

Benefits of Critiquing a ChatGPT Generated Essay: Preliminary Results and Insights

Zoe Long, University of Florida

Zoe Long is a PhD student at the University of Florida. They are interested in research about how to improve engineering education in creative ways. Currently, they are working on how large language models can be utilized to improve engineering education with a focus on the student perspective.

Golbarg Nazari, University of Florida

Golbarg Nazari is a Ph.D. student in Industrial and Systems Engineering at the University of Florida. Her research focuses on AI in education and engineering education, with particular interests in student motivation, creativity, and human-centered approaches to learning. Her work examines how innovative instructional strategies, including GenAI-supported, reflective, and creative activities, can enhance student engagement, deepen conceptual understanding, and promote meaningful learning in engineering education.

Addison Grayce Lee, University of Florida

Addison Lee is an Industrial and Systems Engineering undergraduate student at the University of Florida. Alongside her studies she assists Zoe Long and Dr. Akçali in their research efforts through data organization and analysis. Her interest in the complex nature of artificial intelligence, specifically how its usage can both aid and hinder society, has led to her current research in the usage of large language models in the classroom.

Prof. Elif Akçali, University of Florida

Dr. Elif Akçali is an Associate Professor in the Department of Industrial and Systems Engineering as well as The Michael Durham Professor in Creativity in the Engineering Innovation Institute in the Herbert Wertheim College of Engineering at the University of Florida. She is an industrial engineer, a visual artist, and an explorer of the interplay between thinking and making in the arts and engineering. Her scholarly interests focus on the use of arts-based processes for the development of divergent thinking and critical thinking skills of engineering students.

Benefits of Critiquing a ChatGPT Generated Essay: Preliminary Results and Insights

Abstract

ChatGPT has been increasingly popular with students since it became publicly available in 2022. There are lots of different ways ChatGPT and other similar tools can be utilized to improve higher education, but it is important to make sure students use it carefully. To provide a structured introduction to the capabilities of ChatGPT, we designed an assignment for an undergraduate level supply chain course, asking students to work with ChatGPT to create essays and then to analyze them. Each student prompted ChatGPT for the essays using prompts they created based on the assignment requirements, asking them to analyze 160- to 180-word essays that included citations and references. Students then had to complete a sentence-by-sentence analysis to look for both conceptual and grammatical errors. At the end of the semester, students were then asked to reflect on this experience, to gain an understanding of its impact on students' comprehension of the topics as well as understanding of the capabilities of ChatGPT to aid in their education. This study will look at data from the Spring 2024 semester to investigate how students analyzed ChatGPT's work along with their observations about their experiences with the assignment. There were 38 undergraduate industrial and systems engineering students in the course with 31 consenting to having their submitted assignments and end-of-course survey responses used in research. Ultimately, this paper will outline the general impact this assignment has on students and an analysis of essays written by ChatGPT and students' analyses.

Keywords. ChatGPT; Critical Thinking; Undergraduate; Supply Chain

Introduction

Many different Large Language Models (LLMs) and generative artificial intelligence (AI) models have come out in the past couple of years. ChatGPT is one of the most popular tools among students as it was one of the first free-to-use LLMs. Students are regularly using LLMs such as ChatGPT, Copilot, and Gemini to summarize course content and for assistance on assignments [1]. As students use these tools, it needs to be ensured that they are using them productively for education and well for their future careers. Many engineers, particularly in industrial and systems engineering, are hired at consulting firms after graduation, and some of these firms have had trouble because of mistakes made by AI. One example is Deloitte Australia, who had to partially refund 440,000 Australian dollars (\$290,000) to the Australian government in 2025 for a report they provided that had many AI-generated errors [2]. If students are not trained to look for and correct mistakes made by AI, they could put themselves in precarious situations in their future careers. Students need to know how to be mindful when using ChatGPT and other LLMs.

To this end, an assignment was designed and implemented in a junior/senior level course on inventory control and supply chain management. The assignment asked students to prompt ChatGPT for a short essay about ABC analysis, a key concept of inventory control. The model of ChatGPT available at the time students were completing this assignment was GPT-3.5. As of January 24, 2026, the model of ChatGPT available under their unpaid version is GPT-5.2 which is just the newest of many models that have come out since GPT-3.5. Once students had the ChatGPT generated essay, they were to analyze it sentence by sentence, looking for errors in the information as well as assessing the accuracy of the references given by ChatGPT. At the end of the semester, students were surveyed to understand the impact of the assignment on students' learning as well as their opinions on the use of ChatGPT in their education. This paper aims to explore if and how this assignment can support students' learning of a conceptual technical topic and whether the students' survey responses indicate if and how critical thinking skills (interpretation, analysis, inference, evaluation, explanation and self-regulation) and behaviors (such as examining arguments, identifying assumptions, seeking information, assessing the credibility and relevance of information, drawing supported conclusions) aided in their learning.

