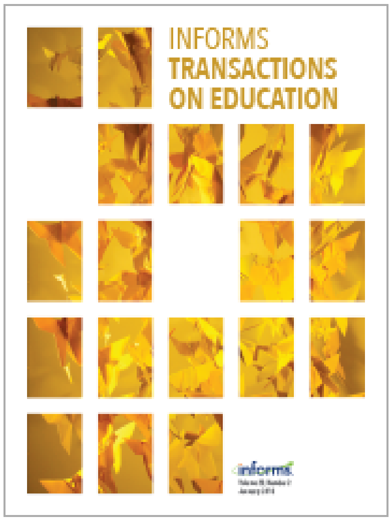


INFORMS Transactions on Education

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Integrating Generative AI into Teaching Linear Programming: Understanding Student-AI Interaction, Limitations, and Perceptions

Yeawon Yoo, Sydney Kim

To cite this article:
Yeawon Yoo, Sydney Kim (2026) Integrating Generative AI into Teaching Linear Programming: Understanding Student-AI Interaction, Limitations, and Perceptions. INFORMS Transactions on Education

Published online in Articles in Advance 18 Aug 2026

. https://doi.org/10.1287/ited.2025.0175

INFORMS TRANSACTIONS ON EDUCATION
Articles in Advance, pp. 1–14
ISSN 1532-0545 (online)
https://pubsonline.informs.org/journal/ited

Integrating Generative AI into Teaching Linear Programming: Understanding Student-AI Interaction, Limitations, and Perceptions

Yeawon Yoo,[a,*] **Sydney Kim**[b]

[a] California State University, Northridge, Northridge, California 91330; [b] University of North Florida, Jacksonville, Florida 32224

*Corresponding author

yeawon.yoo@csun.edu (Y. Yoo), sydney.kim@unf.edu (S. Kim)

https://orcid.org/0009-0008-3781-3781 (Y. Yoo)

Received: August 30, 2025
Revised: April 25, 2026
Accepted: June 15, 2026
Published Online in Articles in Advance: August 18, 2026

https://doi.org/10.1287/ited.2025.0175

Abstract. Generative artificial intelligence (AI) is rapidly reshaping the landscape of higher education. However, how students interact with these tools in quantitative learning contexts remains underexplored. This study examines student-AI interaction patterns, common sources of error in AI-generated solutions, and students' perceptions of generative AI in the context of linear programming. In the assignment, students solved problems manually and then engaged with generative AI, critically reviewing AI-generated outputs through iterations. The results reveal that whereas AI-generated responses often correctly formulated decision variables, objective functions, and constraints, errors frequently arose during the solution stage, particularly in identifying feasible regions and corner points. Students frequently characterized generative AI as conditionally useful: generative AI was regarded as valuable when used appropriately, but insufficient for fostering deeper conceptual learning. Perceived learning benefits were strongly associated with students' perceptions of reliability, overall experience, and intentions for future use. These findings highlight the importance of student-AI interaction in shaping outcomes and underscore the need to develop students' ability to guide and critically evaluate AI-generated solutions. The study offers practical insights for instructors integrating AI into quantitative coursework and highlights directions for future research in AI literacy and instructional design.

Keywords: generative AI in education • pedagogical research • developing critical thinking skills • teaching with technology

1. Introduction

The development of generative artificial intelligence (AI) has made significant impacts on our daily lives and society, with even greater implications on higher education. Whereas traditional AI has long been a valuable tool in education, its applications have typically been limited to predefined, rule-based systems that draw on prebuilt resources (e.g., AI-enabled proctoring and exam monitoring (Slusky 2020, Nigam et al. 2021), predictive analytics in learning management systems (Cavus 2010, Nenkov et al. 2016), plagiarism detection (Chitra and Rajkumar 2016, Sahu 2016), and adaptive feedback systems (Wambsganss et al. 2025)). In contrast, generative AI marks a new era in education with its ability to produce original outputs such as text, images, and music (Marr 2023). This generative capability creates new possibilities for enhancing both teaching and learning in higher education. For example, generative AI helps provide students with personalized learning experiences. As it is capable of producing humanlike text and providing context-aware response in real time, it dynamically generates original educational content, which is tailored to each student's needs (Binhammad et al. 2024, Wu et al. 2025). The capability of generative AI to make real-time adjustments to each student's needs can boost students' engagement, motivation, understanding, and performance (Kasneci et al. 2023, Jauhiainen and Garagorry Guerra 2024, Do et al. 2025, Herrero 2025). In addition to enhancing student learning, generative AI also provides benefits for educators (Lu et al. 2024a). It can assist routine tasks such as drafting course announcements and emails and creating study guides, practice exams, and review materials (Hashem et al. 2024, Moorhouse et al. 2024). Generative AI can also assist in writing grading rubrics, streamlining grading processes, and generating timely personalized feedback (Li et al. 2024, Lu et al. 2024b). Therefore, generative AI enables instructors to put more effort into core instructional activities.

Despite the aforementioned benefits, the use of generative AI also raises challenges and concerns. One major concern is that students are becoming increasingly overreliant on generative AI, which reduces students' opportunities to develop problem-solving skills and critical thinking (Lee et al. 2025, Qu et al. 2025). Higher education is intended to foster intellectual growth by exposing students to new ideas and promoting deep engagement with learning—through critical thinking and independent problem-solving and reasoning (Boyer Commission on Educating Undergraduates in the Research University 1998, Pascarella and Terenzini 2005). However, with the growing availability of generative AI, some students may increasingly rely on generative AI, especially when they are working on assignments and exams, instead of working on their own. Kosmyna et al. (2025) conducted a study comparing brain activity during the timed writing activity across three groups: one using OpenAI ChatGPT, another using any search engines, and the other group without any digital tools. When participants' brain activity was monitored during the writing tasks, the group using ChatGPT demonstrated significantly reduced neural connectivity and engagement in brain areas associated with memory, attention, and executive function relative to the other groups. Moreover, when participants were asked to recall their essays, the group using ChatGPT showed poorer memory recall compared with the other groups. These findings suggest that reliance on AI assistance may negatively affect critical thinking skills and ultimately weaken cognitive abilities.

This issue is also tied to another widely discussed concern in higher education: academic integrity (Stokel-Walker 2022, Kofinas et al. 2025, Walsh 2025). Whereas using generative AI in productive and ethical ways can support learning, the misuse of generative AI—particularly during exams or assignments without critical engagement—raises significant concerns about academic integrity. When students use the AI-generated output for graded assignments, instructors struggle to accurately assess each student's understanding. Although various AI detection tools have been developed, they are not consistently reliable, as both false positives and false negatives can be produced (Elkhatat et al. 2023, Perkins et al. 2024). Hence, this makes it increasingly difficult for educators to distinguish between authentic student work and content produced by generative AI. According to a recent survey of college and university leaders, 59% of respondents reported an increase in student cheating since the large language models (LLMs) were introduced, with 21% saying a significant increase. Additionally, 54% indicated that faculty at their institutions are unable to reliably identify AI-generated content (Flaherty 2025). This finding reinforces the view that generative AI poses a significant challenge to academic integrity within higher education.

An additional concern is that generative AI does not always produce accurate or reliable responses. Recent research highlights that LLMs, which underpin many widely used generative AI tools, often generate confidently stated yet incorrect responses—a phenomenon known as hallucination (Ji et al. 2023, Farquhar et al. 2024, Takita et al. 2025). In particular, as LLMs were originally designed for language-related tasks (e.g., text generation, question answering, translation), they sometimes make mistakes in quantitative reasoning or problem-solving tasks. There exists some research testing the accuracy of generative AI on some reasoning or quantitative tasks (Dao and Le 2023, Li et al. 2023, Wang et al. 2025, Xu et al. 2025). Dao and Le (2023) examined ChatGPT's performance on multiple-choice questions from the Vietnamese National High School Graduation Examination. The study found that as the difficulty level increased, the accuracy of ChatGPT dropped sharply, reaching as low as 10%. Xu et al. (2025) evaluated the performance of seven LLMs, including OpenAI GPT-3, GPT-4, and Google Gemini (previously BARD), using a benchmark of logical reasoning tasks categorized into deductive, inductive, abductive, and mixed-form reasoning settings. Their findings reveal that all models demonstrated specific limitations in logical reasoning, showing evident weakness with inductive reasoning. Notably, ChatGPT underperformed across all reasoning types—deductive, inductive, and abductive. These empirical findings underscore that generative AI, although promising, cannot yet be considered a consistently reliable tool for quantitative or reasoning tasks. Therefore, the combination of students' overreliance on generative AI and its inaccurate outputs further presents a significant concern for learning in quantitative and analytical subjects, as students who depend on these tools without the ability to critically evaluate the correctness of outputs risk internalizing incorrect or flawed information.

