www.ijemst.net

Artificial Intelligence in Mathematics Education: ChatGPT's Capabilities and Limitations in Problem Posing

Çiğdem Arslan
Bursa Uludag University, Türkiye

Hatice Kübra Güler Selek
Bursa Uludag University, Türkiye

To cite this article:

Arslan, C., & Guler Selek, H.K. (2025). Artificial intelligence in mathematics education: ChatGPT's capabilities and limitations in problem posing. *International Journal of Education in Mathematics, Science, and Technology (IJEMST), 13*(5), 1189-1205. https://doi.org/10.46328/ijemst.5712

International Journal of Education in Mathematics, Science and Technology

2025, Vol. 13, No. 5, 1189-1205 https://doi.org/10.46328/ijemst.5712

Artificial Intelligence in Mathematics Education: ChatGPT's Capabilities and Limitations in Problem Posing

Çiğdem Arslan, Hatice Kübra Güler Selek

Article Info

Article History

Received:
7 March 2025
Accepted:
18 August 2025

Keywords

ChatGPT
Flexibility
Geometry
Problem posing

Abstract

This study investigates the ability of ChatGPT to generate mathematical problems creatively, focusing on fluency, flexibility, and originality. The findings indicate that ChatGPT can produce a vast number of problems, demonstrating a high level of fluency. Additionally, it tends to prioritize the most original problem among its generated outputs. In terms of flexibility, the model successfully formulates problems across various mathematical content areas, showing adaptability in different contexts. Despite its fluency, ChatGPT occasionally repeats content or diverges from the given instructions in later stages, highlighting the importance of expert supervision. When used in educational settings, AI-generated problems should be reviewed to ensure their relevance and accuracy. While ChatGPT has potential as a supportive tool in mathematics education, it should not function independently in instructional environments. Another observed limitation is the inconsistency in how ChatGPT categorizes mathematical problems, which may indicate gaps in its conceptual understanding. Furthermore, its inability to generate visual representations of geometric concepts could contribute to misclassifications or errors. These findings suggest that while ChatGPT can enhance creative problem posing in mathematics, its effectiveness depends on careful monitoring and structured implementation within the teaching and learning process.

Introduction

Problem posing is a fundamental process in mathematics education that enhances students' mathematical thinking, conceptual understanding, and creative problem-solving skills (Silver, 1997; Cai & Hwang, 2002). Engaging in problem-posing activities allows students to actively construct their own mathematical problems, rather than merely solving given ones, which in turn increases their metacognitive awareness (Brown & Walter, 2005) and problem-solving abilities (Silver, 1997). Additionally, these activities foster questioning, interpretation, and critical thinking skills, helping students develop a more positive attitude towards mathematics and a deeper comprehension of mathematical concepts (Akay, Soybas & Argun, 2006). Research also suggests that problem posing diversifies students' problem-solving strategies and enhances their engagement in active learning processes (Cai, 2003).

Developing students' problem-posing skills strengthens their mathematical competence by increasing their creative thinking capacities (Singer, Ellerton, & Cai, 2013). The integration of artificial intelligence-supported tools into educational processes has emerged as a growing research area, particularly in exploring their role in students' problem-posing processes. AI-based language models can support mathematical thinking by assisting students in generating various types of problems (Su & Yang, 2023; Segal & Biton, 2024). However, further research is needed to determine the extent to which AI is effective in this process and how it contributes to students' problem-posing skills.

The use of today's artificial intelligence applications is rapidly becoming widespread in every field. The most popular artificial intelligence tool that can also be used as a chatbot is ChatGPT. Among the things that make ChatGPT so popular are that it is interactive and easy to use, free, has a wide application area, and has the ability to produce quality text. At the same time, ChatGPT's ability to understand and interpret the text, as well as its ability to produce text, is among the factors that increase its popularity. The use of ChatGPT may continue to increase in the coming years (Gusteti et. al., 2024). In this context, it has become increasingly important to examine the use of ChatGPT in education from different perspectives, as it continues to gain prominence in our daily lives (Adeshola & Adeola, 2023). Recent bibliometric analyses indicate a significant rise in research exploring ChatGPT's role in educational settings, particularly in medical and higher education, while its application in K-12 contexts remains underexplored (Boral, Mondal, & Saikia, 2024). The increasing interest in ChatGPT's integration into education suggests that AI-driven tools can enhance personalized learning, student engagement, and academic support. Given these developments, exploring the potential of artificial intelligence tools like ChatGPT in fostering creative problem-solving and problem-posing skills is of great value. Enhancing student-machine interaction in these activities can provide new opportunities for developing higher-order thinking skills. Although studies on the role of ChatGPT in problem-solving are rapidly growing (Sánchez-Ruiz et al., 2023; Wardat et al., 2023), its application in problem posing remains an area requiring further investigation. Understanding how AI can contribute to students' ability to generate and structure mathematical problems could open new research avenues in mathematics education.

This study aims to examine the effectiveness of ChatGPT, an AI-based language model, in the problem-posing process in mathematics education. Specifically, it will investigate how ChatGPT contributes to creative problem-posing and how the problems it generates are evaluated in terms of fluency, originality, and flexibility. In this context, in order to make comparisons, the study of Van Harpen and Sriraman (2013) on the creativity of advanced high school students in problem posing activities was taken as a basis. The sub problems are as follows:

- To what extent do the problems generated by ChatGPT meet the criterion of fluency in creative problem posing?
- To what extent do the problems generated by ChatGPT meet the criterion of originality in creative problem posing?
- To what extent do the problems generated by ChatGPT meet the criterion of flexibility in creative problem posing?
- What are the similarities and differences between the problems generated by ChatGPT and the problems posed by the high school students in Van Harpen and Sriraman (2013)'s study, in terms of

flexibility, one of the components of creativity?

