Engineering Educators Bringing the World Together
2025 ASEE Annual Conference & Exposition
Palais des congrès de Montréal, Montréal, QC • June 22–25, 2025
Paper ID #48416

# Generative AI as a Tool for Effective Problem Generation in Engineering

**Runu Proma Das, University of Georgia**
**Ms. Tathyana Moratti, University of Georgia**

PhD student

**Shari Gasper, University of Georgia**

Fourth year Cognitive Science and Psychology Double Major at the University of Georgia, MODEL Undergraduate Research Assistant

**Dr. Beshoy Morkos, University of Georgia**

Beshoy Morkos is an associate professor in the College of Engineering at the University of Georgia where he directs MODEL (Manufacturing Optimization, Design, and Engineering Education Lab) Group. His research group currently explores the areas of system design, manufacturing, and their respective education. His system design research focuses on developing computational representation and reasoning support for managing complex system design through the use of Model Based approaches. The goal of Dr. Morkos' manufacturing research is to fundamentally reframe our understanding and utilization of product and process representations and computational reasoning capabilities to support the development of models which help engineers and project planners intelligently make informed decisions. On the engineering education front, Dr. Morkos' research explores means to improve persistence and diversity in engineering education by leveraging students' design experiences.

# Generative AI as a Tool for Effective Problem Generation in Engineering


## Abstract

The generation of engineering problems is an essential step in an effective problem-solving process. However, interpreting it is difficult due to differences in individuals' knowledge and expertise. A poor approach to problem generation may lead to an ambiguous or incorrect problem solution. Despite its significance in pedagogy, redefining engineering problems is often overlooked particularly, in the context of traditional engineering textbook-based problems, which are well-defined but lack the complexity of real-world problems. This gap restricts students' experience to the uncertainty and complexity present in real-world engineering problems. This study uncovers the potential of using generative AI to redefine engineering problems and overcome their limitations. Chat-GPT, a user-friendly generative AI, is utilized as a problem-generation tool. The study hypothesizes that by integrating AI-generated problems with conventional settings, students will gain a deeper understanding of engineering content and improve their performance. It adopts a mixed-method approach with 14 participants to investigate this hypothesis. Students' performance is evaluated using structured analytical rubrics with a deductive coding scheme. The study identifies problem-processing elements, error execution, and solution accuracy as the major experiences of students. Ultimately, the study concludes that there are significant impacts of the generation of problems on student performance compared with conventional textbook problems. The insights of this research offer valuable guidance for redefining traditional engineering problems.



*Keywords: Engineering Problem generation, Generative AI, Student Performance, Engineering Education*


## 1. Introduction

Engineering problems are a fundamental element of formal education pedagogy. Traditional engineering problems are formed by acquired knowledge and experience. The process of problem formation serves as an essential phase in problem-solving that could directly impact the outcome [1], [2], [3]. A deficient problem-generation approach can lead to hindrances in applying earned knowledge which causes unclear or incomplete problem-solving [4]. Moreover, engineering courses feature well-defined homework problems from traditional textbooks, which differ from the complex challenges encountered in the workplace. Research suggests that aligning academic problems with real-world scenarios is beneficial [5], [6]. However, the process of developing engineering problems that bridge this gap remains underexplored. To overcome pedagogical limitations, generative AI could be used to develop engineering problem generation. Prior research on Gen-AI focused mostly on its solution-producing capabilities [7], [8] and imitating human intelligence [9], [10].With these studies, generative AI has proved to be an educational tool that functions as a peer as a guide, and as an expert for students in their learning process.

This research aims to redefine the creation of engineering problems by utilizing generative AI, particularly ChatGPT. The study assesses student performance by conducting a mixed-method approach that combines both quantitative and qualitative analyses. By adopting a novel approach for creating engineering problems beyond traditional textbook problems, we explore a way to improve student learning outcomes and enhance the essence of engineering education.

This research specifically addresses a major research question illustrated in Figure 1:

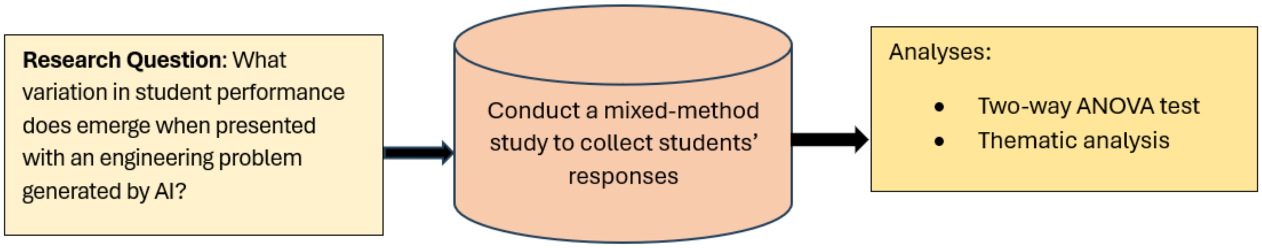


Figure 1: Assessment of Research Question

## 2. Background

### 2.1. Pedagogy of Engineering Problem

Several recent studies focused on reshaping the engineering curriculum, engaging students from their freshman year in recognizing and solving engineering problems [11]. Well-structured problems are traditionally studied with effective communication of problems as an essential component [12]. The shift towards hands-on and outcomes-based education sheds light on gaps in addressing problem types in curricula. Professionals can pinpoint relevant problem states, while non-professionals struggle with this phenomenon [13]. Over the years, Senior Capstone projects in school are intended to provide real-world engineering experience, highlighting the importance of problem identification [14]. However, conventional engineering problems lack the connection with practical engineering expertise. An effective problem structure requires instructor flexibility and appropriate strategy application [15]. Prioritizing authentic engineering problems that enhance essential skills can play a crucial role in improving engineering curricula [16].