As generative AI becomes an inevitable presence in higher education, it is important to integrate it into curricula and to teach students how to use it effectively while critically evaluating the outputs. Despite its importance, there exist significant gaps in the literature. Prior studies have largely focused on the accuracy of LLMs, language-based tasks, and general perceptions of AI. To the best of our knowledge, research has not yet extended to generative AI in the domain of linear programming, which is a cornerstone of the operations research and management science curriculum. Specifically, no prior work has provided empirical evidence that jointly examines students' interactions with generative AI, objective evaluations of generative AI, and students' perceptions of generative AI along with their academic performance. Guided by this gap, this study examines the following research questions:

- What patterns of error and limitation emerge in AI-generated solutions to linear programming problems?
- How do students interact with and perceive generative AI in solving linear programming problems?

The organization of the rest of this paper is as follows. Section 2 provides an overview of the study, including the assignment design, postassignment survey questions, and participant details. Section 3 presents the key findings of the study, focusing on the accuracy of generative AI tools and students' perceptions of these tools. Lastly, Section 4 summarizes the main findings, highlights the contributions, outlines the study's limitations, and suggests directions for future research.

2. Study Design and Methodology

This section outlines the instructional context and methodological framework of the study. We describe the design of the linear programming assignment integrating generative AI, the survey instrument used to capture students' perceptions, and the characteristics of the course and participants in which the study was conducted.

2.1. Assignment Structure

The assignment "Linear Programming with Generative AI" consisted of three basic linear programming questions with two decision variables. The full text of the assignment questions is provided in Appendix A.

The purpose of the assignment was fourfold: (1) to strengthen students' understanding of linear programming concepts, (2) to practice in formulating linear programming models in real-world business contexts, (3) to compare student-generated solutions with those produced by generative AI, and (4) to critically evaluate the accuracy and reliability of AI-generated outputs.

The assignment consists of two parts: manual problem-solving and AI-based problem-solving. In the manual problem-solving part, students were asked to solve three linear programming problems without using generative AI. For each problem, they were asked to provide clear descriptions of the decision variables, objective function, and constraints in words before translating them into mathematical notation. Subsequently, students formulated the linear programming model by expressing the objective function, constraints, and decision variables in formal mathematical notation. To solve the problem manually, students were directed to use the corner point approach, including a graphed feasible region, identification of corner points, and determination of the optimal solution. After completing the manual problem-solving part, students used a generative AI platform of their choice to solve the same problems. Although students designed their own prompts, they were required to ask the AI to provide the linear programming formulation, graph the feasible region, identify the corner points of the feasible region, and find the optimal solution. They were also tasked with refining the responses from AI through at least three iterations of chat to ensure meaningful interaction. Finally, students compared their manual solutions with the AI-generated ones, focusing on differences in accuracy, formulation, and clarity; refer to Appendix A for instructions given to students.

2.2. Postassignment Survey

Following the completion of the assignment, students were asked to complete a survey to share their perceptions of using generative AI in solving quantitative problems. Specifically, students were asked to indicate which generative AI platform they used and respond to the questions about perceived impact on learning (Davis 1989) and critical thinking (Facione 1990, Lee et al. 2025), performance and reliability (Jian et al. 2000, Hoffman et al. 2023), and user experience (Davis 1989, Lu et al. 2022, Hoffman et al. 2023) through a five-point scale, and the survey was open-ended. The full list of questions is shown in Table B.1 in Appendix B. These survey questions collectively capture multiple key dimensions of students' perceptions of generative AI use in quantitative learning contexts.

2.3. Participants

This study was conducted at a regional public institution in an asynchronous online, undergraduate-level operations management course during the spring 2025 semester, with the assignment administered in week 5 of the eight-week term. The course was composed of junior and senior students across various business disciplines. Among the total of 44 students enrolled in the course, 35 students (77%) agreed to participate in the study. Note that participation was voluntary and approved by the institutional review board of the institution.

3. Results

This section presents the empirical findings of the study. We begin by reporting students' choice of generative AI platforms and the accuracy of AI-generated responses across questions and platforms. We then summarize students' perceptions of generative AI use through descriptive statistics and thematic analysis of open-ended responses. Finally, we examine the relationships among perception measures and their associations with academic performance.

3.1. Generative AI Platform Chosen

In this assignment, students were allowed to choose any generative AI platforms they liked. Table 1 summarizes students' use of different AI platforms, showing the number of students who reported using each platform. Among 35 participating students, 25 students (71.4%) reported using OpenAI ChatGPT, making it the most commonly selected tool. Six students (17.1%) reported using Google Gemini, and two students (5.7%) reported using both ChatGPT and Gemini—one used both platforms for the same question, whereas the other used Gemini for one question and ChatGPT for the remaining

Table 1. Generative AI Tool Usage

Tool/category	Number of students (%)
ChatGPT	25 (71.4)
Gemini	6 (17.1)
ChatGPT and Gemini	2 (5.7)
Copilot	1 (2.9)
No response (manual only)	1 (2.9)
Total	35 (100)

Note. Most students used ChatGPT.

two. One student (2.9%) used Microsoft Copilot, and one student (2.9%) completed only the handwritten portion of the assignment; hence, this was recorded as "no response."

3.2. Accuracy of Reported Generative AI Responses

The assignment consisted of three linear programming questions; however, not all students completed every question in the manual problem-solving and AI-based problem-solving components. Specifically, 28 out of 35 students submitted responses for question 1, whereas 27 students completed questions 2 and 3.

Table 2 presents the accuracy of generative AI responses, excluding nonresponses. Correctness of the reported generative AI responses was evaluated by verifying whether the reported responses corresponded to the true optimal solution. The correct response rates were 57.1% for question 1, 25.9% for question 2, and 66.7% for question 3. It is important to note that for students who used both ChatGPT and Gemini, a response was counted as correct if at least one of the platforms produced the correct solution.

The accuracy of generative AI responses by question and platform is summarized in Table 3. Note that the table excludes nonresponses. Additionally, for students who reported using both ChatGPT and Gemini, their responses were counted under both platforms, which explains why the total number of responses is one greater than the total number of students who responded.

Given that over 70% of students reported using ChatGPT, the overall accuracy results primarily reflect ChatGPT-generated responses. The results for other platforms are based on a small number of observations and should therefore be interpreted as descriptive rather than as evidence of systematic performance differences across platforms.

To provide additional context, students' manual problem-solving accuracy ranged from 42.9% to 62.9% across the three questions. Accuracy was highest for question 3 and lowest for question 2, with fewer than half of students answering questions 1 and 2 correctly. These results indicate moderate variation in students' manual solution performance across problem types. For readers interested in how students performed on the manual problem-solving portion of the assignment, a summary of the accuracy of manually solved student responses by question is presented in Table C.1 in Appendix C.

Table 2. Accuracy of Generative AI Responses by Question (Excludes Nonresponses)

Question	Correct (%)	Incorrect (%)	Total responses
Q1	16 (57.1)	12 (42.9)	28
Q2	7 (25.9)	20 (74.1)	27
Q3	18 (66.7)	9 (33.3)	27

3.3. Diagnostic Analysis of AI-Generated Solutions and Student Interaction

To better understand the sources of inaccuracies in the solutions from generative AI, we conducted a diagnostic analysis of incorrect responses. Each response was examined to identify whether errors occurred during model formulation (i.e., defining decision variables, objective function, and constraints) or during the solution process (i.e., identifying the feasible region, determining corner points, and evaluating the objective function).