Student survey responses and student work were qualitatively analyzed to investigate the following research questions (RQs):

RQ1. Do students think that the ChatGPT-based assignment adds to their education in a way other assignments do not?

RQ2. In what ways did the ChatGPT-based assignment help students with their understanding of the topic?

RQ3. How does the proposed ChatGPT-based assignment engage students' critical thinking skills?

The rest of the manuscript is organized as follows. In the next section, literature related to ChatGPT use in education is discussed. Following the related literature is a discussion of the data collection methodology used. Then is data analysis of the students' responses to both the assignment and survey. Next is the results and discussion section, to speak on the outcomes of the data analysis. For instructors who may want to implement something similar in their own courses, we have included recommendations for implementation as well. Finally, the conclusion summarizes the paper along with providing limitations of the research and future directions.

Related Literature

Since the release of ChatGPT in late 2022, the use of it and other generative AI tools (such as Gemini and Copilot) has become popular [3]. This popularity extends into the engineering undergraduate student body as one study found that of a population of undergraduate mechanical engineering students, 32.6% used generative AI tools daily with another 29.1% using them weekly [1]. It is clear: Students are using these tools regularly, and it is engineering educators who need to make sure that they are using them responsibly for their education. Some concerns

have been raised over student's overreliance and dependency on these tools, with another study finding that even when students were told that ChatGPT can make mistakes, most students will agree with incorrect answers if they know the answers are provided by ChatGPT [4].

There exists a number of studies in the literature that investigate how ChatGPT can be incorporated into existing courses to improve students' problem solving and critical thinking skills. For example, one study looked at how to incorporate ChatGPT into a system that is already in place to improve students' ability to solve problems in chemistry [5]. They found that it can be a useful tool but needs to be carefully scaffolded to ensure students can identify the errors properly. Another study found that when correcting an essay written by ChatGPT, students' confidence in multiple critical thinking skills significantly improved [6].

Data Collection

Course. The course in which this research was conducted has the aim of developing students' ability to model and analyze problems relating to supply chain and logistics problems that are common in national and global trade systems. The specific topics of this course include a discussion of the concerns related to introduction to inventory management systems; introduction of a framework for the classification of inventory models; deterministic inventory modeling; stochastic inventory modeling; logistics system design; supply chain management and system design; and demand forecasting.

Topic. One of the first topics addressed to introduce students to inventory management systems is the ABC Analysis. The discussion starts with a presentation of the historical context of the Pareto Analysis and the wide applicability of the 80-20 Rule. Next, the mechanics of distribution value analysis (DVA) are described, and how its results can be used in conjunction with cut-off points to divide items in an inventory system into Class A, B, and C. Also, a demonstrative example is used to show how DVA can be applied and how the classification of items can be done. The students are reminded that the output of the DVA must be examined closely. Finally, discussion concludes pointing out the need for a human decision maker to (i) examine where the cut-off lines are and whether there is a need to adjust the cut-off lines slightly to reclassify some of the items around the cut-off points; and (ii) comb through different classes of items to selectively change the class assignments for some items as necessary.

Assignment. In Spring 2024 semester, instead of asking students to write an essay describing ABC Analysis, the students were tasked with prompting ChatGPT to generate a 160- to 180-word long essay about ABC analysis for inventory management with several citations and references to support it. In addition, students were to analyze the ChatGPT generated response for technical accuracy, how relevant references were, and whether the references were properly cited. Students had access to the slide deck that was used for in-class discussion along with additional detailed course notes provided by the instructor while completing the assignment.

Note that the assignment was especially designed to give students an opportunity to practice and demonstrate their critical thinking and creative problem-solving skills. The assignment asked students to demonstrate their critical thinking skills because they had to analyze and evaluate each sentence of the essay based on their own understanding of the topic as well as the information provided to them in the course to verify each sentence in the ChatGPT output. Further, students also had to decide whether each sentence was correct or not. Similarly, the assignment required students to utilize their creative problem-solving skills because the students had to suggest edits for the sentences that they deemed to be incorrect (see Table A.1 in Appendix A for detailed student instruction used in Spring 2024 semester).

Grading. Student assignments were graded out of a total of 10 points: 4 points for turning in an essay provided by ChatGPT and 6 points for the students' review of ChatGPT output. Also, one point was deducted, (i) if the references were obviously unrelated to ABC Analysis and had not been pointed out by the students, or (ii) if the ChatGPT-generated essay did not include full references, and students did not comment upon it.

Student Survey. In the Spring 2024 semester, no other assignment in this course asked students to utilize ChatGPT. At the end of the semester, students were asked to complete a survey about the assignment (see Table A.2 in Appendix A). For the purposes of this paper, the focus will be on the first question of the survey.