### 2.2. Student Performance

Student performance is a crucial aspect of pedagogical research, as it provides valuable insights into the challenges that students face throughout their educational journeys. Understanding the challenges encountered by the students is essential for developing effective teaching strategies that can enhance overall educational outcomes. The most used student performance measure is cumulative GPA in school. Studies have indicated that GPA is an inadequate measure of a student's performance at the university level [17]. Therefore, various aspects affecting performance are considered. Common studies on student performance examine several factors, including the effects of the learning environment [18], [19], [20]. The impact of different learning contexts (such as online versus in-person learning), various behavioral perspectives (e.g., intellectual, interpersonal, and intrapersonal behaviors), and the influence of different textbook selections [21], [22], [23]**.** Previous research carried out by one of the authors indicates that student motivation plays a critical role in influencing students' performance in senior design capstone courses [24 - 33]. Listed studies have significantly increased our understanding of how different educational settings affect student performance and academic success. However, there remains a notable gap in understanding how various sources of problem generation, including experts, textbooks, and advanced technologies such as artificial intelligence tools, affect students' performance in engineering.

### 2.3. Student Performance with Generative AI

Generative Artificial Intelligence (Gen-AI) represents an advanced form of artificial intelligence that generates content by identifying patterns within pre-trained data [34]. This technology is a combination of autoregressive models, Variational Autoencoders (VAEs), and Generative Adversarial Networks (GANs) [35]. A notable accomplishment in this field is the development of complicated language models, such as OpenAI’s GPT, which significantly enhances both text comprehension and generation capabilities. Gen-AI has made considerable contributions across a variety of domains, particularly in the creation, prediction, and analysis of synthetic data, including applications within engineering education [36], [37], [38].

Generative AI (Gen-AI) tools are advancing education by delivering customized and highly effective learning experiences. They provide virtual simulations, offer personalized responses, and present mathematical problems with varying levels of difficulty [39], [40]. Extensive research has demonstrated that AI-driven instructional technologies significantly enhance analytical thinking, problem-solving abilities, and academic achievement [41]. In computing education, Gen-AI tools deliver real-time programming assistance, including clear code explanations, alternative solutions, and valuable instructional feedback, benefiting both novice and advanced users. Tools such as Codex empower students to become proficient coders by generating code from textual descriptions and annotating existing code [42], [43]. Furthermore, Gen-AI improves essay correction by providing insightful guidance on vocabulary, structure, and logic [44].

While the advantages of Gen-AI in education are substantial, it is important to acknowledge certain limitations. Concerns regarding the quality, accuracy, and timeliness of the information provided remain [45]. Additionally, ChatGPT's potential to facilitate plagiarism, fraud, and academic dishonesty, particularly in assessment practices, raised important concerns among educators. Issues related to data protection, privacy, security, and discrimination further complicate their integration into educational contexts [46], [47] [48]. Despite these challenges, Gen-AI is effectively utilized to enhance student performance [49], [50]. It reduces cognitive load and offers user-friendly features that enable students to manage their cognitive resources efficiently. However, overreliance on Gen-AI can lead to procrastination and memory loss, which can negatively impact academic performance [51].

Building on the transformative role of Gen AI in education, Gen AI has primarily operated as a front-end tool interacting with students through problem-solving assistance, real-time feedback, and personalized learning experiences. However, the back-end role of Gen AI is limited to exploration. With this research, a balanced and moderated approach to using Gen-AI in generating engineering has been explored with student performance.

## 3. Method

### 3.1. Problem Selection and Generation

This study is based on fundamental manufacturing topics such as bending, extrusion, forging, and machining. The selected manufacturing topic problems are sourced from a standard manufacturing textbook. In addition to problem development with Gen AI, ChatGPT is utilized to modify these problems while maintaining the traditional principles, numerical values, and technical accuracy.

The development of Gen AI problems was an iterative process using several structured prompts. For each topic, two variations of AI-integrated problems were created: one with a personalized context (AI-generated) and another with a more detailed background to enhance engagement and understanding (a combination of textbook and AI). The final selection of AI-generated problems was based on expert reviews to confirm the clarity and avoid repetitiveness. The detail of problem generation is illustrated in Figure 3.1.