Overall, generative AI models performed well in the formulation stage; decision variables were correctly defined, and both the objective function and constraints were accurately specified. This suggests that generative AI is generally effective at translating problem descriptions into mathematical models.

However, the majority of errors occurred during the solution stage, particularly in the application of the corner-point method. Many incorrect responses were associated with misidentification of the feasible region, which ultimately led to the selection of invalid corner points that did not satisfy all constraints. For example, in question 3, several AI-generated responses selected the point $(5,5)$ as the optimal solution (see Figure 1). Although the constraint $10x + 25y \geq 350$ was correctly specified, some AI-generated solutions incorrectly represented this constraint as $10x + 25y \leq 350$ when constructing the feasible region. Whereas the point $(5,5)$ satisfies the minimum hour constraints ($x \geq 5$, $y \geq 5$), it does not satisfy the task requirement constraint and therefore is not part of the feasible region. Despite correct model formulation, the failure to accurately

Table 3. Accuracy of Generative AI by Question and by Generative AI Platform

Question	ChatGPT (n/N)	Gemini (n/N)	Copilot (n/N)
Q1	52.2 (11/22)	83.3 (5/6)	0.0 (0/1)
Q2	29.2 (7/23)	0.0 (0/4)	0.0 (0/1)
Q3	75.0 (18/23)	0.0 (0/4)	0.0 (0/1)

Note. n, number of students who reported an accurate generative AI response; N, number of students who used that platform and responded to the question.

Figure 1. Example of an Incorrect AI-Generated Solution for Question 3

Note. The constraint $10x + 25y \geq 350$ is correctly specified; however, the feasible region is incorrectly constructed below the constraint line.

AI Response (excerpt):
Step 3: Optimal Solution
At $(0, 20)$:
$$\text{Revenue} = 3(0) + 5(20) = 100$$
At $(13.33, 0)$:
$$\text{Revenue} = 3(13.33) + 5(0) = 40$$
At $(0, 24)$:
$$\text{Revenue} = 3(0) + 5(24) = 120$$
The maximum revenue is 120, which occurs at $(0, 24)$.
Student Follow-up:
"$(0, 24)$ is not a valid response."
AI Response (excerpt):
"You're right! I made a mistake in the previous steps. Let's go through the problem again and carefully calculate the correct corner points and the optimal solution."
(excerpt shortened for brevity)
Step 4: Evaluate the Objective Function
At $(0, 20)$:
$$\text{Revenue} = 3(0) + 5(20) = 100$$
At $(13.33, 0)$:
$$\text{Revenue} = 3(13.33) + 5(0) = 40$$
The maximum revenue is 100.
Student Follow-up:
"Did you include the origin $(0, 0)$ as one of the corner points?"
AI Response (excerpt):
"You're absolutely right. I missed considering the origin $(0, 0)$ as a possible corner point. Let's reanalyze the problem, including all possible corner points."

represent and evaluate feasibility resulted in an incorrect final answer. As a result, the evaluation of the objective function was often based on incorrect corner points, leading to erroneous solutions. This pattern was consistently observed across multiple student submissions, where correct formulations were followed by incorrect corner-point identification.

These findings indicate that whereas generative AI demonstrates strong capability in model formulation, it is less reliable in executing multistep solution procedures that require geometric interpretation and constraint validation. This distinction highlights that inaccuracies are not primarily because of misunderstanding the problem structure but, rather, because of challenges in systematically applying solution methods such as identifying feasible regions and evaluating corner points.

In addition to analyzing the sources of inaccuracies in AI-generated solutions, we examined students' interactions with generative AI to better understand how prompt design and usage patterns are associated with outcomes.

First, students who obtained correct solutions tended to use more specific prompts. Many students relied on minimal prompts by directly copying the problem statement without additional clarification or guidance, which often resulted in incomplete or incorrect solutions—particularly in identifying feasible regions and corner points.

Second, students who obtained correct solutions also tended to engage in more iterative interaction. Those who refined AI responses through multiple iterations—such as asking follow-up questions about corner points or requesting clarification of intermediate steps—were more likely to identify and correct initial errors. However, iterative prompting did not always guarantee improvement. In some cases, additional interactions introduced new inconsistencies or altered previously correct answers, highlighting limitations in the stability of AI-generated reasoning in multistep quantitative tasks at the time of data collection. This pattern suggests that effective use of generative AI in quantitative problem-solving depends not only on the model's capabilities but also on the user's ability to guide and critically evaluate the interaction process.

To illustrate these interaction patterns, we present a representative example of student-AI interaction (the full student-AI interaction for this example is provided in Figure D.1 in Appendix D).

3.4. Students' Perception on Generative AI

This subsection presents student responses to the post-assignment survey, which explored their perceptions of using generative AI in linear programming. The survey focused on four key areas: *Impact on Learning*, *Critical Thinking*, *Performance and Reliability*, and *User Experience and Future Use*. Table 4 provides descriptive statistics for the five-point scale items, whereas Table 5 summarizes themes and representative quotes from the open-ended responses. The open-ended responses were analyzed using a combination of manual thematic coding, guided by human interpretation, and computational topic modeling implemented in Python.

Table 4. Postassignment Survey Responses on Generative AI Usage in Linear Programming (N = 35)

Category	Survey question (scale)	1	2	3	4	5	Mean (SD)
Impact on Learning	Did using generative AI platforms help you better understand linear programming concepts? (1 = not at all, 5 = very much)	6 (17.1%)	10 (28.6%)	6 (17.1%)	9 (25.7%)	4 (11.4%)	2.86 (1.31)
	Do you feel more confident in your ability to solve linear programming problems after validating answers from generative AI platforms? (1 = not at all, 5 = very much)	5 (14.3%)	5 (14.3%)	8 (22.9%)	10 (28.6%)	7 (20.0%)	3.26 (1.34)
Critical Thinking	Do you feel that using generative AI platforms improved your critical thinking skills? (1 = not at all, 5 = very much)	7 (20.0%)	9 (25.7%)	8 (22.9%)	6 (17.1%)	5 (14.3%)	2.80 (1.35)
	How confident are you in identifying errors in solutions provided by generative AI platforms? (1 = not confident at all, 5 = very confident)	5 (14.3%)	4 (11.4%)	11 (31.4%)	12 (34.3%)	3 (8.6%)	3.11 (1.18)
Performance and Reliability	How accurate were the solutions provided by ChatGPT, Gemini, or other generative AI platforms? (1 = very inaccurate, 5 = very accurate)	0 (0.0%)	7 (20.0%)	12 (34.3%)	12 (34.3%)	4 (11.4%)	3.37 (0.94)
	How reliable do you think generative AI platforms are for solving linear programming problems? (1 = not reliable at all, 5 = very reliable)	2 (5.7%)	9 (25.7%)	12 (34.3%)	9 (25.7%)	3 (8.6%)	3.06 (1.06)
User Experience and Future Use	Do you plan to use generative AI platforms for future quantitative assignments? (1 = definitely not, 5 = definitely yes)	12 (34.3%)	8 (22.9%)	6 (17.1%)	8 (22.9%)	1 (2.9%)	2.37 (1.26)
	What was your overall experience using generative AI for this assignment? (1 = very negative, 5 = very positive)	1 (2.9%)	1 (2.9%)	8 (22.9%)	14 (40.0%)	11 (31.4%)	3.94 (0.97)

Notes. Percentages may not sum to 100% because of rounding. Students reported relatively low gains in understanding linear programming and critical thinking, and they expressed limited intentions to use AI for future assignments. SD, standard deviation.

Table 5. Themes from Open-Ended Survey Responses on Generative AI in Learning Contexts

Theme	*n* (%)	Representative quotes
Impact on learning		
Conditional usefulness	16 (46%)	"I think that generative AI could be helpful if used correctly"; "Helpful for those who are curious, but not necessarily for in depth learning."
Valuable for learning	13 (37%)	"It is a valuable tool for learning because it can show you the correct steps and figure out some mistakes that you missed, which it did for me."
Not valuable	5 (14%)	"It just gives you an answer without explaining."
Ethical concerns	2 (6%)	"I think it is a very valuable tool for learning as long as, it is not used in a manner to plagiarize cheat, or any dishonest manner."
Risk of overdependence	2 (6%)	"It may also have some drawbacks as students may rely on it without actually completing the work hence students may not fully grasp the material."
Performance and reliability		
Quality and accuracy issues	12 (34%)	"Some of the results show limited accuracy, especially during complex mathematical computations."
Need for specificity in prompts	10 (29%)	"You need to be very specific with what you want to be answered."
Limited visual representation	10 (29%)	"Generative AI responses don't provide visuals, like graphs."
Lack of step-by-step explanations	8 (23%)	"There was minimal explanation between the steps and the processes."
Inconsistency in responses	7 (21%)	"I also came to the conclusion that it has multiple answers for the same question which can cause confusion for students who don't understand the material."

