2026 ASEE ANNUAL CONFERENCE June 21–24, 2026 | Charlotte Convention Center, NC
Host Sponsor UNC Charlotte
Paper ID #51654

Generative AI as a Reflection Partner in Quantitative Engineering Education: A Pilot Implementation in a Construction Equipment Course

Dr. Tianjiao Zhao, East Carolina University

Dr. Tianjiao Zhao is an Assistant Professor in the Department of Construction Management at East Carolina University, USA. Her research focuses on integrating artificial intelligence, visualization, and human–computer interaction to improve efficiency, safety, and sustainability in construction management. She is also committed to incorporating emerging technologies into construction education to support data-driven decision-making and workforce development.

Dr. Ri Na, University of Delaware

Ri Na is an Associate Professor in the Department of Civil, Construction, and Environmental Engineering at the University of Delaware. Her teaching and research focus on construction education, digital technologies, and improving student learning through tools such as BIM, AR/VR, and 3D printing.

Xi Lin, East Carolina University

Dr. Xi Lin is an associate professor at East Carolina University, US. Her research focuses on seeking best practices to enhance student engagement and interaction in online learning environments. More information can be found at http://whoisxilin.weebly.com/

Dr. Xi Wang, Drexel University

Dr. Xi Wang is an Associate Teaching Professor at Drexel University. She received her Ph.D. and M.Eng both in Civil Engineering, from the University of Kentucky and Auburn University. She is licensed as a Professional Engineer and LEED Green Associate. She is teaching a range of courses in construction management and will be assisting capstone design projects that directly serve regional construction firms. Her research interests include technology adoption in workforce development in the construction industry, sustainable developments in construction education, and learning motivation for student success in engineering education.

Chen Xia, Western New England University

Generative AI as a Reflection Partner in Quantitative Engineering Education: A Pilot Implementation in a Construction Equipment Course

Abstract

Complex calculations such as earth volume estimation and equipment productivity analysis (e.g., shovel or excavator cycles) often challenge construction management students due to multi-step reasoning, parameter coordination, and unit consistency demands. This study reports a pilot instructional implementation in which generative AI tools (such as Copilot) were positioned as "reflection partners" to support students' step-by-step reasoning and self-diagnostic practices in a Construction Equipment course (N = 33). Students first completed two problem sets—one focusing on earthwork volume and another on equipment productivity—without AI assistance. Afterward, they optionally engaged the AI for diagnostic feedback on reasoning steps, assumptions, and improvement recommendations. Students then revised their work, reflected on the AI feedback, and completed a short post-activity survey. Assignment-level descriptive analyses showed larger observed score shifts in higher-step tasks, while qualitative reflections suggested that many students used AI as a logic-checking and self-explanation scaffold rather than an answer generator. Students also reported a stronger sense of confidence and awareness of how they learn best. As a pilot classroom implementation, the findings are presented as instructional design insights grounded in observed learning patterns rather than as experimental claims of AI impact. While demonstrated in a construction context, this study offers a low-cost instructional framework for integrating generative AI into quantitative engineering education in ways that emphasize metacognitive monitoring and responsible use.

Keywords:
Artificial Intelligence (AI); Generative AI; Metacognition; Quantitative Engineering Education; Reflective Learning; Self-Assessment

Introduction

Quantitative problem-solving is a core competency in construction management education, yet many students struggle with multi-step calculations that require careful parameter selection, unit consistency, and logical sequencing. In a Construction Equipment course, two commonly challenging topics are (1) earthwork haul distance estimation using mass ordinates and mass diagrams and (2) equipment productivity analysis across different equipment types (e.g., dozers, compactors, excavators/hoes, scrapers, and truck hauling cycles). These tasks are cognitively demanding not merely because they involve formulas, but because they require students to coordinate multiple interdependent steps, track assumptions, and detect small errors that can cascade into major numerical discrepancies [1].

At the same time, the rapid adoption of generative AI tools (e.g., ChatGPT, Microsoft Copilot) has created both opportunities and concerns in engineering education. While generative AI can provide instant explanations and feedback, instructors worry about overreliance, academic integrity risks, and students substituting AI output for genuine understanding [2], [3]. A practical pathway that can reduce these risks is to position AI not as an "answer engine," but as a reflection partner that helps students articulate their reasoning, identify missing steps or flawed assumptions, and practice self-checking behaviors aligned with metacognition and self-regulated learning [4], [5].