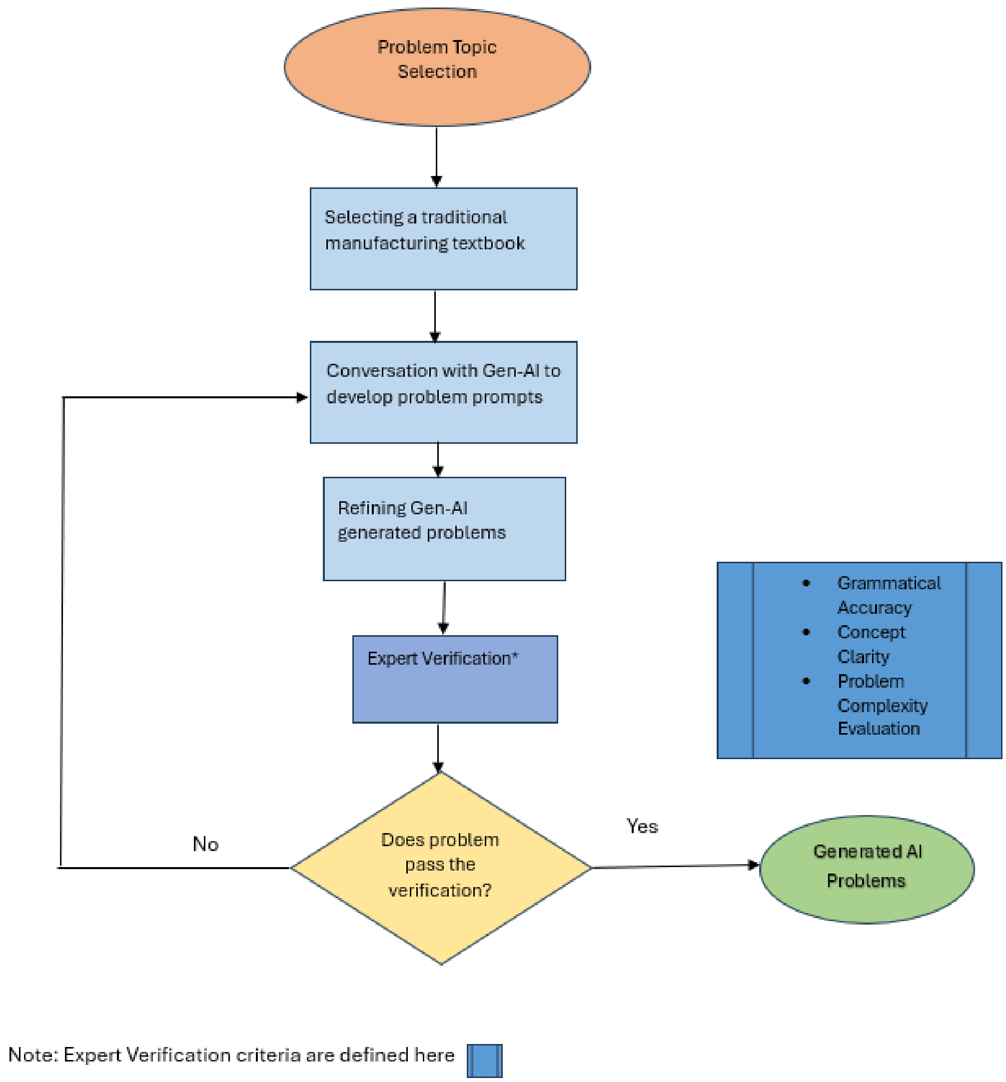


Figure 3.1: Problem Generation Process

## 3.2. Subject of Study

The study was performed with mechanical engineering students from an Advanced Manufacturing Process course. The study was conducted at the end of the semester to ensure the period during students' ability to produce expected solutions for the study. It enabled the researchers to ensure that the participants learned the necessary knowledge for this study. In total, 14 students participated in the study.

Table 3.2: Study Subject Demographic

| | Undergraduate | Graduate |
|---|---|---|
| Male | 8 | 2 |
| Female | 3 | 1 |

### 3.3 Experiment Design

All participants are required to complete all three portions of the study as an integral component. In the first portion of the experiment, the participants are provided a consent form approved by the University of Georgia IRB board and then offered a demographic survey using Qualtrics on the computer screen, which reflects their education level, understanding of manufacturing class, and knowledge of Gen-AI tools in education. Following the survey's completion, participants engaged in a selection process for manufacturing problem topics, where they chose and solved three out of four provided problems (e.g., Bending, Extrusion, Forging, Machining). Continuing with the experiment procedure, participants solved the three selected problems through the pen-and-paper format. Participants were not informed about the origins (e.g., Textbook, AI, Textbook+AI) of the problem generation during the problem selection and solving period; such details were revealed at the end of the experiment. On average, participants will take approximately 35 minutes to complete the problem-solving tasks. As a token of appreciation, participants received compensation of $50 for participating in the study.

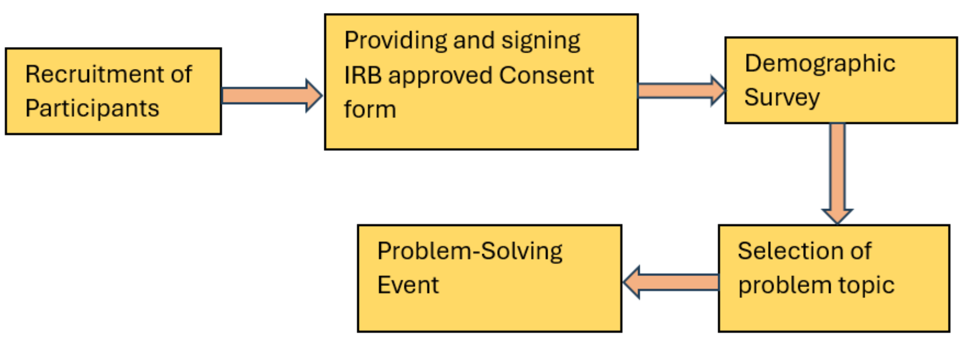


Figure 3.2: Experiment Steps

### 3.4 Quantitative Data Collection and Analysis

With the students' responses, the students' performance towards the several types of problems is evaluated. Literature claims that performance criteria are a set of guidelines or standards that serve as criteria for students' performances or responses. When the intricacy of students' scripts is factored into consideration, the term 'Rubric' becomes relevant [52]. There are two diverse kinds of rubrics for performance assessment, holistic rubric and analytical rubric. Holistic rubric assigns one grade without breaking the performance down into its components, depending on the overall assessment of the student's performance. The analytical rubric breaks down into multiple criteria, and each criterion can be scored separately to provide the student with thorough feedback on various elements of their work [53]. This study followed the analytical rubric but did not provide any feedback to the participants. The participants' scripts were graded in the following rubrics in Table 3.1 to reflect the performance evaluation in this research. The rubric is as below:

Table 3.3: Student Performance Evaluating Rubrics Content Score

| **Content** | **Points** |
|---|---|
| Assumptions | 01 |
| Correct Equation | 01 |
| Correct Solution | 05 |
| Correct Answer | 02 |
| Correct Unit | 01 |

It should be noted that there was a partial grading only for the right solution content. The highest score anyone can achieve for a problem is 10.

