LUMAT
International Journal on Math, Science and Technology Education

Research article

ChatGPT as a tool for posing of mathematical problems by prospective teachers

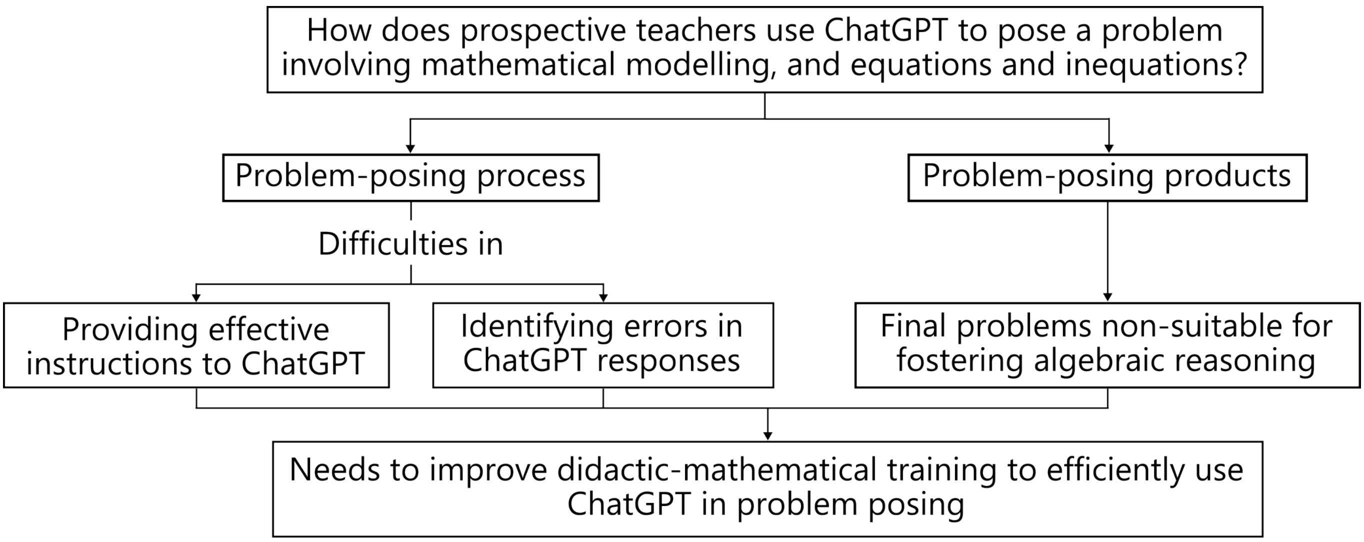
María Burgos and Nicolás Tizón-Escamilla*

University of Granada, Spain

Abstract: Problem posing is a fundamental competence of the mathematics teacher, which motivates the interest in including this activity in teacher training plans. Incorporating problem posing into teaching practice involves considering the tools available to the teacher for this purpose, in particular generative artificial intelligence programs such as ChatGPT. This article analyzes how prospective teachers use ChatGPT to pose a problem involving mathematical modeling and equations and inequalities. The results highlight the difficulties of prospective teachers in providing effective instructions and identifying errors in ChatGPT responses. The problems that were finally proposed by the prospective teachers were, for the most part, not suitable for fostering algebraic reasoning. The joint analysis of the interaction of prospective teachers with ChatGPT and their final problem proposal made it possible to detect shortcomings in relation to their didactic-mathematical knowledge of algebraic reasoning and, in particular, of modeling. We conclude that using ChatGPT for posing mathematical problems has potential in teacher training, but didactic-mathematical training is needed for efficiently using this tool to elaborate meaningful problems.

Keywords: problem posing, algebraic reasoning, teacher training, ChatGPT

Correspondence: tizon@ugr.es



Received 31 June 2025 / Accepted 19 December 2025 / Published 27 January 2026
Published by the University of Helsinki, Finland and LUMA Centre Finland.

1 Introduction

Research on problem posing has experienced significant growth in recent decades, driven by the variety of aspects of mathematics education that converge in this activity (Baumanns & Rott, 2021; Cai & Hwang, 2020; Cai & Leikin, 2020; Liljedahl & Cai, 2021). Student learning improves when they participate in problem invention activities (Silber & Cai, 2021); problem posing enriches students' perception of mathematics, motivates them, enhances their reasoning and problem-solving skills, increases their confidence in mathematics, and contributes to a deeper understanding of mathematical concepts, properties, and procedures (Ayllón et al., 2016; Elgrably & Leikin, 2021; Fernández & Carrillo, 2020).

To effectively introduce problem invention into primary and secondary mathematics classrooms, it is necessary to incorporate problem posing into teacher education programs (Singer et al., 2013). Problem posing not only allows prospective teachers to explore mathematical content in depth and identify their own deficiencies (Yao et al., 2021), but also enhances their analytical capacity and didactic-mathematical knowledge (Malaspina et al., 2019). However, even teachers with years of experience encounter difficulties in posing problems that are meaningful for their students' learning (Mallart et al., 2018).

Considering problem posing in teacher training as a way to engage them in "genuine learning activities" (Singer et al., 2013, p. 5) and bring them closer to the reality of their educational practice requires taking into account the tools available to teachers to support them in this practice. In this regard, the emergence of Artificial Intelligence (AI)-based tools offers teachers a vast array of resources. In particular, recent studies have analyzed the role of generative AI language models, such as the chatbot application ChatGPT, in mathematical problem-solving and its use in teacher education (Getenet, 2024; Pelton & Pelton, 2023; Wardat et al., 2023). These studies highlight both the advantages of using this application, primarily in reducing the generative phases of task resolution, and the need to evaluate the quality (coherence, accuracy, and meaningfulness) of the results beyond their apparent plausibility or credibility (Urban et al., 2024). However, to the best of our knowledge, no research has analyzed the use of ChatGPT for posing mathematical problems in teacher training.

In this article, we explore the implications of using ChatGPT in problem posing as a means to develop and assess the didactic-mathematical knowledge of preservice teachers, focusing on two aspects of algebraic reasoning: modeling and relationships of equality and order. On the one hand, authors such as Zapatera and Quevedo (2021) recommend including in teacher training programs experiences that allow teachers to design tasks to detect and promote algebraic thinking in their future students. On the other hand, according to Hartmann et al. (2023), the development of self-generated modeling problems has great potential to foster this activity. However, there is little research on modeling through problem formulation and the challenges it presents for both students and teachers (Hansen & Hana, 2015; Hartmann et al., 2022, 2023).

In the experiment we describe, 72 preservice teachers were asked to work in teams to utilize ChatGPT in posing a problem incorporating modeling and equations or inequalities, designed for sixth-grade primary school students (11-12 years old). During problem-posing process, they had to decide what instructions to provide to ChatGPT, identify and refine errors in the responses, and adjust the inputs to produce a final problem that met the expected objectives, both in terms of the mathematical knowledge involved and the meaningfulness of the problem. We aim to answer the following research questions:

- What didactic-mathematical knowledge do prospective teachers (PTs) bring into play when interacting with ChatGPT to pose mathematical problems?
- What is the meaningfulness of the final problems posed?
- What criteria do they rely on to justify the adequacy of their final problem?

2 Theoretical framework

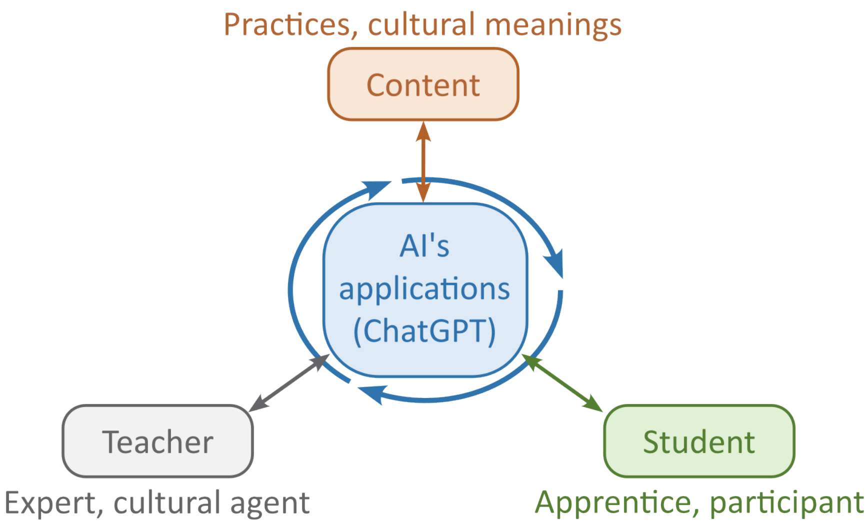
2.1 Didactic-Mathematical Knowledge of Teachers and the Use of ChatGPT

The Didactic-Mathematical Knowledge and Competencies (DMKC) model for mathematics teachers (Godino et al., 2017) provides a framework for guiding mathematics teacher education. It is recognized that teachers must possess both common mathematical knowledge associated with the educational level in which they teach and extended knowledge that enables them to articulate it with higher levels. However, as mathematical content is introduced, teachers must also have didactic-mathematical knowledge that encompasses various facets involved in the teaching and learning of a specific mathematical concept: epistemic (knowledge of the content itself and its institutional meanings), ecological (selection of tasks in accordance with the mandatory institutional curriculum), cognitive (how students learn), affective (ways to address students' interests and respond to their anxiety or indifference), interactional (identification and response to student conflicts and interactions), and mediational (knowledge of the most suitable resources for teaching).