Literature Review

This section begins by addressing key themes concerning creativity in mathematical problem posing. It then explores the role of ChatGPT within the context of education and mathematics education, as well as its relevance to artificial creativity.

Mathematical Problem Posing and Creativity

Posing problems is a valuable method for promoting meaningful learning and exploring an individual's mathematical abilities. By establishing connections between various mathematical concepts and numerical structures, problem posing encourages higher levels of abstraction and reflection. Ayllón et al. (2016) indicate that this process facilitates reasoning and enhances the construction of mathematical knowledge. Numerous researchers (Leung & Silver, 1997; Shuk-Kwan, 1997; Silver, 1994; Silver, 1997; Van Harpen & Sriraman, 2013) have linked problem posing to the development of mathematical knowledge and creativity, highlighting an inherent relationship between problem-posing abilities and levels of creativity and mathematical proficiency.

Daher and Anabousy (2018) found that problem-posing tasks foster students' creativity, particularly in fluency, flexibility, and originality. Their study investigated the flexibility of mathematics teacher candidates in problem posing, demonstrating that technology-supported problem posing, combined with the "what-if-not" strategy, led to broader use of problem-posing strategies. This suggests that integrating technology with creative approaches enhances problem-posing flexibility and promotes innovative mathematical thinking.

Silver (1994) argued that while problem solving has been extensively researched, less attention has been paid to diversifying the sources from which students derive problems. In traditional mathematics instruction, students primarily solve problems provided by teachers or textbooks, with little opportunity to formulate their own mathematical problems. Constructivist learning theories emphasize the importance of student-driven problem posing, which aligns with the increasing integration of artificial intelligence in educational settings. AI-generated problems are now emerging as alternatives to problems posed by teachers, textbooks, and students, necessitating a comparison between human- and AI-generated problem posing.

Shuk-Kwan (1997) explored the relationship between creative thinking and problem posing, noting that fluency is a common trait in both verbal creativity and mathematical problem posing, while flexibility plays a critical role in generating diverse problem types.

This study builds upon Van Harpen and Sriraman's (2013) research, which examined the problem-posing performance of first-year U.S. college students and senior high school students from two regions in China. Students were asked to pose problems in free, structured, and semi-structured formats. The analysis focused on fluency, flexibility, and originality as components of creativity. The study found that students' fluency was lower

than expected, as they often posed problems with incomplete information. Regarding flexibility, most students posed problems related to length and area, with minimal representation of analytical geometry. Originality was measured by identifying problems posed by fewer than 10% of students, revealing that highly original problems were rare. Interviews indicated that students had varying perceptions of originality, and overall, their performance in mathematical problem posing was limited.

While previous research has focused on human-generated problem posing, the potential of AI-powered tools like ChatGPT in fostering mathematical creativity remains underexplored. This study aims to fill this gap by analyzing ChatGPT's problem-posing performance and comparing it to high school students' creativity levels. The findings will contribute to understanding the role of AI in education and its implications for creative mathematical problem posing.

ChatGPT as an AI Tool and Artificial Creativity

In recent years, artificial intelligence has demonstrated remarkable success in processing text, often producing outputs that are nearly indistinguishable from those generated by humans (Bishop, 2021). Nevertheless, it is quite challenging to argue that creative artificial intelligence is not merely an extension of the programmer's creativity. For a system to be considered genuinely creative, it must be capable of fully engaging with its creative environment (Jennings, 2010). Research consistently shows that artificial creativity cannot be equated with human creativity (Esling & Devis, 2020; Runco, 2023). Similarly, Marrone, Taddeo, and Hill (2022) reported that middle school students believed artificial intelligence would never match human creativity. These findings suggest that the creative capacities of AI require further investigation to be meaningfully understood.

One of the artificial intelligence tools is ChatGPT. The use of ChatGPT, a chatbot launched by OpenAI in 2022, in education as well as in daily life, is gaining momentum day by day. The widespread use of artificial intelligence tools has inevitably led to their integration into the field of education, prompting a rapid surge in research on the use of ChatGPT in educational settings (Adeshola & Adeola, 2023; Javaid et al., 2023; Su, Lin & Lai, 2023; Su & Yang, 2023). This rapid increase in studies necessitates a thorough discussion of the advantages and disadvantages of ChatGPT.

There are various studies on the advantages and disadvantages of using ChatGPT in education. From the perspective of teachers, the use of ChatGPT may be useful in the course planning and material design processes, in creating content according to the student's needs, in creating homework and exams, and in evaluating these documents (Javaid et al., 2023; Kasneci et al., 2023; Rasul et al., 2023; Tlili et al., 2023). From the students' perspective, Su, Lin and Lai, (2023) emphasized the importance of providing students with information about the effective use of these tools and ethical issues, rather than banning the use of chatbots as writing assistants. Halaweh (2023) stated that ChatGPT transforms various activities in educational settings such as seeking information, answering specific questions, creating software codes, providing case scenarios for databases and analysis, solving mathematical problems and statistical calculations that provide significant potential for enhancing educational outcomes. Javaid et al. (2023) stated that ChatGPT is an excellent tool for language lessons, that it can help

students improve their language proficiency, and that it will allow children to focus on the subject and improve their communication skills and think critically. They also stated that ChatGPT could help students understand complex problems. However, they stated that it might sometimes provide inaccurate information. The possibility of students having their homework done by ChatGPT and copy-pasting is also among the limitations of ChatGPT.