For quantitative data, a two-way Analysis of Variance (ANOVA) is employed to determine differences across the generation of problems (AI, Textbooks, Textbook+AI problems) and types of problems (Bending, Extrusion, Forging, Machining) to address the research questions. The threshold of significance is identified as the P-value of the analysis as 0.05. In ANOVA, a significant F-value (at p-value<0.05) indicates that there are differences among group means (for example, in problem generation: Textbook, AI, and Textbook + AI). However, it does not specify which specific pairs of means differ. A post-hoc analysis, such as Duncan's Multiple Range Test (DMRT), is employed to identify specific mean differences to specify the significant mean. DMRT calculates the standard error (SE) of the means using the Mean Square Error (MSE) and the sample size. Least Significant Ranges (LSR) are then derived by multiplying the SE by the significant studentized ranges (SSR) at a p-value of 0.05, along with the error degrees of freedom obtained from ANOVA. Pairwise mean differences are compared to the LSR; means are grouped with the same letter when the differences are not significant, while distinct letters are assigned to means that show significant differences. This letter provides a visual representation of the relationships between group means.

## 3.5 Qualitative Data Collection and Analysis

A deductive approach was incorporated to understand the performance of students in depth. The scheme adopted by Grigg and Benson [54] is used to capture each step related to problem-solving. The scheme introduced 54 unique codes, which are classified as Process Elements, Errors, Strategies, and Accuracy in problem-solving. It was created via a problem analysis methodology based on a mathematical education hierarchical structure. The coding method enables its potential to evaluate students' problem-solving techniques in detail, emphasizing the significance of comprehending mistakes and self-correction in educational research. After employing a coding scheme, a thematic approach is also implemented to combine multiple examination flexibility with research questions to strategic codes and themes [55]. All students' responses were fed into the well-known software, NVivo 12 Pro, to extract and code them.

To ensure the trustworthiness of the coding scheme, an interrater reliability test was incorporated. Inter-rater reliability is a critical measure that evaluates the degree of agreement among two or more raters regarding the consistent application of a rating system. While different researchers may reach a consensus on a specific code, their interpretations can vary significantly. Thus, maintaining inter-rater reliability is particularly crucial to ensure process consistency. In this study, two coders engage in coding the data according to predefined coding schemes. A variety of statistical methodologies can be utilized to assess inter-rater reliability [56]. Cohen's Kappa coefficient is employed in our research, which is denoted by the lowercase Greek letter k. Cohen's Kappa serves as a reliable statistic for measuring inter-rater reliability. The coefficient can range from -1 to +1, where a value of +1 signifies complete agreement between raters, while a value of 0 reflects the level of agreement that could be anticipated due to chance. The calculation of Cohen's Kappa is conducted using the formula specified by McHugh (2012) [57]:

$$k = \frac{Po - Pe}{1 - Pe} \tag{3.5.1}$$

Here, $Po$ represents the actual observed agreement, and $Pe$ represents the chance of agreement.

## 4. Result

The result of the study is based on students' responses to several types of generation of problems. The students' responses are evaluated with the rubric mentioned in subsection 3.4 and the coding scheme mentioned in subsection 3.5.

### 4.1 Quantitative Analysis

This study conveys the impact of the generation of problems and the types of problems on student performance in solving engineering tasks. To evaluate student performance, the mean and standard deviations of student responses are considered. Additionally, statistical significance is assessed using the Two-way ANOVA test, and the results are analyzed using Duncan's Multiple Range Test (DMRT) to effectively group means with similar statistical properties.