Generative AI-based systems make available to mathematics teachers a variety of tools and resources to support the teaching and learning of mathematics. Among these models, ChatGPT stands out due to its range of application and widespread use. ChatGPT is a pre-trained generative natural language processing model capable of analysing user instructions, considering the context of the conversation (previous instructions) and generating coherent and relevant texts as a response (Sarrion, 2023). To do so, in 2025, ChatGPT has been trained on an enormous set of data, allowing it to learn language patterns and relationships among words. As a consequence, ChatGPT has been proved to be useful in a variety of domains such as marketing and business, healthcare, or education (Memarian & Doleck, 2023; Ray, 2023; Sarrion, 2023). In particular, within the field of mathematics education, ChatGPT can assist teachers in designing and planning mathematics lessons (Wardat et al., 2023), or be used to enhance teachers' knowledge of mathematical content and their ability to identify and address students' misconceptions (Getenet, 2024).

Nevertheless, ChatGPT also presents certain limitations that should be considered to ensure its appropriate use in mathematics education. Among other weaknesses are: a) deficiencies in understanding how words and reality are related, despite having a deep technical understanding of the meanings of words (it is able to meaningfully rephrase and translate texts), which can lead to responses lacking depth and reflection, particularly in domain-specific areas of knowledge (Farrokhnia et al., 2024; Sarrion, 2023); b) limitations associated with the potentially low quality of the data used in its training (Su & Yang, 2023); c) its inability to assess the credibility of such data, undermining its capacity to evaluate the quality and accuracy of its own responses (Farrokhnia et al., 2024); d) the absence of higher-order thinking abilities, such as analytical and critical thinking (Farrokhnia et al., 2024; Rudolph et al., 2023). Thus, despite the advantages of using ChatGPT in mathematics education, teachers must possess the necessary didactic-mathematical knowledge to assess the relevance of the information generated by ChatGPT before incorporating it into their teaching.

Figure 1. Didactic system: Interaction of teacher, student and content with AI.



Indeed, ChatGPT is a generative AI that learns from instructions but can provide imprecise responses when faced with questions beyond the scope of its training data (Urban et al., 2024). On the one hand, ChatGPT produces incorrect problem solutions when contextual understanding and creativity are required (Getenet, 2024), and the accuracy and effectiveness of its solutions vary depending on the complexity of the equation, input data, and provided instructions (Wardat et al., 2023). On the other hand, ChatGPT does not always generate or interpret a correct solution to a problem, nor does it necessarily employ strategies appropriate for the students it is intended for (Getenet, 2024). As such, teachers, as experts and cultural agents, need knowledge to define precise and clear prompts (instructions) when interacting with ChatGPT, refine errors, and adjust their approach to ensure coherence in responses (teacher-ChatGPT interaction in the didactic system shown in Figure 1). They must also recognize the meaning that ChatGPT attributes to mathematical content (content-ChatGPT interaction in the didactic system shown in Figure 1). Finally, they need to monitor and oversee student interactions with ChatGPT, ensuring that this new collaborative learning context fosters their critical thinking and mathematical competence (student-ChatGPT interaction in the didactic system shown in Figure 1). This makes ChatGPT an optimal tool for developing and assessing the didactic-mathematical knowledge and competencies of preservice teachers.

2.2 Problem Posing in Teacher Education

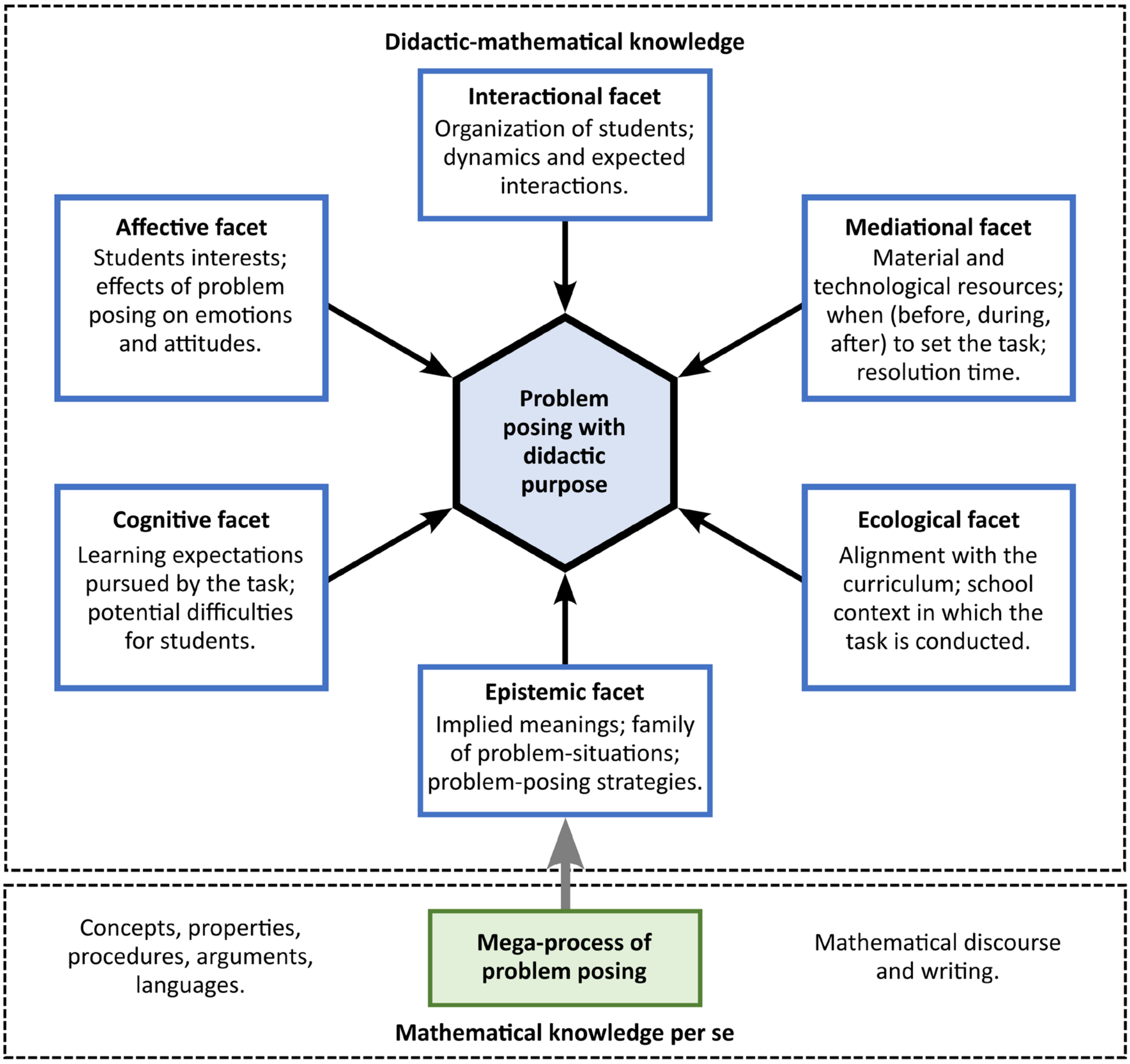
Although different authors have assigned various names to the activity of problem posing, it essentially involves both the formulation of new situations and the reformulation of given problems. One possible approach to categorizing problem-posing methods focuses on the elements that characterize a mathematical problem. According to Malaspina et al. (2015), these elements are: a) information — the quantitative or relational data provided in the problem; b) requirement — what is requested to be found, examined, or concluded (which can be quantitative or qualitative, including graphs and proofs); c) context — whether intra-mathematical or extra-mathematical, which defines the environment or scenario that gives rise to mathematical activity; and d) mathematical environment or structure — which encompasses the mathematical concepts involved or that may be used to solve the problem, along with their properties and relationships.