Rasul et al. (2023) emphasized in their study that higher education educators and students should be careful when using ChatGPT for academic purposes to ensure its ethical, reliable and effective use while they stated five benefits of ChatGPT as the potential to facilitate adaptive learning, provide personalized feedback, support research and data analysis, offer automated administrative services, and aid in developing innovative assessments. As can be seen, in addition to its many positive aspects, the use of ChatGPT in the field of education creates various challenges. These challenges include the fact that some students think that ChatGPT can replace the teacher, allow cheating in assignments or exams, and create disadvantages for students who do not have digital competence or have limited access to digital tools (Adeshola & Adeola, 2023). According to Tlili et al. (2023), positive opinions about ChatGPT were related to it being a new technology in education, while negative opinions were discussed in the context of deep and critical thinking. In this context, they emphasize that ChatGPT should be approached with caution. It has been determined that ChatGPT helps teachers prepare teaching materials and exams, but sometimes cannot provide accurate and quality answers to the questions asked. They stated that quality answers can be obtained when proper questions are asked by thinking critically. The limitations of ChatGPT are that the resources are not subject to accuracy control and that humans are still needed for improvement (Rudolph et al., 2023). ChatGPT already creates a control mechanism by stating that the answers given as footnotes must be checked by users.

In addition to its advantages and disadvantages in literature various studies continue to be carried out on the use of ChatGPT in the field of education. In their study with kindergarten teachers, Su and Yang (2023) determined that teachers' opinions about ChatGPT were not clear. While they stated that the potential benefits of using ChatGPT in education include lesson planning, pedagogical and content knowledge, and gaining twenty-first century skills. They also highlighted concerns about hardware issues, lack of resources, and accuracy. In their study, Tu and Hwang (2023) determined that university students thought that ChatGPT facilitated learning. In their studies, high learning attitude groups used ChatGPT for discussion purposes and for a certain content while low learning attitude group found that they preferred to use it to complete homework and reports. Azmi et al. (2024), in their study with university students, noted that while students found ChatGPT helpful in enhancing their understanding of topics, they expressed concerns about its potential to foster dependency and diminish writing and critical thinking skills. Additionally, students believed that due to ChatGPT's lack of emotional understanding, it cannot replace higher education or the role of instructors.

When the studies conducted in mathematics education are examined in general, especially in mathematics education, Lo (2023) found that ChatGPT could not meet user expectations. Wardat et al. (2023) stated that ChatGPT sometimes gives incomplete or incorrect mathematical answers and that its understanding of geometry is not sufficient. Not being able to draw geometric shapes is an obstacle to ChatGPT providing accurate and adequate answers. Additionally, its performance in mathematical calculations is unreliable (Sánchez-Ruiz et al.,

2023). Urhan, Gençaslan, and Dost (2024) conducted an argumentation study between a university student and ChatGPT on the tangent drawn to the graph of a function and determined that ChatGPT's argumentation based on definitions in a rational context was appropriate. However, ChatGPT could not adequately assist the student in situations requiring reasoning and used contradictory statements. Frieder et al. (2024) stated that contrary to many positive reports in the media about GPT-4 and ChatGPT's exam-solving abilities (a potential case of selection bias), their overall mathematical performance is well below the level of a graduate student. Therefore, although it can be used for problem solving at the undergraduate level, using ChatGPT for problem solving at the graduate level may produce wrong results. However, it was noted that ChatGPT has a more flexible structure and is suitable for all areas of mathematics compared to other digital tools used for mathematical purposes. It was suggested that ChatGPT could serve as an assistant for users with mathematical proficiency, functioning as a search engine or knowledge base to expedite tasks.

Segal and Biton (2024) asked pre-service teachers to use ChatGPT to formulate problems for advanced high school mathematics. Although the pre-service teachers were aware of the need to revise and critically evaluate the problems generated by ChatGPT, they reported satisfaction with its use. Furthermore, the study found that this process contributed to an improvement in the pre-service teachers' technological pedagogical content knowledge. Yunianto, Galic, and Lavicza (2024) investigated students' experiences with completing a GeoGebra-based mathematics and computational thinking task supported by ChatGPT. The study revealed that only a limited number of students were able to successfully create objects in GeoGebra with the assistance of ChatGPT. Students struggled to provide detailed instructions to ChatGPT. Nevertheless, most students found ChatGPT useful but noted that the responses they received required further refinement.

Rane (2023) defined the main difficulty in using ChatGPT as not being sure of the accuracy of the answers given to the problems. He emphasized the need to develop a synergistic relationship between generative artificial intelligence and human intelligence, encouraging the development of powerful problem-solving strategies while leveraging the computational capabilities of artificial intelligence. He stated that ChatGPT can be used in problem solving, conceptual understanding, learning assistance, language translation, collaborative learning, accessibility and ethical issues. However, he emphasized that these processes might create challenges regarding clarity, abstract concepts, mathematical notations and security. He also offered solutions for each of these difficulties and suggested a control mechanism. While he stated that it can be used for each sub-field of mathematics, he also stated that there might be difficulties in its use in these areas, especially in terms of clarity and accuracy. He emphasized that both mathematical and pedagogical competence is required for the use of ChatGPT in problem solving process. Zeng (2023) examined the use of ChatGPT in problem-solving tasks and determined that it is competent in tasks of the type "especially in tasks that involve common sense and knowledge". However, he found that its performance was negatively affected in cases where the text was entered incompletely and there were unclear expressions. As evident from the literature, there are a limited number of studies examining the use of ChatGPT in problem-solving activities. However, no research has been found that explores its application in problem-posing activities.

Given the advantages and limitations of problem-posing research, this study aims to evaluate ChatGPT's

performance in creative problem posing and compare it with human-generated problem creativity. Prior research has largely focused on human-generated problems, leaving a gap in understanding how AI-generated problems align with the components of creativity. AI-powered tools, such as ChatGPT, hold promise for enhancing creative mathematical problem posing, but their limitations and potential must be critically examined. This study contributes to the growing body of research on AI in education, addressing the need for comparative analyses between AI- and human-generated mathematical problems.

