

Investigating the use of artificial intelligence supported tools by preservice middle school mathematics teachers in the numbers and quantities theme

 Merve Çınar¹

 Aykut Bulut¹ 

RESEARCH ARTICLE



 10.65933/zq3xp497

¹ Kırşehir Ahi Evran University,
Türkiye

 Correspondence
aykut.bulut@ahievran.edu.tr

Article History

Received: 21.10.2025

Accepted: 15.12.2025

Published: 22.12.2025

Keywords

Artificial intelligence tools,
ChatGPT, Gemini, Preservice
mathematics teachers, Numbers
and quantities

Abstract

Technology and artificial intelligence (AI) have increasingly influenced teaching and learning processes. With their integration into education, AI-based tools have begun to alleviate instructional challenges faced by teachers and preservice teachers, including lesson planning, activity design, assessment, and mathematical problem posing. One of the central themes of the middle school mathematics curriculum, Numbers and Quantities, has long been emphasized in mathematics education research due to its conceptual complexity and instructional significance. The use of AI tools within this theme is therefore expected to support instruction and instructional design. Accordingly, this study aimed to examine the use of two widely adopted AI tools—ChatGPT and Gemini—by preservice middle school mathematics teachers. The study focused on how these tools were used in relation to the Grade 6 learning outcome defined in the Turkish Century Education Model: solving real-life problems that require the four basic operations involving fractions, decimals, and percentages. A qualitative case study design was employed with 45 fourth-year preservice teachers enrolled in an elementary mathematics teacher education program at a public university. Data were collected through problem-posing and problem-solving tasks and analyzed using descriptive analysis. The findings showed that most problems constructed independently by preservice teachers did not fully meet the specified learning outcome, whereas a lower proportion of misaligned problems was observed in those generated using ChatGPT and Gemini. However, preservice teachers experienced difficulties in writing effective prompts, and AI-generated problems frequently relied on repetitive contexts, outdated numerical values, or incomplete data. These results emphasize the need for critical evaluation and validation of AI-generated outputs for instructional use.

Cite this as

Çınar, M., & Bulut, A. (2025). Investigating the use of artificial intelligence supported tools by preservice middle school mathematics teachers in the numbers and quantities theme. *Journal of Educational Innovations & Practices*, 1, Article e5. <https://doi.org/10.65933/zq3xp497>

1 INTRODUCTION

Technology, which has long been a decisive factor in the historical development of civilizations, has reached an unprecedented level of advancement in the contemporary era. As a result of these progressive developments, knowledge production and dissemination have evolved incrementally, laying the foundation for the rapid expansion of information and communication technologies. Recent technological advancements have given rise to innovative digital content generation techniques, most notably artificial intelligence (AI) systems (Bishop, 2021). Although AI was initially conceptualized primarily in relation to supercomputing infrastructures, its scope has since expanded considerably to encompass embedded computing systems and a wide range of everyday applications (Chen, Chen, & Lin, 2020).

Education represents one of the sectors most profoundly influenced by technological change. Traditional instructional tools, such as chalkboards, notebooks, and printed textbooks, have increasingly been replaced by interactive whiteboards and digital learning resources, while conventional instructional approaches have gradually given way to technology-driven educational practices. This transformation has further evolved with the integration of generative AI tools, marking a new phase in educational innovation. By facilitating learning processes and enhancing learner autonomy, AI technologies have contributed substantial benefits to the education sector, leading to what may be characterized as a modest yet impactful transformation in educational practices (Osman & Ahmed, 2024).

AI systems are designed by drawing inspiration from the functional processes of the human brain, enabling them to perform assigned tasks while continuously improving through accumulated experience. Advanced AI systems, commonly referred to as large language models, are specifically developed to comprehend and generate texts that closely resemble human-written discourse. In fact, AI has demonstrated remarkable success in text processing, to the extent that its outputs are often indistinguishable from those produced by humans (Bishop, 2021; OpenAI, 2024). With the emergence of Generative Pre-trained Transformer (GPT) models, AI systems have become capable of generating text, translating across languages, summarizing content, and performing a wide range of linguistic tasks by leveraging patterns and structures inherent in human language (OpenAI, 2024). Owing to advances in deep learning techniques within the field of natural language processing, GPT-based systems exhibit a strong capacity for producing human-like textual outputs (Brown et al., 2020; Othman, 2023).

Among these systems, the AI-powered conversational agent known as ChatGPT, developed by OpenAI, has emerged as one of the most prominent, influential, and widely adopted tools (Grassini, 2023; OpenAI, 2024). Empirical studies have demonstrated that ChatGPT is capable of generating contextually appropriate, fluent, and original texts, while also exhibiting a flexible and innovative structure (Roose, 2022; Lund & Wang, 2023). Furthermore, analyses of Google Trends data indicate a gradual and consistent increase in the use of ChatGPT within the field of education, both globally and within the national context. Given its ease of access from any location with an internet connection, its availability without direct financial cost, and its more widespread adoption compared to other language processing models, ChatGPT—alongside Gemini—was selected as the focus of the present study.

Designing lesson plans that are effective, goal-oriented, and responsive to students' needs and individual differences constitutes a highly complex and time-intensive process for teachers. One of the primary challenges encountered in this process is the insufficient consideration or integration of core lesson planning components—namely learning objectives, instructional activities, and assessment elements—which are widely regarded as fundamental to instructional design (Sebullen, 2023). In recent years, the incorporation of AI-based tools into educational contexts has introduced new opportunities for addressing these challenges. Beyond lesson

planning, AI has also been shown to exert influence across multiple domains, ranging from student achievement to instructional processes (Temur, 2024).

ChatGPT supports teachers' idea-generation processes by facilitating brainstorming, fostering creative thinking, and enabling rapid access to relevant resources and complex information. Within this context, ChatGPT can function as an extensive knowledge repository and assist lesson planning by generating instructional materials and educational resources tailored to specific teaching goals (Castro, Cachicatari, Aste, & Medina, 2024). Importantly, ChatGPT should not be positioned as a substitute for teachers; rather, it should be employed as a supportive tool that assists educators in identifying more effective instructional strategies and developing well-structured lesson plans (Samala et al., 2024). Moreover, AI-based tools such as ChatGPT may serve as an initial reference point for novice teachers or recent graduates with limited teaching experience and pedagogical content knowledge, particularly during lesson planning and content development processes.

In addition, AI systems are increasingly capable of generating exercises, tests, and even personalized assessment materials aligned with specific learning objectives, as well as automating the grading of assignments and assessments. These functionalities have the potential to reduce teachers' workload and contribute to more efficient use of instructional time (Farrokhnia, Banihashem, Noroozi, & Wals, 2024; Zhai, 2022). Empirical studies have similarly argued that educators can leverage generative AI technologies such as ChatGPT to support students' learning processes (Alshahrani, 2023). Furthermore, evidence suggests that AI-supported personalized learning approaches may lead to improvements in student achievement (Yeniçeri & Kenan, 2025). Despite these advantages, research has also documented potential drawbacks associated with the incorrect or excessive use of AI technologies, including declines in learning quality, reduced academic performance, and weakened communication skills (Chen & Lin, 2024). Another critical concern relates to the possibility of AI tools generating inaccurate or incomplete information. When educators lack sufficient expertise or time to verify AI-generated outputs, monitoring and validating these responses can become a substantial challenge (Samala et al., 2024). Accordingly, it is essential to remain cognizant of the limitations and potential disadvantages of ChatGPT and similar AI-based tools. As the adoption of AI continues to expand and teachers' professional roles evolve, decisions regarding when, how, and to what extent AI should be used must ultimately remain under the control of teachers themselves (Baidoo-Anu & Owusu Ansah, 2023).

Another factor influencing the effective and efficient use of AI systems is the formulation of appropriate prompts. Prompts can be defined, in their simplest form, as the commands or instructions provided to AI models, which guide the generation of specific responses or outputs (YEĞİTEK, 2025). Consequently, producing accurate, consistent, and purpose-driven outputs from AI tools requires the use of clear and well-structured prompts (Şengel, 2024). Open-ended prompts, in particular, may result in unpredictable outcomes, while even minor modifications in wording or alternative lexical choices can lead to substantially different—or potentially undesirable—outputs (Zhou, Yang, Loy, & Liu, 2022). This highlights the necessity for teachers and preservice teachers to possess foundational knowledge and skills related to prompt design in order to integrate AI tools meaningfully and effectively into teaching and learning processes (YEĞİTEK, 2025). Given that AI technologies present both opportunities and risks, their use must be carefully aligned with instructional goals to minimize potential negative consequences (Yılmaz, 2023).