Based on these elements, the following categorization of problem-posing processes can be considered (Burgos et al., 2025):

- *Free or unstructured creation*: There is no base problem-situation, and no indications or restrictions on the elements of the new problem are included.
- *Semi-structured creation*: There is no base problem-situation, but indications or restrictions on the elements of the new problem are provided. For example, the problem may contain given information (data) that must be completed with the requirement (question), or vice versa. The restriction may also be defined in terms of the mathematical environment, such as specifying the content or strategy leading to the solution.
- *Structured creation or variation*: A new problem is posed based on a preexisting one, meaning that the different elements are known. For example, this may involve modifying or adding information, changing the requirement, exchanging information and requirement, or proposing new questions.

Teachers can pose problems, whether in the semi-structured or structured approach, to serve an educational purpose. In this activity, they require didactic-mathematical knowledge to incorporate specific mathematical content (epistemic dimension), align with curricular descriptors for a given educational stage (ecological dimension), achieve a certain level of cognitive demand (cognitive dimension), or address students' interests and real-world experiences (affective dimension), among others (Figure 2). Conversely, posing problems with a didactic purpose serves as a means to develop these knowledge areas, as it requires: reflecting on the overall structure of the problem (its objectives and whether the provided information is sufficient for solving it); analyzing possible solution methods and the mathematical objects and processes involved, as well as their relationships (mathematical structure); and identifying potential difficulties students may encounter and deciding how to address them when designing new problem-situations.

Figure 2. Facets of didactic-mathematical knowledge in problem posing. Adapted from Burgos et al. (2025).



Reflecting English's (2020) concern about the need for a minimum set of criteria to ensure that teachers formulate high-quality problems, Leavy and Hourigan (2022) build on previous models to develop a framework that helps preservice teachers analyze, select, and pose meaningful mathematical problems. These problems, in addition to addressing cognitive demand, consider other aspects such as context, language, alignment with the curriculum, and the variety of strategies or possible solutions. Based on this model, we consider a meaningful problem to meet the following criteria:

1. *Context*. The context is engaging, relatable, and realistic. The context of the task should be "real in the mind of the student": the student should be able to think about it, make sense of it, and recognize the mathematics embedded within it. The problem establishes a connection between mathematics and the student's world, allowing them to use their personal experiences and knowledge to understand the situation.
2. *Language*. The language is clear, and the cultural setting is accessible. The wording and terminology are adapted to the reading level and mathematical background of the intended students. The problem statement is well-structured, with short and precise sentences, and there is coherence between the information provided and the requirement.
3. *Purpose*. The problem has a clear and relevant purpose. Its learning expectations are explicitly conveyed, and it clearly establishes what is expected from the student.
4. *Curricular Coherence*. The problem involves mathematical content appropriate for the students' level, as defined by the curriculum standards. It aligns with the mathematical content (knowledge) that students are expected to learn by solving the task and the objectives (competencies) it aims to develop, while also considering the prior knowledge students need to solve it.
5. *Robustness*. The problem is robust, meaning it incorporates relevant concepts, procedures, and properties of the content and allows for connections between different representational systems.
6. *Demand*. The level of cognitive demand is sufficient to enhance mathematical reasoning. That is, the problem presents a challenge to students—it is not directly solvable through a formula, algorithm, or routine procedure but instead requires students to establish connections with underlying mathematical concepts and properties. It involves pattern recognition, conjecturing, generalization, and communication of mathematical ideas.
7. *Autonomy*. The problem provides students with the opportunity to work independently, make decisions, and justify their chosen strategy. It includes support or guided questions for students who encounter difficulties in solving it.

2.3 Algebraic Reasoning in Primary Education

Various theoretical perspectives and curricular proposals recommend the incorporation of algebraic content from the early educational levels (Kieran, 2022). Algebraic reasoning, understood as "the reasoning engaged in by 5- to 12-year-olds as they build meaning for the objects and ways of thinking to be encountered within the later study of secondary school algebra" (p. 1131), is included in primary education curricula across different countries, through modeling phenomena, recognizing dependencies between variables, identifying regularities, relationships, and properties, and expressing and manipulating symbolic representations (ACARA, 2014; CCSSI, 2015; MEFP, 2022).

Although there is no single definition of mathematical modeling within the educational community, there is a general consensus in recognizing modeling as a process for solving real-world problems using mathematics. According to Ledezma et al. (2024), modeling moves from the extra-mathematical to the intra-mathematical domain, establishing a connection between mathematical content and real-life situations or even other curricular disciplines, which becomes evident when an individual proposes a mathematical model or expression to solve an applied problem.

Given the role of modeling in developing mathematical competence, various authors advocate for promoting modeling activities from the early years of education, beyond its role in fostering algebraic reasoning (Alsina & Salgado, 2021; MEFP, 2022; Stohlmann & Albarracín, 2016). Early mathematical modeling helps "create initial models to analyze, explain, and understand reality through the mathematical knowledge that young students mobilize" (Alsina & Salgado, 2021). According to Borromeo (2018) and Dogan (2020), a well-designed mathematical modeling task presents a problem situation that is:

- *Open*. Interpretable in multiple ways and allowing for diverse solution strategies.
- *Complex*. Classified as a procedure-with-connection or a mathematical construction task, following Stein et al. (1996), meaning it requires students to comprehend and interpret the context, select essential data, and engage in cognitively demanding work.
- *Realistic*. Enabling students to interpret the problem based on their experiences and mathematical knowledge.
- *Authentic*. Relevant to a real-world, non-academic situation, drawing on students' experiences and involving meaningful data and questions.
- *Model-generating*. Requiring students to use mathematics to construct a model that describes and explains the given situation.

Variables, equations, functions, and the operations that can be performed with them serve as mathematical modeling tools for problems originating within mathematics itself (Godino et al., 2015). The first step in modeling involves identifying, designating, and relating the variables that characterize the system to be modeled. Next, working with the model requires the formal manipulation of symbolic expressions that represent system's properties, interpretation, and finally, the generalization of knowledge derived from the algebraic model.

The relationship between mathematical modeling and problem posing is bidirectional. On the one hand, during the modeling process, questions are formulated and reformulated to move from a disordered real-world situation to a well-defined problem statement. Additionally, meta-questions for monitoring, controlling, or critiquing the mathematical model or its results also emerge (Hansen & Hana, 2015). On the other hand, modeling activities can be involved from the very moment a problem is posed. When designing problems based on real-world situations, it is necessary to understand and explore the possibilities offered by the situation, distinguish between relevant and irrelevant information, and establish connections between data to assess whether the given information is coherent and sufficient for solving the problem. In this way, the context of mathematical modeling is particularly relevant for analyzing the actions involved in problem creation (Hartmann et al., 2022, 2023).

3 Method

In line with our research objective, we adopted a qualitative exploratory-descriptive approach (Lodico et al., 2010). The study involved 72 preservice teachers (hereinafter referred to as PTs), third-year students enrolled in a Primary Education degree program at a Spanish university. In the course where this experiment took place, PTs were required to deepen and apply the knowledge acquired in previous courses to analyze, design, and sequence mathematical tasks according to specific content and learning expectations. PTs had received prior training on: Curricular aspects of mathematics in primary education, particularly algebraic reasoning (early algebraic approaches and practices); foundations and strategies for mathematical problem posing, focusing on the didactic-mathematical purpose of problem design; and criteria for problem meaningfulness, as described in Section 2.1. Two theoretical-practical sessions of two hours each were dedicated to each of these topics.

Figure 3. ChatGPT problem-posing task proposed to PTs

Next, we will use ChatGPT to pose meaningful problems aimed at 6th-grade primary school students, with a specific didactic-mathematical purpose: to work jointly on the following two mathematical knowledge areas.

MAT.3.D.2. Mathematical Modeling.
MAT3.D.2.1. Modeling processes based on real-life problems using mathematical representations.

MAT.3.D.3. Relations and Functions.
MAT3.D.3.1. Equality and inequality relationships and the use of the signs < and >. Determination of unknown data (represented by a letter or symbol) in simple expressions related through these signs and the signs =, ≠.

1) Include the final problem statement that fulfills the task.
2) Describe, including the necessary screenshots, the sequence of instructions you give to ChatGPT and its responses to show how you have adapted ChatGPT's answers to arrive at the final problem.
3) Justify why the final problem meets the required conditions (working jointly on MAT3.D.2.1. and MAT3.D.3.1.) and why it is meaningful.

Following the usual structure of the course's practical sessions, PTs worked in teams of four or five to complete the task described in Figure 3. As mentioned earlier, they had already received the necessary training on the mathematical knowledge involved and the characteristics that define a meaningful mathematical problem.