This paper reports a pilot instructional implementation in which students completed an initial attempt of quantitative assignments without AI assistance, then were offered an optional "second chance" submission after using generative AI as a structured reflection partner. The instructional goal was to maintain student agency and honesty while encouraging deliberate practice: students were prompted to explain their steps, request diagnostic feedback on errors and assumptions, and answer short reflection questions to consolidate learning. This study contributes a practical, low-risk instructional design for integrating generative AI into quantitative engineering education, with empirical insights into how task complexity mediates observed learning gains.

Research Questions

To maximize practical relevance while avoiding over-claiming causality in a single-course pilot, we focus on descriptive and design-oriented research questions:

- RQ1 (Performance Patterns): How do score patterns differ between first attempts and optional AI-supported second attempts, especially across assignments with varying step complexity?
- RQ2 (Reflection Experience): How do students describe the role of AI in supporting their reasoning, error detection, and confidence when used as a reflection partner?
- RQ3 (Design Implications): What low-cost, instructor-friendly design elements (e.g., prompts, grading policy, scaffolding) appear to support responsible AI use in quantitative coursework?

Literature Review

Generative AI in Engineering and Higher Education

Early research on generative AI in education highlights both promise and caution. Large language models can support explanation, feedback, and personalized guidance, but they also introduce risks related to hallucinations, bias, and misuse for shortcutting learning [2]. Policy-facing guidance emphasizes the need for human-centered integration, transparency, and capacity-building for educators and students [3], [6], [7], [8], [9], [10]. In practice, effective classroom use tends to involve structured roles for AI (e.g., tutor, coach, mentor), paired with explicit expectations that students remain "human-in-the-loop" reviewers of outputs [10], [11], [12].

Metacognition, Self-Regulated Learning, and Reflective Practice

Metacognition refers to awareness and regulation of one's thinking processes—planning, monitoring, and evaluating understanding [4]. Self-regulated learning extends this view by describing how learners set goals, deploy strategies, and monitor progress [5]. Reflective practice literature further emphasizes learning through iterative reflection on action, particularly in professional domains where complex judgment and assumptions matter [13], [14]. Together, these frameworks suggest that students learn more robustly when they engage in explanation, self-checking, and error diagnosis—not only when they arrive at correct answers.

Feedback, Worked Examples, and Cognitive Load in Quantitative Learning

Feedback is widely recognized as a powerful influence on learning outcomes, but its effectiveness depends on timing, content, and student engagement in interpreting the feedback [15]. Formative assessment research highlights "good feedback practice" that helps learners self-assess, clarify goals, and close gaps between current and desired performance [16]. In quantitative problem-solving, worked examples can reduce extraneous cognitive load and help students acquire schemas before engaging in independent problem solving [17]. Moreover, self-explanation—having students generate explanations while studying solution steps—has been shown to distinguish deeper learning from superficial answer-following behaviors [18].

Positioning AI as a Reflection Partner

A synthesis of the above suggests a theoretically grounded, low-risk instructional approach: use AI to scaffold self-explanation, error detection, and reflective monitoring, rather than to provide final answers. Conceptually, AI can serve as a "diagnostic mirror" that prompts students to check assumptions, units, and step logic, while the instructor sets boundaries and incentives to preserve learner agency. This aligns with structured "AI-coach/AI-tutor" pedagogies that emphasize student oversight and reflective questioning [8], [11] and with broader guidance calling for responsible implementation and verification practices [3].

Methodology

Course Context and Participants

The pilot was implemented in a Construction Equipment course (N = 33 enrolled students) covering earthwork estimation and equipment productivity analysis across multiple equipment types. The instructional intervention was intentionally designed to integrate into the normal course flow and grading structure without introducing additional required assignments. Figure 1 illustrates the overall workflow of the intervention, including an initial independent submission, AI-supported reflection using structured prompts, an optional revision opportunity, and the final recorded score. This study was approved by the Institutional Review Board (IRB) of the authors' university.

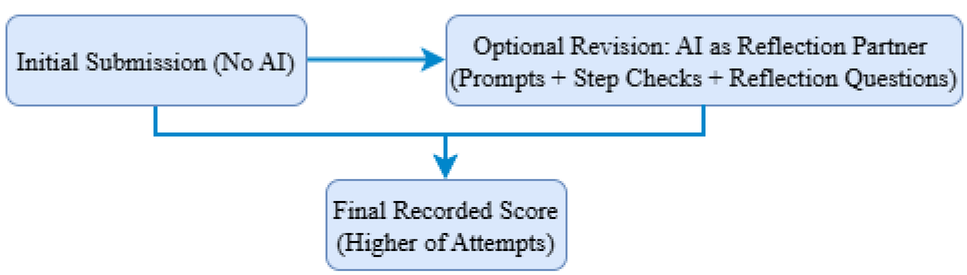

Figure 1. Instructional workflow of the AI-as-Reflection-Partner intervention

Assignment Design and Step Complexity

To reduce procedural barriers, the instructor provided simplified step-by-step instructions derived from textbook examples and lecture materials. The assignments varied in step complexity:

- **HW1 (Earthwork/haul distance calculation):** classic 12-column mass ordinate calculation table → mass diagram → detailed 6-step haul distance calculation (overall perceived as highly complex due to setup + multiple representations).
- **Productivity assignments:**
    - HW2 (Dozer): 7 steps
    - HW3 (Compactor): 8 steps
    - HW4 (Hoe/Excavator cycle): 7 steps
    - HW5 (Scraper): 14 steps
    - HW6 (Truck): 14 steps

This variation enabled a practical exploration of whether AI-supported reflection appears more beneficial when tasks involve longer reasoning chains and higher cognitive load [1].

AI "Reflection Partner" Scaffolding and Prompting

The instructor explicitly modeled how to use AI as a reflection partner rather than an answer generator. Based on the information provided by the students, the instructor used ChatGPT and Copilot, the AI tools most frequently used by the students, to demonstrate prompt engineering. Students were also free to choose other generative AI tools for their actual use. Two structured prompts were demonstrated in class (abridged for paper length, full prompt text is provided in Appendix A for replication.):

Prompt 1 (Guided walkthrough + checks for understanding):

Students submitted the course instructional materials and asked the AI to walk through the example step-by-step, explain what each step means and why it is needed, highlight common mistakes, and ask short comprehension questions before proceeding.

Prompt 2 (Diagnostic feedback on the student's own work):

After uploading their attempted solution, students asked the AI to identify missing steps, parameter errors, or unit mistakes; suggest revisions to improve clarity of reasoning; and provide 1–2 reflection questions to reveal misconceptions. The tone was specified as supportive, like a teaching assistant.

We analyzed 23 student reflections using an inductive thematic approach [19], iteratively refining codes and consolidating themes through repeated reading. To enhance trustworthiness, we used negative-case checking (for example, actively searching for reflections that contradicted dominant themes). This prompting strategy leverages principles of self-explanation and formative feedback [16], [18].

Grading Policy to Reduce Pressure and Encourage Honesty

To lower anxiety and reduce incentives for dishonest shortcutting, the "second chance" submission was an optional bonus opportunity rather than a requirement. When students submitted a second attempt, the course policy recorded the higher of the two scores for the final grade. This design sought to preserve students' sense of autonomy and to frame AI use as a learning support tool, while still encouraging engagement through a meaningful incentive.

Data Sources

1. **Performance data:**
   Student performance data were summarized at the assignment level for both earthwork (HW1) and productivity calculations (HW2–HW6). For each assignment, the instructor recorded initial submission scores and final scores, which reflect the highest score achieved across submissions. Student-level paired scores were retained for each assignment, enabling both assignment-level descriptive summaries and paired-sample inferential testing (t-tests and effect sizes) to examine performance shifts between initial and final submissions.
2. **Student reflections on AI use:**
   A short post-activity survey question was administered. Student written reflections were collected through a course report question (Question 8). The reflection corpus included varied contexts; for this paper, reflections were de-identified and lightly edited to align with the equipment/productivity learning context (e.g., references to "tables," "rates," or "workflow" were reframed to "productivity steps," "productivity rates," "unit checks," etc.) while preserving the original intent and stance (e.g., used AI vs. did not use AI). Representative excerpts are presented as anonymized quotes.

Analysis Approach

- **Quantitative:** Analyses included assignment-level descriptive statistics (means, standard deviations, medians, and distributional patterns) as well as paired-sample t-tests and associated effect sizes (Cohen's d^z) to examine shifts between initial and final submissions. Given the classroom-embedded pilot design, inferential results are interpreted as exploratory rather than causal.
- **Qualitative:** thematic analysis using an inductive approach with light alignment to metacognition and feedback frameworks [19]. Themes were identified by repeated reading and grouping statements describing how AI was used, perceived benefits/risks, and verification behaviors.

Results

RQ1: Performance Patterns Across Assignments

HW1 (Earthwork/haul distance) — largest observed shift and variance reduction

HW1 exhibited the largest shift in both central tendency and score dispersion, consistent with its multi-representation and multi-step workflow (mass ordinate table → mass diagram → haul-distance computation). Table 1 summarizes the assignment-level performance statistics for HW1, showing a substantial increase in mean and median scores from the initial submission to the final recorded score. Figure 2 further illustrates this distributional shift, with scores becoming more tightly clustered after the AI-supported reflection activity.

