Paper ID #51203

Designing Guided AI Systems to Foster Problem-Solving and Reflection in Computing Education

Dr. Mahfuza Farooque, Pennsylvania State University

Mahfuza Farooque is an Associate Teaching Professor in the School of Electrical Engineering and Computer Science at Pennsylvania State University. She earned her Ph.D. in Computer Science from École Polytechnique, France, specializing in proof search and automated reasoning. Her research interests include formal methods, large language models, and the application of machine learning in education and mental health. Dr. Farooque has authored a textbook on discrete mathematics. She also serves as Director of the Master of Engineering program in Computer Science and Engineering. She regularly mentors undergraduate and graduate research projects that have received institutional recognition and funding.

Veer Amish Shah, Pennsylvania State University

Veer Shah is an undergraduate honors student specializing in artificial intelligence at The Pennsylvania State University.

Nikhil Khattar, Pennsylvania State University
Sinjoy Saha
Suman Saha, Pennsylvania State University

Suman Saha received a Ph.D. degree in computer science from Pierre and Marie Curie University, Paris, France, in 2013.

He is an Assistant Teaching Professor at Pennsylvania State University, University Park, PA, USA. Since 2016, he has been involved in academia. Before his current role with Pennsylvania State University, he taught at Illinois Institute of Technology, Chicago, IL, USA, and Illinois State University, Normal, IL, USA.

Dr. Saha received the distinguished William C. Carter Award for his substantial contributions to dependable and secure computing during his doctoral studies. He has held positions at several prestigious National and International Research Labs, including Microsoft, Cambridge, U.K., Harvard University, Cambridge, USA, and the National Institute for Research in Digital Science and Technology (Inria), Paris, France.

Designing Guided AI Systems to Foster Problem-Solving and Reflection in Computing Education

Abstract

Artificial Intelligence (AI) is rapidly transforming education, reshaping how students learn, study, and engage with complex material. While AI tools can enhance efficiency and provide rapid access to information, prior research suggests that unstructured use of generative AI may encourage surface-level learning, answer-seeking behavior, and reduced engagement with underlying reasoning processes [1–4]. This approach bypasses reflection and conceptual understanding, the very processes that enable critical thinking and problem solving. As AI becomes a standard presence in higher education, there is an urgent need to rethink its role not as a substitute for learning, but as a catalyst for deeper engagement and cognitive growth.

This study examines how AI can serve as a guided learning partner that helps students develop reasoning skills through structured, step-by-step problem-solving. Instead of providing direct solutions, the proposed system encourages students to think through each stage of a problem, promoting persistence, self-reflection, and meaningful learning. Drawing on theories of self-regulated learning and cognitive scaffolding, the system is designed to prompt students with targeted questions, explanations, and contextual hints aligned with course objectives.

The prototype integrates retrieval-augmented generation (RAG) and generative AI to deliver personalized, context-aware guidance. Using RAG, the system retrieves relevant course content and examples to ground the AI's responses in the material taught by instructors. This ensures that feedback and hints remain pedagogically aligned, reducing hallucinations and preserving academic integrity. The architecture supports interchangeable large language models (LLMs) and adaptive retrievers to maintain flexibility across disciplines.

The system was implemented in a Discrete Mathematics course with approximately 300 to 350 students at a large public R1 university. A mixed-methods research design was used to examine student engagement, confidence, and learning-related experiences. Quantitative data include system interaction logs and usage analytics, while qualitative data were gathered through surveys and reflective feedback. The analysis focuses on how guided AI interactions influence students' motivation, problem-solving strategies, and self-regulated learning behaviors.

This work contributes to ongoing conversations about the responsible integration of AI in education. The findings suggest that when AI systems are intentionally designed to scaffold reasoning rather than deliver answers, they can complement instruction by reinforcing metacognition and persistence without displacing existing learning practices.

Introduction

Generative artificial intelligence (AI) tools are rapidly becoming embedded in students' learning practices across higher education. In computing and mathematically intensive courses, students increasingly rely on AI systems to obtain explanations, generate solutions, and check their work. While these tools offer efficiency and accessibility, prior research has raised concerns about surface-level learning, overreliance on answer generation, and diminished opportunities for reflection when students bypass the underlying reasoning process [1–3]. These concerns are particularly salient in large-enrollment computing courses, where students often struggle with complex, multi-step reasoning and have limited access to individualized instruction. Research on instructional scaffolding and intelligent tutoring systems has shown that effective support should guide learners through problem solving while preserving learner agency and responsibility [5, 6]. Similarly, work on self-regulated learning emphasizes that students develop a deeper understanding when they actively plan, monitor, and regulate their learning strategies [7–9]. Studies on productive failure further suggest that attempting challenging problems prior to receiving explicit instruction can lead to stronger conceptual understanding when followed by targeted guidance [10, 11]. Motivated by these insights, this work investigates how AI systems can be designed not as automated problem solvers, but as guided learning partners that support reasoning, reflection, and persistence. We present an AI-guided learning platform that combines retrieval-augmented generation with intent-aware progressive scaffolding. Rather than providing direct solutions, the system delivers staged guidance—such as metacognitive prompts, conceptual hints, and high-level structural outlines—and advances only when students demonstrate engagement. To ensure instructional consistency and reduce hallucinations, all responses are grounded in instructor-curated course materials [12, 13].

This study addresses the following research questions:
RQ1: How do students perceive the effectiveness of AI-guided scaffolding in supporting problem solving, conceptual understanding, and learning confidence without providing direct solutions?

RQ2: What patterns of student interaction with the platform emerge from system logs, and to what extent do these patterns reflect iterative, process-oriented engagement?

RQ3: How do students use hints and feedback during problem solving, and what evidence suggests that these scaffolds support persistence and error recovery?

RQ4: How do students integrate the platform into their existing study practices, and do they perceive it as a complementary learning resource rather than a replacement for traditional materials?

We describe the design, implementation, and classroom deployment of this platform in a large-enrollment Discrete Mathematics course at a public R1 university. Analysis of interaction logs and post-deployment surveys indicates that students primarily used the system to reason through problems rather than to obtain answers. Quantitative log data show that most student sessions involved multiple interaction turns, with frequent transitions from incorrect attempts to hint usage followed by revised submissions, suggesting iterative engagement and productive struggle. Survey results further indicate that a strong majority of students perceived the platform as helpful for clarifying reasoning steps, increasing confidence, and supporting independent problem solving. Students also reported using the system alongside lectures, notes, and assignments, reinforcing its role as a supplemental learning resource rather than a substitute. Together, these findings demonstrate that AI systems designed with pedagogical constraints and progressive scaffolding can support meaningful engagement and reflective problem solving in large-scale computing courses, addressing common concerns about overreliance on generative AI while preserving learner agency.

Related Work

This work draws on prior research in three intersecting areas: (1) scaffolded learning and intelligent tutoring systems, (2) self-regulated learning and productive struggle, and (3) applications of generative AI in computing education. For each area, we describe how existing work has approached key challenges, identify gaps that remain unaddressed in the context of AI-assisted learning at scale, and explain how our approach extends or differs from prior solutions. This organization also aligns with our research questions, which examine whether AI guidance can support reasoning without answer disclosure (RQ1), promote iterative engagement and persistence (RQ2–RQ3), and complement students' existing study practices (RQ4). Related classroom work also shows that brief, example-based supplemental resources can increase engagement and support exam preparation in large computing courses [14]

Scaffolded Learning and Intelligent Tutoring Systems

Instructional scaffolding emphasizes that effective support should adapt to a learner's current understanding and gradually fade as competence increases [5, 15]. Intelligent tutoring systems (ITS) operationalize these principles by providing structured hints, feedback, and step-level guidance, and prior syntheses report that well-designed tutoring systems can be highly effective when compared to less interactive approaches [6, 16, 17]. Classic ITS architectures, such as model-tracing tutors, rely on cognitive task analysis to construct explicit domain models that anticipate correct and incorrect reasoning paths, enabling adaptive feedback tailored to misconceptions.