The integration of technology into mathematics instruction is widely recognized as both an essential and indispensable principle (NCTM, 2008). When leveraged appropriately, technological tools can make substantial contributions to mathematics teaching and enrich mathematical content (Baki, Yalıçinkaya, Özpinar, & Uzun, 2009). Furthermore, technology that is aligned with instructional content supports students' deep and comprehensive understanding, enabling them to focus more effectively on problem solving, reasoning, and mathematical justification (NCTM, 2008). In Turkish, the Ministry of National Education (MoNE) has

incorporated AI-based applications into the education system as part of its 2023 vision aimed at enhancing educational quality, a commitment that has been concretized through the updated national curriculum, the Turkish Century Education Model. In this context, the General Directorate of Innovation and Educational Technologies published the *Artificial Intelligence Tools Used in Education: Teacher Handbook* in 2024 to inform and inspire teachers. Additionally, studies in the literature serve as valuable resources by documenting various applications of AI in education and their associated benefits. Collectively, these developments may support teachers and preservice teachers in keeping pace with current innovations and adapting to emerging technologies. The use of AI applications has also been shown to deepen students' mathematical understanding and facilitate the learning of mathematical concepts (Lee & Perret, 2022).

AI-supported tools and systems that contribute to enriched learning experiences in mathematics instruction have been found to promote students' problem-solving skills and foster positive attitudes toward learning (Richard, Vlez, & Van Vaerenbergh, 2022). Moreover, the explicit definition of mathematical rules and the deterministic and verifiable nature of mathematical solutions constitute factors that facilitate the application of AI in mathematics education (Holmes, Bialik, & Fadel, 2019). In Turkish, "Mathematical Problem Solving" and "Working with Mathematical Tools and Technology" are among the five core skill areas targeted in the Middle School Mathematics Curriculum published in 2024 (MoNE, 2024). Accordingly, problem-solving competence occupies a central position among mathematical proficiencies. Developing this competence not only facilitates students' transition to subsequent levels of education but also supports their preparation for real-life problem contexts (Baykul, 2009).

Research has demonstrated that problem-posing activities in mathematics support conceptual understanding, strengthen analytical and synthetic thinking as well as abstraction and generalization skills, and enhance creative and critical thinking abilities (Bayazit & Dönmez, 2017). Furthermore, problem posing and problem solving are often addressed jointly and are regarded as central components of mathematics education and mathematical thinking (Silver, 1997). In this regard, when employed appropriately by teachers, AI technologies can make a meaningful contribution to students' problem-solving processes. Given that today's preservice teachers are tomorrow's educators, the extent to which they can effectively utilize AI-supported tools and emerging technologies represents a critical question. Future mathematics teachers are expected not only to convey mathematical knowledge but also to design AI-supported learning environments and manage these environments effectively.

One of the thematic areas in which problem solving—a core domain-specific skill in mathematics—becomes particularly prominent is the theme of "*Numbers and Quantities*". Within this theme, concepts such as operations with fractions and decimal representations are addressed in conjunction with problem-solving skills, and students are expected to demonstrate progress in these competencies (MoNE, 2024). The results of Osana and Royea's (2011) study, as well as Xie and Masingila's (2017) study indicated substantial deficiencies in preservice mathematics teachers' conceptual understanding, particularly in relation to fractions and their operations. These difficulties were evident in their limited ability to generate mathematically meaningful solution strategies and to express those strategies through appropriate symbolic representations. Furthermore, the abstract nature of the fraction concept and the necessity of relating it to real-life contexts (Yenilmez & Kocaoğlu, 2010), together with the difficulties experienced by both students and teachers in working with multiple representations of fractions (e.g., decimals, percentages), further underscore the instructional significance of this content area (Özdemir, Bayraktar, & Yılmaz, 2017). Accordingly, the purpose of the present study is to examine the use of artificial intelligence tools by preservice middle school mathematics teachers in relation to the learning outcome specified in the Grade 6 "*Numbers and Quantities*" theme of the Turkish Century Education Model: "*solving problems that require the four basic operations involving fractions, decimals, and percentages encountered in real-life situations*". Among AI tools, ChatGPT and Google Gemini—two of the most widely used systems—were

selected for analysis.

In line with this purpose, the research questions guiding the study are as follows:

1. What are the characteristics of preservice middle school mathematics teachers' problem-posing and problem-solving processes within the Numbers and Quantities theme?
2. How do preservice middle school mathematics teachers utilize ChatGPT and Google Gemini within the Numbers and Quantities theme?

2 METHODOLOGY

2.1 Research design

This study adopted a case study design, one of the qualitative research approaches. Case studies aim to examine research problems in a multidimensional manner and are grounded in an in-depth and detailed investigative framework (Patton, 2015; Yin, 2009). They are particularly suited to addressing questions related to *how* and *why* phenomena occur within real-world contexts (Yin, 2009). Accordingly, in the present study, preservice mathematics teachers' problem-posing processes and the ways in which they enacted these processes through the use of ChatGPT and Google Gemini were explored. Data were obtained through document analysis, observations, and field notes, yielding rich descriptive and verbal qualitative data.

2.2 Participants

The study group consisted of 45 preservice teachers enrolled in the fourth year of the Elementary Mathematics Teacher Education program at a public university located in the Central Anatolia region of Türkiye. Using a convenience sampling strategy based on accessibility and suitability for the research purpose, 15 male and 30 female participants were selected. To examine their experiences with AI-based tools, participants were administered a structured problem-posing and problem-solving interview form, and their interactions and processes within ChatGPT and Google Gemini were documented in written form.

The decision to include only fourth-year preservice teachers was intentional. Throughout their undergraduate education, these participants had completed coursework related to technology integration—such as information technologies, instructional technologies and material design, and computer-assisted mathematics instruction—as well as pedagogy- and content-specific courses, including principles and methods of teaching, elementary mathematics teaching, teaching numbers, and mathematical problem solving. Compared to students at earlier grade levels, fourth-year preservice teachers are more academically experienced and closer to professional practice; therefore, they are considered to be better prepared and more competent in engaging with problem-posing and problem-solving processes.

2.3 Data Collection Tools and Procedures

Data were collected using a structured problem-posing and problem-solving interview form consisting of three questions. As presented in Table 1, these questions were designed to address the learning outcome related to problem solving with fractions, decimal representations, and percentages within the Numbers and Quantities theme. In the development of the interview form, relevant literature, the national curriculum, and textbooks were systematically reviewed. Subsequently, content validity was established through expert review by two researchers with expertise in mathematics education. Following these validation procedures, the finalized form was administered to the participants.

TABLE 1
Questions from interview form

Related to	Interview Questions
Preservice teachers' example	MAT.6.1.8. Solve problems involving fractions, decimals, and percentages encountered in real-life situations that require the four basic operations. Formulate a problem based on this learning outcome. Solve the problem you have formulated in detail.
ChatGPT and Gemini examples	Create a problem using the AI programs ChatGPT and Gemini related to the same learning outcome. Solve the problem you created using the AI program.
Views about examples	When you examine the question generated by artificial intelligence, do you think it fully meets the learning outcome MAT.6.1.8.? Explain the reasoning behind your answer.

Prior to the administration of the interview form, participants received a one-hour training session on artificial intelligence. This session covered the fundamental use of ChatGPT and Google Gemini, including an introduction to prompts, the purpose of prompting, and principles for writing effective prompts. Immediately following the training, the interview form was administered, and participants completed the questions within an approximately one-hour period.

Throughout both the training session and the implementation of the interview form, the researcher provided detailed procedural explanations to the preservice teachers. However, in order to avoid any potential guiding or biasing effects during data collection, no additional information beyond the learning outcome statements specified in the Grade 6 mathematics curriculum was shared. This approach was adopted to ensure that participants' responses reflected their own reasoning and interactions with the AI tools rather than externally imposed instructional guidance.

Participants completed the interview forms individually while working at separate computer stations, and no interaction or exchange of information occurred among them during the process. They used the free versions of ChatGPT and Google Gemini; participants with existing accounts accessed the platforms using their current memberships, while those without prior accounts created new ones for the purposes of the study. Throughout the entire process, beginning with the training session, participants' behaviors and interactions were systematically observed, and detailed field notes were recorded by the researcher.