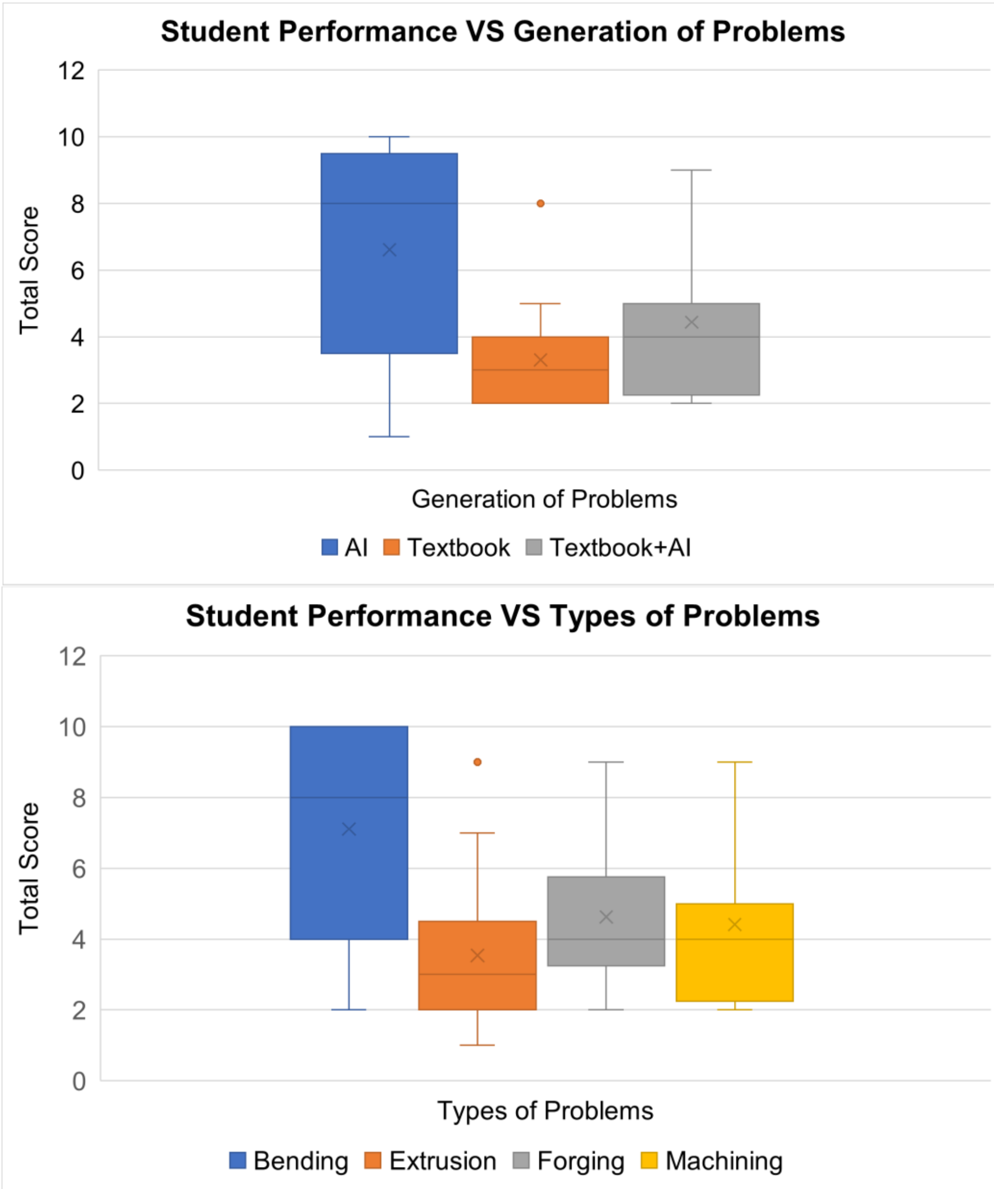


Figure 4.1: Student Performance with both Generation of Problems and Types of Problems

### 4.1.1 Impact of Problem Generation and Types of Problem

Considering two independent variables in the study, the two-way ANOVA test reveals a significant difference (p-value= 0.003) in student performance across all problem generations (e.g., AI, Textbook, Textbook + AI). AI-generated problems have the highest performance score (6.62±2.68), significantly higher than Textbook problems (3.31±1.70) and Textbook + AI problems (4.44±2.28).

A significant difference ($p<0.05$) is observed among the types of problems (e.g., Bending, Extrusion, Forging, Machining). Bending problems had the highest score (7.11±3.17), placing them in group "a" in the mean comparison analysis.

Noticeably, no significant interaction effect is found between problem generation source and types of problems. Despite both variables influencing performance independently, their combined effect does not differ considerably from the impacts identified independently. The summary of the analyses is listed in Table 4.1.

Table 4.1: Summary of Results for Different Problem Generation and Types of Problems

| **Category** | **Subcategory** | **Mean** | **SD** | **p-value** | **Mean Comparison** |
|---|---|---|---|---|---|
| Generation of problems | AI | 6.62 | 2.68 | 0.003 | AI (6.62a) |
| | Textbook | 3.31 | 1.70 | | Textbook (3.31c) |
| | Textbook+AI | 4.44 | 2.28 | | Textbook +AI (4.44b) |
| Types of Problems | Bending | 7.11 | 3.17 | 0.01 | Bending (7.11a) |
| | Extrusion | 3.54 | 2.25 | | Extrusion (3.54b) |
| | Forging | 4.63 | 2.13 | | Forging (4.63b) |
| | Machining | 4.42 | 2.43 | | Machining (4.42b) |
| Interaction of Both | - | - | - | 1 | - |

## 4.2 Qualitative Analysis

Grigg and Benson's coding scheme [54] is implemented in the study to get a deeper understanding of students' responses to the problems. Two coders individually coded the students' responses to avoid bias and increase reliability. 17 codes from the parent codes are used in this study. Three major themes are identified, such as 'Problem Processing' where students' cognitive ability elements to problem-solving are involved, 'Error Code' represents students' computing abilities and comprehension of engineering principles, and 'Accuracy Code' is based on the correctness of final answers. The detailed coding scheme is illustrated in Figure 4.2:

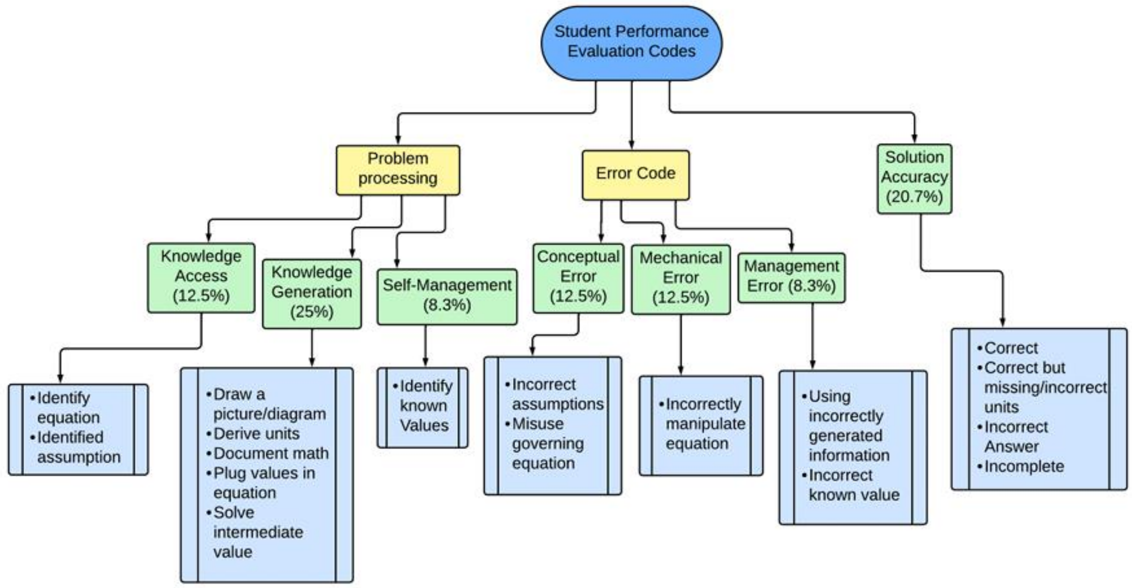


Figure 4.2: Coding Scheme Used in Student Performance

To prevent unexpected findings and enhance the credibility of the used coding scheme, an inter-rater reliability test is conducted. The codes from Figure 4.2 are applied in this process. The Cohen Kappa coefficient is calculated using the equation provided in Section 3. The expected agreement (Pe) is 0.95, while the observed agreement (Po) is 0.98. Substituting these values into the equation yields a Kappa coefficient of 0.71, indicating strong agreement between the two coders. This result is also consistent with the Kappa coefficient of the parent coding scheme.