Furthermore, existing studies indicate that ChatGPT struggles with higher-order mathematical reasoning (Wardat et al., 2023; Urhan et al., 2024). While it can assist with problem posing, its accuracy and consistency require human oversight (Segal & Biton, 2024). The potential of ChatGPT in mathematical education depends on how well it is integrated into instructional practices, ensuring that AI-generated content aligns with pedagogical goals (Rane, 2023). Future research should explore strategies to refine ChatGPT's capabilities in mathematical problem posing and its role in enhancing students' problem-posing skills through guided use.

Method

Research Design

The study was conducted using a qualitative research design. The problems generated by ChatGPT in response to prompts were evaluated in a manner similar to the methodology employed by Van Harpen and Sriraman (2013). The total number of valid problems produced by a student was considered as fluency score, and the total number of categories contained in valid problems was considered as flexibility score. Flexibility and fluency scores do not have to be the same. In terms of originality, evaluation is made within the group, taking into account the rarity among all the answers given.

Data Collection

In this research, data was collected in January 2024 using the ChatGPT 3.5 free version. The reason for using the free version of ChatGPT is that it is accessible to everyone at any time without requiring membership. The study's data was gathered through the semi-structured problem-posing instructions developed by Van Harpen and Sriraman (2013). The prompt used was:

"There is a triangle and its inscribed circle. Make up as many problems as you can that are in some way related to this situation and suitable for 11th and 12th grade students. The problems could also be real-life problems. Again, do not limit yourself to the problems you have seen or heard of - try to think of as many possible and challenging mathematical problems as you can."

Boral, Mondal and Saikia (2024) conducted a review of studies involving the use of ChatGPT in education and found that most of the research has focused on medical education. Among the 30 studies they analyzed, only one addressed primary education and another considered general school education, underscoring the scarcity of research at the K–12 level. In light of this gap, the present study adopts the problem-posing task designed by Van

Harpen and Sriraman (2013), targeting senior high school and first-year college students, as a reference framework for evaluating ChatGPT-generated problems.

Since the free version of ChatGPT does not yet have an image detection feature, the given figure was described, and ChatGPT was asked to pose problems appropriate to the given situation. The participants in the Van Harpen and Sriraman (2013) study were senior high school and first-year college students from two different regions in the USA and China. Thus, ChatGPT was instructed to answer according to the levels of 11th and 12th grade students. Following this initial instruction, the commands given to ChatGPT and the questions asked are provided in the appendix.

Data Analysis

The problems that ChatGPT posed in accordance with the given commands and the answers it gave to the questions about the problems it posed were analyzed by the content analysis method. In order to compare them with the results obtained from the study of Van Harpen and Sriraman (2013), the problems posed by ChatGPT were also analyzed in the context of content area, originality, fluency and flexibility. Examples of the problems posed by ChatGPT according to content areas are given in Table 1.

Table 1. Content Area

Content	Examples of problems
Geometry concept	P1. Given a triangle with sides of lengths a,b, and c, express the radius r of the inscribed circle in terms of a, b, and c.
Algebra and trigonometry	P2. If R is the circumradius and r is the inradius of a triangle, prove that $R \geq 2r$.
Optimization and calculus	P10. Among all triangles with a fixed area, find the dimensions of the triangle that minimize its perimeter. What is the minimum perimeter in terms of the area?
Coordinate geometry	P15. Derive the equation of the tangent to the incircle of a triangle at a given point (x_0,y_0) on the circle.
Real life applications	P4. An architect is designing a triangular park. If the park has an inscribed circle with a desired radius of 20 meters, and the sides of the triangle are 50 meters, 65 meters, and 80 meters, find the area of the park.
Proof and logical reasoning	P7. Prove that the area of a triangle A is equal to rs, where r is the inradius and s is the semi-perimeter of the triangle.
Interdisciplinary connections	P14. A thin uniform triangular plate has vertices at $(0,0)$,$(a,0)$, and $(0,b)$. If the plate has a uniform density ρ, find the coordinates of its center of mass.
Challenge problems	P9. For an equilateral triangle with side length s, find the ratio of the area of the inscribed circle to the area of the triangle.
Hands- on and visual	P10. Among all triangles with a fixed area, find the dimensions of the

Content	Examples of problems
explorations	triangle that minimize its perimeter. What is the minimum perimeter in terms of the area?
Number theory	P16. Show that if a triangle has an integer inradius and all side lengths are integers, then the triangle is a right-angled triangle.

The data obtained from ChatGPT was analysed by two researchers based on the analysis framework used by Van Harpen and Sriraman (2013), and common codes were tried to be obtained by comparing the analyses made.

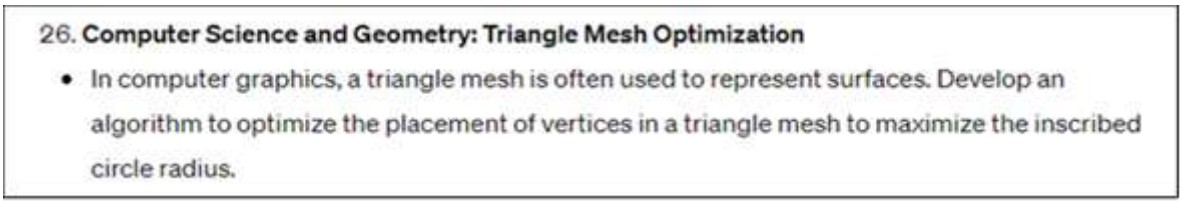
Results

Analysis of the Problems Posed by ChatGPT in the Context of Fluency

As a result of the first directive directed to ChatGPT, the tool produced 10 different problems. When it was asked to pose the problem two more consecutive times, it posed 10 problems each time. It has been observed that the problems in each new group are more complex than those in the previous group. ChatGPT stated that an infinite number of different problems can be posed to suit the given situation.