2.4 Data Analysis Procedure

Creswell's (2017) steps for qualitative data analysis procedure followed in the current study; all data were transcribed verbatim, scanned and prepared for analysis. Then were imported to MAXQDA 2024 qualitative data analysis software. Before the coding process, all the transcriptions and observation notes were read without an analytical perspective in order to get general sense of data. In coding, any meaningful part of the statement, such as a sentence, a few sentences, or a paragraph regarding their problem creating and solving process, has been coded. Similarities and relationships between codes were also examined.

After coding procedure was completed and the open coding method appeared to be saturated, a second coder, who has expert in mathematics education, coded one participant's documents. In initial analysis, there was an approximately 80% agreement on the coding statements and codes. Afterwards they made a consensus on the codes until ensuring at least 90% agreement.

3 FINDINGS

Findings are organized in the order of interview form, firstly participants' questions and solutions, then AI questions and solutions, finally participants' view about the AI performance. The letter P and a number refer to

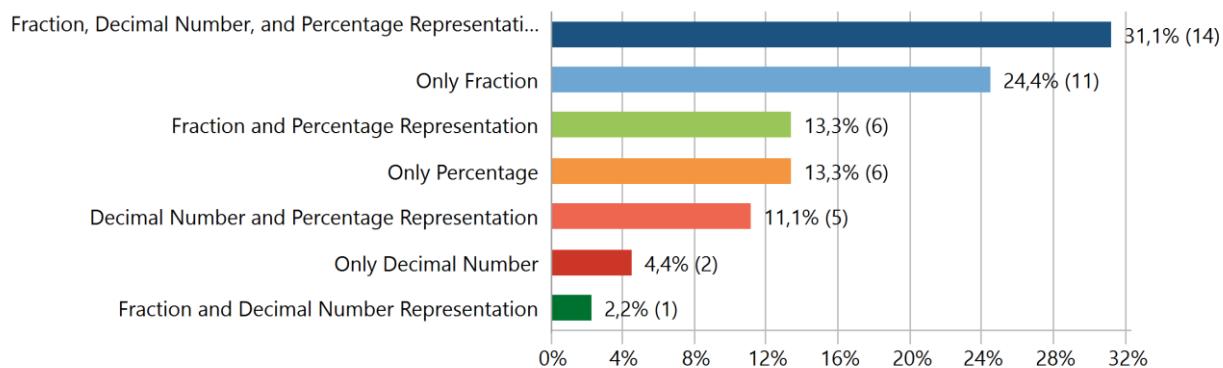
the participants in the quotations below that have been abbreviated. The letters "CQ" and "GQ" stand for ChatGPT question and Gemini Question, respectively. To give an example, P4_CQ means Preservice Teacher 4's responses from the ChatGPT question.

3.1 Problems Posed by Preservice Teachers

In the analysis of the data, the initial focus was on examining whether the preservice teachers accurately interpreted the specified learning outcome. The learning outcome required participants to construct a real-life problem that simultaneously incorporated fractional, decimal, and percentage representations. Based on the findings, Figure 1 illustrates the extent to which the problems posed by the preservice teachers included the use of fractions, decimals, and percentages, indicating whether they employed all three representations or only a subset of them in their problem formulations.

FIGURE 1

Participants using fraction, decimal number and percentage representation in their problems



As shown in Figure 1, 14 preservice teachers constructed problems that incorporated all three representations—fractions, decimals, and percentages—simultaneously. In contrast, 19 preservice teachers relied on a single representation, while 12 participants employed only two of the three representations. Accordingly, only 14 preservice teachers demonstrated a full alignment with the specified learning outcome, whereas the remaining 31 participants did not fully capture the requirements of the learning outcome and consequently produced partially aligned or incomplete problem formulations. These findings indicate that some preservice teachers in this study exhibited deficiencies in conceptual understanding and failed to systematically attend to all targeted concepts specified in the learning outcomes when constructing mathematical problems.

Let's look at some sample problems, only fraction; "Ayşe has some pens. She gave $\frac{1}{3}$ of them to Ali and $\frac{1}{5}$ of the rest to Beren. Since she has 16 pens left, how many pens did she have in total at the beginning?" (P10). Only decimals, P20 wrote this problem:

"Sevim went to the market and bought three bags of chips costing 23.25 TL each, a box of fruit juice costing 20.75 TL, and two kilograms each of peppers and tomatoes costing 7.5 and 5.25 TL per half kilogram, respectively. Sevim had 192.50 TL. How much change will she receive?"

Another example is, P35's problem was suitable for only percentage representation;

"Sibel goes shopping to buy herself a pair of shoes, a shirt, and a pair of pants. The prices at the store she visits are as follows:

Shoes: 1000 TL

Shirt: 400 TL

Pants: 600 TL

Since it coincides with the store's end-of-season sale, there is a 20% discount on items over 500 TL, and an additional 10% discount on the total price. How much did Sibel pay?"

When it was considered all representations; fractions, decimals and percentages, this example illustrates that situation: "Hatice goes to the bakery to buy pastries for breakfast and sees that there is a 20% discount when she buys at least one kilogram. How much does she have to pay when she buys $\frac{9}{6}$ kg of pastries, which cost 50.80 TL per kg?" (P14). Examples of real-life problems that only two of these three presentations address are as follows: P40 wrote for fractions and decimals, "A water tank is 25% full. If 60 liters of water are added to this tank, it will be $\frac{2}{5}$ full. How many liters of water were in the tank initially?" (P40), whereas P39 fulfilled decimal and percentage representations;

"The Kahraman family is traveling from City A to City B. When the Kahraman family has traveled 20% of the way, they still have 6 hours of travel time left. When the Kahraman family has traveled 0.75 of the remaining distance, there are 100 km left in total. How many kilometers is the total distance, and how many hours did it take?"

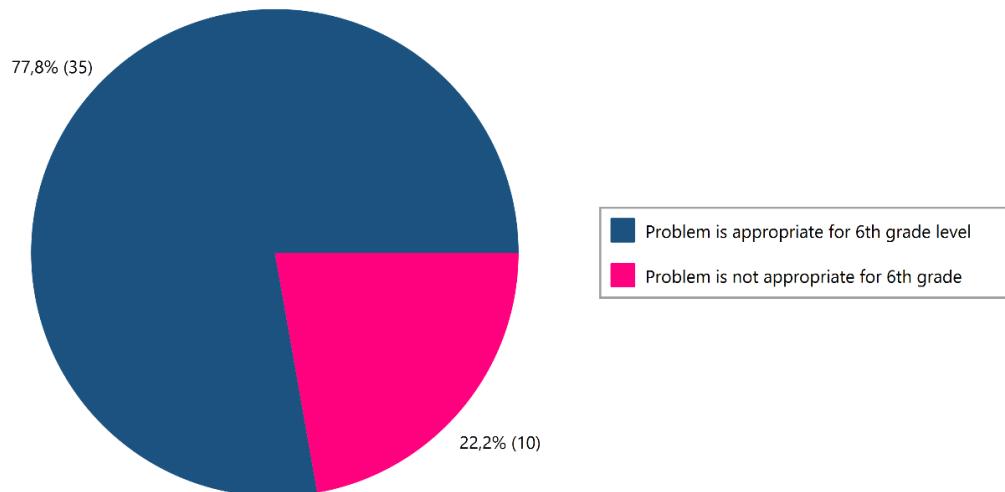
Although the solutions to the problems written by all teacher candidates were correct, according to the findings of the study, the problems created by 35 teacher candidates were appropriate for the 6th grade level, while the problems created by 10 teacher candidates were not appropriate for the 6th grade level (see Figure 2). To give an example, P34 created this problem;

"Elif will give fruit to her siblings Efe and Ecrin from a fruit basket containing 8 apples and 12 oranges. She knows that Efe does not like apples. She will give 20% of the total fruit to Efe and 25% of the remaining fruit to Ecrin. Elif wants equal amounts of apples and oranges. Based on the information given, what is the ratio of the remaining apples to the remaining fruit after this distribution?"

In addition to the concept of percentage, the problem also included the concept of ratio and asked for the ratio of two numbers. The concept of ratio is a topic covered in the 7th grade curriculum and is not appropriate for a 6th grade student's level.