Table 1. Summary Statistics for HW1 Earthwork Assignment (Percent Scores)

Statistic	**Initial Submission**	**Final Recorded Score**	**Improvement (Final – Initial)**
Mean (%)	64.73	80.14	+15.41
Standard Deviation	23.22	9.40	—
Median (%)	62.50	80.00	+11.30
Minimum (%)	0.00	58.40	—
Maximum (%)	100.00	100.00	—

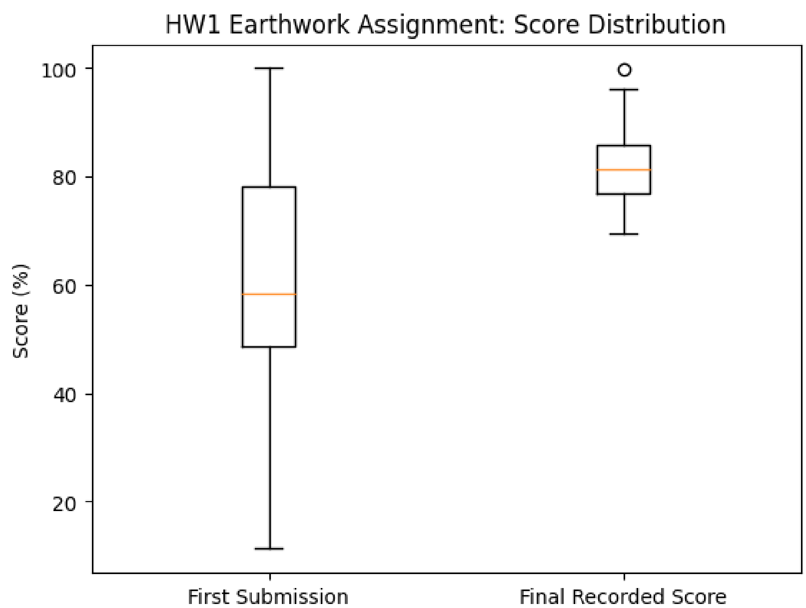

Figure 2. Distribution of HW1 earthwork assignment scores before and after the AI-supported reflection activity (percent scale)

A moderate paired-change effect was further observed for HW1 (Cohen's $d_z = 0.598$). A paired-sample t-test indicated that the increase from initial to final scores was statistically significant ($t_{(32)} = 3.43$, $p = .002$), accompanied by a notable reduction in score dispersion. Beyond improvement in central tendency, the narrowing of the score distribution suggests that structured reflection and error-checking may help reduce large procedural errors in complex, multi-step calculations *[1], [16]*. Because this implementation occurred within the normal progression of the course, the observed improvements likely reflect a combination of factors, including structured reflection, renewed attention to instructor feedback, and extended engagement with the material. Within this broader instructional context, the AI-supported reflection component can be understood as one contributing element that may have supported students' diagnostic and monitoring processes.

HW2–HW6 (Productivity assignments) — improvement concentrated in high-step tasks

Improvements were non-uniform across productivity assignments and aligned closely with step complexity. Assignments with longer reasoning chains (HW5–HW6, 14 steps each) exhibited markedly larger gains than shorter, more streamlined workflows (HW2–HW4, 7–8 steps). Table 2 summarizes assignment-level performance statistics, including mean scores before and after the reflection activity, expressed as percentages of the maximum possible score.

While shorter workflows already showed relatively high initial performance, the larger improvements observed in high-step assignments suggest that AI-supported reflection may be particularly beneficial when students must coordinate numerous interdependent steps, assumptions, and unit conversions within a single problem.

Table 2. Initial vs. Final Performance by Assignment (Percent Scores)

Assignment	**Topic**	**Steps**	**Max points**	**Initial score mean (sd)**	**Final score mean (sd)**	**Mean change (pp)**	**t(32)**	***p***	**Cohen's d^z**
HW2	Dozer productivity	7	30	89.85 (23.41)	92.01 (19.96)	+2.16	1.35	= .187	0.235
HW3	Compactor productivity	8	20	93.92 (18.82)	94.50 (14.80)	+0.58	1.00	= .325	0.174
HW4	Hoe/Excavator productivity	7	20	86.61 (15.91)	89.85 (13.86)	+3.24	3.83	< .001	0.667
HW5	Scraper productivity	14	60	75.10 (28.81)	86.85 (22.52)	+11.75	3.50	= .001	0.609
HW6	Truck productivity	14	20	75.85 (16.82)	87.86 (12.73)	+12.01	5.06	< .001	0.880