Notes. n denotes the number of responses associated with each theme; percentages do not sum to 100% because individual students could report multiple themes. Students identified both benefits and limitations of using AI in learning contexts.

3.4.1. Measurement Reliability. Before presenting the survey results, internal consistency reliability was assessed for each multi-item construct using Cronbach's alpha. Given that each construct was measured using two items, Cronbach's alpha was computed alongside interitem correlations, which provide an appropriate assessment of reliability for short scales. The results indicate acceptable to strong internal consistency for most constructs. Specifically, Impact on Learning ($\alpha = 0.7470$) and User Experience and Future Use ($\alpha = 0.7246$) demonstrated acceptable reliability, whereas Performance and Reliability showed excellent internal consistency ($\alpha = 0.9271$). These results suggest that the items within these constructs are reasonably consistent in capturing their intended concepts. In contrast, the Critical Thinking construct exhibited relatively low internal consistency ($\alpha = 0.3931$), reflecting both the limited number of items and the relatively weak correlation between them. This suggests that the two items under this construct—perceived improvement in critical thinking skills and confidence in identifying errors in AI-generated responses—may capture related but distinct aspects of students' experiences. Accordingly, results associated with the Critical Thinking construct are interpreted with caution, and the two items are considered individually in subsequent analyses rather than as a unified scale.

3.4.2. Impact on Learning. Whereas students generally did not perceive generative AI as significantly improving their understanding of linear programming concepts (mean = 2.86), their confidence in solving such problems showed a modest increase after validating their answers with AI (mean = 3.26). To further explore students' experiences, an open-ended question asked: "Do you think generative AI platforms are valuable tools for learning? Why or why not?" Responses revealed a range of perspectives. The most common response was that its usefulness is conditional (46%), with students emphasizing that generative AI can be helpful when used appropriately or to supplement the existing knowledge, but less effective for deeper conceptual learning. A substantial proportion of students (37%) considered generative AI to be generally valuable, particularly in reinforcing understanding and assisting in the identification of errors. In contrast, a smaller group (14%) perceived generative AI as not valuable, pointing to its tendency to provide answers without sufficient explanation. Additionally, a minority of students raised concerns related to academic integrity (6%), whereas another group (6%) raised concerns about overdependence, noting that reliance on AI could limit deeper understanding.

3.4.3. Critical Thinking. Students reported moderate confidence in their ability to identify errors in AI-generated responses (mean = 3.11). Despite this moderate confidence, students expressed relatively low agreement that the use of generative AI improved their critical thinking skills (mean = 2.80). This suggests that whereas students feel moderately capable of evaluating AI outputs, they are less convinced that interacting with AI substantially enhances their critical thinking skills.

3.4.4. Performance and Reliability. The majority of students perceived AI-generated solutions as generally or somewhat accurate (mean = 3.37) and expressed similarly moderate views of reliability (mean = 3.06). The open-ended responses to the question, "What limitations or weaknesses did you notice in the responses from generative AI platforms?" offer insight into the limitations they encountered. The most frequently reported limitation was quality and accuracy (34%), with students noting that AI-generated responses were sometimes inaccurate. Another common limitation was the need for specificity in prompts (29%), as students emphasized that useful responses often required carefully phrased input. The same proportion of students (29%) reported limited or missing visual representations, which they considered important for understanding. In addition, students highlighted the lack of step-by-step explanations (23%), reporting that responses often presented final answers without sufficient explanation of intermediate steps. Lastly, some students (21%) identified inconsistencies across outputs, observing that identical prompts yielded different outputs across attempts.

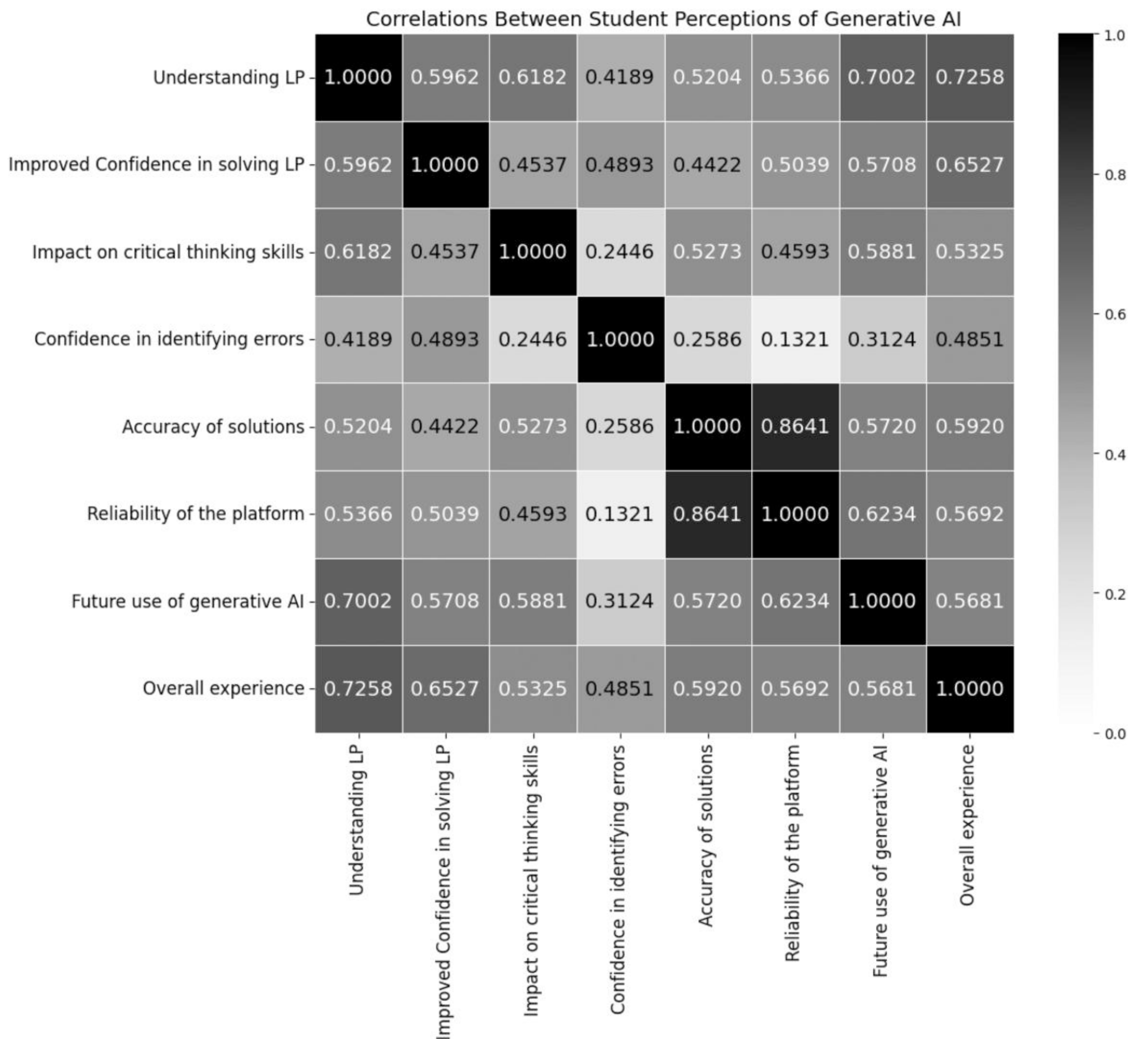
3.4.5. User Experience and Future Use. Students reported a positive overall experience with generative AI tools (mean = 3.94). This indicates a generally favorable perception toward generative AI-assisted learning experiences in quantitative assignments. However, despite this moderately positive experience, students' intentions to use generative AI for future quantitative assignments were considerably lower (mean = 2.37), with 57.2% of respondents indicating negative intentions toward future use.