To address the research questions, we analyzed the reports submitted by the PTs in response to the assessment task (Figure 3). We examined:

1. The sequence of prompts given by PTs to ChatGPT to pose or modify problems, along with their descriptions of the decisions made. The analysis focused on: a) the type of instruction provided to ChatGPT, identifying key terms such as pose, design, construct, modify, or improve; b) whether and how didactic-mathematical guidelines were introduced, particularly whether PTs used terms related to the demanded mathematical knowledge or meaningfulness as characteristics that ChatGPT should consider in problem creation; c) whether PTs requested a solution from the AI and their reasoning behind it.
2. The alignment of the final problems with the intended didactic-mathematical purpose. A problem was considered a modeling problem if it led to the generation of a model and met at least three of the four remaining characteristics (open-ended, complex, realistic, authentic), as defined by Borromeo (2018) and Dogan (2020).
3. The degree to which the problems met the indicators of meaningfulness, including: a relatable and realistic context; clear language and accessible terminology; a clearly stated and relevant purpose; curricular coherence; robustness by incorporating the expected mathematical content; a sufficient level of cognitive demand; and encouragement of independent problem-solving. The assessment was categorized as full, partial (meeting only some defining characteristics), or absent. This evaluation considered both the problem statement and the mathematical practices required to solve it. If PTs had requested ChatGPT to generate a solution, we analyzed the AI's response, identifying the mathematical content and its complexity relative to the intended student age group. Otherwise, the researchers solved the proposed problems themselves.
4. The justifications provided by PTs regarding the meaningfulness of the final problem and its didactic-mathematical adequacy.

This analysis was conducted independently by the authors and later cross-checked. In cases of discrepancies regarding the fulfilment of meaningfulness indicators, a joint review was carried out.

4 Results

4.1. Problem-posing process with ChatGPT

In this section, we analyze the process followed by the PTs in creating problems up to their final product, highlighting both the decisions that influenced the instructions given to ChatGPT and the didactic-mathematical knowledge involved in their decisions. In Table

1, we summarize the types of purposes addressed by PTs in their prompts during the problem-posing process.

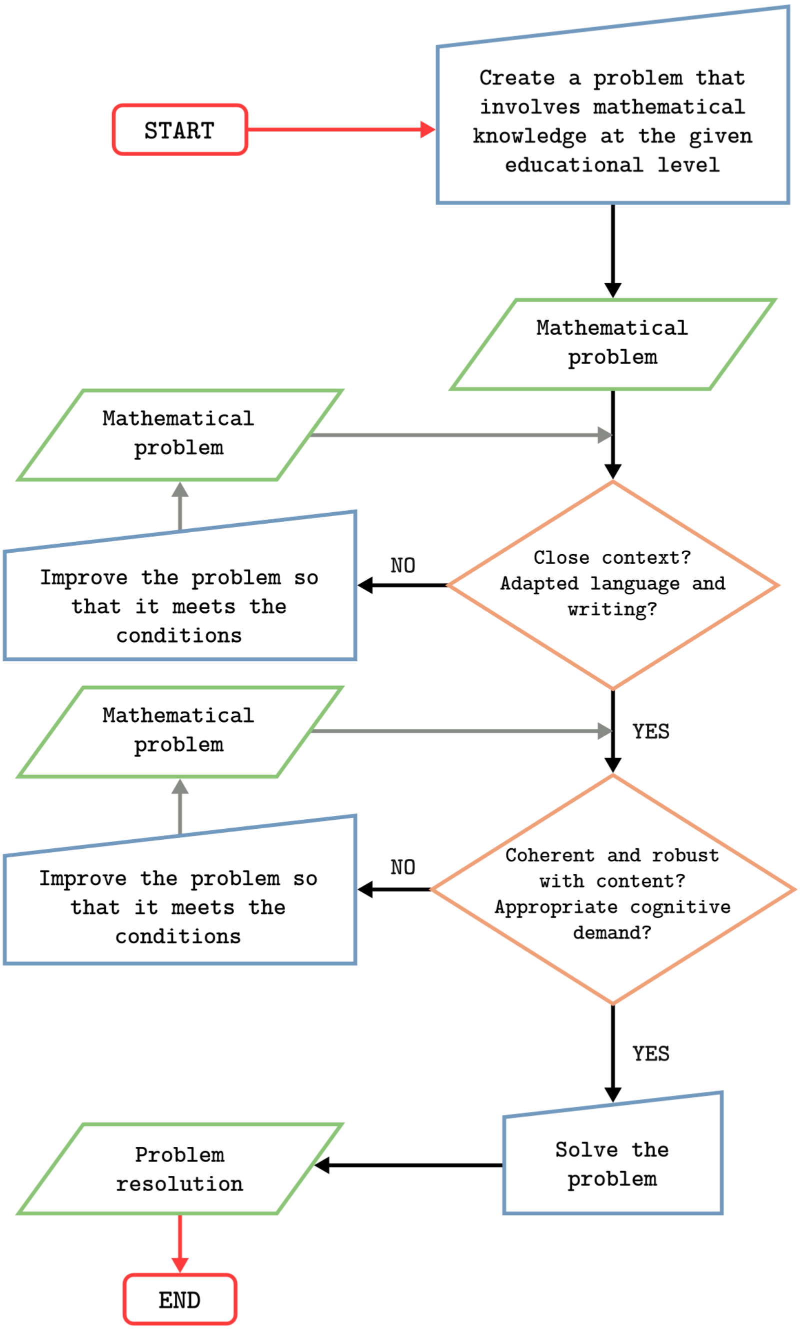
Table 1. Prompts' purposes during the problem-posing process by each team

Team	Prompt's purpose									
	Math knowledge	Educational context	Meaningfulness							Solution
			Context	Language	Purpose	Curricular coherence	Robustness	Demand	Autonomy	
T1	✓	✓	✓			✓		✓		✓
T2	✓	✓				✓		✓		
T3	✓	✓				✓		✓		✓
T4	✓	✓	✓	✓	✓	✓	✓	✓		✓
T5	✓	✓		✓	✓	✓	✓	✓		
T6	✓	✓	✓	✓	✓	✓	✓	✓		✓
T7	✓	✓				✓				
T8	✓	✓	✓	✓	✓	✓	✓	✓		✓
T9	✓	✓				✓		✓		✓
T10	✓	✓								
T11	✓	✓	✓	✓	✓	✓	✓	✓		
T12	✓	✓	✓	✓	✓	✓	✓	✓		✓
T13	✓	✓	✓			✓	✓			
T14	✓	✓								✓
T15	✓	✓				✓				
T16	✓	✓		✓	✓					✓

Figure 4 illustrates the most frequent process followed by PTs when creating problems. The type of instruction provided to ChatGPT is shown in blue, the output generated by the application in green, and the questions posed by users to assess the adequacy of the output and determine the next prompt, if necessary, in orange.

The average number of prompts given by the teams to ChatGPT before deciding on their final product is five, which is also the most frequent number of them. Most teams (12 out of 16) request ChatGPT in their first prompt to pose (other terms used include "design" or "construct") a problem that jointly addresses the required mathematical knowledge areas, which they introduce verbatim as it appears in the curricular regulations. In this first prompt, they include the grade level (6th grade) and, in half of the cases, the students' age (11–12 years) or their educational context (when this information is omitted, it becomes the subject of the next prompt). These instructions reflect a special focus on both the epistemic and ecological dimensions of didactic-mathematical knowledge.

Figure 4. Problem-posing process using ChatGPT



On two occasions, PTs do not copy the required mathematical knowledge verbatim in their first prompt but rather interpret it and additionally request that the task meet a set of characteristics (usually too generic for ChatGPT to interpret accurately) that ensure its meaningfulness: that it be robust and suitable for establishing relationships between different representation systems (epistemic facet), challenging (cognitive facet), and set in a familiar and realistic context (affective facet). It is observed that when meaningfulness characteristics are introduced in the first prompts and ChatGPT is then asked to pose a problem that meets them, its output tends to be a teacher-oriented description of an activity sequence, including objectives, materials, student instructions, and timing. This leads the PTs to request modifications or reformulate this output as the mathematical problem they would propose to their students.

Another two teams begin by explaining to ChatGPT what it should interpret as algebraic reasoning (T4) or by situating it in a specific role: “Imagine you are a primary school mathematics teacher. Your goal is to pose problems for sixth-grade students with a specific didactic-mathematical purpose” (T10). They then proceed to explain what constitutes a meaningful task.

Subsequent prompts given to ChatGPT aim to improve the meaningfulness of the problem proposed by the AI. Most frequently, the PTs first focus on affective or cognitive aspects, providing instructions related to the problem’s context and language: seven teams request that ChatGPT "modify" or "pose a new problem based on the previous one" to meet characteristics such as "using a familiar and realistic context," while five of these teams also ask that the language be better adapted to the students' educational level by using short and concrete sentences. Later, they mostly focus on aspects related to curricular coherence (ecological dimension), problem robustness (epistemic dimension), or cognitive demand (cognitive dimension), including prompts where they again ask ChatGPT to "modify" the problem to "involve conceptual and procedural aspects, consider relevant relationships and properties," "pose a challenge for students," or "establish relationships between different representation systems."