Improvements across HW2–HW6 were non-uniform and appeared to align with procedural step complexity. Paired-sample t-tests indicated statistically significant gains for higher-complexity assignments (HW4, HW5, and HW6), while lower-step assignments (HW2 and HW3) did not demonstrate statistically significant shifts. Specifically, HW4 showed a moderate effect ($t_{(32)}$ = 3.83, $p < .001$, $d^z = 0.667$), HW5 demonstrated a moderate effect ($t(32) = 3.50$, $p = .001$, d^z = 0.609), and HW6 exhibited a large effect ($t_{(32)}$ = 5.06, $p < .001$, $d^z = 0.880$). In contrast, HW2

($t_{(32)} = 1.35$, $p = .187$, $d_z = 0.235$) and HW3 ($t_{(32)} = 1.00$, $p = .325$, $d_z = 0.174$) did not show statistically significant paired improvements.

This pattern reinforces the descriptive alignment between step complexity and observed gains. Assignments requiring longer procedural chains and sustained coordination of intermediate variables exhibited stronger paired effects, whereas shorter workflows with already high initial performance levels showed limited change. While exploratory in nature, these results further support the interpretation that structured diagnostic reflection may be particularly impactful in higher-complexity quantitative tasks.

To further examine whether improvement patterns aligned with task complexity, Figure 3 visualizes the relationship between the number of calculation steps and mean score improvement across productivity assignments.

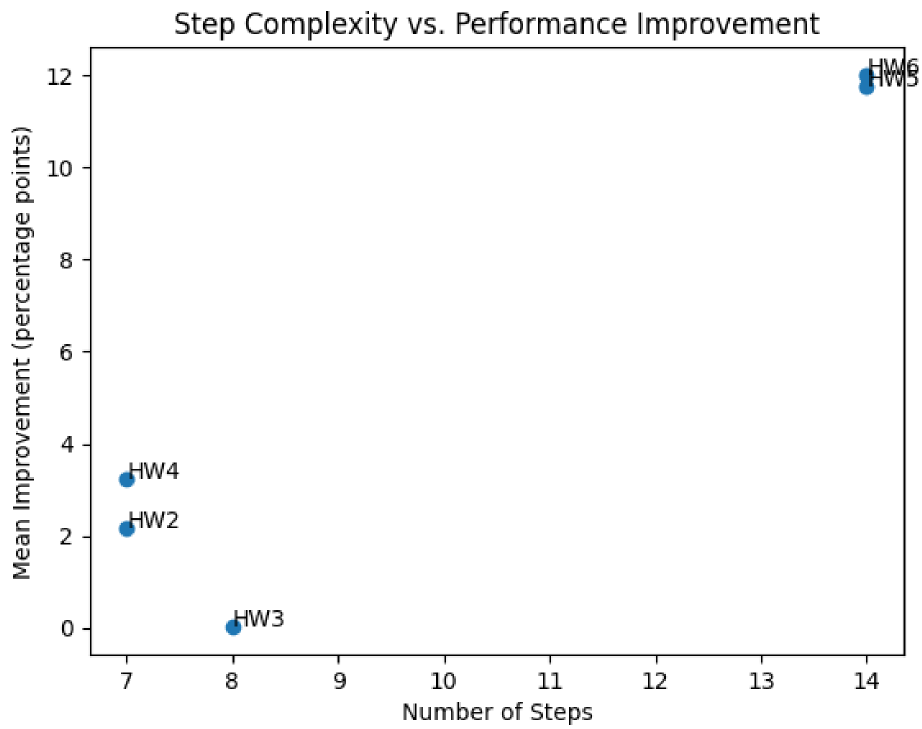

Figure 3. Relationship between assignment step complexity and mean improvement from initial to final recorded scores (percentage points)

RQ2: Student Reflections (Themes)

A total of 23 student reflections were analyzed using an inductive thematic approach. Of these, 18 students reported using generative AI during the assignment revision process, while 5 reported no AI use and relied primarily on independent manual problem solving. Across reflections, students described a spectrum of AI engagement, ranging from targeted diagnostic use to deliberate non-use.

Theme 1 — Error Identification and Debugging ("finding where I went wrong")

Many students described AI as particularly useful for identifying missing steps, incorrect parameter selections, or unit conversion errors in multi-step productivity calculations. Students noted that explaining their reasoning to AI helped surface subtle mistakes that were difficult to detect when focusing only on final numerical results.

- “After I wrote out my steps for the scraper productivity, the AI pointed out I mixed minutes and hours in one sub-step. I would not have noticed because the final number still looked reasonable.”
- “It helped me check cycle time assumptions and spot which step I copied from the example incorrectly.”

This theme aligns with formative feedback practices that help learners identify gaps and revise strategies [16].