Study Data. Data collection and analysis protocols were reviewed and approved as exempt by the Institutional Review Board at the institution where this study was completed (Protocol Number: suppressed to preserve anonymity). In Spring 2024, 38 students were enrolled in the course, and, at the end of the semester, 31 students gave their consent for their data to be included in the study whose assignment submissions and responses to the survey questions were entered as data for the study. One of the students who submitted the assignment did not complete the survey submission. That is, there are 31 assignment submissions but only 30 survey responses in the data. Participants' information was deidentified and randomly assigned a number between 1 and 31. In the remainder of this document, respondents will be referenced with a number between 501 and 531.

Data Analysis

Study data was analyzed in several stages:

Gains in Understanding. To investigate the research question, "Do students report gains in their understanding of the material due to this assignment?" responses were coded independently by three coders considering three coding options: (i) positive; (ii) mixed (somewhat positive and negative); and (iii) negative. The coders discussed independent coding results, and coding discrepancies were managed by discussing interpretations and justifications to reach consensus.

For instance, a participant (526) stated, "Doing this assignment helped me understand the topic a little better. It basically complemented what I had already learned, by saying the same thing in a

different way it helped me understand better.” This response was assessed to be positive by all of the coders. Similarly, another participant (511) said, “For me, having to review the ABC analysis for the assignment helped reinforce my knowledge of the system. I don't think I found the actual reading and editing of ChatGPT to be very helpful in this case though.” This response was assessed to be somewhat positive and negative by all the coders, since while the participant thought that the assignment contributed to their understanding of the topic by requiring them to review the material, but they did not think that the editing process was very helpful.

Enhancing Critical Thinking. To investigate the research question, “Do students think that the assignment provided them with an opportunity to practice critical thinking skills?” responses were coded independently by two coders (ZL,GN) considering two coding options: (i) yes; (ii) no. A response that lists one or more processes (e.g., examining the validity of an argument, conducting research to find evidence, interpreting the results of the research, making a reasoned judgement about the argument) related to critical thinking received the code “yes” and “no” otherwise. These processes were chosen based on the definition of critical thinking in the literature provided by Facione based on the findings of the Delphi project, which outlined the aspects of critical thinking as being interpretation, analysis, inference, evaluation, explanation and self-regulation [7].

For instance, a participant (516) stated, “Looking for errors in the work required me to get an understanding of the topic beyond what the essay mentions or even beyond what is taught in class, I did some research on the analysis to verify whether some statements were correct or not. Since I did supplemental research, I gained a greater understanding of the topic. As well I reviewed the noted multiple times and therefore gained a deeper understanding on the topic.” This response was assessed to reflect critical thinking, as the participant states that to verify if the information in some of the sentences in the essay were accurate, they felt the need to conduct supplemental research and enhanced their understanding. In doing so, the participants described the aspects of critical thinking, evaluation, analysis, and self-regulation as they reviewed the essay and assessed that they needed more knowledge to confirm the truth of the information the essay provides. In contrast, for instance, another participant (517) noted, “This chat gbt assignment helped me understand the things I did not ABC analysis.” This response was assessed to include no explicit expression on the practice of critical thinking and would receive a code of “no”, although the participant notes that their understanding improved.

Thematic Analysis. Participant responses were qualitatively analyzed by three researchers (EA, ZL, GN) to identify thematic content. The researchers coded the data independently and met to discuss and agree upon the definitions and codes associated with the themes observed in the data.

Analysis of ChatGPT Output. The ChatGPT essays were studied by one researcher (EA), who was also the instructor for the course in Spring 2024, in detail to gain a deeper understanding of common mistakes in ChatGPT-generated essays using thematic content analysis. All ChatGPT-generated essays were read several times, and types of information that essays included were

identified. A subset of these types of information was examined in further detail to understand what types of errors exist and how they could lead to misunderstandings.

Analysis of Student Analysis of ChatGPT Output. Student analyses of ChatGPT generated essays were also examined by one researcher (EA) to assess to what extent the students can detect inaccuracies in ChatGPT generated essays.

Results and Discussion

Gains in Understanding. In this study, 77% of the responses (23 out of 30) received a positive code, as these participants stated that the assignment contributed to their understanding of the topic in some way. The remaining 23% of the responses (7 out of 30) received a mixed (positive with some limitations, such as credibility concerns and difficulty validating information) code as these participants also noted the assignment furthered their understanding of the topic but noted some drawbacks also. In this data set, there were no responses that received a negative score, i.e., there were not participants who noted only drawbacks.