3.5. Relationships Among Students' Perceptions of Generative AI

Beyond the descriptive statistics, which focus on the responses to each question, correlation analyses were conducted to examine relationships across students' perceptions (see Figure 2). Spearman's rank correlation coefficients (r_s) and their associated p-values were computed; the full results are reported in Table E.1 in Appendix E. Consistent with Cohen (2013) guidelines, correlations with absolute values above 0.50 were interpreted as strong correlations. All reported relationships are correlational and do not imply causal or mediating effects.

First, the extent to which students felt they understood linear programming was strongly associated with other perceptions. Specifically, students who reported

Figure 2. Correlation Results Among Students' Perceptions of Generative AI



Note. Perceived understanding LP and confidence in solving LP were strongly associated with overall experience and the intention for future use.

higher levels of understanding also indicated stronger confidence in solving problems ($r_s = 0.5962$, $p < 0.001$), higher self-reported improvement in critical thinking skills ($r_s = 0.6182$, $p < 0.001$), and higher ratings of accuracy ($r_s = 0.5204, p = 0.001$) and reliability ($r_s = 0.5366$, $p < 0.001$). Most notably, perceived understanding was very strongly correlated with both intentions for future use ($r_s = 0.7002, p < 0.001$) and overall experience ($r_s = 0.7258, p < 0.001$). These results suggest that students' trust in and willingness to adopt generative AI are closely associated with their perceived learning benefits (i.e., understanding the concepts well).

Moreover, confidence in solving LP after validating AI outputs was positively associated with perceptions of reliability ($r_s = 0.5039, p = 0.002$), future use ($r_s = 0.5708, p < 0.001$), and overall experience ($r_s = 0.6527$, $p < 0.001$). These findings suggest that students who reported greater confidence also tended to report higher levels of perceived reliability, more positive overall experiences, and stronger intentions to use generative AI in the future. This relationship may partly reflect underlying differences in student ability, whereby higher-performing students tend to report greater confidence rather than indicating a directional or causal effect. This pattern is consistent with self-determination theory (Deci and Ryan 2012), which identifies perceived competence as a key psychological need supporting intrinsic motivation.

Lastly, given the low internal consistency of the Critical Thinking measure, the associated items are interpreted individually rather than as a unified construct. In particular, confidence in identifying errors in AI-generated responses was positively associated with perceived accuracy ($r_s = 0.5273$, $p = 0.001$), intentions for future use ($r_s = 0.5881$, $p < 0.001$), and overall experience ($r_s = 0.5325, p < 0.001$).

In contrast, average ratings of perceived improvement in critical thinking were modest in the descriptive results.

This pattern suggests that whereas students reported moderate confidence in evaluating AI-generated outputs, such confidence does not necessarily translate into broader perceived gains in critical thinking skills. Rather, students who felt more capable of identifying errors in AI responses also tended to report higher levels of trust in AI and greater willingness to adopt it in the future.

More broadly, whereas descriptive statistics indicated only moderate enthusiasm overall, the correlational evidence suggests that students who perceived greater educational value—particularly in terms of conceptual understanding and confidence in solving LP problems—were more likely to express intentions for continued use of generative AI.

3.6. Relationship Between Students' Perceptions of Generative AI and Academic Performance

In addition to the relationship between students' perceptions, we calculated Spearman correlation coefficients with associated p-values between students' perceptions of generative AI and their academic performance, focusing on scores from this assignment and the final exam in linear programming.

With regard to their performance in this assignment, moderate positive correlation was observed between students' confidence in identifying errors in AI-generated solutions and their performance on the assignment. Specifically, students' responses to the question "How confident are you in identifying errors in the solutions from generative AI platforms?" were positively correlated with assignment scores ($r_s = 0.4328$, $p = 0.0106$). This finding suggests that students who reported greater confidence in evaluating the accuracy of AI outputs tended to perform better. It is plausible that these students were more critically engaged with the material and possessed stronger analytical and metacognitive abilities, enabling them to use AI tools more effectively without overrelying on them.

With regard to final exam performance, only one survey item demonstrated a statistically significant relationship at the 10% level. Students who gave higher ratings on the question "Did using generative AI platforms help you better understand linear programming concepts?" tend to do better in the final exam ($r_s = 0.2998$, $p = 0.0849$). Whereas this result does not meet the conventional threshold for significance ($p < 0.05$), it indicates a potential association between students' perceived conceptual benefit from AI tools and their performance on the assessments.

No statistically significant correlations were found for the remaining survey items. Detailed results of the correlation analyses are provided in Table F.1 in Appendix F.

4. Discussion and Conclusion

This study examined how students engaged with generative AI in the context of a linear programming assignment, providing empirical insights into both the performance of AI-generated solutions and students' perceptions of generative AI. The findings highlight that outcomes vary depending not only on the capabilities of generative AI but also on how students frame prompts, engage in iterative interaction, and critically evaluate AI-generated responses in solving quantitative problems.

From a performance of the generative AI standpoint, generative AI frequently produced incorrect solutions, with fewer than 50% of the reported AI-generated responses being accurate. Accuracy varied notably across both questions and platforms, reflecting inconsistencies in how generative AI solves linear programming. These findings align with broader concerns in the literature regarding the limitations of generative AI in addressing quantitative and reasoning-intensive problems.

To better understand the sources of these inaccuracies, we conducted a diagnostic analysis of AI-generated solutions and student interactions. The results indicate that errors were concentrated in the solution stage, particularly in identifying feasible regions and corner points. Importantly, these inaccuracies were closely linked to how students interacted with generative AI, including specificity and iteration of prompts. Students who relied on minimal prompts were more likely to retain incorrect solutions, whereas those who engaged in more structured and iterative interactions—such as refining prompts and questioning AI-generated outputs—were better able to identify and correct errors. These students were therefore more likely to obtain accurate final solutions. These findings suggest that the performance of generative AI is not solely determined by the model itself but also depends on how users guide and critically evaluate its outputs.

Regarding the survey results, students generally perceived AI-generated solutions as only moderately accurate and expressed hesitation about using generative AI for future quantitative assignments, reflecting concerns about its long-term value in problem-solving. At the same time, students viewed generative AI as conditionally useful—particularly for verifying solutions and clarifying basic concepts—while noting that it was less effective for learning new material. Students also identified several limitations, including the need for highly specific prompts, limited visualizations, insufficient step-by-step explanations, and inconsistencies across responses. These perceptions reinforce the importance of how generative AI is used in practice, suggesting that its effectiveness depends not only on model capability but also on the quality of user interaction and prompt design.

A particularly noteworthy and more objective contribution of this study lies in the correlational findings of students' perceptions. Perceived learning benefits—particularly, conceptual understanding of LP and confidence in solving LP—were strongly associated with trust (accuracy and reliability), overall experience, and intentions for future use. At the same time, perceptions of accuracy and reliability were closely associated with the foundation of students' trust, with both measures showing strong links to future intent to adopt generative AI.

Beyond these empirical findings, this study also offers important implications for how generative AI can support the teaching and learning of linear programming. Whereas this study focused on introductory linear programming problems using graphical methods, similar instructional approaches could be extended to other classes of optimization problems. For example, future assignments could examine how students use generative AI to formulate and solve discrete optimization problems, such as binary or integer programming, as well as network optimization, infeasibility diagnosis, and sensitivity analysis (e.g., interpreting shadow prices and reduced costs). In addition, generative AI may support more advanced learning tasks, such as identifying modeling errors and generating solver-based implementations in tools such as Excel and Python. These extensions would allow instructors and students to better understand how generative AI can support not only getting solutions but also the development of deeper conceptual understanding and analytical reasoning across a broader range of optimization topics.

Building on these insights, the proposed framework and assignment design can be extended beyond linear programming to other areas of operations research and management science (e.g., simulation, supply chain analytics, game theory), where similar dynamics of student trust, perceived learning outcomes, and adoption may be observed. Future research could further explore structured and pedagogically guided uses of generative AI, particularly in examining how students' AI literacy in quantitative topics can be developed through targeted instructional design.