Despite their success, traditional ITS often face scalability constraints in large, rapidly evolving courses. Many approaches require substantial domain engineering, in which instructors encode problem-solving rules, common errors, and remediation strategies. This becomes costly when content changes across semesters or when courses include diverse problem types that resist straightforward formalization. In addition, classical ITS frequently delivers feedback at predetermined granularities, which may not match students' needs: some learners benefit from high-level strategic prompts, while others require detailed procedural guidance.

To reduce authoring cost, more recent work has explored alternative approaches such as example-tracing tutors, which allow instructors to author worked examples rather than full cognitive models [17]. Constraint-based modeling further reduces procedural modeling by representing domain knowledge as correctness constraints. However, these approaches still require substantial upfront investment and domain-specific customization, limiting portability across courses and institutions.

Our approach addresses these limitations by integrating retrieval-augmented generation (RAG) with instructor-curated materials to provide pedagogically grounded guidance without requiring explicit cognitive models or constraint sets. Rather than anticipating all solution paths through manual encoding, the system dynamically retrieves relevant course content and uses it to ground AI-generated scaffolding aligned with instructor terminology and instructional intent. This design is directly motivated by **RQ1**: supporting reasoning without direct answer disclosure while remaining scalable for large-enrollment courses.

Self-Regulated Learning and Productive Struggle

Theories of self-regulated learning (SRL) highlight learners' active role in planning, monitoring, and regulating cognition and motivation [7–9]. SRL includes metacognitive processes such as goal setting, strategy selection, self-monitoring, and adaptive help seeking. Within technology-enhanced learning environments, research on help seeking shows that students may misuse hints or rely excessively on assistance unless systems scaffold productive engagement [18, 19]. For example, students may request hints prematurely without attempting problems independently, resulting in shallow processing and reduced opportunities for learning.

Complementing SRL research, studies on productive failure and productive struggle demonstrate that carefully designed difficulty, followed by targeted guidance, can improve conceptual understanding and transfer [10, 11]. However, this balance is delicate: too little support can produce frustration and disengagement, whereas premature assistance can undermine effortful processing.

Prior work has investigated mechanisms that encourage productive help seeking and appropriate struggle. Some systems delay hint availability, limit hint frequency, or require learners to demonstrate effort (e.g., submitting an attempt) before accessing assistance [18, 20]. Adaptive hint systems also vary feedback specificity based on student performance, providing more explicit support after repeated errors (i.e., contingent tutoring) [21, 22]. Nevertheless, many approaches operate at coarse levels of adaptation (e.g., step-level) and may not distinguish between concept clarification and procedural problem-solving needs. A student asking "What does it mean for a function to be bijective?" requires different support than a student asking "How do I prove this function is bijective?" mid-proof, yet many systems treat these requests similarly [23].

Our system advances this line of work by implementing progressive scaffolding that balances support with learner autonomy. Students do not advance to more explicit scaffolding levels without demonstrating engagement at prior levels (e.g., attempting an intermediate step or requesting a targeted hint). In addition, intent-aware response generation distinguishes between conceptual and procedural queries, providing definitional explanations when appropriate while offering metacognitive prompts and strategic guidance during problem solving. These design choices connect directly to **RQ2** and **RQ3**, which examine iterative engagement, persistence, and students' perceptions of reflection and confidence under scaffolded AI support.

Generative AI in Computing Education

Recent work on generative AI in education highlights both opportunities for scalable instructional support and risks related to hallucination, curriculum misalignment, and academic integrity [1–3]. Large language models (LLMs) such as ChatGPT can generate explanations and solutions across domains, but they are trained on broad corpora rather than course-specific materials. As a result, responses may conflict with instructor terminology, notational conventions, or the problem-solving approaches emphasized in a particular course. In mathematically precise domains such as discrete mathematics, these inconsistencies can increase confusion rather than support learning. Related work has also used LLMs to generate microlearning materials such as quizzes, flashcards, and mini-lessons, while emphasizing the need for instructor review for accuracy and alignment [24].

Empirical work in computing education suggests that unstructured use of AI tools (e.g., ChatGPT or code-generation assistants) can shift effort away from core reasoning activities and encourage answer-seeking behavior [4, 25], whereas recent guided GenAI systems have begun exploring personalized, interactive support for targeted computing concepts [26]. These patterns raise concerns about learning quality and academic integrity, particularly when AI tools optimize for solution delivery rather than reasoning support.

Retrieval-augmented generation (RAG) has emerged as a strategy for grounding AI outputs in vetted instructional materials, improving factuality and alignment with course content [12, 13]. In educational contexts, RAG-based systems can improve consistency with course terminology and reduce hallucination. However, grounding alone does not guarantee pedagogical appropriateness: students may still use grounded AI responses to bypass effortful problem solving if the system provides direct solutions.

Our approach extends this line of work by integrating RAG with intent-aware progressive scaffolding explicitly designed to support learning processes rather than answer delivery. The system uses retrieved course context to generate metacognitive prompts, conceptual hints, and high-level structural guidance while avoiding complete solution disclosure. Further, instructor-facing logging enables oversight of how students engage with AI support and supports iterative refinement of the knowledge base. This positioning directly addresses **RQ4** by examining whether students view the platform as a complementary learning resource rather than a replacement for established study practices.

System Design and Methodological Framework

This section describes the design and operational framework of the AI-guided learning platform used in this study. The platform was deployed in a large-enrollment undergraduate Discrete Mathematics course at a public R1 university. Because this paper is a research-to-practice study focused on classroom implementation, the methodology emphasizes how the system was designed, deployed, and instrumented to support scaffolded problem solving at scale. We present the system architecture, a representative interaction flow, and the pedagogical and technical design principles that guided development. The platform logs fine-grained interaction traces; in this paper, we report descriptive, session-level log measures to characterize real-world use, while

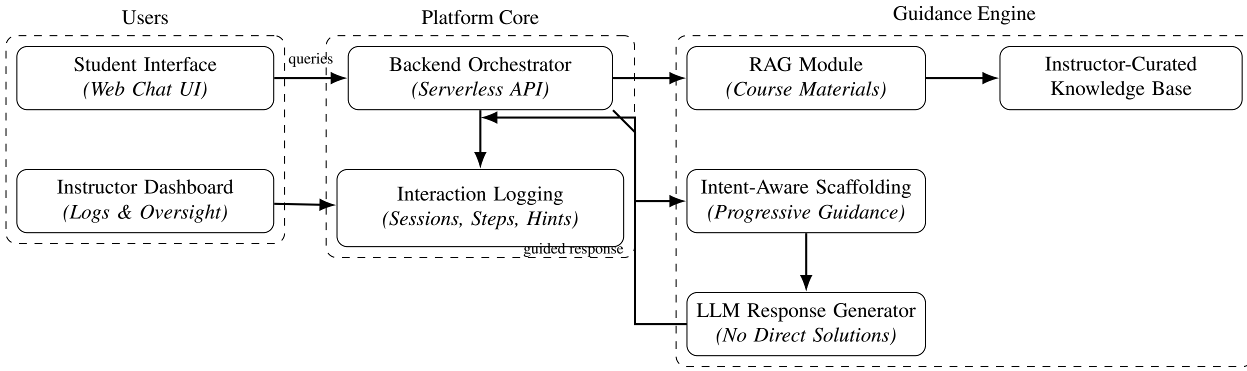


Figure 1: System architecture of the AI-guided learning platform. Student queries are processed by a serverless backend that retrieves instructor-aligned materials, applies intent-aware progressive scaffolding, and generates guided responses without revealing direct solutions. Interactions are logged to support instructor oversight and analysis of classroom use.

richer sequence- and time-based behavior modeling is reserved for future work.