Analysis of the Problems Posed by ChatGPT in the Context of Originality

After ChatGPT posed 30 problems, the following question was asked: "Which one of the problems that you posed is the most original one? And why?". In the context of originality, ChatGPT identified the 26th problem, which was related to computer science, as the most original. This problem is presented in Figure 1.


26. **Computer Science and Geometry: Triangle Mesh Optimization**

- In computer graphics, a triangle mesh is often used to represent surfaces. Develop an algorithm to optimize the placement of vertices in a triangle mesh to maximize the inscribed circle radius.

Figure 1. Screenshot of Problem 26 Posed by ChatGPT

ChatGPT mentioned that it selected this problem due to its uniqueness and interdisciplinarynature. It highligted other elements such as interdisciplinary connections, optimization challenges, real world applications and critical thinking. When asked to propose additional original problems, it noted that the field of mathematics is vast and countless problems could be posed, and subbequently suggested six more problems. However, the problems it posed were about reasoning, reasoning and argument and inquiry rather than requiring mathematical calculations. A few examples are as follows:

> "Modify parameters: Change the parameters in the problems, such as the side lengths of the triangle, the radius of the inscribed circle, or the angles. This can lead to a variety of problem variations."
> "Introduce Constraints: Impose specific conditions or constraints on the triangle, such as having one

angle fixed, or certain relationships between the side lengths."

"Technology integration: Incorporate technology tools like dynamic geometry software or computational tools to explore and analyze geometric relationships dynamically."

Analysis of the Problems Posed by ChatGPT in the Context of Flexibility

The 30 problems generated by ChatGPT in line with the questions posed to them were classified in terms of content areas by two researchers. At the same time, ChatGPT was asked to classify the problems it encountered. Table 2 summarizes the classifications made by researchers and ChatGPT.

Table 2. Classification of Problems

Categories	The first group (1-10)		The second group (11-20)		The third group (21-30)	
	Researchers	ChatGPT	Researchers	ChatGPT	Researchers	ChatGPT
Geometry concepts	1, 4, 7, 9	1, 3, 6, 7, 8	17, 20	13, 15, 18	29	23, 28
Proof	2, 3, 7, 8	7, 8	11, 12, 16			23
Algebra	2	2, 4, 5, 9		11, 12, 17, 19		22, 25
Trigonometry	3, 8	2, 4, 5, 9	12	11, 12, 17, 19		22, 25
Real life	4	4, 5	17	14, 16		27, 29
Calculus	5, 10	10	18	24	26, 27	26
Analytic Geometry	6	7	13, 14, 15	13, 15	21	
Interdisciplinary			14	14		26, 27, 29
Probability			19		23	
Complex numbers					22	
Differential Equations					24	
Challenge		9		20		
Number Theory				16		
Hands-on and visual explorations		10				

The researchers found that problems 25, 28 and 30 generated by ChatGPT were irrelevant to the provided figure. Although problem 25 requires knowledge of calculus and real life, it is not related to drawing in any way. Problem 28 involves topology; however, solving it does not require an inner tangent circle, making it disconnected from the visual prompt. Similarly, while problem 30 relates to the area of a triangle and has real-life context, it, too, lacks relevance to the depicted figure.

In addition, ChatGPT did not assign problems 21 and 30 to any specific category. Among the 30 posed problems, it was observed that the last 10 were more complex and required advanced mathematical knowledge. Another notable finding is that three of the last 10 problems were unrelated to the given figure. The problems mentioned

are shown in Figure 2.

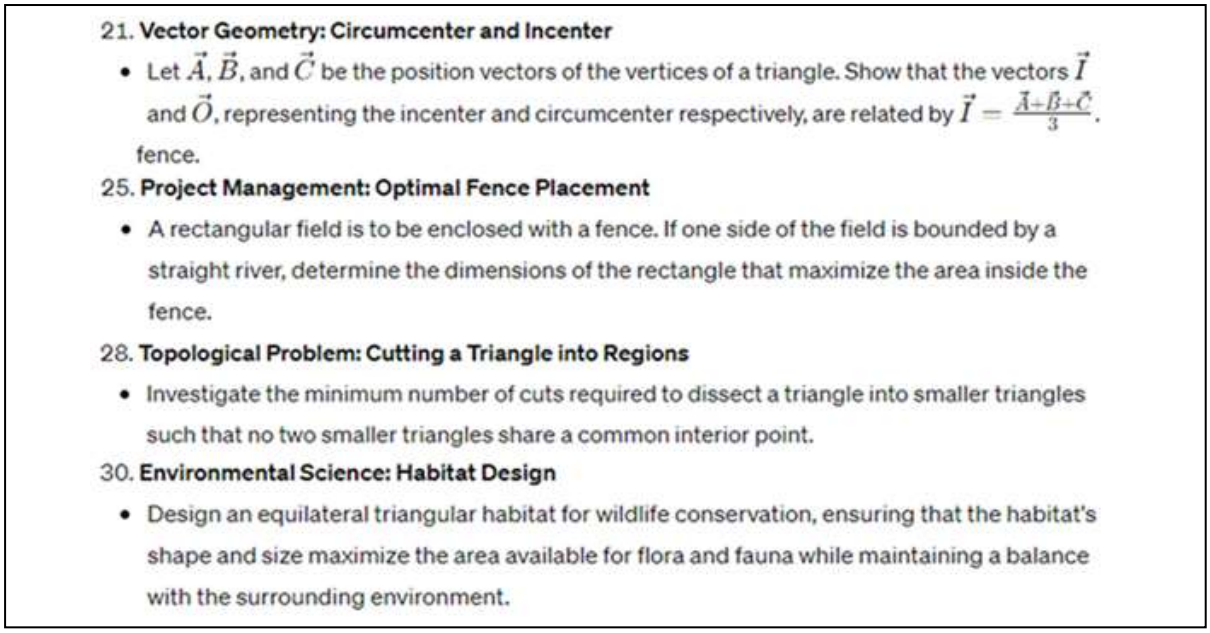


Figure 2. Screenshot of Problems 21, 25, 28 and 30 Generated by ChatGPT

Flexibility-Based Comparison of Problems by ChatGPT and Students

In the study of Van Harpen and Sriraman (2013), 10 categories were created in the classification of the problems posed by the students: analytical geometry, lengths, area, angles, transformation, use of auxiliary figures, three-dimensional (3D), probability, proof and others. In the current study, ChatGPT was asked to classify the 30 problems it generated according to these same categories. ChatGPT's classification and corresponding percentages are presented in Table 3.