## 5. Discussion

### 5.1 Addressing Research Question

The results of the ANOVA test clearly demonstrate significant variation in student performance based on the different generations of problems. Specifically, significant variations in total scores highlight the strong impact of problem generation on overall student performance. Duncan's Multiple Range Test (DMRT) strengthens the finding, revealing that AI-generated problems consistently excel over textbook problems. Moreover, from the coding scheme analysis, students have encountered less errors in AI-generated problems leading to their higher scores. This finding highlights that AI-generated problems could have more elements to address problem-solving strategies. However, it remains unrevealed whether the participants prefer the structure of AI-generated problems.

To capture a detailed understanding of the qualitative results, interrater reliability of the implemented coding scheme is conducted. The reliability of the coding process is confirmed by Cohen's Kappa Coefficient of 0.71, demonstrating strong agreement between coders and confirming the credibility of the coding framework employed in the study. The consistency of the coding scheme underlines the robustness of our qualitative analysis. From the coding analysis, it is evident that knowledge application strategies (Knowledge access + Knowledge generation) are the most utilized approach taken by participants for this study. Moreover, the coding scheme revealed that the participants mostly encountered conceptual and mechanical errors, which underscores the critical need to solidify students' fundamental knowledge before they engage in problem-solving tasks. Additionally, in the types of problems, 'Bending' problems achieved the highest performance scores. Importantly, no significant interaction effect was found between the methods of problem generation and the types of problems (p-values > 0.05). This finding confirms that while both factors independently influence performance, their combined effect does not alter or undermine performance outcomes.

In summary, the findings of this study support that AI-generated problems significantly enhance student performance. By mitigating the prevalence of common errors and fostering effective problem-solving strategies, AI-generated problems can have a powerful impact on improving educational outcomes. Integrating AI-generated problems with traditional textbooks into curricula could empower students to build stronger foundational knowledge and develop the essential skills necessary for tackling complex problem-solving tasks.

### 5.2 Limitations of the study

While the study presents several valuable insights, it also has limitations that need to be addressed. Firstly, the study relies on a small sample size, which may restrict the generalizability of the findings. A larger and more diverse sample size could significantly enhance the robustness of ANOVA analysis and the overall impact of the study. Additionally, the data lacks balance in data due to the flexibility of students' selection of problem topics, potentially influencing the outcomes. Although four fundamental manufacturing topics were examined, investigating additional engineering topics could produce different results across various generation sources. Moreover, the study focuses particularly on evaluating students' performance in response to problems, without exploring aspects such as student creativity or collaboration strategies which could be explored in the future. Each problem topic was associated with a specific solution related to all types of generation sources. Incorporating a broader array of manufacturing topics with multiple solution options may offer richer insights for future research.

## 6. Conclusion

This study evaluates student performance on AI-generated versus traditionally generated engineering problems with four manufacturing topics. While some aspects of student understanding and problem clarity emerged in this study, the primary focus remained on student performance evaluation. The analytical rubric was used for performance evaluation, and a two-way ANOVA test was applied to determine statistical significance, accompanied by a coding scheme. The knowledge application strategies empower students to leverage their foundational knowledge effectively on specific problems. The results revealed significant variations in student performance depending on the problem-generation methods and the types of problems used. AI-generated problems are well dealt with across all types of problems due to their distinct contextual structure. Additionally, the research highlighted the complementary roles of textbook problems. Textbook problems are beneficial for developing basic conceptual knowledge, while incorporating AI-generated problems could enhance application skills and overall performance. A combined approach can represent a promising strategy for cultivating a comprehensive understanding of engineering concepts and improving problem-solving abilities. While the study focused on a particular area, it has opened multiple areas for future exploration. Future studies could investigate the cognitive and metacognitive dimensions of AI-generated problems, including their influence on some problem-solving strategies, such as student engagement and critical thinking. During this research, participants do not have access to AI for problem-solving assistance. AI intervention in problem-solving events may foster students' creativity and collaboration, which indicates the future direction of this research. Exploring different engineering disciplines could strengthen the implications of AI in education.

**References:**


[1] M. Dinar and J. J. Shah, "Enhancing Design Problem Formulation Skills For Engineering Design Students."

[2] Michael A. Eierman and George C.Pholip, "The Task of Problem Formulation," *Int J Inf Technol Decis Mak*, vol. 2, no. 03, pp. 353–372, 2003.

[3] C. D. Schunn, P. B. Paulus, J. Cagan, and K. Wood, "Final report from the NSF innovation and discovery workshop; The scientific basis of individual and team innovation and discovery," 2006.

[4] D. A. Cowan, "Developing a Process Model of Problem Recognition," 1986.

[5] J. A. M. Boulet, A. Lumsdaine, and J. F. Wasserman, "The Transition from Textbook Problems to Realistic Problems."

[6] A. R. Sloboda, "The effect of context on student perceptions of homework-style problems in engineering," in *ASEE Annual Conference and Exposition, Conference Proceedings*, American Society for Engineering Education, Jun. 2019. doi: 10.18260/1-2--33385.

[7] R. Rodriguez-Echeverria, J. D. Gutierrez, J. M. Conejero, and A. E. Prieto, "Analysis of ChatGPT Performance in Computer Engineering Exams," *Revista Iberoamericana de Tecnologias del Aprendizaje*, vol. 19, pp. 71–80, 2024, doi: 10.1109/RITA.2024.3381842.