Table I Frequency distribution of Gain in Understanding codes

Code	Count
Positive	23
Mixed	7
Negative	0

Critical Thinking. Study data indicated that 63% (19 out of 30) of responses discussed activities completed in doing the assignment that allowed them to practice their critical thinking skills based on them receiving the code “yes”. The other 37% (11 out of 30) of participants received the code “no” because they did not discuss activities that would allow them to practice their critical thinking skills in their survey responses. These findings showed that the assignment given to students was able to make them practice their critical thinking skills, mainly through evaluation as students were asked to look through the essays and find errors. An example of how this critical thinking also could help their understanding was given by participant (512) when they said “[the assignment] helped further my understanding in the same way that teaching others helps me learn a topic rather than studying on my own”. Participants who made no mention of activities that require critical thinking focused on either the support ChatGPT provided for their learning through offering a varied explanation to that given in class or simply stated that having to review the material more helped in their learning.

Thematic Analysis. We conducted a thematic analysis [8] of students’ open-ended reflections describing how ChatGPT generated essay affected their understanding of ABC analysis and their perceptions of ChatGPT as a learning tool. Six interrelated themes were identified:

Understanding Technical Topic. Most students described that the assignment improved or reinforced their understanding of ABC Analysis. Students frequently reported that the activity clarified key concepts, reinforced recall, or provided an alternative explanation and examples that complemented course instruction. For many students, the activity promoted learning by requiring them to revisit the essay and clearly differentiate accurate concepts from misconceptions.

Understanding ChatGPT Capabilities/Limitations. The theme described students' calibration of trust in ChatGPT. Students reported that ChatGPT could be effective for generating explanations and offering alternative wording, but several emphasized that the tool can be unreliable especially regarding citations, references, and factual precision. These reflections indicate that the assignment provided an early opportunity for students to calibrate ChatGPT's capabilities by learning when it can support understanding and when its responses must be verified, given risks such as fabricated citations and unsubstantiated claims.

Competency. To interpret students' reflections through a motivational lens, we drew on Self-Determination Theory (SDT) [9, 10], which describes competence as a learner's sense of effectiveness and mastery when engaging with a task. In SDT, competence refers to a learner's sense of effectiveness and mastery when engaging with a task; in our data, students expressed increased confidence, deeper understanding, or reinforced mastery after completing the activity. Accordingly, competence-related statements were observed in all responses (30 out of 30). Therefore, this assignment appeared to support competence by helping students feel more capable of explaining ABC analysis and distinguishing between correct and incorrect statements. Notably, students described competence gains both as learning new content and as solidifying or confirming knowledge they already had.

Autonomy. To further interpret students' reflections through an SDT lens, we examined autonomy, defined as a learner's sense of choice, volition, and self-directed use of ChatGPT or self-directed learning. In our data, students highlighted controlling the interaction (e.g., requesting rephrases, adjusting the level of detail, or iterating prompts to obtain alternative explanations). Others described autonomy through actions beyond the generated essay, such as independently researching uncertain points, returning to lecture notes, and forming their own conclusions rather than accepting ChatGPT's output at face value. Together, these responses suggest that autonomy was enacted when students positioned themselves as decision-makers in evaluating and refining information.

Critical Thinking. In many responses, students described behaviors consistent with evaluation and self-regulation, such as verifying claims line-by-line, checking references, comparing content to notes, researching uncertain statements, and searching for errors. These students often described stronger perceived learning gains, suggesting that competence was supported not simply by exposure to an explanation, but through an evaluative process that required students to justify correctness. Additionally, autonomy appeared to interact with critical thinking, as students

who reported greater control (e.g., prompt iteration, independent follow-up research, or forming their own conclusions) were more likely to describe verification behaviors. Overall, our analysis suggests that competence and autonomy are not only outcomes of AI-assisted learning but also appear to develop alongside critical thinking practices such as verification, comparison to course materials, and error-checking that enable students to use ChatGPT more responsibly, critically, and effectively.

Analysis of ChatGPT Output. Based on the thematic content analysis of the student submissions, following observations can be made:

What types of information are in ChatGPT generated essays? A typical ChatGPT generated essay in the study data included statements about (i) the role of ABC Analysis in inventory system management; (ii) how ABC Analysis categories items into three groups of A, B, and C; (iii) the characteristics of different classes of items; (iv) the relationships between ABC Analysis and Pareto Analysis and/or 80/20 rule; (v) how ABC Analysis help facilitate the allocation of inventory-related resources and attention to different classes of items; and (vi) how ABC Analysis can help optimize operational performance and efficiency.

None of the essays mentioned the potential benefit of scanning through the three classes to examine if any items are misclassified, which is an important practical consideration for ABC Analysis that points to a potential weakness. It is not clear if this issue did not appear in the essays in the study data either because the word limit was too limiting or because ChatGPT did not have sufficient access to any resources that addressed the issue of potential misclassification of items.