Although the problems posed by ChatGPT during the interaction process with the different teams were adapted to the required curricular level, they cannot be considered solid in terms of mathematical content, since they do not develop modeling (they are not open-ended, are of low cognitive demand or relevance, and do not lead to the generation of mathematical models) or do not involve algebraic knowledge (the use of relationships is arithmetic rather than algebraic and does not motivate the determination of unknown data). In this regard, only in half of the cases do the PTs notice, as T4 does, that “the content is not adequately addressed” and request a revision ("remember," "change," "clarify") of the given problem to ensure it promotes modeling and the use of equality and inequality relationships with the symbols $<$, $>$, $=$, and $\neq$. On rare occasions, PTs are more precise and clearly define the goal of the prompt. For example, T11, in response to the lack of content robustness, insists that "it must allow the student to create a mathematical model," while T13 instructs ChatGPT to "change the problem statement so that the student has to find the [functional] relationship" (Figure 5).

T13 then asks ChatGPT to “not provide the unknown [literal symbol for the variable].” In this regard, it is observed that in a significant portion of the problems generated with ChatGPT during successive interactions with all groups, literal symbols are assigned to unknown values (Figure 5) as part of the information provided in the problems. Additionally, ChatGPT confuses variables with unknowns and equations with inequalities. However, no team noticed this error on the part of the AI, reflecting deficiencies in their didactic-mathematical knowledge of algebraic reasoning within the epistemic dimension (different uses of literal symbols).

Figure 5. Prompt of T13 to improve robustness in relation to content

ChatGPT

Of course! Here is a problem that integrates the MAT3.D.3.1 knowledge more clearly:

Problem: Laura and Marta are organizing a bicycle race in the park. Laura knows she is faster than Marta, but she doesn't know how much faster. They decide to have a race to find out. After the race, Laura discovers that she took 3 minutes less than Marta to complete the course. If Marta's time is represented by the letter "M", how long did Marta take to complete the race?

This problem involves a clear inequality relationship (Laura is faster than Marta) and the determination of an unknown value (Marta's time), represented by a letter in an expression related through the "<" (less than) sign. Students will need to use the provided information to solve an equation and find the unknown value "M".

Anonymous

Change the statement so that the student has to find the relationship between the times Laura and Marta took to complete the race.

ChatGPT

Sure, here is the modified problem:

Problem: Laura and Marta are organizing a bicycle race in the park. Laura knows she is faster than Marta, but she doesn't know how much faster. They decide to have a race to find out. After the race, Laura discovers that she took 3 minutes less than Marta to complete the course. If Laura's time is represented by the letter "L" and Marta's time by the letter "M", what is the relationship between Laura's and Marta's times in completing the race?

Ten teams requested improvements related to cognitive demand, with instructions such as T6 stating, "I need the problem to be more complex," or T12 saying, "pose a simpler problem". Four of these teams, in an effort to enrich the problem's cognitive demand, provided ChatGPT with the definition of a mathematical construction problem (Stein et al., 1996) and then asked it to generate a problem with those characteristics. In these cases, the problems posed by ChatGPT did not fully reach this level, although they were considered procedural problems with connections, according to Stein et al. (1996).

Nine teams asked ChatGPT to solve the proposed problem as a way to check its "coherence with the mathematical content" it was supposed to involve. This was the final prompt in seven cases and an intermediate prompt in two others, either to validate the proposal except for avoiding terms like "mathematical model" in the problem statement (T6) or to recognize that the problem was "not sufficiently challenging" (T9) and request an increase in cognitive demand to the construction level (involving didactic-mathematical knowledge in the cognitive dimension).

In many of the problems posed during the process (as well as in half of the final proposed problems), it is not clearly established what the student is expected to do—the requirement is vague, or there is a lack of coherence between the provided information and the expected response (see Figures 8 and 9). However, only two teams identified these deficiencies in the problems generated during the posing process and requested changes such as adding missing data or clarifying the question to improve them. These results highlight shortcomings in the cognitive facet of PT's didactic-mathematical knowledge.

Once the interaction with ChatGPT is completed, PTs establish their final problem. In seven cases, this does not match the last proposal generated by ChatGPT, as they introduce changes to the information and requirements to avoid having the problem statement assign literal symbols to unknowns or variables (Figure 5), make the language and terminology more accessible to students, request justification for the answer, or remove the student instructions that ChatGPT includes in the problem descriptions. In three instances, the teams expanded the requirement to include aspects such as the use of <, >, =, ≠ symbols or the inclusion of tabular representation. For example, T8 added, "represent the relationship between correct answers and points in a table" to their final problem, which was not originally provided by ChatGPT (Figure 9), justifying their decision as follows:

> Since this task posed by ChatGPT also does not meet the criterion of using different representation systems, we added a section to the problem in which students must create a table representing the proportional relationship between points and correct answers.

4.2. Products of problem posing with ChatGPT

To assess the adequacy of the final problems generated by the PTs, we first analyzed whether they included the required mathematical knowledge.

As shown in Table 2, six of the proposed problems (by T2, T3, T5, T6, T11, and T12) fulfill the didactic-mathematical purpose, meaning they promote modeling and involve algebraic aspects. Regarding the specific characteristics of mathematical modeling, the final problems, although realistic (interpretable based on students' experiences and mathematical knowledge), are generally not open-ended (only in four cases does a part of the problem allow for different strategies). The problems require students to interpret the context but do not prompt them to reflect on which data are relevant for their resolution. Additionally, while the scenarios may be engaging for students by connecting to their personal experiences, in nearly half of the final problems the questions are not considered relevant in the sense that they do not lead to or generate significant mathematical knowledge (Figures 9 and 10). The most authentic problems are those that allow for the construction of models that describe the situation, such as analyzing and generating budgets (Figure 6), studying costs based on variable units, or determining the minimum expenditure based on the number of guests (as a variable or within a range) at a party.

Table 2. Mathematical characteristics of the problems posed by each team

Team	Mathematical model					Relations. Unknow data		
	Open	Complex	Realistic	Authentic	Model-generating	Equation	Inequation	Functional relation
T1		✓	✓				✓	
T2		✓	✓	✓	✓			✓
T3	✓	✓	✓	✓	✓			✓
T4	✓	✓	✓					
T5		✓	✓	✓	✓			✓
T6		✓	✓					✓
T7			✓					
T8		✓	✓	✓	✓	✓		
T9						✓		
T10		✓	✓		✓	✓		
T11	✓	✓	✓	✓	✓	✓		
T12	✓	✓	✓	✓	✓			✓
T13			✓	✓				✓
T14			✓					
T15		✓	✓					
T16		✓	✓					

These are modeled through functional relationships (linear or affine) in four cases or through equations (of the form $Ax + Bx + C = D$, where $A, B, C,$ or D can be 0) defined over integers (Figure 7). The not known values are either unknowns (in equations and inequalities) or variables (in functions), depending on the model involved.

Figure 6. Final problem proposed by T12. Early modeling.

> Pablo and Marta are organizing a candy stall in their neighborhood to raise funds for a school trip. Both are trying to decide on different prices to attract customers.
>
> Pablo sells each bag of candy for €1.50 and offers a special discount for those who buy more than 4 bags: for each additional bag purchased after the fourth, the price per bag is reduced by €0.25.
>
> Marta, on the other hand, has a fixed price of €1.25 per bag of candy but also offers a discount: for every 5 bags purchased, customers receive one additional bag for free.
>
> They are trying to decide which option, Pablo's or Marta's, is more convenient, considering that they can buy between 1 and 10 bags of candy.
>
> Which option (Pablo's or Marta's) will allow them to raise more money for the school trip? Create a graph to solve the problem and justify your answer.

In contrast, five of the proposed problems do not involve determining unknown data in equality and inequality relationships and are instead solved through comparisons and operations with natural numbers.

Figure 7. Final problem proposed by T11. Early modeling.

> María is facing an exciting and delicious challenge: With €15 in her pocket, we will discover all the possible combinations of chocolates and candies she can buy. Additionally, which combination will allow her to get the most sweets? Get ready to dive into a world of flavor and fun while learning math in a delicious way!
>
> Each chocolate costs €2, and each candy costs €1.
>
> How many different combinations of chocolates and candies can she buy?
>
> Which combination allows her to buy the maximum number of sweets within her budget?
>
> Justify your answer.

Next, Table 3 presents the criteria of meaningfulness met by the proposed final problems.