Table 3. Problem Classification Based on Van Harpen and Sriraman's (2013) Framework

Van Harpen and Sriraman (2013)	Problems	Frequency
1.Analytic Geometry	6, 13, 15, 23	4 (%13,3)
2. Lengths	1, 2, 10, 16, 26	5 (%16,7)
3. Area	4, 7, 9, 10, 14, 20, 26	7 (%23,3)
4. Angles	3, 8, 11, 12, 26	5 (%16,7)
5. Transformation	26	1 (%3,3)
6. Use of auxiliary figures	1, 3, 6, 7, 8, 10, 13, 14, 15, 23, 28	11 (%36,7)
7. 3D	26	1 (%3,3)
8. Probability	19, 29	2 (%6,7)
9. Proof	7, 8, 23	3 (%10)
10. Others	5, 17, 18, 21, 22, 24, 25, 27, 30	9 (%30)

In Van Harpen and Sriraman's (2013) study, the topics where students posed the most problems were area and length, while ChatGPT's were use of auxiliary figures and area category. While US Students and Shanghai students did not pose any problems in the transformation and proof content areas, ChatGPT posed one problem

in the transformation field and three problems in the proof field. The students from Jiaozhou posed problems similar to ChatGPT in both content areas. In the study of Van Harpen and Sriraman (2013), it was observed that all three groups produced very few problems in the analytical geometry content area. On the other hand, ChatGPT has posed more problems in the field of analytic geometry.

In ChatGPT's own classification, unlike the classification made by Van Harpen and Sriraman (2013), there are categories such as "Interdisciplinary connections, Algebra, Trigonometry, Challenge problems, Hands-on and visual exploration, Number theory, Real life, Optimization and calculus". For example, while ChatGPT placed problem 5 as trigonometry and algebra in its own classification, it placed the same problem in the other category of Van Harpen and Sriraman (2013). Similarly, it placed problem 27, which was in the other category, in the real life and interdisciplinary category in its own classification. Although the analysis made by the researchers is not exactly the same as ChatGPT, the categories determined are similar or even more than ChatGPT's categories. When these comparisons were evaluated, it was seen that ChatGPT posed problems in different content areas than the students in the Van Harpen and Sriraman (2013) study. These findings can be interpreted as ChatGPT being more flexible in problem posing.

Discussion

The aim of this study was to reveal how creative (artificial creativity of) ChatGPT is in posing mathematical problems in terms of fluency, flexibility and originality. Since ChatGPT states that it can pose an infinite number of problems, it has been concluded that its fluency capacity is quite high. It has been observed that ChatGPT chooses the most original one among the problems it poses, by referring to the definition of originality. In terms of flexibility, it was concluded that ChatGPT was able to pose problems in different content areas and that the problems it posed were more flexible than the students in the study of Van Harpen and Sriraman (2013), which was referenced in the current study.

Although ChatGPT stated that it could pose endless problems, the problems it posed in the later stages are either problems that addressed similar content areas or problems that deviated from the given instructions. This necessitates that the problems posed by ChatGPT be checked by a researcher or teacher. If problem posing is to be done with students during the teaching process, with support from ChatGPT, the problems posed should be examined by an expert, and the student should not be left alone with the artificial intelligence tool. This result emphasizes the importance of ethics, reliability and effective use, which Rasul et al. (2023) pointed out. In addition, the statements of the participants in the study of Adeshola and Adeola (2023) reveal that the possibility of artificial intelligence replacing the teacher is unlikely. As Esling and Devis (2020) emphasized the importance of human-machine interaction in creativity, ChatGPT can be utilized in creative problem-posing activities, not independently, but under the guidance of an expert or teacher.

If not used under the supervision of an expert or teacher, ChatGPT may be off task and give incorrect answers. For example; although it included the 14th Problem in "area" content area, when the question was examined, it was seen that it is suitable for the analytical geometry content area. In fact, while placing the 6th Problem it posed

in the category of geometric concepts in its first classification; according to Van Harpen and Sriraman (2013), when wanted to classify it, ChatGPT placed it in the analytical geometry category. This situation can be considered as an indication that ChatGPT contradicts itself. In this context, Javaid et al. (2023) stated, it should also be noted that users may sometimes provide inaccurate information. There are studies in the literature showing that ChatGPT may give incorrect answers and that it should be used carefully, especially in matters related to higher-level thinking (Tlili et al., 2023; Wardat et al., 2023; Urhan et al., 2024; Rane, 2023). This research also yielded results that support previous studies in the literature. Wardat et al. (2023) stated that ChatGPT sometimes gives incomplete or incorrect mathematical answers and that its understanding of geometry is not sufficient. In this study, the reason why the problems posed by ChatGPT in the following steps are not suitable for the desired situation may be that geometric shapes cannot be drawn in the free version of ChatGPT.

Conclusion

This study provides insights into the potential and limitations of ChatGPT in generating mathematically creative problems, evaluated through the lenses of fluency, flexibility, and originality. The findings demonstrate that although ChatGPT possesses a high fluency capacity, its performance tends to decline in later stages, with some problems deviating from the initial instructions or losing relevance to the given visual prompt. While the model exhibited flexibility by generating problems across various mathematical content areas and showed the capacity to select more original responses, inconsistencies in its classifications and occasional misinterpretations raise concerns regarding its autonomous use.