How do ChatGPT generated essays characterize different classes of items? A type of information in the ChatGPT essays that could lead to misconceptions and misunderstandings related to the characterizations of Class A, B, and C items. ChatGPT generated essays in the study data included 1, 2, or 3 sentences to describe the characteristics of different classes of items. For instance, a submission (505) included only one sentence that read, “Typically, items are categorized into three groups: A, representing high-value items with lower frequency; B, moderate-value items with moderate frequency; and C, low-value items with high frequency,” which presented attributes for the items that belong to all three classes. Similarly, for instance, another submission (524) included two sentences that stated, “Category A comprises high-value items that contribute significantly to overall costs but represent a small portion of the inventory. Category B includes moderate-value items, while Category C involves low-value items that constitute a large proportion of the inventory.” Table II summarizes the instances encountered in the study data. For example, 4 of the essays had only one sentence in the essay, where a sentence about B and C Class items was preceded as well as followed by two sentences about Class A items. Similarly, for example, 7 of the essays had two sentences on the characteristics of the items in different classes. While for 6 of these essays, the sentence mentioned all classes A, B, and C, the other one mentioned only classes A and C.

Table II The frequency distribution of the number of sentences on the properties of Class A, B, and C items along with the encountered patterns and associated counts for each pattern.

Number of sentences	Count	Pattern (Count)
1	7	ABC (6) AC (1)
2	20	A, BC (15) AC, B (2) ABC, A (1) ABC, BC (1) ABC, ABC (1)
3	4	A, BC, A (4)

The challenge this class of sentences posed relates to the technical ambiguity of the language, which may lead to misconceptions and misunderstandings. For instance, ABC Analysis relies on the unit cost of an item as well as its annual usage volume, the multiplication of which provides a measure for the “annual consumption value” which is a measure of the item’s contribution to the value of the inventory moved in the system. In “Class A includes high-value items, Class B comprises items of moderate value, and Class C consists of lower-value items.” (502), it is not clear what the term “value” refers to exactly. The sentence “Category A comprises high-value items with a significant impact on total inventory costs, while Category C includes low-value items with minimal impact.” provides a bit more context suggesting that the value of an item is related to its impact on total inventory costs which is in contrast to “Group A comprises high-value items with a significant impact on overall costs and profits, while Group C consists of low-value items with minimal impact.” (508) that mentions overall costs and profits. However, none of these examples refer to the item’s “usage volume” or “usage frequency.” In fact, in the entire data set for the study, only “Class A items are high-value goods with low frequency, B items are moderate in both value and usage, while C items are low in value but have high usage frequency.” (506) makes an explicit reference to usage frequency. In “Category A comprises high-value items with a significant impact on overall costs but a lower frequency of occurrence.” (526) for instance, it is unclear what “frequency of occurrence” refers to. Hence, it can be concluded that ChatGPT generated essays lack the clarity and precision needed for technical writing. This, in turn, leads to two issues. First and foremost, it seems lack of context, insufficient information or language that can be interpreted in multiple ways may lead to ambiguity which might prove to be difficult for a beginning learner to detect and resolve. As a result, a response that can erroneously be classified as being correct/precise by a student can reinforce their understanding of the topic incorrectly/impactly, and, hence, lead to misconceptions and misunderstandings. Furthermore, while some students (7 out of 31) had only one such sentence to analyze and edit if they could discern the ambiguity, few others (3 out of 31) had to analyze 3 such sentences. That is, the number of such sentences implies varying workloads for the students.

How are the benefits of using ABC Analysis captured? Another common theme in ChatGPT generated essays in the study data related to the sentences that state claims about the potential impact or benefits of ABC analysis in operations management. ChatGPT generated essays included 2 or 3 sentences on this aspect. For instance, a submission (509) included three sentences that read, “Originating from the Pareto Principle, ABC Analysis ensures a focus on the most critical aspects of inventory control, optimizing resources and enhancing decision-making. Researchers such as Cox and Blackstone (2004) highlight its effectiveness in achieving cost-efficiency and improved supply chain performance. Implementing ABC Analysis empowers organizations to streamline inventory management, reduce costs, and enhance overall operational effectiveness.”

According to these sentences ABC Analysis helps with “optimizing resources,” “enhancing decision-making,” “achieving cost-efficiency,” “improved supply chain performance,” “streamline(d) inventory management,” “reduce(d) costs,” and finally “enhance(d) overall operational effectiveness.” While some of these claims can be justified by thinking through to see if a link can be built between outcomes of the ABC Analysis and the claimed impact or benefit, most of these claims are overarching generalizations that could be said for any inventory control or supply chain practice that could help improve operational efficiency somehow. Furthermore, there is some redundancy too. Perhaps a single sentence such as “Originating from the Pareto Principle, ABC Analysis ensures a focus on the most critical items in an inventory management system, which could ultimately lead to enhanced operational effectiveness.”