Overview of System Architecture

Figure 1 illustrates the architecture of the AI-guided learning platform. The system comprises five primary components: (i) a student-facing interaction interface, (ii) an instructor-facing dashboard, (iii) a backend orchestration and logging layer, (iv) a retrieval-augmented grounding module, and (v) an intent-aware scaffolding and response generation engine. Together, these components operationalize progressive scaffolding, instructional alignment, and instructor oversight while remaining scalable for large-enrollment courses.

Students interact with the system through a web-based chat interface implemented using React and Next.js. Each interaction is associated with a specific assignment problem and session identifier to preserve contextual continuity across turns. The interface displays the current scaffolding step using a progress indicator and adapts guidance based on student attempts and follow-up questions.

Instructors access a separate dashboard that provides visibility into student–system interactions at the level of individual problems. Interaction data are stored in Amazon DynamoDB and indexed by anonymized identifiers, enabling efficient retrieval of interaction histories for instructional monitoring and iterative refinement.

Interaction Flow

Figure 2 illustrates a representative interaction. An interaction begins when a student submits a problem-related query through the web interface. The backend orchestrator validates the session, assigns an anonymized identifier and timestamp, and forwards the query to the retrieval-augmented grounding module.

The grounding module performs semantic retrieval over instructor-curated course materials,

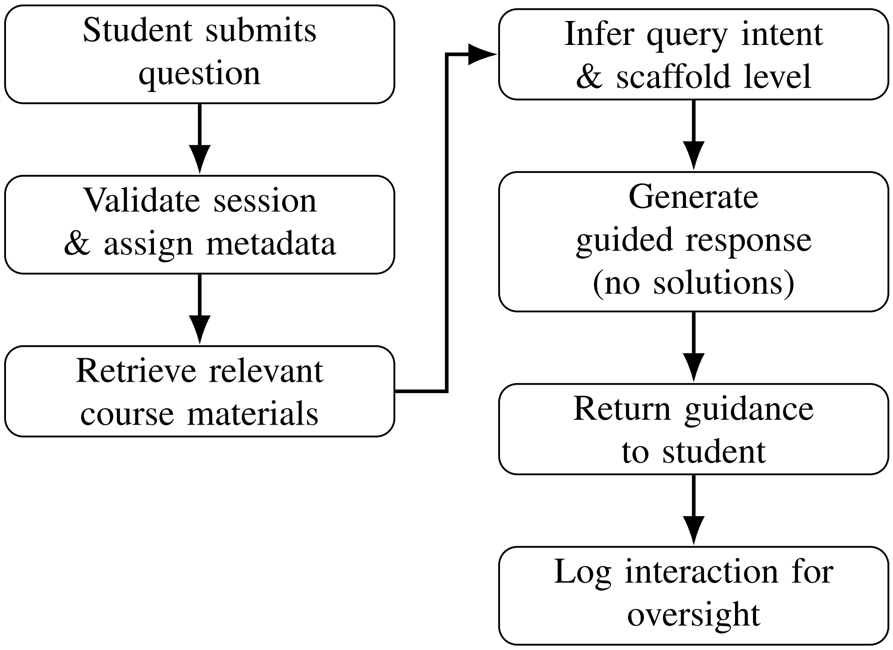


Figure 2: Interaction flow for the AI-guided learning platform. Queries are validated, grounded in instructor-curated materials, processed through intent-aware progressive scaffolding, and returned as guided responses. Interactions are logged for instructional oversight and analysis of classroom use.

including lecture notes, selected textbook excerpts, and worked examples. Retrieved passages are assembled into a contextual evidence bundle that anchors subsequent guidance in course-specific terminology and instructional intent.

Using the retrieved context and inferred intent, the system constructs a scaffolded prompt that constrains the language model to course-aligned materials and enforces a progressive scaffolding policy. Depending on the query type and prior interaction context, the system generates metacognitive prompts, conceptual hints, or high-level structural guidance while explicitly avoiding direct solution disclosure.

In this study, *scaffolding engagement* refers to structured, interactive guidance that requires students to actively attempt problem solving while the system adjusts the level of support based on observed effort and need.

Design Principles

The platform was guided by five pedagogical and technical design principles intended to support learning while protecting academic integrity.

1. **Pedagogical Alignment with Course Content.** AI responses are grounded in instructor-curated materials to ensure consistency with course terminology, notation, and problem-solving conventions.
2. **Progressive Scaffolding over Direct Solutions.** Building on prior work on productive

struggle and self-regulated learning [1, 27], the system favors incremental prompts and requires observable engagement before providing more explicit support.

3. **Intent-Aware Response Generation.** Queries are categorized by intent (e.g., conceptual clarification versus procedural guidance), and the form and granularity of guidance are adjusted accordingly.

4. **Transparency in AI Feedback.** Guided responses include brief pedagogical rationales to clarify the purpose of prompts and encourage reflective problem solving.

5. **Instructor Oversight and Academic Integrity.** Interaction logs support instructor monitoring and iterative refinement while discouraging misuse of the system.

Survey Instrument and Data Collection

To capture student perceptions, we administered a post-use survey at the end of the instructional period. A total of 88 students completed the survey on a voluntary basis. The survey consisted primarily of Likert-scale items assessing perceived helpfulness of step-by-step prompts, instructional alignment, clarity of guidance, engagement-related constructs (reflection, persistence, confidence), and overall perceived effectiveness. Response scales used five ordered options (e.g., strongly disagree to strongly agree, or very unhelpful to very helpful), depending on the item.

The survey also included two open-ended questions asking students to describe (i) aspects of the platform they found most helpful and (ii) suggested improvements for future versions. These responses were analyzed to identify recurring themes. Log data were collected automatically for all platform users during deployment and are used in this paper to report descriptive, session-level engagement measures; more detailed sequence- and time-based modeling is left for future work. Survey responses were provided by 88 students; open-ended comments were submitted by 86 respondents; interaction logs include all platform users (104 students); and engagement-level analyses in the Results section use the matched subset of 54 students.

Results

This section reports findings from three complementary sources: (1) post-use student surveys, (2) system interaction logs, and (3) open-ended student feedback. Results are organized around practical classroom adoption questions: whether students experienced the platform as helpful for reasoning (not answer delivery), whether observed usage reflected iterative engagement, and whether students perceived the system as accurate and instructionally aligned.