FIGURE 2

Level of participants' problems



Similarly, when the solutions to the problems created by the teacher candidates in this study were checked for

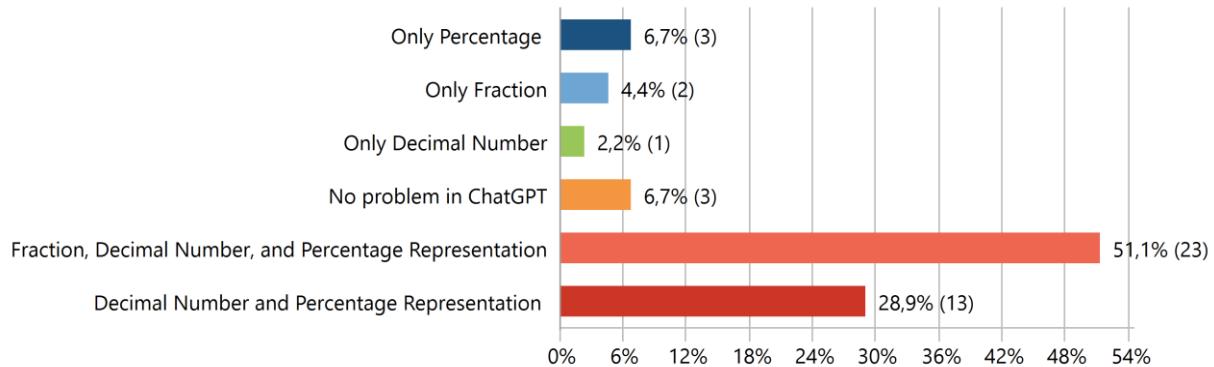
curriculum compliance, the same number was encountered. While the solutions of 10 participants were not compliant with the 6th grade curriculum, the solutions of 35 students were compliant with the 6th grade curriculum. The non-compliant solutions were mostly those made using ratios and proportions. This situation constitutes another indicator of the pre-service teachers' conceptual shortcomings. Had they been fully aware of which topics belong to the 6th grade curriculum and which are introduced at the 7th grade level, they would not have employed such solution strategies, or they would have selected solution approaches more appropriately aligned with the intended grade level.

3.2 Problems Created in ChatGPT

The problems generated through ChatGPT were similarly analyzed by first examining whether they met the specified learning outcome. Based on the findings, a greater proportion of the problems generated via ChatGPT incorporated all three representations—fractions, decimals, and percentages—with a total of 23 such problems identified. In contrast, the number of problems that included only a single representation was limited to six.

FIGURE 3

Fraction, decimal number and percentage representation in ChatGPT problems



An examination of Figure 3 indicates that, in addition to the 23 questions fully aligned with the specified learning outcome, there were also 13 questions that included only decimal and percentage representations, a proportion that cannot be considered negligible. An analysis of the prompts used to generate these problems revealed a noteworthy pattern: the prompts were either identical or highly similar. For instance, a closer comparison of the prompts entered by P1 and P26, along with the corresponding problems generated by ChatGPT, illustrates this similarity;

Prompt P1: MAT.6.1.8 Create a problem related to the learning outcome of being able to solve problems involving the four operations with fractions, decimals, and percentages encountered in real-life situations.

P1_CQ: Zeynep went to a stationery store and purchased school supplies.

The total price of the notebooks was 120 TL, and a 25% discount was applied.

She paid 45.75 TL for the pens.

The total price of the eraser and ruler was $3/5$ of the discounted price of the notebooks. Since Zeynep paid 200 TL at the cash register at the end of her shopping,

a) How much change did the cashier give Zeynep?

b) How much did Zeynep pay for all the items?

Prompt P26: MAT.6.1.8 Create a problem for me related to the learning outcome of being able to solve problems

involving the four operations with fractions, decimals, and percentages encountered in real-life situations, and solve this problem.

P26_CQ: Elif went shopping at a supermarket. The items she purchased are as follows:

1.5 kg of cheese costs 72.40 TL per kilogram.

750 grams of olives cost 64 TL per kilogram.

She bought 2.5 liters of milk. It costs 18.25 TL per liter.

There was a 10% discount applied to the total price of the items she bought.

How much did Elif pay at the end of her shopping trip?

Despite the near-identical nature of the prompts, the problems generated by ChatGPT differed in their structure and representational coverage. Notably, while the problem generated for P1 fully satisfied the specified learning outcome, the corresponding problem produced for P26 lacked fractional representations. A similar pattern was observed across the remaining 12 questions. These findings suggest that, even when identical or highly similar prompts are provided, ChatGPT does not consistently generate problems that fully align with the intended learning outcome. Conversely, when problems containing only a single representation—such as percentages only, decimals only, or fractions only—were examined, the results indicated that this limitation was primarily attributable to the structure and specificity of the prompts. For instance, this issue can be observed by examining the prompt entered by P29 and the resulting problem generated by ChatGPT, which focused exclusively on the fraction concept:

Prompt P29: Prepare and solve a problem related to fractions for 6th grade.

P29_CQ: Ayşe divided a cake into 8 equal pieces to share with her friends.

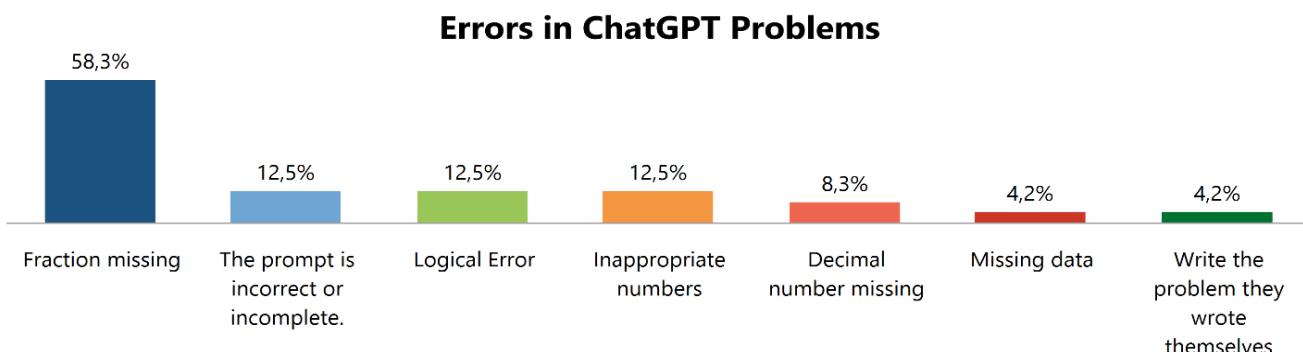
She ate $\frac{3}{8}$ of the cake.

Her brother ate $\frac{1}{4}$ of the cake.

She gave the rest of the cake to her neighbors. How much of the cake did she give to her neighbors?

For three participants, no problems were generated via ChatGPT, as these preservice teachers relied exclusively on Gemini to construct the problems. Of the problems produced by ChatGPT, 21 were deemed appropriate, whereas 24 exhibited various deficiencies (see Figure 4). The relatively high number of problems that included only decimal and percentage representations parallels the observed absence or underrepresentation of fractional components. Another noteworthy finding concerns the nature of the inadequacies observed in some ChatGPT-generated problems. Specifically, certain problems were rendered unsuitable due to logical inconsistencies, the use of inappropriate numerical values, or the presence of missing information, all of which undermined the mathematical coherence of the tasks.

FIGURE 4
Distribution of errors in inappropriate ChatGPT problems



As an illustrative example, in the problem generated for P31, the quantities and costs associated with the pizzas were not represented using whole numbers. Consequently, this led to a non-integer result for the number of people, rendering the problem mathematically inconsistent:

Prompt P31: Learning outcome: Solving problems involving fractions, decimals, and percentages encountered in real-life situations that require the four basic arithmetic operations. ... This learning outcome is for 6th grade students. Write and solve problems for them.

P31_CQ: Zeynep and her family ordered pizza over the weekend. There were three different pizzas on the menu:

Small pizza: Serves 1.5 people (i.e., 1 whole and 1/2) and costs 54 TL

Medium pizza: Serves 2.25 people and costs 72 TL

Large pizza: Serves 3.5 people and costs 98 TL

Zeynep ordered 2 small pizzas, 1 medium pizza, and 1 large pizza. The pizza restaurant offers a 20% discount on orders of 200 TL or more. Zeynep's mother also wants to use a 20 TL discount coupon. This coupon is valid on the discounted price.

Question 1: How much will Zeynep's family pay at the counter for these pizzas?

Question 2: How many people can these pizzas serve in total

Another participant, P44, encountered an issue with ChatGPT due to missing data:

Prompt P44: The problem I want you to create should be related to the 6th grade, Unit 1, Learning Outcome 8. This learning outcome is: 'Solving problems involving fractions, decimals, and percentages encountered in real-life situations that require the four basic arithmetic operations.' Create and solve a problem related to this learning outcome.