Whereas the findings offer important pedagogical implications, several limitations should be considered when interpreting the results. First, no separate control group was included, as all students completed the same AI-integrated assignment; this decision reflected ethical and fairness considerations because excluding some students from the AI-integrated experience could have placed them at either an advantage or a disadvantage in the remainder of the course. However, the absence of a control group limits the ability to draw causal inferences regarding the impact of generative AI assignment on learning outcomes or its comparative effectiveness relative to traditional instructional approaches. Future research could address this limitation through the use of parallel course sections or counterbalanced designs, enabling comparisons between AI-integrated and traditional instructional approaches while maintaining instructional equity. Such designs would allow for a more rigorous evaluation of the causal impact of generative AI on student learning and performance in quantitative courses. Second, the study was not designed to track how students revised their original responses after viewing generative AI outputs. Future research could address this by collecting both pre-AI and post-AI responses, enabling a deeper analysis of how students adopt, modify, or resist AI-generated suggestions and how these interactions affect learning outcomes. Third, the study did not fully control student-AI interactions, such as prompt design and iteration depth, which likely influenced both the accuracy of AI-generated responses and students' perceptions of the tool. Subsequent research could investigate how different types of prompting (e.g., generic versus structured prompts) affect the quality of AI outputs and the development of students' prompting skills. Lastly, specific model versions were not standardized in this study. Whereas the primary focus of this research concerns student interaction and evaluative judgment within an authentic instructional context, future studies should examine how newer model iterations perform under comparable pedagogical conditions and whether improvements in model capability meaningfully alter student engagement and learning outcomes.

Acknowledgments
We would like to express our sincere gratitude to the students who completed the assignment and generously shared their honest reflections on their experiences with generative AI. Their thoughtful participation made this study possible. We also thank the editor and anonymous reviewers for their constructive feedback and helpful suggestions, which significantly improved the quality and clarity of this manuscript.

Appendix A. Assignment Instructions and Questions

A.1. Assignment Instructions Provided to Students for Completing the Manual and Generative AI Components of the Task

1. *Define the key components of linear programming:* Provide a clear description of the decision variables, objective function, and constraints in plain language.

2. *Formulate the model:* Write the linear programming model, including the objective function, constraints, and decision variables, in mathematical notation.

3. *Solve the model:* Graph the feasible region, identify the corner points, and determine the optimal solution manually.

4. *Solve the problem with generative AI:* Use a generative AI tool (e.g., ChatGPT, Bard) to solve the problem. Ask the AI to (a) provide the linear programming formulation, (b) find the corner points of the feasible region, and (c) determine the optimal solution. Refine the AI's response through at least three iterations. Share your interaction with the AI using one

of the following methods: (a) share a link to the conversation, (b) upload screenshots of the chat (as a zip file or included in a single document), or (c) or copy and paste the conversation into a document (DOC or PDF format).

5. *Record the AI's final solution*: Document the AI's final linear programming formulation and its optimal solution, including the corner points of the feasible region.

6. *Compare and validate*: Compare the AI's solution with your original formulation and solution. Verify if the AI's response is accurate. If there are any errors or inconsistencies, identify them and explain why they are incorrect.

A.2. Questions

Question 1: Printing Office Problem

Royce manages a printing office in Camarillo using two types of machines: a high-speed printer and a binding machine. Each report requires 1.5 hours on the high-speed printer and 0.5 hour on the binding machine. Each brochure requires one hour on the high-speed printer and one hour on the binding machine. The high-speed printer is available for 20 hours per week, and the binding machine is available for 24 hours per week. Reports generate $3 in revenue each, whereas brochures generate $5 in revenue each. How many of each should Royce produce to maximize revenue? What is the maximum revenue?

Question 2: Lemonade Stand Problem

Rylee plans to open a lemonade stand in Ventura, California, where they will sell two types of lemonade: small and large. A large lemonade requires 12 ounces of water and 2 ounces of powder, whereas a small lemonade requires 8 ounces of water and 1 ounce of powder. Rylee has 600 ounces of water available and two cans of powder, each containing 45 ounces. Rylee expects to sell at least 10 large lemonades. Small lemonades are priced at $1 each, whereas large lemonades cost $1.25 each. How many of each type should Rylee prepare to maximize revenue? What is the maximum revenue?

Question 3: Dolphin Company Hiring Problem

The Dolphin Company needs to hire two substitute employees to complete a specified number of tasks. Employee A can complete 10 tasks per hour and earns $20 per hour, whereas employee B can complete 25 tasks per hour and earns $30 per hour. Both employees must work at least five hours. The company requires a minimum of 350 tasks to be completed. How many hours should each employee work to minimize total wages while meeting the task requirement? What is the minimum total wage?

Appendix B. Postassignment Survey Questions

Table B.1. Postassignment Survey Questions

Category	Survey question
Platform Usage	Q1: Which generative AI platform did you use to complete this assignment?
Impact on Learning	Q2-1: Did using generative AI platforms help you better understand linear programming concepts? (1 = not at all, 5 = very much)
	Q2-2: Do you feel more confident in your ability to solve linear programming problems after validating answers from generative AI platforms? (1 = not at all, 5 = very much)
	Q2-3: Do you think generative AI platforms are valuable tools for learning? Why or why not? (open-ended)
Critical Thinking	Q3-1: Do you feel that using generative AI platforms improved your critical thinking skills? (1 = not at all, 5 = very much)
	Q3-2: How confident are you in identifying errors in solutions provided by generative AI platforms? (1 = not confident at all, 5 = very confident)
Performance and Reliability	Q4-1: How accurate were the solutions provided by ChatGPT, Gemini, or other generative AI platforms? (1 = very inaccurate, 5 = very accurate)
	Q4-2: How reliable do you think generative AI platforms are for solving linear programming problems? (1 = not reliable at all, 5 = very reliable)
	Q4-3: What limitations or weaknesses did you notice in the responses from generative AI platforms? (open-ended)
User Experience and Future Use	Q5-1: Do you plan to use generative AI platforms for future quantitative assignments? (1 = definitely not, 5 = definitely yes)
	Q5-2: What was your overall experience using generative AI for this assignment? (1 = very negative, 5 = very positive)

Appendix C. Overview of Manually Solved Student Responses

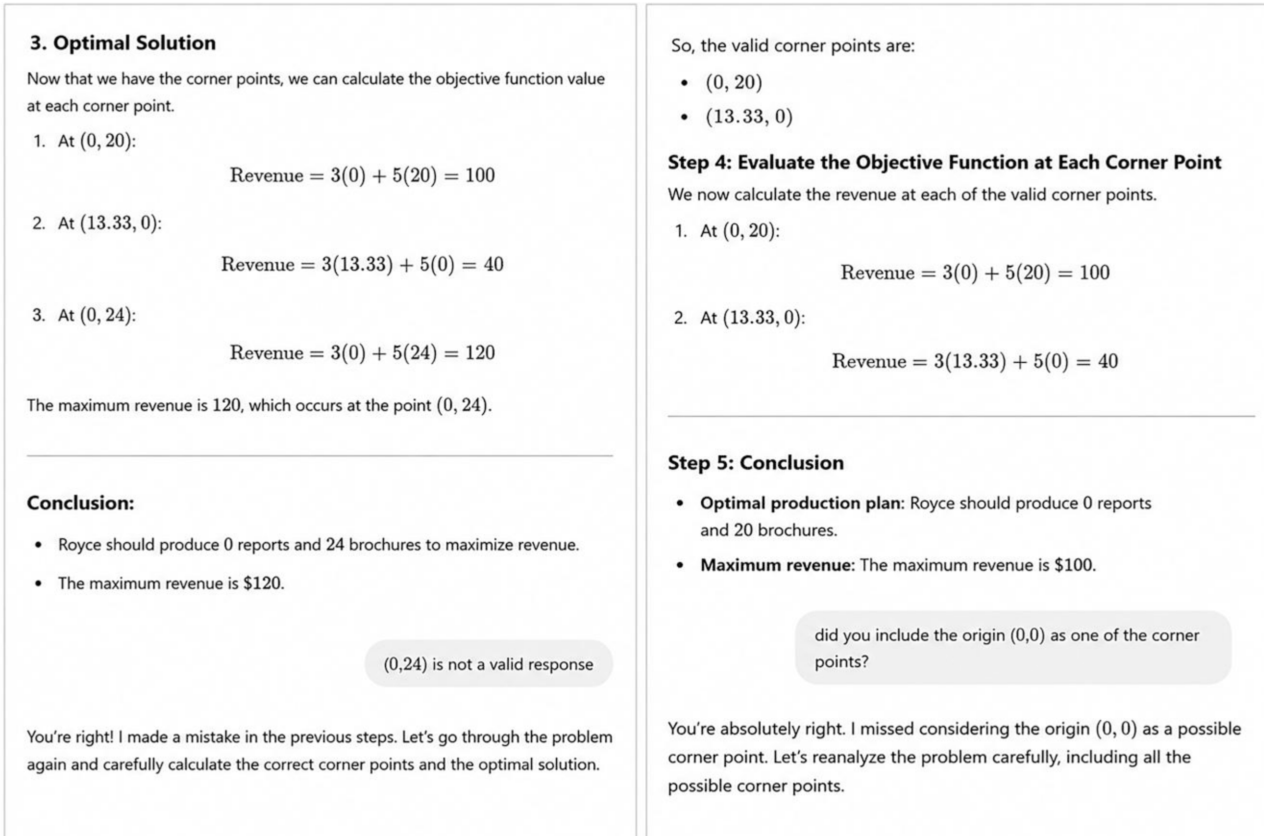
The assignment consisted of three linear programming questions; however, not all students completed every question in the manual problem-solving component. Specifically, 30 out of 35 students submitted responses for questions 1 and 2, whereas 29 students submitted a response for question 3.