How reliable are the citations and references? One of the objectives of the assignment was to help students understand that when prompted ChatGPT has the capability to provide a list of citations that may appear to be legitimate to a person who is not sufficiently familiar with the field [11], however, either the citations do not exist or they exist but are not related to the topic. In the study data, numerous instances were encountered. For instance, there were citations that simply did not exist (e.g., 526 provided a citation and a full reference a journal publication that does not exist) or were in another language (e.g. 512 and 523 cited Pareto’s book in French) or were related to Activity-Based Costing method which shares the same acronym as the ABC Analysis (e.g., 526 cited Cooper and Kaplan’s paper) or were to fundamental textbooks related that are related to supply chain but do not discuss ABC Analysis (e.g., 512 cited Chopra and Meindl’s textbook).

While citations that did not exist may have been relatively easy to verify, those that existed but the participants may not have had the means to access and/or may have had the access but not the means to verify the appropriateness of the citation by reading the entire source were more difficult to do so.

A Preliminary Exploration of Student Analysis of ChatGPT Output. A subset (11 out of 31) of the student analyses of ChatGPT generated essays has been studied in detail. That is, one coder

(EA), the instructor for the course, analyzed each sentence in each of the ChatGPT generated essays and suggested edits. Next, student- and coder-suggested edits were compared for overlap.

For instance, one participant (526), whose essay had 9 sentences (see Table B.1 in Appendix B), noted that 2 of the sentences were not accurate and suggested edits, whereas the coder suggested edits for 8 of the sentences. Hence, this participant could not identify the ambiguities and the impreciseness in the language. Also, the coder marked one of the suggested edits of the participant as being incorrect as well. Furthermore, this student's response to the survey prompt read, "Doing this assignment helped me understand the topic a little better. It basically complemented what I had already learned, by saying the same thing in a different way it helped me understand better." which received a code of "no" for critical thinking. For this participant, the submitted work missed a lot of the inaccuracies in the essay, and the response to the survey prompt did not show an engagement level that demonstrated critical thinking.

Similarly, for example, participant (521), who had an essay with 8 sentences (see Table B.2 in Appendix B), noted that 4 of the sentences were correct and 4 of them were only partially correct and suggested edits. They thought that one was correct but noted that the content extended beyond the scope of the discussion in the class. The coder also noted that 4 of the sentences were correct and agreed with 2 of the suggested edits but suggested additional edits for the other two. For the last sentence, the coder noted that the sentence was problematic and suggested editing, agreeing with the student that the statement made an overarching statement that should be edited carefully. This student's response in the survey reads, "I think this assignment definitely deepened my understanding of ABC analysis. I understood it fairly well before the assignment, but when I used ChatGPT, I felt incentivized to try and prove ChatGPT wrong and therefore worked harder to understand the concept so that I could find errors. I was then able to find errors, which was exciting because I am a big supporter of ChatGPT and love using it," which received a code of "yes" for critical thinking. For this participant, the submitted work captured a lot of the inaccuracies in the essay, and the suggested edits demonstrated a nuanced understanding of the concept. Furthermore, the corresponding response to the survey prompt showed an engagement level that demonstrated critical thinking.

Recommendations for Implementation

While the student work in the study data provides a direct assessment of student learning, the analysis of ChatGPT generated essays as well as the analysis of the student provided responses require a substantial effort, even for a relatively short essay of 160- to 180-words long. The idea of asking students to study a piece of technical writing to identify potential ambiguities and inaccuracies appears to be an effective exercise. However, as with any assignment that demands creative problem solving from the student, assessment of student work places a big burden on the instructor. To this end, the instructor may create the essay in real-time on the screen in the classroom or create it in advance and edit as necessary to add more or remove some of the

ambiguities and inaccuracies in the essay, i.e., design the assignment as suitable, and share with the class as part of the assignment instructions.

When the study data was collected, the students were asked to experiment and find an appropriate prompt but were not required to capture this exploratory process and submit it as part of the assignment. It is well known that prompt engineering is becoming an increasingly important aspect associated with productive and effective engagement with generative AI tools. Hence, it would be useful for students to submit the prompts they used to obtain the essays for assignments similar to the one investigated in this paper to see how the wording of the prompt influences the quality of the LLM generated essays.

Conclusions

This study offered some evidence on the benefits of assignments that use ChatGPT as a tool for students to use analysis to reinforce their understanding. Most participants reflected on the assignment on having had a positive impact on their learning, partially answering RQ1. In answer to RQ3, participants' survey responses were analyzed to understand the impact the assignment had on their critical thinking skills. It was found that over 60% of participants discussed critical thinking skills they had to engage in completing the task that aided in their overall understanding of the topic. To answer RQ2, a thematic analysis was conducted on participants' responses to understand the impact of the assignment on students. This open-ended analysis identified six themes that will be useful to look deeper at with further research. Then an analysis of the responses that participants received from ChatGPT was completed to understand the quality of information ChatGPT provides students.