P44_CQ: Various products on a farm are sold at different fractions, decimals, and percentages. The farm owner earns income from the sale of three types of products:

3/5 of the first type of product was sold.

0.75 of the second type of product was sold.

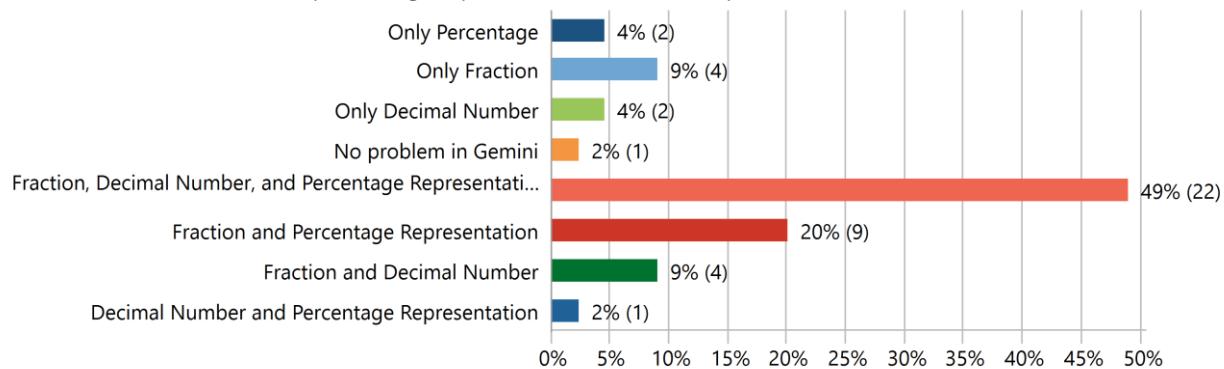
A total of 40% of the third type of product was sold.

There are 1000 kg of products in total on the farm. The total income generated from the sale of each type of product is 2400 TL. The farm owner wants to calculate how much of each type of product was sold and how much income was generated from each.

In the problem generated by ChatGPT, although the total number of products and the total revenue were provided, the quantities sold for each individual product were not specified, rendering the problem unsolvable. In this particular case, the solution produced by ChatGPT was based on conjectural assumptions rather than determinable data. Overall, the findings indicate that the solutions generated by ChatGPT were mathematically correct when the numerical information provided in the problems was complete and accurate. However, in cases involving missing or incorrect data, the corresponding solutions relied on assumptions, resulting in speculative rather than definitive answers.

3.3 Problems Created in Gemini

The problems generated via Gemini were also analyzed in terms of their alignment with the specified learning outcome. The findings indicate that a greater number of Gemini-generated problems incorporated all three representations—fractions, decimals, and percentages—with a total of 22 such problems identified. In contrast, the number of problems that included only a single representation was eight.

FIGURE 5*Fraction, decimal number and percentage representation in Gemini problems*

An examination of Figure 5 reveals that, in addition to the 22 questions fully aligned with the specified learning outcome, there were also 14 questions that incorporated only two of the three representations, a proportion that cannot be regarded as negligible. Similar to the findings for ChatGPT, an analysis of the prompts used to generate problems in Gemini revealed that these prompts were either identical or highly similar. For example, a comparison of the prompts entered by P11 and P37, together with the corresponding problems generated by Gemini, illustrates this pattern:

Prompt P11: MAT.6.1.8. Solve problems involving fractions, decimals, and percentages in real-life situations that require the four basic operations.

Can you create a problem based on this learning outcome?

P11_GQ: Mrs. Ayşe bought $2\frac{1}{2}$ kilograms of apples, 1.75 kilograms of oranges, and 3 kilograms of tangerines, 20% of which were rotten, from the market. Apples cost 12 TL per kilogram, oranges cost 10 TL per kilogram, and tangerines cost 8 TL per kilogram.

After separating the rotten mandarins, Mrs. Ayşe mixed all the remaining fruit together to make a fruit salad. Accordingly:

- How many kilograms of good mandarins did Mrs. Ayşe use?
- Find the total amount Mrs. Ayşe paid for all the fruit she bought.
- What is the total weight of the fruit salad Mrs. Ayşe prepared?

Prompt P37: MAT.6.1.8. Can you create and solve problems that require the four operations involving fractions, decimals, and percentages encountered in real-life situations, in line with the learning outcome?

P37_GQ: Mrs. Ayşe used $\frac{2}{5}$ of her savings to buy a washing machine and 30% of her remaining savings to buy an iron. Mrs. Ayşe initially had 2500 TL. How much money does she have left?

As observed in the ChatGPT findings, although the prompts were nearly identical, the resulting problems generated by Gemini differed in terms of representational completeness. Specifically, while the problem generated for P11 fully satisfied the specified learning outcome, the problem produced for P37 lacked decimal representations. A similar pattern was evident in the remaining questions in which one representation was missing. These findings indicate that, even when identical or highly similar prompts are used, neither Gemini nor ChatGPT consistently generates problems that fully align with the intended learning outcome. Another notable finding concerns problems that included only a single representation—fractions only, decimals only, or percentages only—which appeared to stem directly from the structure and specificity of the prompts. For instance, this issue can be illustrated by examining the prompt entered by P5 and the resulting problem generated by Gemini, which focused exclusively on the percentage concept;

Prompt P5: Can you create a problem related to percentage displays encountered in real-life situations?

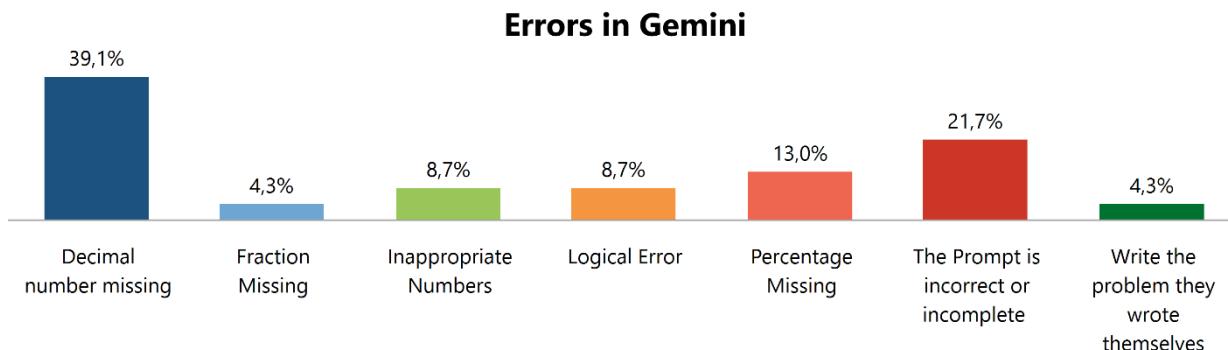
P5_GQ: Ayşe bought a novel for 150 TL with a 20% discount at a bookstore. If she had wanted to buy the same novel online, she would have received a 15% discount.

Questions:

- What is the full (list) price of the novel in TL?
- How much would Ayşe have paid if she had bought it online?
- If Ayşe had bought the book online instead of at the bookstore, how much would she have saved or lost?

For one participant, no problem was generated because the participant reported being unable to access Gemini. Among the problems generated via Gemini, 22 were considered appropriate, whereas 23 exhibited various deficiencies (see Figure 6). The relatively high number of problems incorporating only two representations helps to explain the prevalence of representational incompleteness observed in single-representation problems. In addition, inaccuracies in the prompts entered by preservice teachers constituted one of the contributing factors to these outcomes. Another noteworthy finding concerns Gemini's generation of flawed problems due to logical inconsistencies, the use of inappropriate numerical values, or missing data, which in turn compromised the mathematical validity of the tasks.

FIGURE 6
Distribution of errors in inappropriate Gemini problems



One example of this situation is when the total purchase amount exceeds the money we have on hand, as in the question created by P42, and Gemini calculates a negative result:

Prompt P42: Can you create a problem related to the learning outcome "solving problems involving fractions, decimals, and percentages encountered in real-life situations that require the four basic arithmetic operations"?

P42_CQ: Aunt Ayşe bought $2\frac{1}{2}$ kilograms of tomatoes, 1.75 kilograms of cucumbers, and 5 kilograms of potatoes, 80% of which were potatoes, from the market. Tomatoes cost 8 TL per kilogram, cucumbers cost 6 TL per kilogram, and potatoes cost 5 TL per kilogram. Since Aunt Ayşe gave the vendor 50 TL, how much change should she receive?