Theme 2 — Enhanced Step-by-Step Reasoning and Self-Explanation

Students frequently emphasized that articulating each calculation step to AI forced them to organize their thinking more clearly. This process promoted deeper engagement with the logic of productivity formulas and reinforced understanding of why specific steps were required, rather than merely how to execute them.

- “When I had to explain each step, I realized I couldn’t justify one of my conversion factors. Writing it out made me catch it.”
- “The AI didn’t just give the result—I used it to walk through each step like a TA and answer quick check questions.”

This resembles self-explanation effects reported in worked-example literature [17], [18].

Theme 3 — Confidence and reduced anxiety (without removing responsibility)

Several reflections indicated that AI functioned as a second-opinion tool that increased confidence without replacing student responsibility. Students commonly reported attempting the problem independently first, then using AI to verify assumptions, check logic, and confirm consistency with lecture materials.

- “I tried the problem first, then used AI to confirm I wasn’t missing a step. It helped me feel more confident before submitting.”
- “It’s like tutoring—useful when you’re stuck, but you still have to understand it.”

This matches the intended “reflection partner” framing and aligns with responsible use guidance emphasizing student agency and oversight [3], [10].

Theme 4 — Learner Agency, Verification, and Preference for Independent Problem Solving

In contrast to students who actively engaged AI as a reflection partner, a minority expressed skepticism or a preference for independent problem solving. One student noted that they “did not use any AI tools for this assignment” and felt that working through the steps independently reinforced their understanding. Another student emphasized that AI “can be wrong” and therefore chose to verify all steps manually or avoid AI altogether. These responses highlight that AI was not universally perceived as necessary or beneficial, underscoring the voluntary and learner-controlled nature of the reflection activity.

- "I double-checked the numbers myself because AI can be wrong. I mainly used it to check logic and units."
- "AI helped me find sources of mistakes, but I still verified with lecture slides and the step sheet."

This is important because concerns about hallucinations and misleading outputs are central in generative AI adoption discussions [2].

Discussion

Ethical and responsible use of generative AI.

The instructional design framed generative AI not merely as a tool to be regulated, but as a bounded pedagogical role within a structured learning ecology. By explicitly requiring independent first attempts, encouraging verification behaviors, and limiting AI-supported revision to an optional bonus opportunity, the intervention operationalized responsible AI use as a design feature rather than a compliance rule. This distinction is important: instead of positioning AI primarily as a risk to academic integrity, the course structure embedded human oversight, self-explanation, and verification into the workflow. In this sense, responsible AI use emerged not from surveillance or prohibition, but from deliberate instructional scaffolding. Such framing may help shift classroom discourse from "whether AI should be allowed" to "how AI roles can be pedagogically constrained."

Why the impact appears larger for high-step assignments

The alignment between step complexity and observed performance shifts invites interpretation through a cognitive load perspective. In this study, the number of procedural steps functioned as a pragmatic indicator of task complexity, particularly in assignments requiring sustained coordination of parameters, units, and intermediate results. From this standpoint, longer reasoning chains may heighten the likelihood of cumulative minor errors, making externalized diagnostic dialogue—whether human or AI-mediated—potentially valuable. While cognitive load was not directly measured, the pattern observed is consistent with theoretical expectations regarding monitoring demands in multi-step quantitative tasks. Future work incorporating validated load instruments could further clarify these mechanisms.

Pedagogical value: AI as structured formative feedback rather than answer production

Beyond immediate score changes, the pedagogical contribution of this study lies in reframing AI as a structured formative assessment mechanism embedded within quantitative problem solving. Rather than functioning as a shortcut to solution production, AI was positioned as a dialogic prompt for self-explanation, error detection, and assumption articulation. In quantitative engineering contexts—where minor procedural inconsistencies can cascade into significant discrepancies—such targeted structured diagnostic dialogue may cultivate habits of verification that align with professional engineering judgment. Importantly, the instructional value does not depend on AI's correctness per se, but on the cognitive act of explanation and cross-checking triggered by the interaction.

Design features that strengthened pedagogical feasibility

These observations directly inform RQ3 by highlighting several low-cost, instructor-friendly design elements that supported responsible AI use in this course. Several design elements make this approach practical and defensible for engineering education contexts:

1. **Optional bonus resubmission** lowered pressure and reduced incentives to misuse AI, while still creating a meaningful reason to engage.
2. **"Higher of two scores" policy** maintained fairness and encouraged improvement without penalizing experimentation.
3. **Prompting that requires explanation and reflection questions** reinforced the student's role as the primary problem solver, reducing the "AI did my homework" perception.
4. **Instructor-provided step scaffolds** ensured that AI was not the only support; students anchored reasoning in course materials.