Study data was collected in January 2024 when GPT-3.5 was the publicly available version. Since then, ChatGPT's capabilities are reported to have been improved substantially. Hence, if the data collection were to be repeated, the issues identified by the thematic content analysis of the ChatGPT generated essays reported in this paper may no longer be relevant. However, the analysis still offers some benefits as it provides ample evidence on how ambiguous and imprecise language can be in essays generated by LLMs on technical topics, as these LLMs rely on publicly available resources for training, and inaccuracies that may exist in such resources might get propagated and find their way on students' desks and lead to misconceptions and misunderstandings rather than helping students learn a topic deeply.

References

1. H. Akolekar, P. Jhamnani, V. Kumar, V. Tailor, A. Pote, A. Meena, K. Kumar, J.S. Challa, D. Kumar, "The role of generative AI tools in shaping mechanical engineering education from an undergraduate perspective," *Scientific Reports*, Vol. 15, No. 9214, Mar. 2025, <https://doi.org/10.1038/s41598-025-93871-z>

2. R. McGuirk. "Deloitte Australia to partially refund \$290,000 report filled with suspected AI-generated errors." *AP News*. <https://apnews.com/article/australia-ai-errors-deloitte-ab54858680ffc4ae6555b31c8fb987f3> (accessed January 22, 2026).
3. D. Baidoo-Anu, L. Owusu Ansah, "Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning," *Journal of AI*, Vol. 7, pp. 52-62, Dec. 2023, DOI:[10.61969/jai.1337500](https://doi.org/10.61969/jai.1337500)
4. L. Ding, T. Li, S. Jiang, A. Gupad, "Student's perceptions of using ChatGPT in a physics class as a virtual tutor," *International Journal of Educational Technology in Higher Education*, Vol. 20, No. 63, Dec. 2023 <https://doi.org/10.1186/s41239-023-00434-1>
5. B. Exintaris, N. Karunaratne, E. Yuriev, "Metacognition and critical thinking: Using ChatGPT-generated responses as prompts to critique in a problem-solving workshop (SMARTCHEMPer)," *Journal of Chemical Education*, Vol. 100, pp. 2972–2980, July. 2023, <https://doi.org/10.1021/acs.jchemed.3c00481>
6. Y. Guo, D. Lee, "Leveraging ChatGPT for enhancing critical thinking skills," *Journal of Chemical Education*, Vol. 100, pp. 4876-4883, July. 2023, <https://doi.org/10.1021/acs.jchemed.3c00505>
7. P.A. Facione, "Critical thinking: What it is and why it counts," *Insight Assessment*, 2015.
8. J. Saldana, *The Coding Manual for Qualitative Researchers*, 4th ed. SAGE Publications Ltd., 2021.
9. C.P. Niemiec, R.M. Ryan, "Autonomy, competence, and relatedness in the classroom: Applying self-determination theory to educational practice," *Theory and Research in Education*, Vol. 7, No. 2, pp. 133–144, 2009
10. R.L. Kajfez, H.M. Matusovich, "Competence, autonomy, and relatedness as motivators of graduate teaching assistants," *Journal of Engineering Education*, Vol. 106, No. 2, pp. 245–272, 2017
11. Anonymous. (2025) "Why ChatGPT creates non-existing scientific citations (and how I learned this the hard way)." *Medium*. <https://medium.com/@reviewraccoon/why-chatgpt-creates-non-existing-scientific-citations-and-how-i-learned-this-the-hard-way-894c864b92d8> (accessed January 24, 2026).

Appendix A: Data Collection Tools

Table A.1. Detailed student instructions.

Task 1. If you do not have one already, please create an account on ChatGPT. Ask ChatGPT to create a 160–180-word essay describing the ABC Analysis for inventory management and ask ChatGPT to support the essay with appropriate references from the literature. You may experiment with a couple of different prompts until you obtain an essay that you like. Please copy and paste the response you obtain in the box below.

Task 2. Recall our in-class discussion and review the lecture notes on ABC analysis in inventory management. Revisit the essay you obtained from ChatGPT. Read this essay carefully, and for each sentence in this essay. Answer the following questions:

1. Verify each sentence by answering the two following questions:
 - a. Based on your understanding of the course material, is this sentence correct? Please note that if a sentence is partially correct, then your answer to this question should be no or partially.
 - b. If any references are provided by ChatGPT to support the argument in this sentence, do these references exist and are they relevant?
2. For any sentence in the ChatGPT generated essay that you could not verify in Step 1 above, suggest an edited version of the sentence with improved accuracy.
3. Please copy and paste your sentence-by-sentence analysis and improvement process in the box below.

Table A.2. Student survey prompts.