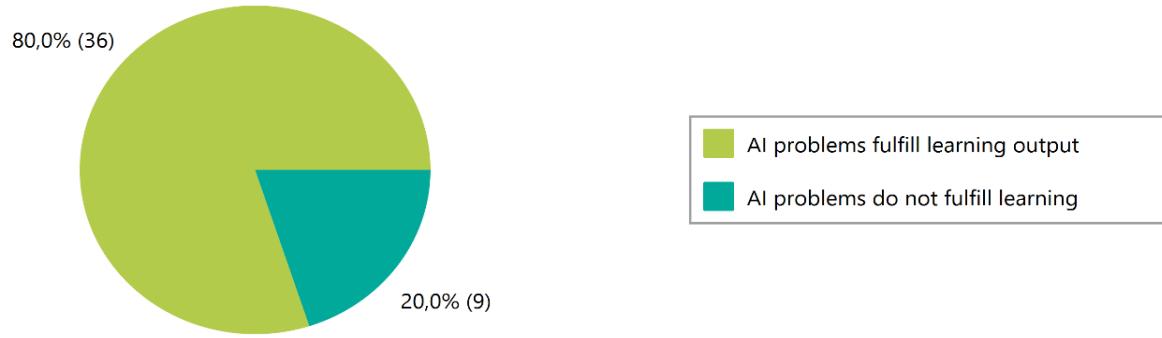
Based on the findings, the solutions generated by Gemini were mathematically accurate when the numerical information provided in the problems was complete and correct. However, in cases involving incorrect or insufficient data, the solutions relied on assumed numerical values, resulting in conjectural rather than definitive outcomes.

3.4 Preservice Teachers' Views regarding AI Tools

According to the findings, nine of the preservice teachers reported that the problems generated using AI tools

did not meet the specified learning outcome, whereas 36 participants indicated that the generated problems were aligned with the learning outcome (see Figure 7). However, a closer examination revealed that four of the nine problems identified by participants as misaligned were, in fact, fully consistent with the learning outcome and exhibited no substantive deficiencies.

FIGURE 7
Participants' views about AI problems



For example, P23 articulated a specific expectation regarding the structure of AI-generated problems: "I expected a problem posed as a coherent whole with interrelated questions, but both [tools] presented the questions in multiple separate items. ... I wanted a single, integrated problem ... sometimes the questions were not clearly expressed." This statement indicates that P23 expected all three representations specified in the learning outcome to be integrated within a single, cohesive problem rather than distributed across multiple items. Similarly, P36 explained the perceived misalignment with the learning outcome as follows: "While writing the problem, fractional numbers were used at a level that is not appropriate for sixth-grade students. The questions are good, but the solution would be challenging for students at that age." P24 expressed a comparable concern: "ChatGPT constructed a problem that students might find difficult. Since both the fraction and the monetary values were fractional, this could confuse students." In a similar vein, P16 highlighted issues related to the numerical values used in the problems: "The flow of the problem is good; however, the numbers provided are not ones students are likely to encounter in daily life. This weakens the connection to real-life contexts." Notably, the issue raised by P16 was observed in more than half of the problems generated by ChatGPT and Gemini. In many cases, the numerical values were unrealistic or implausible in everyday contexts, and the prices used were not up to date. Interestingly, when participants requested that prices be updated, both ChatGPT and Gemini acknowledged this feedback and revised the problems by replacing the values with more realistic figures. A related and noteworthy pattern concerned the repetition of characters and contextual settings in AI-generated problems. For instance, in Gemini-generated problems, a character named "Ayşe Teyze" appeared repeatedly—shopping at a market for fruits and vegetables in 15 different problems. Similarly, in ChatGPT-generated problems, characters such as Ayşe or Zeynep frequently appeared shopping at stationery stores. This repetition was also highlighted by a participant: "...they offered many of us very similar problems; even when our prompts differed, the range of problems they could generate in line with the learning outcome was quite limited" (P34).

Some discrepancies were also observed between participants' evaluations and the researchers' analyses. For example, P18 believed that ChatGPT had made an error in solving the problem, while considering the problem generated by Gemini to largely meet the learning outcome. However, further analysis showed that ChatGPT's solution was mathematically correct, whereas the problem generated by Gemini lacked decimal representations,

indicating a misjudgment on the part of the participant. In contrast, some participants provided accurate evaluations. For instance, P22 stated: *"Gemini actually used percentages and decimals, whereas ChatGPT was insufficient in this regard. The problem lacked comprehensiveness and used only fractions."*

Seven participants reported that they found Gemini-generated problems to be more comprehensive and successful overall. Illustrative comments include: *"I really liked it; I found Gemini especially more successful"* (P44). *"While ChatGPT focused on a single topic, Gemini created a more detailed problem. In my opinion, it was more effective... if I were to use one, I would prefer the problem generated by Gemini"* (P43). *"The problem generated by ChatGPT was sufficient... Gemini's problem was also sufficient—indeed, it was even more comprehensive"* (P42). There were also participants who emphasized the clarity and quality of both the problems and the solutions generated by AI tools. For example, P20 remarked: *"...the solution was very clear and sufficient. I particularly liked the step-by-step explanation. I think it adequately meets the learning outcome."* Similarly, P36 noted: *"...the way the questions were solved with clear explanations was good."* On the other hand, P38 stated that ChatGPT provided more detailed and well-structured solutions, whereas Gemini's solutions were more superficial, and also noted that the visual designed by ChatGPT was more appealing. In contrast, P33 expressed dissatisfaction with the visual support offered by AI tools: *"I asked it to generate a visual, but it did not... it was inadequate in terms of visual support."*

4 DISCUSSION

An examination of the study findings indicates that preservice teachers did not fully grasp the specified learning outcome, as evidenced by the problems they constructed themselves. While approximately one-third of the participants were able to generate problems aligned with the learning outcome, nearly two-thirds failed to produce problems that fully satisfied its requirements. This decline in preservice teachers' problem-posing competence has also been documented in previous studies. For instance, Yıldız (2014) reported low levels of problem-posing success among preservice mathematics teachers, while Işık and Kar (2011) similarly found that preservice teachers demonstrated limited success in posing problems involving multiple representations. Researchers in the literature have emphasized the importance of engaging preservice teachers in problem-posing activities (English, 2020; Kılıç, 2015). Özdemir-Yıldız (2019) further identified deficiencies in preservice teachers' problem-posing abilities, noting common errors such as constructing problems with missing information or disregarding given instructions. Accordingly, reducing these shortcomings requires preservice teachers to critically examine not only the problems they construct themselves but also those generated by AI tools they employ.

With respect to AI-generated problems, ChatGPT was examined first. The findings revealed that approximately half of the problems generated by ChatGPT fully incorporated fraction, decimal, and percentage representations, whereas nearly one-third included only decimal and percentage representations. A particularly noteworthy finding was that the prompts used to generate these problems were often identical or highly similar. This suggests that even when appropriate prompts are entered, the outputs produced by AI tools must be carefully reviewed by users. Furthermore, in cases where ChatGPT-generated problems did not meet the learning outcome, the deficiencies were largely attributable to incomplete or incorrect prompts provided by participants. These results underscore the importance of both crafting well-structured prompts and systematically evaluating AI-generated outputs. Supporting these findings, Wardat, Tashtoush, AlAli, and Jarrah (2023) emphasized that the accuracy of ChatGPT's responses is highly dependent on prompt quality and that its integration into mathematics education requires careful consideration. Similarly, Pilavci (2024) argued that high-quality prompts significantly enhance the effectiveness of AI models.

Another important finding of the study was that ChatGPT occasionally generated problems containing incorrect numerical values or incomplete information. Detecting and addressing such errors depends largely on users'

awareness and their ability to critically evaluate AI-generated content. In parallel with these findings, Çilingir and Akdemir-Tırıcı (2025), as well as Dasari et al. (2024), recommended that AI-based applications such as ChatGPT be implemented under teacher guidance, noting that guided use is more likely to enhance learning outcomes.

Similar patterns were observed for Gemini. The results showed that Gemini, like ChatGPT, predominantly generated problems that included either all three representations—fractions, decimals, and percentages—or at least two of them. Consistent with the ChatGPT findings, the prompts associated with problems meeting these criteria were often identical or highly similar, once again highlighting the critical role of prompt formulation and output verification. In Gemini-generated problems, the primary reason for missing representations appeared to be the model's tendency to combine concepts in different pairings, even when the same prompts were used. Compared to ChatGPT, Gemini produced fewer problems involving missing data or incorrect numerical values. Nevertheless, as with ChatGPT, effective use of Gemini requires users to remain vigilant and to continuously apply their own evaluative judgment. This conclusion aligns with existing literature emphasizing the necessity of filtering AI-generated content through critical thinking and pedagogical scrutiny (Lee & Zhai, 2024; Şimşek, 2025).