Table 3. Criteria of meaningfulness for each team's final problem

Team	Context	Language	Purpose	Curricular coherence	Robustness	Demand	Autonomy
T1	**			*		*	*
T2	**	**	**	*	*	*	*
T3	**	**	**	**	**	**	*
T4	**		*	*	*	*	
T5	**	*		**	*	*	
T6	**	*		*			
T7	**	**	**				
T8	**	**		**	*	*	
T9			*	*		*	
T10				*	*	*	*
T11	**	**	**	**	*	**	*
T12	**	**	**	**	**	**	*
T13	**	**	*	*	*	*	*
T14	**	**	**				
T15	**				*		
T16	**	**	**				

Note. ** indicates full compliance with the criterion, * indicates partial compliance with the criterion.

It is observed that only five of the final problems proposed by PTs at least partially meet the expected characteristics to be considered meaningful for working on the required mathematical knowledge.

The context of the proposed problems is generally engaging, relatable to students, and, except in two cases, realistic. The most common contexts include sports competitions (jumping contests, bicycle races, football championships), games or fairground attractions (Figure 8), and buying and selling candy or sweets. Less frequently, the problems are set in mathematical riddle contests (school fairs) or "treasure hunts". The language and terminology used are appropriate in nine of the 16 final problems. In the remaining cases, the problem statement is not well-structured (T1, T4), the questions are unclear (T4, T5, T6), or there is no coherence between the requirement and the provided information (T9, T10, T15). Figure 8 presents the final problem proposed by T6. This team, like T4 and T5, provided intermediate prompts requesting changes in terminology or reordering of the statement, but these improvements were not reflected in the final problem.

Figure 8. Final problem proposed by T6

> Imagine you are a playground designer tasked with creating a play area for a new park. You want to ensure that the distribution of play equipment is fair for all children, regardless of their age or physical abilities.
>
> In your design, you have decided to include three main play areas: a swing area, a slide zone, and a climbing area. To ensure safety and fun for everyone, each play area will have a maximum capacity of users at the same time.
>
> The swing area can accommodate up to 20 children swinging simultaneously. The slide zone can have up to 15 children sliding at the same time. The climbing area can have up to 10 children climbing simultaneously.
>
> You want to distribute the space fairly among the play areas so that all children have an opportunity to enjoy them while ensuring that none of the maximum capacities are exceeded.
>
> Describe how you would organize the distribution of children in each play area, using mathematical relations to show how the number of children in each play area relates to the maximum capacity. Then, determine how many children can participate in each play area simultaneously while maintaining fairness and respecting the maximum capacity of each play area.

Only in half of the proposed problems is it clearly stated what the student is expected to do or how they should respond to the questions. In other cases, situations are presented where it may be difficult for students to understand that no value satisfies the given conditions (for example, $5 < x < 4$, with x as a natural number, in the problem proposed by T1) and that this is the expected answer. Additionally, some problems lead to solutions where only an integer value makes sense as a response, but the equation modeling the problem yields a rational solution. For example, the final problem proposed by T8 (Figure

9) involves solving an equation $x + 2x + 3 = 17$, where x represents the number of riddles Pedro answered correctly, but its solution is not an integer.

Figure 9. Final problem proposed by T8

> At the school fair, Ana and Pedro participated in a math riddle contest. Ana solved 3 more than twice the number of riddles that Pedro solved. In total, they solved 17 riddles. If each correct answer earned them 5 points, how many points did each of them score? Additionally, if the first-place prize was 100 points, who won and by how many points? Represent the relationship between solved riddles and earned points in a table.

In general, the proposed problems can be considered appropriate for the students' educational level. Those that partially satisfy the curricular coherence indicator are problems that, while suitable for sixth-grade students, only address one of the two mathematical concepts they were expected to learn through the task (they are usually not modeling problems, as seen in Figure 8). Problems that do not meet this criterion only require operating with or comparing known natural numbers (Figure 10).

Figure 10. Final problem proposed by T7. Directly solvable.

> At a sports store, a long jump contest is being organized for sixth-grade students. The goal is for each participant to jump as far as possible from a starting mark to a finish line. It is known that Marta jumped three times the distance that Jorge jumped, and Raúl's jump distance was twice Marta's distance. If Jorge's jump was 3 meters, how far did each of the other children jump (in meters)? How far did Marta jump? How far did Raúl jump? Who jumped farther, Marta or Raúl? Who jumped less, Jorge or Marta? If Marta's jump distance had been 9 meters, how far would Raúl have jumped?

Regarding problem robustness, we observe that most problems only partially meet this criterion, as they do not encourage students to connect or apply relevant properties, nor do they prompt them to establish relationships between different representation systems.

Problems where the level of cognitive demand is appropriate (e.g., Figure 7) correspond to the procedures with connections level (Stein et al., 1996). The eight problems that partially meet the demand criterion are those that, while not directly solvable using a formula or routine procedure, do not require students to establish connections, make conjectures, or generalize. In five other cases, the proposed problems are directly solvable through arithmetic operations. For example, T7 considers the problem generated by ChatGPT in Figure 10 to be a modeling problem, even though it is directly solvable. During the creation process, this team requested that ChatGPT include questions 3, 4, and 5 (comparison between jumps), which had not been proposed by the application in earlier stages. Finally, only seven problems partially foster student autonomy (Figures 6 and 7) as they allow students to decide and justify their strategy but do not provide support or guided questions to assist those who may struggle.

An interesting finding regarding the instructions given to ChatGPT is that the teams which proposed the most meaningful problems provided prompts related to curricular coherence and cognitive demand. Additionally, they modified the final problem proposed by ChatGPT, changing the information (e.g., removing indications about assigning letters to unknowns or making the language more accessible) or the requirements (e.g., requesting justification for the answer or including sections to guide the resolution). Beyond this, no clear relationship was observed between the meaningfulness of the final problem and the sequence of instructions provided to ChatGPT.

4.3. Justification of the adequacy of the final problem

All teams considered that the final problems they posed were suitable for working on the required mathematical knowledge. In this regard, ChatGPT frequently accompanied its proposed problems with a description of how they incorporated modeling or relationships, which served as justification for the PTs. This demonstrates, particularly in the case of teams that posed problems not truly allowing for developing modeling, a limited understanding of what modeling entails. For example, T7 relied on the information provided by ChatGPT (Figure 11), which was added to their problem (Figure 10), to justify their decision:

> According to ChatGPT, this problem meets all the necessary requirements, and it is correct because, as it clearly explains, in this exercise, we first consider the jump lengths based on an unknown, x, which represents Jorge's jump length, and from there, we must create a modeling system that leads us to determine his teammates' jump lengths, so we are finally provided with all the necessary information (Jorge's jump length) to solve the problem.

Figure 11. ChatGPT's justification of the required mathematical knowledge in conversation with T7

1. Modeling with different mathematical representations: Children must represent the problem's information using equations or mathematical expressions.
2. Equality and inequality relations: Equality and inequality relations are used to establish the relation between the distances jumped by each child.
3. Use of symbols: Children must use mathematical symbols to express the relations between the distances jumped by each participant.

On the other hand, the majority of teams (13) justified the meaningfulness of the problem they proposed by providing a detailed description of how the problem met the different characteristics. They primarily highlighted curricular coherence, the relatability of the context, and the appropriateness of the language. Only in two instances did teams identify limitations, specifically regarding the length of the problem statement or the lack of relationships between different representations (T6, Figure 12). This aspect is generally considered satisfied when the problem encourages translation into symbolic language.

Figure 12. Justification of the meaningfulness of the problem shown in Figure 8 (T6)

Additionally, the problem is meaningful as it meets various characteristics that define it as such:

- It is coherent with the content that the student is intended to learn (as justified earlier), while also taking into account the prior knowledge the student needs to solve the problem.
  - The task is solid, as it involves both conceptual and procedural aspects (It uses conceptual terms that the student must know, such as mathematical relationships or equity, and the student must deduce the procedures to follow), and it poses a challenge to the student (since it cannot be solved directly using a formula, as might be the case with a problem involving proportionality).
  - The context presented is real in the student's mind, as it concerns a children's playground.
  - At the end of the statement, it clearly sets out what is expected of the student (It describes how you would organize the children in each game, using words and numbers to show how the number of children relates to the limit of each game. Then, determine how many children can play in each game at the same time, keeping things fair and respecting the limits of each game.)
  - The language is somewhat adapted to the level of the students the task is directed at, although it could be more suitable. The sentences used are precise and short (The swings can have up to 20 children at once, the slides up to 15, and the climbing wall up to 10).
  - Finally, it also allows establishing relations between different systems of representation (for this, it should specify requesting different systems of representation of the results, such as in the form of graphs or diagrams, for example).

Only one team considered that the problem posed by ChatGPT could not be deemed meaningful, stating: "It does not develop meaningful learning", "it is not a challenge for students", and "it is too lengthy, as the AI is unable to unify sections" (T5).