Table C.1. Accuracy of Manually Solved Student Responses by Question, Including Nonresponses

Question	Correct (%)	Incorrect (%)	No response (%)	Total
Q1	17 (48.6)	13 (37.1)	5 (14.3)	35 (100)
Q2	15 (42.9)	15 (42.9)	5 (14.3)	35 (100)
Q3	22 (62.9)	7 (20.0)	6 (17.1)	35 (100)

Appendix D. Sample Student-AI Interaction

Figure D.1. Full Student-AI Interaction Corresponding to the Example in Section 3.3



3. Optimal Solution

Now that we have the corner points, we can calculate the objective function value at each corner point.

1. At $(0, 20)$:
$$\text{Revenue} = 3(0) + 5(20) = 100$$
2. At $(13.33, 0)$:
$$\text{Revenue} = 3(13.33) + 5(0) = 40$$
3. At $(0, 24)$:
$$\text{Revenue} = 3(0) + 5(24) = 120$$

The maximum revenue is 120, which occurs at the point $(0, 24)$.

Conclusion:

- Royce should produce 0 reports and 24 brochures to maximize revenue.
- The maximum revenue is $120.

(0,24) is not a valid response

You're right! I made a mistake in the previous steps. Let's go through the problem again and carefully calculate the correct corner points and the optimal solution.

So, the valid corner points are:

- $(0, 20)$
- $(13.33, 0)$

Step 4: Evaluate the Objective Function at Each Corner Point

We now calculate the revenue at each of the valid corner points.

1. At $(0, 20)$:
$$\text{Revenue} = 3(0) + 5(20) = 100$$
2. At $(13.33, 0)$:
$$\text{Revenue} = 3(13.33) + 5(0) = 40$$

Step 5: Conclusion

- **Optimal production plan**: Royce should produce 0 reports and 20 brochures.
- **Maximum revenue**: The maximum revenue is $100.

did you include the origin (0,0) as one of the corner points?

You're absolutely right. I missed considering the origin $(0, 0)$ as a possible corner point. Let's reanalyze the problem carefully, including all the possible corner points.

Appendix E. Full Pairwise Correlation Results Between Students' Perceptions of Generative AI

Table E.1. Pairwise Correlations Among Student Perceptions of Generative AI ($n = 35$)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) Understanding LP	—							
(2) Improved Confidence	0.5962***							
	(<0.001)	—						
(3) Critical Thinking	0.6182***	0.4537**						
	(<0.001)	(0.0060)	—					
(4) Error Identification	0.4189*	0.4893**	0.2446					
	(0.0120)	(0.0030)	(0.1570)	—				
(5) Accuracy	0.5204**	0.4422**	0.5273**	0.2586				
	(0.0010)	(0.0080)	(0.0010)	(0.1340)	—			
(6) Reliability	0.5366***	0.5039**	0.4593**	0.1321	0.8641***			
	(<0.001)	(0.0020)	(0.0060)	(0.4490)	(<0.001)	—		
(7) Future AI Use	0.7002***	0.5708***	0.5881***	0.3124†	0.5720***	0.6234***		
	(<0.001)	(<0.001)	(<0.001)	(0.0680)	(<0.001)	(<0.001)	—	
(8) Overall Experience	0.7258***	0.6527***	0.5325***	0.4851**	0.5920***	0.5692***	0.5681***	
	(<0.001)	(<0.001)	(<0.001)	(0.0030)	(<0.001)	(<0.001)	(<0.001)	—

Note. Values represent Spearman's correlation coefficients, and p-values are shown in parentheses below each coefficient.
$^{\dagger}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Appendix F. Full Correlation Results Between Students' Perceptions of Generative AI and Their Performance Outcomes

Table F.1. Confidence in Identifying AI Errors Was Positively Correlated with Assignment Performance

Category	Survey question	Assignment	Final exam
Impact on Learning	Did using generative AI platforms help you better understand linear programming concepts?	0.0587 (0.7414)	0.2998† (0.0849)
	Do you feel more confident in your ability to solve linear programming problems after validating answers from generative AI platforms?	0.0460 (0.7960)	0.0273 (0.8781)
Critical Thinking	Do you feel that using generative AI platforms improved your critical thinking skills?	0.1379 (0.4367)	0.0107 (0.9523)
	How confident are you in identifying errors in solutions provided by generative AI platforms?	0.4328* (0.0106)	0.1884 (0.2858)
Performance and Reliability	How accurate were the solutions provided by ChatGPT, Gemini, or other generative AI platforms?	0.0162 (0.9276)	−0.1718 (0.3311)
	How reliable do you think generative AI platforms are for solving linear programming problems?	−0.1766 (0.3179)	−0.1550 (0.3814)
User Experience and Future Use	Do you plan to use generative AI platforms for future quantitative assignments?	−0.0666 (0.7082)	0.0123 (0.9449)
	What was your overall experience using generative AI for this assignment?	0.1849 (0.2951)	0.0300 (0.8664)

Note. Values represent Spearman's correlation coefficients, and p-values are shown in parentheses below each coefficient.
**$p < 0.10$; **$p < 0.05$.

References

Binhammad MHY, Othman A, Abuljadayel L, Al Mheiri H, Alkaabi M, Almarri M (2024) Investigating how generative AI can create personalized learning materials tailored to individual student needs. *Creative Ed.* 15(7):1499–1523.

Boyer Commission on Educating Undergraduates in the Research University (1998) Reinventing undergraduate education: A blueprint for America's research universities. Report, State University of New York at Stony Brook, Stony Brook.

Cavus N (2010) The evaluation of learning management systems using an artificial intelligence fuzzy logic algorithm. *Adv. Engrg. Software* 41(2):248–254.

Chitra A, Rajkumar A (2016) Plagiarism detection using machine learning-based paraphrase recognizer. *J. Intelligent Systems* 25(3):351–359.

Cohen J (2013) *Statistical Power Analysis for the Behavioral Sciences* (Routledge, New York).

Dao XQ, Le NB (2023) Investigating the effectiveness of ChatGPT in mathematical reasoning and problem solving: Evidence from the Vietnamese national high school graduation examination. Preprint, submitted June 10, https://arxiv.org/abs/2306.06331.

Davis FD (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quart.* 13(3):319–340.

Deci EL, Ryan RM (2012) Self-determination theory. Van Lange PAM, Kruglanski AW, Higgins ET, eds. *Handbook of Theories of Social Psychology* (Sage Publications, London), 416–436.