[8] M. L. Tsai, C. W. Ong, and C. L. Chen, "Exploring the use of large language models (LLMs) in chemical engineering education: Building core course problem models with Chat-GPT," *Education for Chemical Engineers*, vol. 44, pp. 71–95, Jul. 2023, doi: 10.1016/j.ece.2023.05.001.

[9] A. Gilson *et al.*, "How Does ChatGPT Perform on the United States Medical Licensing Examination? The Implications of Large Language Models for Medical Education and Knowledge Assessment," *JMIR Med Educ*, vol. 9, 2023, doi: 10.2196/45312.

[10] D. M. Katz, M. J. Bommarito, S. Gao, and P. Arredondo, "GPT-4 passes the bar exam," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 382, no. 2270, Apr. 2024, doi: 10.1098/rsta.2023.0254.

[11] M. Kaushik, P. Baligar, and G. Joshi, "Formulating An Engineering Design Problem: A Structured Approach," 2018.

[12] B.-A. Schuelke-Leech, "A Problem Taxonomy for Engineering," *IEEE Transactions on Technology and Society*, vol. 2, no. 2, pp. 105–105, Apr. 2021, doi: 10.1109/tts.2021.3072213.

[13] D. H. Jonassen, "Instructional Design Models for Well-Structured and Ill-Structured Problem-Solving Learning Outcomes," 1997.

[14] Akili Waddah, "Frontiers in Education 2015 : launching a new vision in engineering education : proceedings : October 21-24, 2015, Camino Real Hotel & Conference Center, El Paso, Texas," IEEE, 2015.

[15] A. H. Schoenfeld, "Teaching Problem-Solving Skills," 1980.

[16] F. Miyara, "Session 8D4 I t m " I a N Engineering Problem Formulation for Mathematics Courses."

[17] E. E. Noftle and R. W. Robins, "Personality Predictors of Academic Outcomes: Big Five Correlates of GPA and SAT Scores," *J Pers Soc Psychol*, vol. 93, no. 1, pp. 116–130, Jul. 2007, doi: 10.1037/0022-3514.93.1.116.

[18] C. James Kado and E. Kames, "Examining the Effect of Design Stimuli on Perception of Peer Contribution in Design Teams."

[19] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977.

[20] F. Mozaffar, C. Chen, B. Morkos, and J. Ma, "Development of a Manufacturing Assessment Survey to Promote Entrepreneurial Mindset in Engineering," in *2023 ASEE Annual Conference & Exposition*, 2023.

[21] T. Easton, "Online Versus In Person Student Learning Outcomes."

[22] A. K. van den Ham and A. Heinze, "Does the textbook matter? Longitudinal effects of textbook choice on primary school students' achievement in mathematics," *Studies in Educational Evaluation*, vol. 59, pp. 133–140, Dec. 2018, doi: 10.1016/j.stueduc.2018.07.005.

[23] F. L. Oswald, N. Schmitt, B. H. Kim, L. J. Ramsay, and M. A. Gillespie, "Developing a Biodata Measure and Situational Judgment Inventory as Predictors of College Student Performance," *Journal of Applied Psychology*, vol. 89, no. 2, pp. 187–207, Apr. 2004, doi: 10.1037/0021-9010.89.2.187.

[24] E. Kames, "A Mixed Methods Analysis of Motivation Factors in Senior Capstone Design Courses."

[25] D. Shah, M. Clark, E. Kames, and B. Morkos, "Development of a Coding Scheme for Qualitative Analysis of Student Motivation in Senior Capstone Design," in *ASME International Design Engineering Technical Conference, Anaheim*, 2019.

[26] D. Shah, E. Kames, Pelham Jaslynn, and B. Morkos, "Design Evaluations in Educational Settings : A Neuroscientific Study of Incentivized Test / Retest on Student Performance," *J Comput Inf Sci Eng*, vol. 22, no. June, pp. 1–10, 2022, doi: 10.1115/1.4053801.

[27] M. Clark, D. Shah, E. Kames, and B. Morkos, "Developing an Interview Protocol for an Engineering Capstone Design Course."," in *ASME International Design Engineering Technical Conferences & Computers and Information in Engineering Conferences*, 2019.

[28] D. Shah, B. Morkos, and E. Kames, "Impact of Educators Changing Student Motivation: A Study of Transient Factor Correlation and Orthogonality," in *2021 ASEE Virtual Annual Conference Content Access*, 2021.

[29] D. Shah, E. Kames, and B. Morkos, "Neurocognitive Effects of Incentivizing Students to Improve Performance Through Repeat Attempts in a Design Setting," in *International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*, 2021, pp. 1–12.

[30] E. Kames, B. Morkos, and A. Bessette, "The implementation of an intervention plan to improve student motivation and performance in mechanical engineering senior design capstone," *International Journal of Engineering Education*, vol. 35, no. 3, pp. 779–794, 2019.

[31] E. Kames, D. D. Shah, and B. Morkos, "A longitudinal study exploring motivation factors in cornerstone and capstone design courses," in *ASEE Annual Conference and Exposition, Conference Proceedings*, 2018.

[32] Elisabeth. Kames, Devanshi. Shah, M. Clark, and B. Morkos, "A Mixed Methods Analysis of Motivation Factors in Senior Capstone Design Courses."

[33] P. Lynch, S. Sangelkar, G. Demeo, and B. Morkos, "The ICDM methodology improving undergraduate engineering student motivation, satisfaction, and performance," in *2017 IEEE Frontiers in Education Conference (FIE)*, IEEE, 2017, pp. 1–7.

[34] U. Mittal, S. Sai, V. Chamola, and Devika, "A Comprehensive Review on Generative AI for Education," *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3468368.