Discussion

This study investigates how a group of pre-service mathematics teachers use ChatGPT to pose problems for sixth-grade students, focusing on modeling and determining unknown data in equality and inequality relations. The use of ChatGPT allows teachers to transition from a semi-structured problem-creation scenario — where they must formulate a problem based on a given didactic-mathematical purpose — to a structured problem-creation scenario, in which they must analyze the adequacy of a problem that, in principle, already incorporates the mathematical content, and decide what modifications are needed and how to request them from the AI to ensure its meaningfulness.

In line with the findings of Alsina et al. (2024) and Zapatera and Quevedo (2021), PTs faced difficulties in posing meaningful problems that foster algebraic reasoning. Although most problems involved the study of relationships (equations, inequalities, functions), only a little more than one-third were modeling problems. Generally, PTs failed to recognize some essential characteristics of a problem to qualify as a modeling task, particularly the need for it to be open-ended. Therefore, these results point out that, even with the support of a tool such as ChatGPT, pre-service teachers exhibit limitations in their competence to pose algebraic problems. Indeed, ChatGPT can provide responses lacking depth and reflection when the topic at hand is related to a specific domain of knowledge such as mathematics (Farrokhnia et al., 2024; Getenet, 2024). Therefore, limited didactic-mathematical knowledge on the part of pre-service teachers may lead them to perform superficial analyses of ChatGPT's responses and to accept as final products problems that are not sufficiently meaningful to foster their students' learning.

Analyzing both the instructions provided to ChatGPT (the process) and the meaningfulness and justification of the final problem (the product) allowed us to identify which areas of didactic-mathematical knowledge PTs applied during the problem-posing process and to detect their shortcomings. We observed that most problems proposed by ChatGPT during the instructional sequence reflect an inadequate use of algebraic concepts. ChatGPT tends to pose problems where literal symbols are assigned to unknown values as part of the problem statement and often confuses equations with inequalities and unknowns with variables. Additionally, as noted in previous studies (Getenet, 2024), when requested to solve the proposed problems, the application may provide incomplete or incorrect solutions. These deficiencies in ChatGPT went unnoticed by PTs, an aspect which together with their limitations to identify the absence of the essential features to consider their problem as modeling tasks, reveal an insufficient didactic-mathematical knowledge in the epistemic dimension. In the ecological dimension, PTs correctly assess the appropriateness of the problems for the educational level but fail to recognize curricular deficiencies in terms of the required mathematical knowledge. From a cognitive perspective, in some cases, they do not perceive the routine nature and low challenge level of the task, nor do they reflect on whether students can clearly understand what is expected of them to solve the problem. Moreover, they exhibit a partial view of autonomy, which does not include support or guidance for students who may need it. However, in the affective aspect, they recognize which contexts might be engaging for students and strive to present problems in a relatable and motivating manner.

With respect to the prompt sequences provided by the pre-service teachers, we did not find a significant relationship between following any given series of instructions and succeeding in obtaining a meaningful problem that fosters algebraic reasoning. Nevertheless, it is noteworthy that, of the three proposed problems with the highest degree of meaningfulness, in two cases the teachers provided prompts including a description of almost all of the meaningfulness criteria and requested that the problem be modified in accordance with those criteria. Conversely, the three teams that did not provide any instruction aimed at improving curricular coherence ended up proposing problems that neither involved the use of relations or functions nor incorporated mathematical modeling. Consequently, although the results of this study do not allow us to infer an optimal prompt sequence, we believe that providing a detailed description of the main aspects of the didactical-mathematical knowledge involved in the problem-posing activity (very likely unknown to ChatGPT given its training data) is essential to obtain a meaningful and pertinent mathematical problem.

Finally, as was mentioned in the Results section, pre-service teachers used the arguments provided by ChatGPT to justify why their problems involved both functional relationships or (in)equality and modeling, even when those arguments were erroneous or limited. This fact not only reveals deficiencies in the epistemic facet of pre-service teachers' didactical-mathematical knowledge, but also indicates that they seem to ascribe to ChatGPT deep didactical-mathematical knowledge of the mathematical content, as well as higher-order thinking skills such as analytical thinking and reasoning. These results underscore the importance of raising pre-service teachers' awareness of the limitations of ChatGPT.

Conclusions

Problem posing engages teachers in authentic formative activities that promote a deep understanding of mathematical concepts, properties, and procedures. However, it also involves specific didactic knowledge about the quality of the problems posed, how they respond to certain educational objectives, and what difficulties students may encounter in solving them. In this way, it serves as a powerful tool for assessing and developing teachers' professional knowledge.

The results of our study highlight both the potential and the challenges of using ChatGPT for problem posing. We also consider its relevance for teacher educators, as it enables not only the fostering of pre-service teachers' competence in problem posing, but also the identification and assessment of the didactical-mathematical knowledge they mobilize throughout the prompts sequence leading to the final formulation of the problem. ChatGPT is a tool that can generate problem statements tailored to the specific interests and needs of teachers, enhancing their autonomy and engagement (Pelton & Pelton, 2023; Wardat et al., 2023). Nevertheless, given ChatGPT's deficiencies in handling mathematical content (Getenet, 2024; Wardat et al., 2023), especially algebraic content, its apparent lack of training in didactic-mathematical knowledge (Figure 2), and the absence of high-order thinking skills (Farrokhnia et al., 2024), teachers must make informed decisions to "guide" ChatGPT in generating meaningful problems. In this sense, ChatGPT tends to produce "credible" responses to achieve its objectives: the reader is satisfied and "rewards" it, or the reader does not notice that the response is incorrect and it "avoids punishment" (Urban et al., 2024). Authors such as Cooper (2023) even point to the possibility that ChatGPT may position itself as the ultimate epistemic authority, assuming its own outputs as truth without well-substantiated evidence to support them. This implies that pre-service teachers must actively supervise and evaluate the information they receive from ChatGPT, whether it is the problem statement or the description of its characteristics. This requires, on the one hand, being aware of ChatGPT's intrinsic limitations and weaknesses and, on the other hand, developing meta-didactic-mathematical knowledge, enabling them to assess the didactic suitability (Godino et al., 2017) of the final product.

Regarding limitations and future research directions, we consider it necessary to address three aspects: comparing the outcomes of problem creation by the same participants with and without ChatGPT support; providing didactic suitability guidelines to guide reflection and evaluation of problem adequacy across different dimensions; and creating discussion spaces where pre-service teachers can share and critically assess their proposals.

Research ethics

Author contributions

M.B.: Conceptualization, Methodology, Formal analysis, Writing - Original draft, Writing - Review and editing.

N.T-E: Conceptualization, Methodology, Formal analysis, Writing - Review and editing.

All authors have read and agreed to the published version of the manuscript.

Artificial intelligence

Artificial intelligence was not used in writing the article.

Funding

None.

Institutional review board statement

Not necessary.

Informed consent statement

Not applicable.

Data availability statement

The data supporting the results of this study will be made available by the authors upon request.

Acknowledgements

Research conducted as a part of the research Project PID2022-139748NB-100 funded by MCIN/AEI/10.13039/501100011033/ and FEDER, with support of the Research Group FQM-126 (Junta de Andalucía, España).

Conflicts of Interest

The authors declare no conflicts of interest.

References

Alsina, Á., Pincheira, N., & Delgado-Rebolledo, R. (2024). The professional practice of designing tasks: How do pre-service early childhood teachers promote mathematical processes in early algebra? *ZDM – Mathematics Education*, *56*(6), 1197–1210. https://doi.org/10.1007/s11858-024-01636-1

Alsina, Á., & Salgado, M. (2021). Introduciendo la Modelización Matemática Temprana en Educación Infantil: Un marco para resolver problemas reales. *Modelling in Science Education and Learning*, *14*(1), 33. https://doi.org/10.4995/msel.2021.14024

Australian Curriculum, Assessment and Reporting Authority (ACARA). (2014). *Foundation to year 10 curriculum: Number sense and algebra (ACMSPO24).*

Ayllón, M. F., Gallego, J. L., & Gómez, I. A. (2016). La actuación de estudiantes de educación primaria en un proceso de invención de problemas. *Perfiles Educativos*, *38*(152), 51–67. https://doi.org/10.22201/iisue.24486167e.2016.152.57588

Baumanns, L., & Rott, B. (2021). Rethinking Problem-Posing Situations: A Review. *Investigations in Mathematics Learning*, *13*(2), 59–76. https://doi.org/10.1080/19477503.2020.1841501

Borromeo, R. (2018). *Learning How to Teach Mathematical Modeling in School and Teacher Education*. Springer International Publishing. https://doi.org/10.1007/978-3-319-68072-9

Burgos, M., Tizón-Escamilla, N., & Chaverri, J. (2025). A model for problem creation: Implications for teacher training. *Mathematics Education Research Journal*, *37*(1), 55–84. https://doi.org/10.1007/s13394-023-00482-w