Do TD, Shafqat UB, Ling E, Sarda N (2025) Paige: Examining learning outcomes and experiences with personalized AI-generated educational podcasts. Yamashita N, Evers V, Yatani K, Ding X(S), Lee B, Chetty M, Toups-Dugas P, eds. *CHI'25: Proc. 2025 CHI Conf. Human Factors Comput. Systems* (Association for Computing Machinery, New York), 1–12.

Elkhatat AM, Elsaid K, Almeer S (2023) Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. *Internat. J. Ed. Integrity* 19(1):1–16.

Facione P (1990) Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. Report, American Philosophical Association, Newark, DE.

Farquhar S, Kossen J, Kuhn L, Gal Y (2024) Detecting hallucinations in large language models using semantic entropy. *Nature* 630(8017):625–630.

Flaherty C (2025) AI and threats to academic integrity: What to do. *Inside Higher Ed* (May 20), https://www.insidehighered.com/news/students/academics/2025/05/20/experts-weigh-everyone-cheating-college.

Hashem R, Ali N, El Zein F, Fidalgo P, Khurma OA (2024) AI to the rescue: Exploring the potential of ChatGPT as a teacher ally for workload relief and burnout prevention. *Res. Practice Tech. Enhanced Learn.* 19:23.

Herrero F (2025) Generative AI's impact on highly personalized learning. *AMCIS 2025 Proc.*, vol. 4 (Association for Information Systems, New York).

Hoffman RR, Mueller ST, Klein G, Litman J (2023) Measures for explainable AI: Explanation goodness, user satisfaction, mental models, curiosity, trust, and human-AI performance. *Frontiers Comput. Sci.* 5:1096257.

Jauhiainen JS, Garagorry Guerra A (2024) Generative AI and education: Dynamic personalization of pupils' school learning material with ChatGPT. *Frontiers Ed.* 9:1288723.

Ji Z, Lee N, Frieske R, Yu T, Su D, Xu Y, Ishii E, Bang YJ, Madotto A, Fung P (2023) Survey of hallucination in natural language generation. *ACM Comput. Surveys* 55(12):1–38.

Jian JY, Bisantz AM, Drury CG (2000) Foundations for an empirically determined scale of trust in automated systems. *Internat. J. Cognitive Ergonomics* 4(1):53–71.

Kasneci E, Seßler K, Küchemann S, Bannert M, Dementieva D, Fischer F, Gasser U, et al. (2023) ChatGPT for good? On opportunities and challenges of large language models for education. *Learn. Individual Differences* 103:102274.

Kofinas AK, Tsay CHH, Pike D (2025) The impact of generative AI on academic integrity of authentic assessments within a higher education context. *British J. Ed. Tech.* 56(6):2522–2549.

Kosmyna N, Hauptmann E, Yuan YT, Situ J, Liao XH, Beresnitzky AV, Braunstein I, Maes P (2025) Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task. Preprint, submitted June 10, https://arxiv.org/abs/2506.08872.

Lee HP, Sarkar A, Tankelevitch L, Drosos I, Rintel S, Banks R, Wilson N (2025) The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. Yamashita N, Evers V, Yatani K, Ding X(S), Lee B, Chetty M, Toups-Dugas P, eds. *CHI'25: Proc. 2025 CHI Conf. Human Factors Comput. Systems* (Association for Computing Machinery, New York), 1–22.

Li X, Yang Z, Zhang W, Yang Z (2024) ChatGPT and teacher human-machine collaboration for personalized teaching-taking poetry writing teaching as an example. *2024 13th Internat. Conf. Ed. Inform. Tech. (ICEIT)* (IEEE, Piscataway, NJ), 7–11.

Li PH, Lee HY, Cheng YP, Starčič AI, Huang YM (2023) Solving the self-regulated learning problem: Exploring the performance of ChatGPT in mathematics. Huang YM, Rocha T, eds. *Innovative Tech. Learn. ICITL 2023*, Lecture Notes in Computer Science, vol. 14099 (Springer, Cham, Switzerland), 77–86.

Lu J, Schmidt M, Lee M, Huang R (2022) Usability research in educational technology: A state-of-the-art systematic review. *Ed. Tech. Res. Development* 70(6):1951–1992.

Lu J, Zheng R, Gong Z, Xu H (2024a) Supporting teachers' professional development with generative AI: The effects on higher order thinking and self-efficacy. *IEEE Trans. Learn. Tech.* 17:1267–1277.

Lu Q, Yao Y, Xiao L, Yuan M, Wang J, Zhu X (2024b) Can ChatGPT effectively complement teacher assessment of undergraduate students' academic writing? *Assessment Evaluation Higher Ed.* 49(5):616–633.

Marr B (2023) The difference between generative AI and traditional AI: An easy explanation for anyone. *Forbes* (July 24), https://www.forbes.com/sites/bernardmarr/2023/07/24/the-difference-between-generative-ai-and-traditional-ai-an-easy-explanation-for-anyone/.

Moorhouse BL, Wan Y, Wu C, Kohnke L, Ho TY, Kwong T (2024) Developing language teachers' professional generative AI competence: An intervention study in an initial language teacher education course. *System* 125:103399.

Nenkov N, Dimitrov G, Dyachenko Y, Koeva K (2016) Artificial intelligence technologies for personnel learning management systems. *2016 IEEE 8th Internat. Conf. Intelligent Systems (IS)* (IEEE, Piscataway, NJ), 189–195.

Nigam A, Pasricha R, Singh T, Churi P (2021) A systematic review on AI-based proctoring systems: Past, present and future. *Ed. Inform. Tech.* 26(5):6421–6445.

Pascarella ET, Terenzini PT (2005) *How College Affects Students: A Third Decade of Research*, vol. 2 (Jossey-Bass, San Francisco).

Perkins M, Roe J, Postma D, McGaughran J, Hickerson D (2024) Detection of GPT-4 generated text in higher education: Combining academic judgement and software to identify generative AI tool misuse. *J. Acad. Ethics* 22(1):89–113.

Qu X, Sherwood J, Liu P, Aleisa N (2025) Generative AI tools in higher education: A meta-analysis of cognitive impact. Yamashita N, Evers V, Yatani K, Ding X(S), eds. *CHI EA'25: Proc. Extended Abstr. CHI Conf. Human Factors Comput. Systems* (Association for Computing Machinery, New York), 1–9.

Sahu M (2016) Plagiarism detection using artificial intelligence technique in multiple files. *Internat. J. Sci. Tech. Res.* 5(4):111–114.

Slusky L (2020) Cybersecurity of online proctoring systems. *J. Internat. Tech. Inform. Management* 29(1):56–83.

Stokel-Walker C (2022) AI bot ChatGPT writes smart essays—Should professors worry? *Nature*, ePub ahead of print December 9, https://doi.org/10.1038/d41586-022-04397-7.

Takita H, Kabata D, Walston SL, Tatekawa H, Saito K, Tsujimoto Y, Miki Y, Ueda D (2025) A systematic review and meta-analysis of diagnostic performance comparison between generative AI and physicians. *NPJ Digital Medicine* 8(1):175.

Walsh J (2025) Everyone is cheating their way through college. *New York Magazine* (May 7), https://nymag.com/intelligencer/article/openai-chatgpt-ai-cheating-education-college-students-school.html.

Wambsganss T, Janson A, Söllner M, Koedinger K, Leimeister JM (2025) Improving students' argumentation skills using dynamic machine-learning–based modeling. *Inform. Systems Res.* 36(1):474–507.

Wang R, Wang R, Shen Y, Wu C, Zhou Q, Chandra R (2025) Evaluation of LLMs for mathematical problem solving. Preprint, submitted May 30, https://arxiv.org/abs/2506.00309.

Wu F, Dang Y, Li M (2025) A systematic review of responses, attitudes, and utilization behaviors on generative AI for teaching and learning in higher education. *Behav. Sci.* 15(4):467.

Xu F, Lin Q, Han J, Zhao T, Liu J, Cambria E (2025) Are large language models really good logical reasoners? A comprehensive evaluation and beyond. *IEEE Trans. Knowledge Data Engrg.* 37(4):1620–1634.