Perceived Helpfulness of Guided Reasoning

Figure 3 summarizes student ratings of how helpful the platform's step-by-step Prompts were for guiding reasoning without directly providing answers. Responses were concentrated in the positive range: the largest group selected *Somewhat helpful* (about 50 students), followed by *Very helpful* (about 18 students). A smaller portion selected *Neither helpful nor unhelpful* (about 14 students), and comparatively few students selected negative options (*Somewhat unhelpful* or *Very* *unhelpful*). From an implementation perspective, this distribution matters because the platform was intentionally constrained to provide scaffolded prompts rather than final solutions. The predominance of positive ratings suggests that students valued process-oriented support, while the presence of neutral responses is consistent with selective use (e.g., students choosing support only when needed).

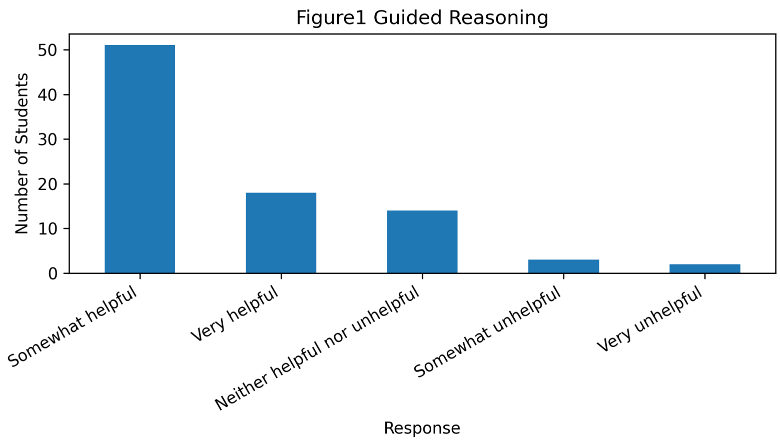
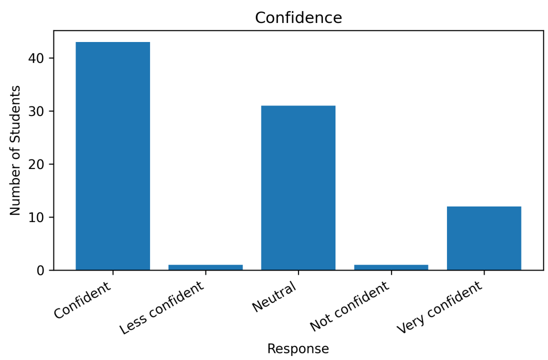


Figure 3: Student perceptions of the helpfulness of step-by-step prompts for guiding reasoning without directly providing answers.

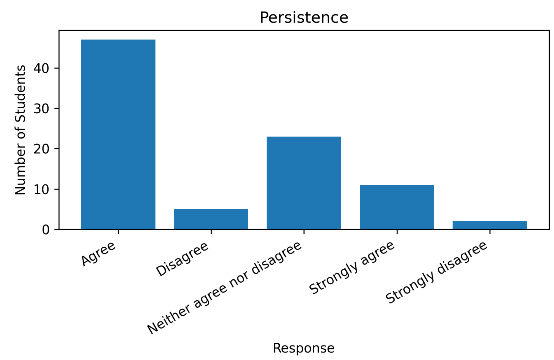
Reflection, Persistence, and Confidence

Figure 4 summarizes student responses related to reflection, persistence, and confidence following use of the platform. For confidence, responses were distributed primarily across the *confident*, *neutral*, and *very confident* categories, with comparatively few students indicating reduced confidence. This distribution indicates that most students reported stable or increased confidence after interacting with the system, while negative responses were rare. Responses related to persistence showed a similar pattern. A majority of students selected *agree* or *strongly agree* when asked whether the platform helped them keep working when problems were difficult. Neutral responses were also common, while disagreement was reported by a smaller subset of students. Strong negative responses were infrequent. Reflection-related responses indicate that students varied in the extent to which the platform prompted them to reflect on their problem-solving process. Most responses fell within the *somewhat*, *neutral*, or *a great deal* categories, suggesting that reflective engagement occurred for many students but was not uniformly experienced. Very few students reported no reflective benefit.

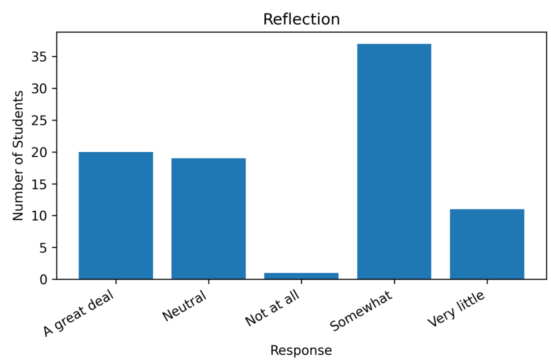
Across all three constructs, responses clustered in the neutral-to-positive range, with limited strongly negative responses. These distributions indicate that the platform was not widely perceived as detrimental to confidence, persistence, or reflection, while also revealing variation in how strongly students experienced these effects.



(a) Confidence



(b) Persistence



(c) Reflection

Figure 4: Student perceptions of confidence, persistence, and reflection after interacting with the AI-guided platform.

Accuracy and Instructional Alignment

Figures 5 and 6 report student perceptions of the accuracy and alignment of the platform's guidance with course materials and the consistency of hints and examples with what was taught in the course. Responses were largely favorable. In Figure 5, most students rated the platform as *Mostly Accurate* (45) or *Very Accurate* (20), while 18 selected *Neutral* and 5 selected *Sometimes Inaccurate*. In Figure 6, most students agreed that hints and examples were consistent with course instruction (*Agree*: 58; *Strongly Agree*: 11), with few students indicating disagreement (*Disagree*: 5; *Strongly Disagree*: 2) and no neutral selections.

From a research-to-practice standpoint, these results matter because classroom-facing AI tools are frequently criticized for hallucinations or misalignment with instructor expectations. While these are perception-based measures, the observed distributions suggest that most students experienced the platform as instructionally aligned, with a small minority reporting accuracy concerns that motivate targeted refinement (e.g., tightening prompt templates, expanding instructor-curated examples, or adding lightweight validation for high-risk topics).

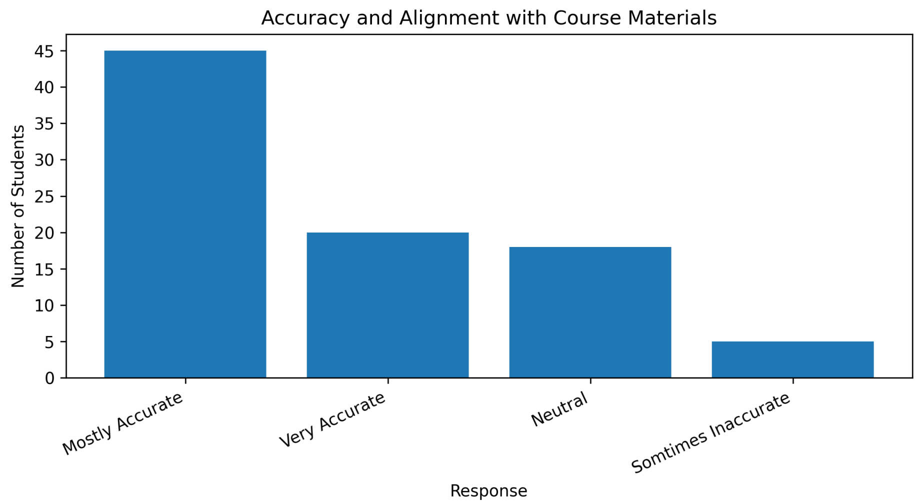


Figure 5: Student perceptions of the accuracy and alignment of the platform's guidance with course materials.