Participants in the study also evaluated the problems generated by AI tools and shared their perceptions of ChatGPT and Gemini. The findings suggest that some preservice teachers lacked sufficient understanding of the learning outcome itself, as several participants reported misalignment even when the generated problems fully met the learning outcome criteria. On the other hand, participants were also able to identify instances in which AI tools produced logical errors or incomplete numerical information, leading them to conclude that such tools should not be relied upon uncritically and require further refinement. Similar conclusions were reached by Barros (2024) and Kehoe (2023), who reported that ChatGPT tends to produce errors in multi-step or complex mathematical problems and therefore requires further development in the mathematics domain. Kehoe (2023) likewise emphasized the necessity of verifying ChatGPT's responses. Another salient observation was that both ChatGPT and Gemini frequently generated highly similar problems, relied on repetitive contexts, and used numerical values that were often unrealistic or outdated. Both ChatGPT and Gemini consistently relied on outdated numerical values in the problems they generated, with the numbers used being far removed from current real-life conditions. Given that these AI tools are designed to be continuously updated, this finding was unexpected and warrants careful consideration. This issue has also been highlighted by Lee and Zhai (2024), who noted that ChatGPT may occasionally rely on inaccurate or outdated data, resulting in inconsistencies. They cautioned users against accepting AI-generated outputs without question. In the present study, when preservice teachers explicitly prompted AI tools to correct errors or update values, both ChatGPT and Gemini revised their outputs accordingly, which may be considered a positive development. However, this also underscores the necessity for users to recognize such issues and to actively intervene through appropriate prompting.

Finally, preservice teachers expressed several positive views regarding ChatGPT and Gemini. These views primarily centered on the perception that AI-generated problems were more comprehensive than those they constructed themselves and that the solutions were presented in a detailed, step-by-step manner. Notably, a majority of participants who expressed evaluative opinions reported that Gemini-generated problems were more comprehensive and of higher quality overall. These findings suggest that, when used critically and under appropriate pedagogical guidance, generative AI tools may serve as valuable supports in mathematics instruction and teacher education.

5 PRACTICAL IMPLICATIONS

This study examined how preservice middle school mathematics teachers employed AI-supported tools in

relation to a specific curriculum-based learning outcome. By focusing on participants' problem-posing practices, their interactions with ChatGPT and Gemini, and their evaluations of AI-generated problems, the study contributes to the limited body of research that directly investigates the practical use of AI tools in instructional contexts. At the same time, it adds to the growing literature on the integration of artificial intelligence within mathematics education. In this regard, the present study is expected to contribute particularly to research that examines the use and implementation of AI-based applications such as ChatGPT and Gemini. Moreover, it offers practical insights into instructional approaches for integrating artificial intelligence into educational contexts, thereby contributing to the growing body of applied research on the pedagogical use of AI in education.

The findings indicate that preservice teachers need to adopt a cautious, critical, and controlled approach when formulating prompts and interpreting the outputs generated by AI tools. Rather than accepting AI-generated content at face value, participants are required to filter and evaluate these outputs before use. This result aligns with and reinforces conclusions reported in related studies. Moreover, the study highlights the need to further develop preservice mathematics teachers' problem-posing processes in alignment with specified learning outcomes, thereby contributing to the literature on problem posing and problem solving in mathematics education. Moreover, the findings obtained within the Numbers and Quantities theme may serve as a foundation for research in other mathematical domains. For instance, focusing on visual enrichment in geometry or on modeling activities in data handling may provide new perspectives on how ChatGPT and Gemini can be used to design instructional tasks and learning activities across different areas of mathematics.

Another important implication of the study is the identified need for targeted training on the pedagogically appropriate use of AI tools. The results suggest that preservice mathematics teachers would benefit from systematic instruction on how to effectively and critically integrate AI technologies into instructional planning and problem design. This finding represents an important consideration for researchers and curriculum designers involved in teacher education programs. Overall, the present study is expected to contribute to the preparation and professional development of future mathematics teachers by informing more informed, reflective, and pedagogically grounded uses of AI-supported tools in mathematics instruction.

6 LIMITATIONS AND FUTURE DIRECTIONS

This study was conducted with preservice middle school mathematics teachers and was limited to the *Numbers and Quantities* theme. Therefore, no claim of generalizability is intended. While similar results may be obtained with comparable participant groups, it is recommended that future studies be conducted with different participant profiles and across other specific mathematical content areas, followed by comparative analyses of the findings. In addition, conducting a similar investigation with in-service mathematics teachers would provide valuable insights and contribute meaningfully to the existing literature.

Previous research indicates that preservice teachers who are less familiar with technological and AI-based tools may be inclined to avoid using such tools or may employ them inaccurately or incompletely when they do attempt to use them (Lo, 2023). Accordingly, there is a clear need for the design of courses or training programs that focus on the effective and high-quality use of AI tools. Ensuring that both preservice teachers and practicing teachers are knowledgeable and competent in writing effective prompts emerges as a critical requirement for the pedagogically sound integration of AI technologies into mathematics education.

DECLARATIONS

Author Contributions

MC: Conceptualization, Methodology, Formal analysis, Writing – Original Draft, Resources

AB: Conceptualization, Methodology, Supervision, Writing – Review & Editing, Visualization, Software

Conflict of Interest

Statement (no conflict): The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics Committee Approval

Statement: Ethical approval for this study was obtained from the Kırşehir Ahi Evran University Social and Humanities Sciences Scientific Research and Publication Ethics Committee (Date: 05.11.2025, Decision No: 2025/17/20).

Funding / Financial Support

Statement (no funding): This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

AI Declaration

During the preparation of this work, the author(s) used ChatGPT in order to improve language clarity and to check for grammatical errors. After using this tool, the author(s) reviewed and edited the content and take full responsibility for the content of the publication.

Data Availability Statement

Available on request (example): The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgments

The findings of this study were presented by the corresponding author and second author at the 7th International Symposium on Turkish Computer and Mathematics Education (TÜRKBİLMAT-7, 2025).

REFERENCES

Alshahrani, A. (2023). The impact of ChatGPT on blended learning: Current trends and future research directions. *International Journal of Data and Network Science*, 7(4), 2029–2040.

Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, 7(1), 52–62. <https://doi.org/10.61969/jai.1337500>

Baki, A., Yalçınkaya, H. A., Özpinar, İ., & Uzun, S. Ç. (2009). İlköğretim matematik öğretmenleri ve öğretmen adaylarının öğretim teknolojilerine bakışlarının karşılaştırılması. *Turkish Journal of Computer and Mathematics Education*, 1(1), 65–83.

Barros, C. (2024). ChatGPT in the resolution of a math exam: Results obtained in Portuguese and in English language. In *International Conference on Lifelong Education and Leadership for All (ICLEL 2023)* (pp. 37–47). Atlantis Press.

Bayazit, İ., & Dönmez, E. S. (2017). Öğretmen adaylarının problem kurma becerilerinin orantısal akıl yürütme gerektiren durumlar bağlamında incelenmesi. *Turkish Journal of Computer and Mathematics Education*, 8(1), 130–160.

Baykul, Y. (2009). *İlkokulda matematik öğretimi*. Pegem Yayıncılık.

Bishop, J. M. (2021). Artificial intelligence is stupid and causal reasoning will not fix it. *Frontiers in Psychology*, 11, Article 513474. <https://doi.org/10.3389/fpsyg.2020.513474>

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... Amodei, D. (2020). Language

models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
<https://doi.org/10.48550/arXiv.2005.14165>

Castro, R. A. G., Cachicatari, N. A. M., Aste, W. M. B., & Medina, M. P. L. (2024). Exploration of ChatGPT in basic education: Advantages, disadvantages, and its impact on school tasks. *Contemporary Educational Technology*, 16(3), Article ep511.

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, 8, 75264–75278.

Chen, J. J., & Lin, J. C. (2024). Artificial intelligence as a double-edged sword: Wielding the POWER principles to maximize its positive effects and minimize its negative effects. *Contemporary Issues in Early Childhood*, 25(1), 146–153.

Creswell, J. W. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. SAGE Publications.

