education. *Scientific Programming*, *2021*, Article 1996904. https://doi.org/10.1155/2021/1996904

Xu, X., Qiao, L., Cheng, N., Liu, H., & Zhao, W. (2025). Enhancing self-regulated learning and learning experience in generative AI environments: The critical role of metacognitive support. *British Journal of Educational Technology*, *56*(5), 1842–1863. https://doi.org/10.1111/bjet.13599

Yang, X., Liu, X., & Gao, Y. (2025). The impact of generative AI on students' learning: A study of learning satisfaction, self-efficacy and learning outcomes. *Educational Technology Research & Development*, *73*(1), 55–78. https://doi.org/10.1007/s11423-025-10540-8

Zhang, Q., Zuo, J., & Yang, S. (2025). Research on the impact of generative artificial intelligence (GenAI) on enterprise innovation performance: a knowledge management perspective. *Journal of Knowledge Management*, *29*(7), 2238–2257. https://doi.org/10.1108/JKM-10-2024-1198

Zhang, S., Gao, Q., Sun, M., Cai, Z., Li, H., Tang, Y., & Liu, Q. (2022). Understanding student teachers' collaborative problem solving: Insights from an epistemic network analysis (ENA). *Computers & Education*, *183*, Article 104485. https://doi.org/10.1016/j.compedu.2022.104485

Zimmerman, B. J. (2000). Chapter 2-attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). Academic Press. https://doi.org/10.1016/B978-012109890-2/50031-7