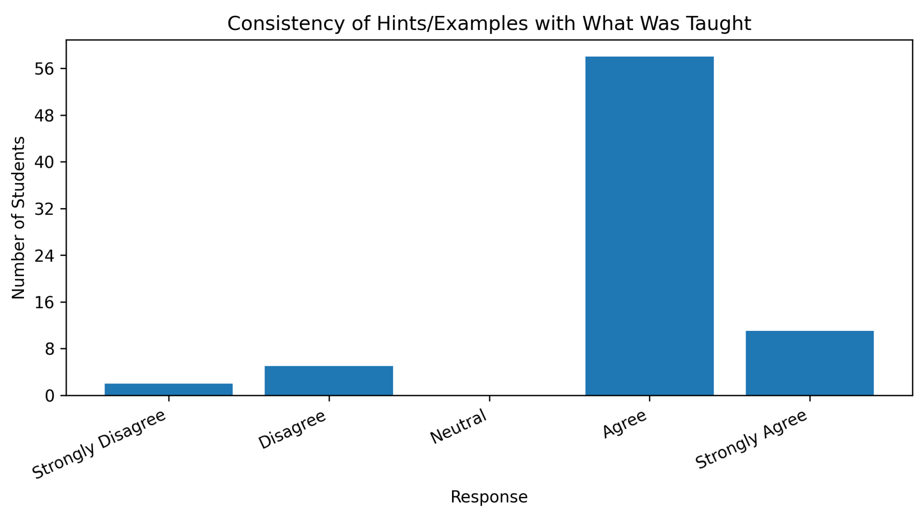


Figure 6: Student perceptions of whether hints and examples were consistent with what was taught in the course.

Effectiveness Relative to Usual Study Methods

Figure 7 summarizes student ratings of the platform's overall effectiveness compared to their usual study methods. Responses were concentrated in the neutral-to-positive range. The largest group selected *Somewhat effective* (about 44 students), followed by *Very effective* (about 15 students) and *Neither effective nor ineffective* (about 18 students). Fewer students rated the platform negatively (*Somewhat ineffective*: about 10; *Very ineffective*: 1). Notably, very few students rated the platform as ineffective.

Because the response options in Figure 7 are framed as effectiveness categories, the interpretation here follows the scale shown in the figure: most students viewed the platform as at least moderately effective, with a meaningful neutral group and limited negative ratings. In a large-enrollment setting, this is consistent with the platform functioning as a supplemental resource that some students use intensively and others use selectively.

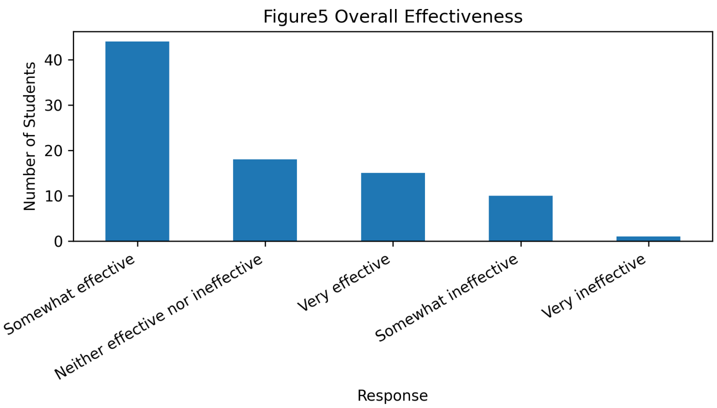


Figure 7: Student perceptions of the platform's effectiveness compared to usual study methods.

Observed Engagement from Interaction Logs

Survey findings are complemented by behavioral evidence from interaction logs. Across 104 unique students, the platform recorded 335 sessions and 1,460 interaction events (Table 1). Sessions typically involved multiple turns (median = 3; mean = 4.36), indicating that many students engaged in iterative exchanges rather than single-turn interactions.

Table 1: Summary of interaction log metrics (N=104 unique students; 335 sessions; 1,460 events). Percentages are computed over sessions.

Metric	Value
Unique students	104
Sessions	335
Total interactions	1,460
Median interactions per session	3
Mean interactions per session	4.36
Sessions with $\geq$1 hint request	30.45%
Sessions with $\geq$1 step answered	43.28%
Sessions with overall answer submitted	65.07%
Sessions with Attempt$\rightarrow$Hint$\rightarrow$Attempt	17.91%
Sessions with Incorrect$\rightarrow$Hint$\rightarrow$Correct	12.24%

Figure 8 shows the distribution of interactions per session. Most sessions cluster between 1 and 5 interactions, with a smaller number extending to higher counts, reflecting occasional sustained use.

To avoid ambiguity, the percentages discussed below are log-derived session-level rates that appear both in Table 1 and in the visualization in Figure 9. Overall answers were submitted in 65.07% of sessions. Step-level answers occurred in 43.28% of sessions and at least one hint request occurred in 30.45% of sessions, indicating that intermediate scaffolding features were used in a substantial share of sessions. Figure 9 also highlights revision-oriented sequences consistent with scaffolded problem solving: 17.91% of sessions followed an Attempt→Hint→Attempt pattern and 12.24% followed an Incorrect→Hint→Correct sequence. These trajectories indicate that, in many sessions, students used hints to revise attempts rather than treating hints as endpoints.

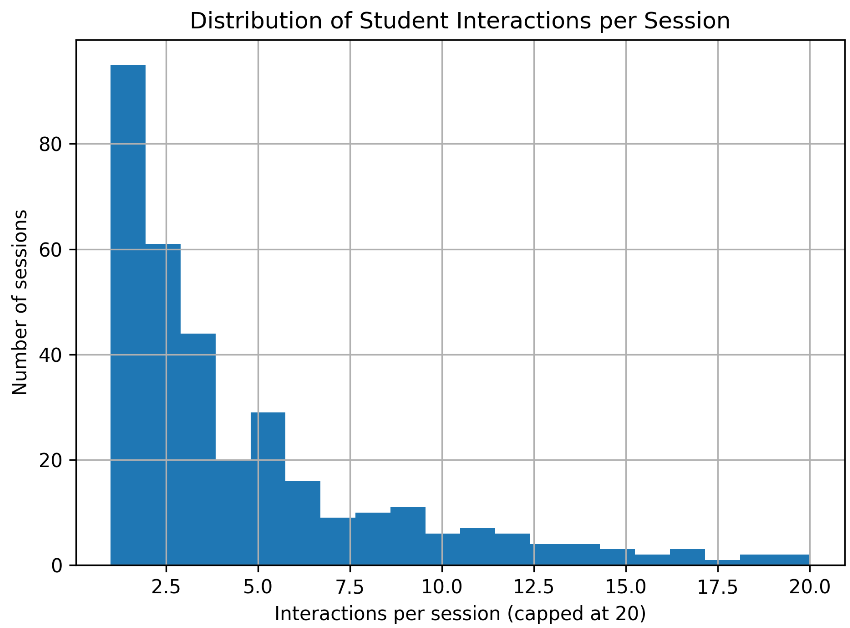


Figure 8: Distribution of interactions per session (capped at 20 for readability). Most sessions involve multiple interactions, indicating iterative engagement.