Çilingir, E., & Akdemir-Tirsi, A. (2025). Yapay zekâ tabanlı eğitim planları: Bir öğrenme çıktıları üzerinden MEB Maarif Modeli ile karşılaştırmalı bir inceleme (Sözlü bildiri). In *II. İstanbul Bilim ve Sanat Merkezleri Çocuk Kongresi* (pp. 93–97). İstanbul İl Millî Eğitim Müdürlüğü.

Dasari, D., Hendriyanto, A., Sahara, S., Suryadi, D., Muhaimin, L. H., Chao, T., & Fitriana, L. (2024). ChatGPT in didactical tetrahedron, does it make an exception? A case study in mathematics teaching and learning. *Frontiers in Education*, 8, Article 1295413.

English, L. D. (2020). Teaching and learning through mathematical problem posing: Commentary. *International Journal of Educational Research*, 102, Article 101451.

Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*, 61(3), 460–474.

Grassini, S. (2023). Shaping the future of education: Exploring the potential and consequences of AI and ChatGPT in educational settings. *Education Sciences*, 13(7), Article 692.

Haleem, A., Javaid, M., & Singh, R. P. (2022). An era of ChatGPT as a significant futuristic support tool: A study on features, abilities, and challenges. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, 2(4), Article 100089.

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. The Center for Curriculum Redesign. <https://curriculumredesign.org/wp-content/uploads/AIED-Book-Excerpt-CCR.pdf>

Işık, A., & Kar, T. (2011). Öğretmen adaylarının sözel ve görsel temsillere yönelik kurdukları problemlerin analizi. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 30(30), 39–49.

Kehoe, F. (2023). Leveraging generative AI tools for enhanced lesson planning in initial teacher education at post-primary. *Irish Journal of Technology Enhanced Learning*, 7(2), 172–182.
<https://doi.org/10.22554/ijtel.v7i2.124>

Kılıç, C. (2015). Analyzing pre-service primary teachers' fraction knowledge structures through problem posing. *Eurasia Journal of Mathematics, Science and Technology Education*, 11(6), 1603–1619.

Lee, I., & Perret, B. (2022, June). Preparing high school teachers to integrate AI methods into STEM classrooms. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 36, No. 11, pp. 12783–12791).

Lee, G. G., & Zhai, X. (2024). Using ChatGPT for science learning: A study on pre-service teachers' lesson planning. *IEEE Transactions on Learning Technologies*, 17, 1643–1660.
<https://doi.org/10.1109/TLT.2024.3401457>

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13(4), 1–15.

Lund, B. D., & Wang, T. (2023). Chatting about ChatGPT: How may AI and GPT impact academia and libraries? *Library Hi Tech News*, 40(3), 26–29.

Millî Eğitim Bakanlığı. (2024). *Ortaokul matematik dersi öğretim programı*. MEB.

National Council of Teachers of Mathematics. (2008). *Principles and standards for school mathematics*. Author.

OpenAI. (2024). *Introducing ChatGPT-4.0*. <https://openai.com>

Osana, H. P., & Royea, D. A. (2011). Obstacles and challenges in preservice teachers' explorations with fractions: A view from a small-scale intervention study. *The Journal of Mathematical Behavior*, 30(4), 333–352. <https://doi.org/10.1016/j.jmathb.2011.07.001>

Osman, S. A., & Ahmed, Z. E. (2024). Navigating AI integration: Case studies and best practices in educational transformation. In Z. Ahmed, A. Hassan, & R. Saeed (Eds.), *AI-enhanced teaching methods* (pp. 240–267). IGI Global.

Othman, A. (2023). Demystifying GPT and GPT-3: How they can support innovators to develop new digital accessibility solutions and assistive technologies. *Nafath*, 8(22).

Özdemir, B. G., Bayraktar, R., & Yılmaz, M. (2017). Sınıf ve matematik öğretmenlerinin kavram yanılışlarına ilişkin açıklamaları. *Trakya Üniversitesi Eğitim Fakültesi Dergisi*, 7(2), 284–305.

Özdemir-Yıldız. (2019). *Matematik öğretmen adaylarının problem kurma becerilerinin incelenmesi ve problem kurma hakkındaki görüşlerinin belirlenmesi* (Yüksek lisans tezi). Gazi Üniversitesi.

Patton, M. Q. (2015). *Qualitative evaluation and research methods*. SAGE Publications.

Pilavcı, S. (2024). Yapay zekâ gelişimi ve GPT. In E. Şengel (Ed.), *Teoriden uygulamaya yapay zekânın temelleri: Geçmiş, gelecek ve güçlü promptlar* (pp. 1–8). Serüven Yayınevi. <https://dspace.mudanya.edu.tr/server/api/core/bitstreams/f5a65c61-c958-4930-b424-7e61e0f7f8c4/content>

Richard, P. R., Vlez, M. P., & Van Vaerenbergh, S. (2022). *Mathematics education in the age of artificial intelligence: How artificial intelligence can serve mathematical human learning*. Springer Nature.

Roose, K. (2022). The brilliance and weirdness of ChatGPT. *The New York Times*. https://hawksites.newpaltz.edu/fdc/files/2023/02/The_Brilliance_And_Weirdness_O.pdf

Samala, A. D., Zhai, X., Aoki, K., Bojić, L., & Zikic, S. (2024). An in-depth review of ChatGPT's pros and cons for learning and teaching in education. *International Journal of Interactive Mobile Technologies*, 18(2), 96–117.

Sebullen, M. T. (2023). Lesson planning challenges of pre-service teachers. *Cognizance Journal of Multidisciplinary Studies*, 3(5), 19–29.

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

Şengel, E. (2024). *Teoriden uygulamaya yapay zekânın temelleri: Geçmiş, gelecek ve güçlü promptlar*. Serüven Yayınevi.

Şimşek, N. (2025). Integration of ChatGPT in mathematical story-focused 5E lesson planning: Teachers and pre-service teachers' interactions with ChatGPT. *Education and Information Technologies*, 1–72. <https://doi.org/10.1007/s10639-024-13258-x>

Temur, S. (2024). Yapay zekânın eğitim sistemine entegrasyonunun potansiyel faydaları. *Bati Anadolu Eğitim Bilimleri Dergisi*, 15(3), 2621–2656. <https://doi.org/10.51460/baebd.1541524>

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), Article em2286.

Xie, J., & Masingila, J. O. (2017). Examining interactions between problem posing and problem solving with prospective primary teachers: A case of using fractions. *Educational Studies in Mathematics*, 96, 101–118. <https://doi.org/10.1007/s10649-017-9760-9>

YEĞİTEK. (2025). *Yapay zekâ ve eğitim: Öğretmenler için uygulamalı prompt mühendisliği ve üretken araçlarla yenilikçi öğrenme stratejileri* (M. Canlı, Gen. Yay. Koord.; S. H. Eral, Ed.). Millî Eğitim Bakanlığı. https://yegitek.meb.gov.tr/meb_iys_dosyalar/2025_07/01154003_yz_promptrehberi.pdf

Yeniçeri, Ü., & Kenan, A. (2025). Öğretmen adaylarının yapay zekâya yönelik tutum ve kaygılarının çeşitli değişkenler açısından betimsel analizi. *Ege Eğitim Dergisi*, 26(2), 226–245.

Yenilmez, K., & Kocaoğlu, T. (2010). Beşinci sınıf öğrencilerinin kesir problemlerinde yaptıkları hatalar ve kavram yanılışları. *Kastamonu Eğitim Dergisi*, 14, 129–146.

Yıldız, Z. (2014). *Matematikte problem kurma çalışmalarının öğretmen adaylarının problem kurma becerilerine ve üstbilişsel farkındalık düzeylerine etkisi* (Doktora tezi). Marmara Üniversitesi.

Yılmaz, A. (2023). Fen bilimleri eğitiminde dijital uygulamalar, yapay zekâ ve akıllı yazılımlar: Tehditler ve fırsatlar. In A. Akpınar (Ed.), *Matematik ve fen bilimleri üzerine araştırmalar II* (pp. 1–20). Özgür Yayıncılar.

Yin, R. K. (2009). *Case study research: Design and methods* (4th ed.). SAGE Publications.

Zhai, X. (2022). *ChatGPT user experience: Implications for education*.
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4312418

Zhou, K., Yang, J., Loy, C. C., & Liu, Z. (2022). Learning to prompt for vision-language models. *International Journal of Computer Vision*, 130(9), 2337–2348. <https://doi.org/10.1109/CVPR52688.2022.01631>