Cai, J., & Hwang, S. (2020). Learning to teach through mathematical problem posing: Theoretical considerations, methodology, and directions for future research. *International Journal of Educational Research*, *102*, 101391. https://doi.org/10.1016/j.ijer.2019.01.001

Cai, J., & Leikin, R. (2020). Affect in mathematical problem posing: Conceptualization, advances, and future directions for research. *Educational Studies in Mathematics*, *105*(3), 287–301. https://doi.org/10.1007/s10649-020-10008-x

Cooper, G. (2023). Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence. *Journal of Science Education and Technology*, *32*(3), 444–452. https://doi.org/10.1007/s10956-023-10039-y

Core State Standards Initiative [CCSSI]. (2015). *Common Core State Standards for Mathematics.*

Dogan, M. F. (2020). Evaluating Pre-Service Teachers' Design of Mathematical Modelling Tasks. *International Journal of Innovation in Science and Mathematics Education*, *28*(1), 44–59. https://doi.org/10.30722/ijisme.28.01.004

Elgrably, H., & Leikin, R. (2021). Creativity as a function of problem-solving expertise: Posing new problems through investigations. *ZDM – Mathematics Education*, *53*(4), 891–904. https://doi.org/10.1007/s11858-021-01228-3

English, L. D. (2020). Teaching and learning through mathematical problem posing: Commentary. *International Journal of Educational Research*, *102*, 101451. https://doi.org/10.1016/j.ijer.2019.06.014

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. https://doi.org/10.1080/14703297.2023.2195846

Fernández, M. E., & Carrillo, J. (2020). Uma aproximação à forma como os alunos do ensino primário formulam problemas. *Revista de Educação Matemática*, *17*, e020015. https://doi.org/10.37001/remat25269062v17id257

Getenet, S. (2024). Pre-service teachers and ChatGPT in multistrategy problem-solving: Implications for mathematics teaching in primary schools. *International Electronic Journal of Mathematics Education*, *19*(1), em0766. https://doi.org/10.29333/iejme/14141

Godino, J. D., Giacomone, B., Batanero, C., & Font, V. (2017). Enfoque Ontosemiótico de los Conocimientos y Competencias del Profesor de Matemáticas. *Bolema: Boletim de Educação Matemática*, *31*(57), 90–113. https://doi.org/10.1590/1980-4415v31n57a05

Godino, J. D., Neto, T., Wilhelmi, M. R., Etchegaray, S., & Lasa, A. (2015). Algebraic reasoning levels in primary and secondary education. In K. Krainer & N. Vondrová, *Proceedings of the Ninth Congress of the European Society for Research in Mathematics Education* (pp. 426–432). CERME.

Hansen, R., & Hana, G. M. (2015). Problem Posing from a Modelling Perspective. In F. M. Singer, N. F. Ellerton, & J. Cai (Eds), *Mathematical Problem Posing: From Research to Effective Practice* (pp. 35–46). Springer New York. https://doi.org/10.1007/978-1-4614-6258-3_2

Hartmann, L.-M., Krawitz, J., & Schukajlow, S. (2022). The process of modelling-related problem posing—A case study with preservice teachers. In C. Fernández, S. Linares, Á. Gutiérrez, & N. Planas (Eds), *Proceedings of the 45th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 355–362). PME.

Hartmann, L.-M., Krawitz, J., & Schukajlow, S. (2023). Posing and Solving Modelling Problems—Extending the Modelling Process from a Problem Posing Perspective. *Journal Für Mathematik-Didaktik*, *44*(2), 533–561. https://doi.org/10.1007/s13138-023-00223-3

Kieran, C. (2022). The multi-dimensionality of early algebraic thinking: Background, overarching dimensions, and new directions. *ZDM – Mathematics Education*, *54*(6), 1131–1150. https://doi.org/10.1007/s11858-022-01435-6

Leavy, A., & Hourigan, M. (2022). The Framework for Posing Elementary Mathematics Problems (F-PosE): Supporting Teachers to Evaluate and Select Problems for Use in Elementary Mathematics. *Educational Studies in Mathematics*, *111*(1), 147–176. https://doi.org/10.1007/s10649-022-10155-3

Ledezma, C., Rodríguez-Nieto, C. A., & Font, V. (2024). The Role Played by Extra-mathematical Connections in the Modelling Process. *Avances de Investigación En Educación Matemática*, *25*, 81–103. https://doi.org/10.35763/aiem25.6363

Liljedahl, P., & Cai, J. (2021). Empirical research on problem solving and problem posing: A look at the state of the art. *ZDM – Mathematics Education*, *53*(4), 723–735. https://doi.org/10.1007/s11858-021-01291-w

Lodico, M. G., Spaulding, D. T., & Voegtle, K. H. (2010). *Methods in educational research: From theory to practice*. Wiley.

Malaspina, U., Mallart, A., & Font, V. (2015). Development of teachers' mathematical and didactic competencies by means of problem posing. In K. Krainer & N. Vondrová (Eds), *Proceedings of the Ninth Congress of the European Society for Research in Mathematics Education* (pp. 2861–2866). Charles University in Prague, Faculty of Education and ERME.

Malaspina, U., Torres, C., & Rubio, N. (2019). How to Stimulate In-Service Teachers' Didactic Analysis Competence by Means of Problem Posing. In P. Liljedahl & L. Santos-Trigo (Eds), *ICME-13 Monographs* (pp. 133–151). Springer International Publishing. https://doi.org/10.1007/978-3-030-10472-6_7

Mallart, A., Font, V., & Diez, J. (2018). Case Study on Mathematics Pre-service Teachers' Difficulties in Problem Posing. *EURASIA Journal of Mathematics, Science and Technology Education*, *14*(4). https://doi.org/10.29333/ejmste/83682

Memarian, B., & Doleck, T. (2023). ChatGPT in education: Methods, potentials, and limitations. *Computers in Human Behavior: Artificial Humans*, *1*(2), 100022. https://doi.org/10.1016/j.chbah.2023.100022

Ministerio de Educación y Formación Profesional [MEFP]. (2022). Real Decreto 157/2022, de 1 de marzo, por el que se establecen la ordenación y las enseñanzas mínimas de la Educación Primaria. *Boletín Oficial Del Estado*, *52*(I), 24386–24504.

Pelton, T., & Pelton, L. F. (2023). Adapting ChatGPT to support teacher education in mathematics. In E. Langran, P. Christensen, & J. Sanson (Eds), *Proceedings of Society for Information Technology & Teacher Education International Conference* (pp. 1662–1670). Association for the Advancement of Computing in Education (AACE).

Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, *3*, 121–154. https://doi.org/10.1016/j.iotcps.2023.04.003

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), 1–22. https://doi.org/10.37074/jalt.2023.6.1.9

Sarrion, E. (2023). *Exploring the power of ChatGPT. Applications, Techniques, and Implications*. Apress Berkeley, CA. https://doi.org/10.1007/978-1-4842-9529-8

Silber, S., & Cai, J. (2021). Exploring underprepared undergraduate students' mathematical problem posing. *ZDM – Mathematics Education*, *53*(4), 877–889. https://doi.org/10.1007/s11858-021-01272-z

Singer, F. M., Ellerton, N., & Cai, J. (2013). Problem-posing research in mathematics education: New questions and directions. *Educational Studies in Mathematics*, *83*(1), 1–7. https://doi.org/10.1007/s10649-013-9478-2

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal*, *33*(2), 455–488. https://doi.org/10.3102/00028312033002455

Stohlmann, M. S., & Albarracín, L. (2016). What Is Known about Elementary Grades Mathematical Modelling. *Education Research International*, *8*(5), 1–9. https://doi.org/10.1155/2016/5240683

Su, J., & Yang, W. (2023). Unlocking the Power of ChatGPT: A Framework for Applying Generative AI in Education. *ECNU Review of Education*, *6*(3), 355–366. https://doi.org/10.1177/20965311231168423

Urban, M., Děchtěrenko, F., Lukavský, J., Hrabalová, V., Svacha, F., Brom, C., & Urban, K. (2024). ChatGPT improves creative problem-solving performance in university students: An experimental study. *Computers & Education*, *215*, 105031. https://doi.org/10.1016/j.compedu.2024.105031

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), em2286. https://doi.org/10.29333/ejmste/13272

Yao, Y., Hwang, S., & Cai, J. (2021). Preservice teachers' mathematical understanding exhibited in problem posing and problem solving. *ZDM – Mathematics Education*, *53*(4), 937–949. https://doi.org/10.1007/s11858-021-01277-8

Zapatera, A., & Quevedo, E. (2021). The Initial Algebraic Knowledge of Preservice Teachers. *Mathematics*, *9*(17), 2117. https://doi.org/10.3390/math9172117