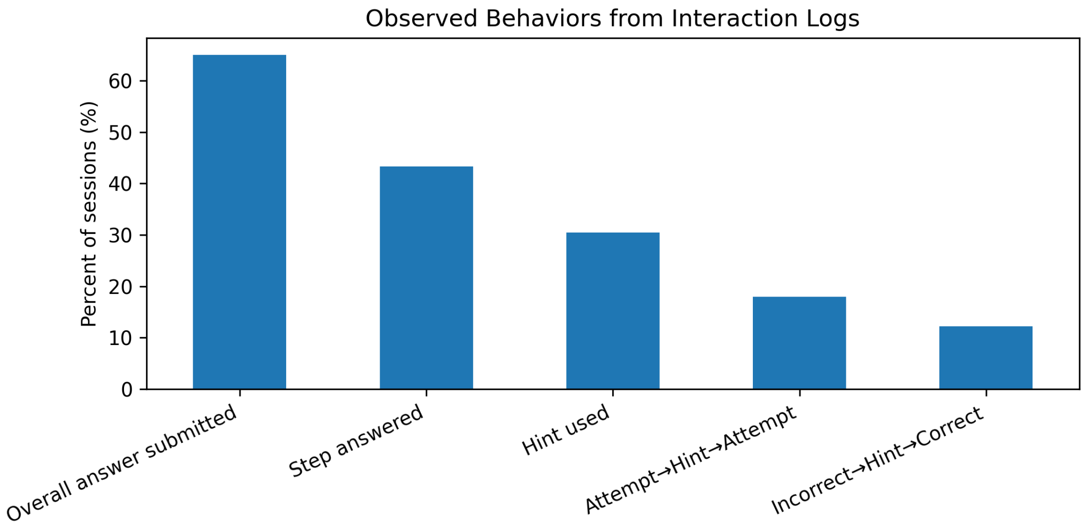


Figure 9: Observed behaviors from interaction logs. Percentages are calculated over sessions and show scaffolded engagement, including hint use and revision patterns.

Engagement-Level Differences in Perceptions and Feedback

To examine whether perceptions varied with observed usage, survey responses were matched to logs for 54 students and grouped into low (1–4 interactions), medium (5–8), and high (9+) engagement categories. Table 2 reports the percentage of students within each engagement group who gave positive, neutral, or negative ratings on three key survey measures.

Across all engagement levels, guided reasoning helpfulness was overwhelmingly positive. Notably, the high-engagement group remained strongly positive on guided reasoning (94.1% positive) and reported high willingness to recommend the platform (100% positive). Ratings of overall effectiveness showed more variation: compared to low-engagement students, medium- and high-engagement groups included a larger share of neutral or negative ratings. In a classroom deployment, this pattern is plausible because higher exposure increases opportunities to notice limitations (e.g., UX friction, latency, or occasional mismatch with expectations) while still valuing the core scaffolded support.

Table 3 provides complementary evidence from open-ended comments. Requests related to UI/usability and more examples were common across all groups but were most frequent among high-engagement students. This suggests that students who used the tool most often were also the most specific about how to improve it, which is useful for iterative refinement after initial deployment.

Table 2: Survey perceptions by engagement level (matched respondents, $n = 54$). Values are percentages within each engagement group.

Measure (Positive/Neutral/Negative)	**Low**	**Medium**	**High**
Guided reasoning helpfulness	100.0/0.0/0.0	94.1/0.0/5.9	94.1/0.0/5.9
Overall effectiveness vs usual methods	75.0/10.0/15.0	52.9/17.6/29.4	52.9/29.4/17.6
Likelihood of recommending platform	95.0/0.0/5.0	88.2/0.0/11.8	100.0/0.0/0.0

Table 3: Frequency of selected comment themes by engagement level (matched respondents, $n = 54$). Values are percentages within each group.

Theme	**Low**	**Medium**	**High**
Process-oriented guidance	45.0	58.8	70.6
Persistence support	0.0	11.8	17.6
UI/usability improvement requests	45.0	41.2	64.7
Requests for more examples	45.0	47.1	58.8
Speed/responsiveness requests	15.0	29.4	29.4

Qualitative Insights from Open-Ended Feedback

Open-ended survey responses ($N = 86$) provide additional context for what students would keep the same or improve. Figure 10 summarizes the distribution of improvement themes. The most frequent response was *nothing / keep the same* (about 43%), indicating broad satisfaction among respondents who chose to comment. The next most common theme focused on *UI/layout/usability* (about 36%), followed by requests for *more examples/explanations* (about 15%). Smaller proportions mentioned *speed/performance* (about 14%) and *accuracy/correctness* (about 12%), while *personalization/adaptivity* was rarely requested.

Table 4 provides representative comments illustrating how students experienced the platform in practice. Many comments emphasize process-oriented support (understanding intermediate steps) and persistence support (continuing when stuck), which triangulates with the quantitative findings on guided reasoning (Figure 3) and the revision patterns seen in log data (Figure 9).

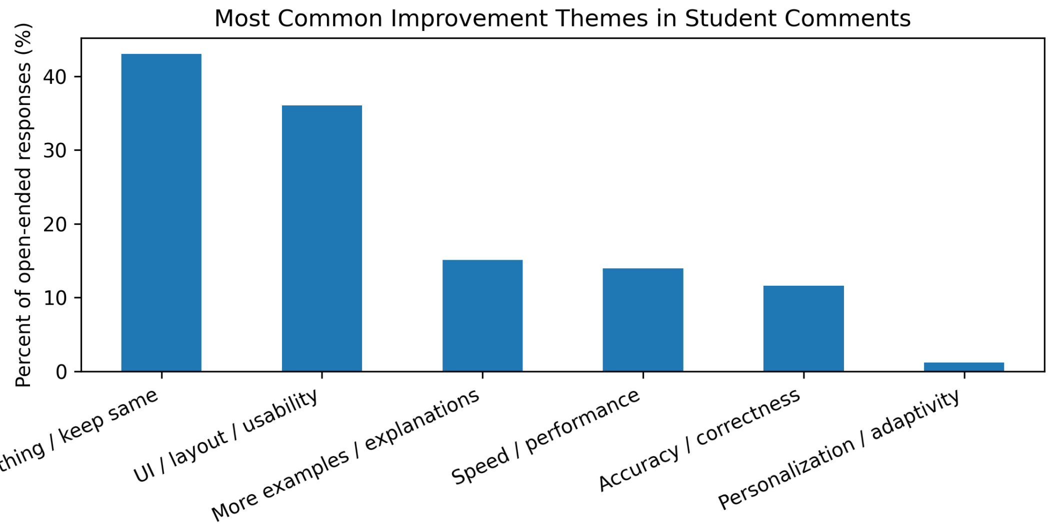


Figure 10: Distribution of improvement themes in open-ended student feedback ($N = 86$).

Table 4: Summary of qualitative themes identified from open-ended student responses.

Theme	**Representative Student Comment**
Process-oriented guidance	"It helped me understand the steps instead of just giving the answer."
Persistence support	"It kept me going when the problems were hard."
Non-intrusive support	"I could choose when to use it, which I liked."

Synthesis Across Data Sources

Across survey responses, interaction logs, and open-ended feedback, results consistently suggest that students used the platform as a scaffold for reasoning rather than a shortcut for answer retrieval. Survey responses indicate positive perceptions of guided reasoning support and generally favorable views of accuracy and instructional alignment (Figures 3–6). Log data show multi-turn sessions and revision-oriented sequences following hints (Figures 8 and 9), consistent with iterative engagement. Finally, qualitative comments suggest that improvement requests centered more on interface and examples than on rejecting the platform's pedagogical approach. Together, these findings provide classroom-grounded evidence that a constrained, scaffolded AI design can support students' problem-solving processes in a large-enrollment course.