No.	Prompt
1	In Homework Assignment 1, you were asked to (1) have ChatGPT generate an essay for you on ABC analysis and (2) review it line by line for the accuracy of the content of this essay and verify the references provided by ChatGPT. How did this assignment impact your understanding of ABC analysis? How did it, if it did, support and/or complement your understanding of the topic?
2	How did Homework Assignment 1 impact your understanding of ChatGPT's potential to help you learn technical topics? In your opinion, what are some shortcomings of ChatGPT? In your opinion, what are some strengths of ChatGPT?
3	How else might ChatGPT be used to support and/or complement teaching and learning of the material in EIN4343 Inventory and Supply Chain Systems?

4	What are some additional topics that could have benefited from ChatGPT-based assignments to support and/or complement teaching and learning of the material in EIN4343 Inventory and Supply Chain Systems?
---	--

Appendix B: Sample Student Work

Table B.1. Student Analysis of ChatGPT Output: 526

ABC analysis is a widely used inventory management technique that categorizes items into three groups based on their significance. yes 1. Developed by H. Ford Dickie in the 1950s, it classifies inventory items as A, B, or C, reflecting their respective impact on overall inventory costs. <i>No; The ABC analysis was developed by Pareto, and it was then further developed by H. Ford Dickie in the 1950s</i> 2. Group A comprises high-value items that contribute significantly to overall costs but represent a small proportion of the total inventory. yes 3. Group B consists of moderately important items, while Group C includes low-value items that constitute a large portion of the inventory but contribute less to overall costs. yes 4. This classification enables businesses to prioritize resources and attention. Yes 5. Effective implementation of ABC analysis allows for better control over high-value items, minimizing the risk of stockouts or overstocking. yes 6. Studies such as the one by Pareek and Sirohi (2012) emphasize its relevance in optimizing inventory management strategies. <i>No; The study mentioned above is not real, but it is true that ABC analysis optimizes inventory management strategies. ABC analysis can help businesses better manage their inventory levels and make better decisions about which items to stock and how to manage them.</i> 7. By focusing efforts on managing Group A items more efficiently, companies can achieve cost savings and improve overall supply chain performance. <i>yes; The reference provided does not exist and was given as an example, but the sentence above is still relevant.</i> 8. In conclusion, ABC analysis remains a valuable tool in inventory management, aiding organizations in making informed decisions to enhance efficiency and reduce costs. yes
--

Pareek, V., & Sirohi, S. (2012). A study of inventory management system in selected small and medium enterprise (SME) sector in Mumbai. International Journal of Latest Trends in Engineering and Technology, 2(1), 192-197 ,

The reference provided does not exist and was given as an example.

Table B.2. Student Analysis of ChatGPT Output: 521

ABC Analysis, a cornerstone of inventory management, categorizes items based on their significance to overall business performance (Anderson & White, 1997).

This sentence is fully correct. However, the work cited does not actually exist! This really surprised me since it sounds so real. There is a work titled How the 'ABC' inventory system can save time, money, and manpower from 1979 by an L. G. Anderson, but these are not even the same initials given by ChatGPT.

1. This classification divides inventory into three categories: A, B, and C, representing high, medium, and low-value importance items, respectively.

Partially correct. A more accurate version of the sentence would swap the word "value" for importance.

2. This strategic approach allows businesses to allocate resources effectively, prioritizing items with a substantial impact on revenue.

This sentence is fully correct.

3. Rooted in the Pareto Principle, the method acknowledges that approximately 80% of effects result from approximately 20% of causes (Pareto, 1896). ,

I would say this sentence is partially correct. I would add the word "approximately" to better reflect how the ABC method works. You do not always set the line at exactly 20% This citation actually is correct, referencing Pareto's work from 1896. It is a relevant citation to introduce his principle.

4. Category A items, while constituting a small portion of inventory, contribute significantly to revenue.

This sentence is fully correct.

5. Category B items hold moderate importance, and Category C items, though low in value, are crucial for operational efficiency are of little importance.

This end of this sentence is completely wrong. We definitely did not learn that Category C items "are crucial for operational efficiency." I would remove the red text and change it to saying the Category C items are of little importance.

6. Focusing the most efforts on Category A items enables companies to optimize working capital and minimize carrying costs, fostering a well-balanced inventory; Nonetheless, Category B items require occasional attention and Category C items require the minimum level of tracking.

I would say this sentence is partially correct but some more detail could be added. Companies should focus most on Category A, but should still give some occasional attention to Category B.

7. Focusing on efficient inventory management through ABC Analysis not only enhances financial performance but also provides a strategic framework for businesses to adapt and respond to dynamic market demands, fostering long-term sustainability.

This sentence is a bit past the scope of what we learned in class this week, but I think it is correct. I would think a company that is well organized through the ABC method would be able to shift dynamically if needed, which would add to their longevity.

References:

- Silver, E. A., & Peterson, R. (1985). Decision Systems for Inventory Management and Production Planning. John Wiley & Sons.
- Pareto, V. (1971). Manual of Political Economy. Macmillan.