These features align with broader guidance for responsible GenAI adoption in education that emphasizes transparency, human oversight, and learning-centered design [3], while acknowledging known challenges [2].

Implementation checklist for instructors

- Decompose complex quantitative problems into explicit, step-by-step workflows.
- Provide students with worked examples and simplified calculation summaries prior to independent work.
- Demonstrate in class how AI can be prompted as a reflection partner (e.g., explaining steps, identifying common mistakes, and asking comprehension-check questions).
- Require students to articulate their reasoning before consulting AI.
- Position AI-supported revision as optional and low-stakes (e.g., bonus resubmission with the higher score recorded).
- Encourage verification of AI feedback using course materials, unit checks, and engineering judgment.
- Collect short reflections to reinforce metacognitive awareness and responsible AI use.

Beyond the specific procedural checklist presented above, this study suggests broader implications for quantitative engineering education. First, positioning generative AI as a structured reflection partner may help instructors reframe AI integration from a compliance-based policy issue to a pedagogical design opportunity. Rather than policing AI use, educators can proactively design bounded roles that preserve student agency while leveraging AI's diagnostic capabilities. Second, the alignment between task complexity and observed improvement patterns indicates that AI-supported reflection may be strategically targeted toward high-load, multi-step quantitative tasks where error propagation is common. Finally, the structured prompting approach demonstrated here offers a scalable model that can be adapted across engineering domains without requiring additional instructional time or technological infrastructure.

Conclusions

This study contributes a structured instructional design model for integrating generative AI as a reflection partner within quantitative engineering coursework. Within this pilot implementation, observed performance patterns and student reflections indicate that such positioning may support step-by-step reasoning, error detection, and metacognitive monitoring. Paired performance analyses indicate that the largest improvements occurred on high-step, high-complexity assignments, consistent with cognitive load expectations. Student reflections further indicate that many learners used AI as a second-opinion tool for logic and verification rather than as a direct answer generator. While limited in scope, this study offers a low-cost, instructor-friendly framework for integrating generative AI into quantitative engineering education in ways that emphasize metacognitive growth and reflective learning.

Limitations & Future Work

This study has several limitations typical of classroom pilots:

1. **Single course, single instructor, small sample size (N = 33):** results may not generalize to other contexts.
2. **Optional resubmission and self-selection:** students who used the second-chance opportunity may differ systematically (motivation, time availability).
3. **Non-experimental design:** the classroom-embedded pilot design limits causal interpretation of observed performance shifts. Other factors (time on task, re-engagement, peer help) may contribute.
4. **Reflection data adaptation:** reflections were de-identified and lightly edited to align with the equipment productivity learning context, which may reduce fidelity to original assignment wording.
5. **Absence of direct cognitive load measurement**: the study did not include direct measures of cognitive load. Although task complexity was discussed using a cognitive load lens, future research incorporating validated load instruments would provide stronger empirical grounding for mechanism-level claims.

Future work can strengthen evidence by (a) using structured reflection rubrics to quantify metacognitive behaviors, (b) adding retention checks (e.g., quiz/exam items later in the semester), and (c) comparing prompt variants (e.g., AI as tutor vs. AI as checker) under controlled conditions.

References

[1] J. Sweller, “Cognitive Load During Problem Solving: Effects on Learning,” *Cogn. Sci.*, vol. 12, no. 2, pp. 257–285, Apr. 1988, doi: 10.1207/s15516709cog1202_4.

[2] E. Kasneci, K. Sessler, F. Fischer, U. Gasser, and G. Groh, “ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education,” 2023.

[3] F. Miao, *Guidance for generative AI in education and research*. Paris, France: UNESCO, 2023.

[4] J. H. Flavell, “Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry.,” *Am. Psychol.*, vol. 34, no. 10, pp. 906–911, Oct. 1979, doi: 10.1037/0003-066X.34.10.906.

[5] B. J. Zimmerman, “Becoming a Self-Regulated Learner: An Overview,” *Theory Pract.*, vol. 41, no. 2, pp. 64–70, May 2002, doi: 10.1207/s15430421tip4102_2.

[6] T. Zhao, R. Chance, C. Buckhalter, and G. Wang, “Impact of ChatGPT on Student Writing in Construction Management: Analyzing Literature and Countermeasures for Writing Intensive Courses,” in *Proceedings of 60th Annual Associated Schools*, May 2024, pp. 339–348.

[7] T. Zhao, G. Wang, R. Chance, and C. Buckhalter, “Board 66: Impact of ChatGPT on Student Writing in Construction Management: A Study of Applied Risks,” in *2024 ASEE Annual Conference & Exposition Proceedings*, Portland, Oregon: ASEE Conferences, Jun. 2024, p. 48363. doi: 10.18260/1-2--48363.