Discussion

This study examined the use of an AI-guided learning platform in a large-enrollment undergraduate Discrete Mathematics course at a public R1 university. Rather than centering on direct measures of learning outcomes, the discussion focuses on how students perceived the platform, how they engaged with its guidance during problem solving, and how it fit within their broader study practices. Drawing on survey responses, system interaction logs, and open-ended feedback, this section interprets the results in relation to the research questions outlined earlier and considers what these findings suggest for the careful and responsible integration of AI-supported scaffolding in computing and mathematically intensive courses.

RQ1: Supporting Reasoning Without Direct Answer Provision

Research Question 1 focused on whether students experienced the AI-guided scaffolding as supportive of reasoning and conceptual understanding in the absence of direct answer provision. The findings provide consistent evidence that students perceived the platform's guidance as helpful despite, and in some cases because of, its deliberate avoidance of final answers. Survey results show that a clear majority of students rated the step-by-step prompts as helpful or very helpful, indicating that students valued guidance that emphasized how to approach problems rather than what the final solution should be.

Qualitative feedback further clarifies how students interpreted this form of support. Many comments highlighted the platform's role in helping students work through intermediate reasoning steps and clarify their thinking, rather than serving as a source of completed solutions. This suggests that students recognized and accepted the system's instructional boundaries, engaging with it as a reasoning aid rather than an answer-generating tool.

The presence of neutral responses across several survey items warrants careful interpretation. Rather than indicating dissatisfaction, these responses appear to reflect selective and context-dependent use of the platform. Students reported turning to the system when they encountered difficulty and relying less on it when they felt confident in their understanding. Such patterns are consistent with self-regulated learning, in which learners actively decide when external support is necessary and when independent problem solving is sufficient.

Taken together, these findings indicate that the platform was perceived as supporting reasoning without displacing student responsibility for solution development. By providing structured prompts while withholding final answers, the system appears to have functioned as a cognitive scaffold that preserved learner agency and was experienced by students as supporting engagement with the problem-solving process.

RQ2: Evidence of Iterative and Process-Oriented Engagement

Research Question 2 examined whether students' observed interactions with the platform reflected iterative, process-oriented engagement rather than one-time answer seeking. The interaction log data provides converging evidence that student use of the system commonly involved multiple exchanges within a session. With a median of three interactions per session and an average exceeding four interactions, students frequently engaged in more than a single submission or query before concluding a session.

Beyond overall interaction counts, the distribution of session lengths indicates that many students returned to the system repeatedly within the same problem-solving episode. While shorter sessions were common, a substantial number of sessions extended across several turns, suggesting that students were using the platform to refine their thinking over time rather than to obtain immediate confirmation or closure. This pattern is notable in the context of a large-enrollment course, where opportunities for sustained, individualized interaction are typically constrained.

These observed behaviors suggest that students engaged with the platform through interactive exchanges during problem solving, revisiting prompts and responding to feedback across multiple steps. Rather than treating the system as a transactional resource, students' interaction patterns reflect continued engagement with the problem-solving process itself. Taken together, the log data indicate that the platform supported iterative engagement in practice, aligning with its design goal of emphasizing process-oriented interaction over single-turn answer delivery.

RQ3: Persistence, Error Recovery, and Productive Struggle

Research Question 3 focused on whether the platform supported persistence and error recovery during challenging problem-solving situations. Evidence from both interaction logs and student self-reports indicates that many students engaged with the system in ways consistent with continued effort following difficulty. A substantial proportion of sessions included the use of intermediate scaffolding, with approximately one third of sessions involving at least one hint request and over 40 percent involving step-level guidance. These patterns suggest that students frequently sought support during the problem-solving process rather than relying solely on final answer submission.

More detailed examination of interaction sequences provides further insight into how students responded to difficulty. A notable share of sessions followed structured trajectories such as Attempt followed by Hint followed by Attempt, as well as Incorrect followed by Hint followed by Correct. These sequences indicate that students often used system feedback to revise their work after encountering errors, rather than disengaging or terminating the session. The presence of these patterns across sessions suggests that the platform supported continued engagement after initial failure. Student survey responses and open-ended feedback align with the behavioral evidence. Many students reported that the platform helped them continue working through problems they found challenging. Qualitative comments frequently referenced sustained effort and the ability to regroup after mistakes, particularly when guidance was available at intermediate steps.

Taken together, the results indicate that the platform supported persistence and error recovery during problem solving by offering guidance at points of difficulty while maintaining student responsibility for revising and completing solutions. These findings are consistent with characteristics of productive struggle, in that students persisted through difficulty while retaining responsibility for revising and completing their solutions.

RQ4: Integration Into Existing Study Practices

Research Question 4 examined how students positioned the platform within their broader study practices and whether it was perceived as a complement to, rather than a replacement for, traditional learning resources. Survey results indicate that most students rated the platform as *somewhat effective* or *very effective*, with a substantial neutral group and relatively few students rating it as ineffective. This response pattern suggests that students did not perceive the platform as competing with established resources such as textbooks, lecture notes, or tutoring.

Open-ended feedback provides additional detail about how students incorporated the platform into their learning routines. Many students described using the system selectively, particularly when working through challenging steps of a problem, while continuing to rely on traditional materials for review, practice, or confirmation of understanding. These descriptions indicate that students integrated the platform alongside existing resources rather than substituting it for them.

Patterns observed across engagement levels further support this interpretation. Even among students who interacted frequently with the system, perceptions of the guided prompts remained positive, and critical feedback focused primarily on interface-related issues or requests for additional examples. Concerns about overreliance or inappropriate substitution of study practices were notably absent from student comments.

Taken together, these findings suggest that the platform was incorporated into students' study practices as a supplementary resource that supported problem solving without displacing other forms of instruction or study.

Implications for Responsible AI Integration in Large Classes

The findings across the four research questions highlight the central role of intentional pedagogical design in AI-supported learning environments. Grounding AI responses in instructor-curated materials, enforcing progressive scaffolding, and requiring observable student engagement before advancing support enabled the system to provide individualized guidance at scale without encouraging answer-seeking behavior.

From a Research-to-Practice perspective, this work illustrates how AI systems can be incorporated into large-enrollment courses in ways that reinforce reasoning, persistence, and learner agency. When designed with clear instructional constraints, AI tools can operate as structured learning partners rather than automated problem solvers. The results offer practical guidance for instructors who wish to adopt AI technologies while maintaining instructional alignment, supporting productive student engagement, and upholding academic integrity.

Limitations and Future Directions

This study has several limitations that point to important directions for future work. First, the reported findings emphasize student perceptions and observed interaction patterns rather than direct measures of learning outcomes. While engagement, persistence, and error recovery are well-established precursors to learning, future studies should more directly examine how AI-guided scaffolding influences conceptual understanding and performance over time, including impacts on course assessments and longer-term retention.

Second, the study was conducted within a single Discrete Mathematics course at one research-intensive institution, which may limit generalizability. Replication across courses, disciplines, and institutional contexts will be necessary to assess the transferability of the observed patterns and to understand how course structure and student populations shape the use of scaffolded AI support.

Finally, although the platform is instrumented to capture fine-grained interaction data, the present analysis focused on aggregate and session-level metrics to prioritize classroom deployment and student experience. Future work will explore richer behavioral signals from interaction logs, such as temporal patterns of hint usage, revision sequences following feedback, and changes in response latency within sessions, to more deeply characterize how students engage in scaffolded reasoning.