These findings highlight that although ChatGPT holds promise as a supportive tool in creative mathematical tasks, it cannot substitute the role of educators. Instead, it should be integrated into educational settings under the supervision of experts to ensure the accuracy and relevance of the problems it generates. In line with previous studies, this research reaffirms the need for careful monitoring when using AI tools in education, particularly in tasks requiring higher-order thinking and conceptual understanding. Further research is recommended to explore how AI-supported tools like ChatGPT can be effectively and ethically utilized in mathematics classrooms, especially in fostering students' creative and metacognitive skills.

Recommendations

Given the findings of this study, several recommendations can be made. First, educators aiming to incorporate ChatGPT into mathematics classrooms should do so under guided supervision to ensure the appropriateness and accuracy of the problems generated. Teacher mediation is especially critical when using AI in creative tasks, to prevent students from being misled by incorrect or irrelevant outputs.

Second, developers of AI-based language models should consider improving the tools' ability to interpret visual stimuli and align generated problems with specific instructional objectives. For instance, enhancing geometry-related functionalities could help overcome current limitations in tasks involving figures or spatial reasoning.
Lastly, future research could expand the scope of this study by involving diverse student populations and

conducting longitudinal analyses. Comparative studies between student-generated and AI-generated problems can also offer deeper insights into the role of artificial intelligence in fostering mathematical creativity.

References

Adeshola, I., & Adepoju, A. P. (2024). The opportunities and challenges of ChatGPT in education. *Interactive Learning Environments, 32*(10), 6159-6172. https://doi.org/10.1080/10494820.2023.2253858

Akay, H., Soybas, D., & Argun, Z. (2006). Problem posing experiences and using open-ended questions in mathematics teaching. *Kastamonu Education Journal*, *14*(1), 129-146.

Ayllón, M. F., Gómez, I. A., & Ballesta-Claver, J. (2016). Mathematical thinking and creativity through mathematical problem posing and solving. *Journal of Educational Psychology-Propósitos y Representaciones*, *4*(1), 195-218. http://dx.doi.org/10.20511/pyr2016.v4n1.89

Azmi, A., Maakip, I., Voo, P., Mohd Nawi, N.H., Kiflee, D.N.A., Halik, M., Marshall, S., & Quah, W.B. (2024). Beyond the bot: ChatGPT's influence on student learning. *International Journal of Education in Mathematics, Science, and Technology (IJEMST), 12*(6), 1488-1503. https://doi.org/10.46328/ijemst.4435

Bishop, J.M. (2021). Artificial intelligence is stupid and causal reasoning will not fix it. *Frontiers in Psychology, 11*. https://doi.org/10.3389/fpsyg.2020.513474

Boral, K., Mondal, K.K., & Saikia, A. (2024). Exploring the landscape: A bibliometric analysis of 30-most cited articles on ChatGPT in education. *International Journal of Education in Mathematics, Science, and Technology (IJEMST), 12*(6), 1583-1602. https://doi.org/10.46328/ijemst.4443

Brown, S. I., & Walter, M. I. (2005). *The art of problem posing*. Psychology Press.

Cai, J. (2003). Singaporean students' mathematical thinking in problem solving and problem posing: an exploratory study. *International Journal of Mathematical Education in Science and Technology*, *34*(5), 719-737.

Cai, J., & Hwang, S. (2002). Generalized and generative thinking in US and Chinese students' mathematical problem solving and problem posing. *The Journal of Mathematical Behavior*, *21*(4), 401-421.

Daher, W.,& Anabousy, A. (2018). Flexibility of pre-services teachers in problem posing in different environments. *In: Singer, F. (eds) Mathematical Creativity and Mathematical Giftedness*. ICME-13 Monographs. Springer, Cham. https://doi.org/10.1007/978-3-319-73156-8_9

Esling, P., & Devis, N. (2020). Creativity in the era of artificial intelligence. arXiv preprint arXiv:2008.05959.

Frieder, S., Pinchetti, L., Griffiths, R. R., Salvatori, T., Lukasiewicz, T., Petersen, P., & Berner, J. (2024). Mathematical capabilities of CHATGPT. *Advances in Neural Information Processing Systems*, 36.

Gusteti, M.U., Rasli, A.M., Wulandari, S., Mulyati, A., Hayati, R., Azmi, K., Elza, S.S., & Resi, N. (2024). Bibliometric study on ChatGPT. *International Journal of Education in Mathematics, Science, and Technology (IJEMST)*, *12*(6), 1435-1450. https://doi.org/10.46328/ijemst.4447

Halaweh, M. (2023). ChatGPT in education: Strategies for responsible implementation. *Contemporary Educational Technology*, *15*(2), ep421. http://doi.org/10.30935/cedtech/13036

Javaid, M., Haleem, A., Singh, R. P., Khan, S., & Khan, I. H. (2023). Unlocking the opportunities through ChatGPT tool towards ameliorating the education system. *BenchCouncil Transactions on Benchmarks,*

Standards and Evaluations, 3(2). https://doi.org/10.1016/j.tbench.2023.100115

Jennings, K. E. (2010). Developing creativity: Artificial barriers in artificial intelligence. *Minds and Machines, 20*, 489-501. https://doi.org/10.1007/s11023-010-9206-y

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. https://doi.org/10.1016/j.lindif.2023.1022744

Leung, S. S., & Silver, E. A. (1997). The role of task format, mathematics knowledge, and creative thinking on the arithmetic problem posing of prospective elementary school teachers. *Mathematics Education Research Journal*, *9*(1), 5-24.