[35] R. K. Gatla, A. Gatla, P. Sridhar, D. G. Kumar, and D. S. Naga Malleswara Rao, "Advancements in Generative AI: Exploring Fundamentals and Evolution," in *1st International Conference on Electronics, Computing, Communication and Control Technology, ICECCC 2024*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/ICECCC61767.2024.10594003.

[36] C. Chen, T. O. Olajoyegbe, and B. Morkos, “The imminent educational paradigm shift: How artificial intelligence will reframe how we educate the next generation of engineering designers,” in *ASEE Annual Conference and Exposition, Conference Proceedings*, American Society for Engineering Education, Jun. 2020. doi: 10.18260/1-2--35326.

[37] A. Johri, “Artificial intelligence and engineering education,” Jul. 01, 2020, *John Wiley and Sons Inc*. doi: 10.1002/jee.20326.

[38] V. Pursnani, Y. Sermet, M. Kurt, and I. Demir, “Performance of ChatGPT on the US fundamentals of engineering exam: Comprehensive assessment of proficiency and potential implications for professional environmental engineering practice,” *Computers and Education: Artificial Intelligence*, vol. 5, Jan. 2023, doi: 10.1016/j.caeai.2023.100183.

[39] J. Qadir, “Engineering Education in the Era of ChatGPT: Promise and Pitfalls of Generative AI for Education,” in *IEEE Global Engineering Education Conference, EDUCON*, IEEE Computer Society, 2023. doi: 10.1109/EDUCON54358.2023.10125121.

[40] W. Y. Hwang and I. Q. Utami, “Using GPT and authentic contextual recognition to generate math word problems with difficulty levels,” *Educ Inf Technol (Dordr)*, Sep. 2024, doi: 10.1007/s10639-024-12537-x.

[41] T. Sri Tulasi and S. B. Inayath Ahamed, “Artificial Intelligence Effects on Student Learning Outcomes in Higher Education,” in *Proceedings of 9th International Conference on Science, Technology, Engineering and Mathematics: The Role of Emerging Technologies in Digital Transformation, ICONSTEM 2024*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/ICONSTEM60960.2024.10568868.

[42] J. Finnie-Ansley, P. Denny, B. A. Becker, A. Luxton-Reilly, and J. Prather, “The robots are coming: Exploring the implications of OpenAI codex on introductory programming,” in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Feb. 2022, pp. 10–19. doi: 10.1145/3511861.3511863.

[43] M. Amoozadeh *et al.*, “Trust in Generative AI among Students: An exploratory study,” in *SIGCSE 2024 - Proceedings of the 55th ACM Technical Symposium on Computer Science Education*, Association for Computing Machinery, Inc, Mar. 2024, pp. 67–73. doi: 10.1145/3626252.3630842.

[44] L. Wang, R. Song, W. Guo, and H. Yang, “Exploring prompt pattern for generative artificial intelligence in automatic question generation,” *Interactive Learning Environments*, 2024, doi: 10.1080/10494820.2024.2412082.

[45] Patrício Maria and Gonçalves Bruno, “ChatGPT: Systematic Review of Potentials and Limitations in Education,” *Springer Nature*, vol. 933, 2024, doi: 10.1007/978-3-031-54256-5.

[46] M. O. M. Al Ghatrifi, J. S. S. Al Amairi, and M. M. Thottoli, “Surfing the technology wave: An international perspective on enhancing teaching and learning in accounting,” *Computers and Education: Artificial Intelligence*, vol. 4, Jan. 2023, doi: 10.1016/j.caeai.2023.100144.

[47] M. Farrokhnia, S. K. Banihashem, O. Noroozi, and A. Wals, “A SWOT analysis of ChatGPT: Implications for educational practice and research,” *Innovations in Education and Teaching International*, vol. 61, no. 3, pp. 460–474, 2024, doi: 10.1080/14703297.2023.2195846.

[48] X. Yang, A. Floyd, L. Smith, and B. Morkos, “Analysis of Engineering Undergraduates’ Confidence with Hands-on Tasks – Preparation for Collaborative Manufacturing Environments in the Era of Industry 4.0 ,” ASEE, Jun. 2023.

[49] M. Abdaljaleel *et al.*, “A multinational study on the factors influencing university students’ attitudes and usage of ChatGPT,” *Sci Rep*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-52549-8.

[50] N. A. Dahri *et al.*, “Investigating AI-based academic support acceptance and its impact on students’ performance in Malaysian and Pakistani higher education institutions,” *Educ Inf Technol (Dordr)*, Oct. 2024, doi: 10.1007/s10639-024-12599-x.

[51] M. Abbas, F. A. Jam, and T. I. Khan, “Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students,” *International Journal of Educational Technology in Higher Education*, vol. 21, no. 1, Dec. 2024, doi: 10.1186/s41239-024-00444-7.

[52] J. Arter and J. McTighe, *Scoring Rubrics in Classroom*. Sage, 2001.

[53] Made Frida Yuiila, “Assessing Performance Using Scoring Rubrics: An Overview,” *Kajian Linguistik dan Sastra*, vol. 17, no. 33, pp. 172–182, 2005.

[54] S. J. Grigg and L. C. Benson, “A coding scheme for analysing problem-solving processes of first-year engineering students,” *European Journal of Engineering Education*, vol. 39, no. 6, pp. 617–635, Nov. 2014, doi: 10.1080/03043797.2014.895709.

[55] V. Clarke and V. Braun, “Thematic analysis,” May 04, 2017, *Routledge*. doi: 10.1080/17439760.2016.1262613.

[56] M. Banerjee, M. Capozzoli, L. McSweeney, and D. Sinha, “Beyond kappa: A review of interrater agreement measures,” *Canadian Journal of Statistics*, vol. 27, no. 1, pp. 3–23, Mar. 1999, doi: 10.2307/3315487.

[57] M. McHugh, “Interrater reliability: the kappa statistic,” *Lessons in biostatistics,* vol. 3, no. 22, pp. 276–282, Aug. 2012.