[8] X. Lin, T. Zhao, S. W. Schmidt, and S. Zhou, “Using AI as a Learning Tool Through Simulation Interviews to Enhance Adult Learning,” *Adult Learn.*, p. 10451595251345274, Jun. 2025, doi: 10.1177/10451595251345274.

[9] T. Zhao *et al.*, “ChatGPT in Engineering Education: Revolutionizing Writing, Navigating Ethics, and Promoting Responsible Use,” in *2025 ASEE Annual Conference & Exposition Proceedings*, Montreal, Quebec, Canada: ASEE Conferences, Jun. 2025, p. 56085. doi: 10.18260/1-2--56085.

[10] X. Lin, T. Zhao, and S. Zhou, “Using AI for Simulation Interviews to Enhance Adult Learning,” in *2025 5th International Conference on Artificial Intelligence and Education (ICAIE)*, Suzhou, China: IEEE, May 2025, pp. 1–5. doi: 10.1109/ICAIE64856.2025.11158060.

[11] E. R. Mollick and L. Mollick, “Assigning AI: Seven Approaches for Students, with Prompts,” *SSRN Electron. J.*, 2023, doi: 10.2139/ssrn.4475995.

[12] K. Sands, X. Wang, and T. Zhao, “Use of AI in a Graduate Construction Estimating Course,” in *2025 ASEE Annual Conference & Exposition Proceedings*, Montreal, Quebec, Canada: ASEE Conferences, Jun. 2025, p. 57755. doi: 10.18260/1-2--57755.

[13] D. A. Schön, *The Reflective Practitioner*. Routledge, 2017. doi: 10.4324/9781315237473.

[14] T. Zhao, X. Lin, and R. Na, “Integrating AI in Construction Estimation Education: A Comparative Study of Togal AI and Bluebeam Revu 20,” *Eur. J. Educ.*, vol. 60, no. 4, p. e70287, Dec. 2025, doi: 10.1111/ejed.70287.

[15] J. Hattie and H. Timperley, “The Power of Feedback,” *Rev. Educ. Res.*, vol. 77, no. 1, pp. 81–112, Mar. 2007, doi: 10.3102/003465430298487.

[16] D. J. Nicol and D. Macfarlane-Dick, “Formative assessment and self-regulated learning: a model and seven principles of good feedback practice,” *Stud. High. Educ.*, vol. 31, no. 2, pp. 199–218, Apr. 2006, doi: 10.1080/03075070600572090.

[17] J. Sweller and G. A. Cooper, “The Use of Worked Examples as a Substitute for Problem Solving in Learning Algebra,” *Cogn. Instr.*, vol. 2, no. 1, pp. 59–89, Mar. 1985, doi: 10.1207/s1532690xci0201_3.
[18] M. T. H. Chi, M. Bassok, M. W. Lewis, P. Reimann, and R. Glaser, “Self-Explanations: How Students Study and Use Examples in Learning to Solve Problems,” *Cogn. Sci.*, vol. 13, no. 2, pp. 145–182, Apr. 1989, doi: 10.1207/s15516709cog1302_1.
[19] V. Braun and V. Clarke, “Using thematic analysis in psychology,” *Qual. Res. Psychol.*, vol. 3, no. 2, pp. 77–101, Jan. 2006, doi: 10.1191/1478088706qp063oa.

Appendix A. AI Prompt Examples

A.1 Prompt 1 (Guided walkthrough + checks)

You are a Reflection Partner for Haul Distance Calculations. I’m a student in a Construction Equipment course. I’ve uploaded the slides and the Excel spreadsheet, which include:

1. Lecture slides for this topic, and the Excel spreadsheet based on which to generate the mass diagram,
2. Simplified step-by-step assignment instructions that summarize the calculation steps at the end of the slides.

We did the haul distance of Section 3 (haul 3) example in class, which is included in the slides.

Please walk me through the calculation step by step. For each step, explain what it means, why it’s needed, and what common mistakes students often make. Ask me short questions to check if I understand each step before moving to the next.

A.2 Prompt 2 (Diagnostic + reflection questions)

Next, I’ll upload my own assignment (two calculation problems which are specified at the end of the slides) and related given materials. When I upload it, please:

1. Identify any missing steps, parameter errors, or unit conversion mistakes,
2. Suggest how I could revise my work to make the reasoning clearer, and
3. Provide 1–2 short reflection questions to help me think about what I might have misunderstood.

Keep your tone supportive and instructive, like a teaching assistant helping me prepare for an exam.