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13(4), 410. https://doi.org/10.3390/educsci13040410

Marrone, R., Taddeo V., & Hill, G. (2022). Creativity and Artificial Intelligence—A Student Perspective. *Journal of Intelligence 10*, 65. https://doi.org/10.3390/jintelligence10030065

Rane, N. (2023). *Enhancing Mathematical Capabilities through ChatGPT and Similar Generative Artificial Intelligence: Roles and Challenges in Solving Mathematical Problems*. http://dx.doi.org/10.2139/ssrn.4603237

Rasul, T., Nair, S., Kalendra, D., Robin, M., de Oliveira Santini, F., Ladeira, W. J., Sun, M., Day, I., Rather, R. A., & Heathcote, L. (2023). The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied Learning and Teaching*, *6*(1), 41–56. https://doi.org/10.37074/jalt.2023.6.1.299

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?. *Journal of Applied Learning & Teaching*, *6*(1). https://doi.org/10.37074/jalt.2023.6.1.9

Runco, M. A. (2023). AI can only produce artificial creativity. *Journal of Creativity, 33*(3), 100063.

Sánchez-Ruiz, L. M., Moll-López, S., Nuñez-Pérez, A., Moraño-Fernández, J. A., & Vega Fleitas, E. (2023). ChatGPT challenges blended learning methodologies in engineering education: A case study in mathematics. *Applied Sciences*, 13(10), 6039. https://doi.org/10.3390/app13106039

Segal, R. & Biton Y. (2024). The contribution that utilizing generative AI for problem posing makes to pre-service high school mathematics teachers' TPACK. *International Journal of Education in Mathematics, Science, and Technology (IJEMST)*, 12(6), 1559-1582. https://doi.org/10.46328/ijemst.4591

Shuk-Kwan, S. L. (1997). On the role of creative thinking in problem posing. *ZDM Mathematics Education*, 29(3), 81-85.

Silver, E. A. (1994). On mathematical problem posing. *For the Learning of Mathematics*, 14(1), 19-28.

Silver, E. A. (1997). Fostering creativity through instruction rich in mathematical problem solving and problem posing. *ZDM–Mathematics Education*, *29*(3), 75-80.

Singer, F. M., Ellerton, N., & Cai, J. (2013). Problem-posing research in mathematics education: New questions and directions. *Educational Studies in Mathematics*, *83*, 1-7.

Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing*

Writing, 57. https://doi.org/10.1016/j.asw.2023.100752

Su, J., & Yang, W. (2023). Powerful or mediocre? Kindergarten teachers' perspectives on using ChatGPT in early childhood education. *Interactive Learning Environments*, 1-13. https://doi.org/10.1080/10494820.2023.2266490

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using Chatbots in education. *Smart Learning Environments*, 10(1), 15. https://doi.org/10.1186/s40561-023-00237-x

Tu, Y.F. & Hwang, G.J. (2023). University students' conceptions of ChatGPT supported learning: a drawing and epistemic network analysis, *Interactive Learning Environments*, 1-25. https://doi.org/10.1080/10494820.2023.2286370

Urhan, S., Gencaslan, O., & Dost, S. (2024). An argumentation experience regarding concepts of calculus with ChatGPT. *Interactive Learning Environments*, 1-26. https://doi.org/10.1080/10494820.2024.2308093

Van Harpen, X. Y., & Sriraman, B. (2013). Creativity and mathematical problem posing: an analysis of high school students' mathematical problem posing in China and the USA. *Educational Studies in Mathematics*, *82*(2), 201-221.

Wardat, Y., Tashtoush, M. A., AlAli, R., & Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *Eurasia Journal of Mathematics, Science and Technology Education*, 19(7). https://doi.org/10.29333/ejmste/13272

Yunianto, W., Galic, S., & Lavicza, Z. (2024). Exploring computational thinking in mathematics education: Integrating ChatGPT with GeoGebra for enhanced learning experiences. *International Journal of Education in Mathematics, Science, and Technology (IJEMST), 12*(6), 1451-1470. https://doi.org/10.46328/ijemst.4437

Zeng, F. (2023). *Evaluating the Problem Solving Abilities of ChatGPT*. McKelvey School of Engineering Theses & Dissertations. 849. https://openscholarship.wustl.edu/eng_etds/8499

Author Information

Çiğdem Arslan
https://orcid.org/0000-0001-7354-8155
Bursa Uludag University
Türkiye
Contact e-mail: *arslanc@uludag.edu.tr*

Hatice Kübra Güler Selek
https://orcid.org/0000-0002-6262-8421
Bursa Uludag University
Türkiye

Appendix. ChatGPT Questions

1. There is a triangle and its inscribed circle. Make up as many problems as you can that are in some way related to this situation and suitable for 11th and 12th grade students. The problems could also be real-life problems. Again, do not limit yourself to the problems you have seen or heard of- try to think of as many possible and challenging mathematical problems as you can.
2. Could you pose more different problems related to this situation and suitable for 11th and 12th grade students?
3. We are wondering if you could pose more different problems related to this situation. Don't forget they must be suitable for 11th and 12th graders.
4. According to you, how many different problems could be posed related to the given situation?
5. Do you think all of these are suitable for 11th and 12th-grade students?
6. Do you think original problems can be produced other than what you have written and suitable for 11th and 12th graders?
7. Which one of the problems that you posed is the most original one? And why?
8. Which one of the problems that you posed is the most creative one? And why?
9. How did you distinguish between creativity and originality when choosing these problems?
10. Which one of the problems that you posed is both the most original and creative one? And why?
11. If you were asked to classify the 30 problems you created on the given situation, how would you classify them?
12. If you were asked to classify the 30 problems you created on the given situation, according to analytical geometry, lengths, area, angles, transformations, involving auxiliary figures, three-dimensional, probability, proofs and others categories?