Conclusion

This work examined how an interactive, AI-guided learning platform functioned in a large-enrollment Discrete Mathematics course, with a focus on supporting reasoning, persistence, and learner autonomy. Survey results, interaction logs, and qualitative feedback together indicate that students valued step-by-step guidance that emphasized problem-solving processes rather than direct answer generation. Behavioral evidence showed iterative engagement and productive use of hints, suggesting that the system supported persistence and recovery from difficulty without encouraging shortcut behavior.

Importantly, students perceived the platform as a complementary learning resource rather than a replacement for existing study practices, reflecting its non-intrusive and flexible design. Neutral survey responses were interpreted as selective engagement consistent with self-regulated learning, highlighting learner agency rather than disengagement. Taken together, these findings demonstrate that intentionally designed AI systems can be integrated responsibly into large courses to reinforce reasoning and metacognitive engagement. Future work will explore longitudinal impacts on learning outcomes, refine personalization strategies, and examine transferability across disciplines.

References

[1] E. Kasneci, K. Seßler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, and G. Groh, “Chatgpt for good? on opportunities and challenges of large language models for education,” *Learning and Individual Differences*, vol. 103, p. 102274, 2023.

[2] M. N. Giannakos, R. Azevedo, M. Siadaty, and P. Dillenbourg, “The promise and challenges of generative artificial intelligence in education,” *Behaviour & Information Technology*, 2024.

[3] P. Denny, S. Gulwani, N. T. Heffernan, T. Käser, S. Moore, A. N. Rafferty, and A. Singla, “Generative ai for education (gaied): Advances, opportunities, and challenges,” *arXiv preprint arXiv:2402.01580*, 2024.

[4] I. Joshi, R. Budhiraja, H. Dev, J. Kadia, M. O. Ataullah, S. Mitra, D. Kumar, and H. D. Akolekar, “Chatgpt in the classroom: An analysis of its strengths and weaknesses for solving undergraduate computer science questions,” in *Proceedings of the ACM Technical Symposium on Computer Science Education (SIGCSE TS)*. ACM, 2024.

[5] D. Wood, J. S. Bruner, and G. Ross, "The role of tutoring in problem solving," *Journal of Child Psychology and Psychiatry*, vol. 17, no. 2, pp. 89–100, 1976.

[6] K. VanLehn, "The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems," *Educational Psychologist*, vol. 46, no. 4, pp. 197–221, 2011.

[7] B. J. Zimmerman, "Becoming a self-regulated learner: An overview," *Theory Into Practice*, vol. 41, no. 2, pp. 64–70, 2002.

[8] P. R. Pintrich, "The role of goal orientation in self-regulated learning," in *Handbook of Self-Regulation*, M. Boekaerts, P. R. Pintrich, and M. Zeidner, Eds. Academic Press, 2000, pp. 451–502.

[9] P. H. Winne and A. F. Hadwin, "Studying as self-regulated learning," in *Metacognition in Educational Theory and Practice*, D. J. Hacker, J. Dunlosky, and A. C. Graesser, Eds. Lawrence Erlbaum Associates, 1998, pp. 277–304.

[10] M. Kapur, "Productive failure in learning the concept of variance," *Instructional Science*, vol. 40, no. 4, pp. 651–672, 2012.

[11] K. Loibl and N. Rummel, "Knowing what you don't know makes failure productive," *Learning and Instruction*, vol. 34, pp. 74–85, 2014.

[12] Y. Gao, Y. Xiong, X. Gao, K. Jia, J. Pan, Y. Bi, Y. Dai, J. Sun, M. Wang, and H. Wang, "Retrieval-augmented generation for large language models: A survey," *arXiv preprint arXiv:2312.10997*, 2023.

[13] Z. Li, Z. Wang, W. Wang, K. Hung, H. Xie, and F. L. Wang, "Retrieval-augmented generation for educational applications: A systematic survey," *Computers and Education: Artificial Intelligence*, vol. 8, p. 100417, 2025.

[14] S. Saha and M. Farooque, "Enhancing student engagement through example-based short videos," in *2025 IEEE Frontiers in Education Conference (FIE)*. IEEE, 2025, pp. 1–9.

[15] M. T. H. Chi, "Active-constructive-interactive: A conceptual framework for differentiating learning activities," *Topics in Cognitive Science*, vol. 1, no. 1, pp. 73–105, 2009.

[16] J. R. Anderson, A. T. Corbett, K. R. Koedinger, and R. Pelletier, "Cognitive tutors: Lessons learned," *Journal of the Learning Sciences*, vol. 4, no. 2, pp. 167–207, 1995.

[17] A. T. Corbett, K. R. Koedinger, and J. R. Anderson, "Cognitive tutoring: Using intelligent tutors to improve instruction," *Educational Psychologist*, vol. 36, no. 4, pp. 235–243, 2001.

[18] V. Aleven, B. M. McLaren, I. Roll, and K. R. Koedinger, "Help helps, but only so much: Research on help seeking with intelligent tutoring systems," *International Journal of Artificial Intelligence in Education*, vol. 26, no. 1, pp. 205–223, 2016.

[19] I. Roll, V. Aleven, B. M. McLaren, and K. R. Koedinger, "Can help seeking be tutored? searching for the secret sauce of metacognitive tutoring," in *Proceedings of the International Conference on Artificial Intelligence in Education (AIED)*. IOS Press, 2007, pp. 203–210.

[20] R. Baker, A. T. Corbett, and V. Aleven, "More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing," in *Proceedings of the International Conference on Intelligent Tutoring Systems*, 2008.

[21] R. Luckin and B. Du Boulay, "Ecolab: The development and evaluation of a vygotskian design framework," *International Journal of Artificial Intelligence in Education*, vol. 10, pp. 198–220, 1999.

[22] H. A. Wood and D. Wood, "Help seeking, learning and contingent tutoring," *Computers & Education*, vol. 33, no. 2-3, pp. 153–169, 1999.

[23] V. Aleven, B. McLaren, I. Roll, and K. Koedinger, "Toward meta-cognitive tutoring: A model of help seeking with a cognitive tutor," *International Journal of Artificial Intelligence in Education*, vol. 16, pp. 101–128, 2006.

[24] S. Saha, F. Rahbari, F. Sadique, S. K. C. Velamakanni, M. Farooque, and W. J. Rothwell, "Next-gen education: Enhancing ai for microlearning," in *2025 ASEE Annual Conference & Exposition*, 2025.

[25] M. I. H. Shihab, C. Hundhausen, A. Tariq, S. Haque, Y. Qiao, and B. Mulanda, “The effects of github copilot on computing students’ programming effectiveness, efficiency, and processes in brownfield programming tasks,” in *Proceedings of the ACM Technical Symposium on Computer Science Education (SIGCSE TS)*. ACM, 2025.

[26] W. Petula, A. Joshi, P. Tu, A. Somasundar, and S. Saha, “Deliverc: Teaching pointers through genai-powered game-based learning,” in *Proceedings of the 57th ACM Technical Symposium on Computer Science Education V. 1*, 2026, pp. 838–844.

[27] I. Roll, V. Aleven, B. M. McLaren, and K. R. Koedinger, “Improving students’ help-seeking skills using metacognitive feedback in an intelligent tutoring system,” *Learning and Instruction*, vol. 21, no. 2, pp. 267–280, 2011.