

THE ROLE OF CHATBOTS AS RESOURCES IN MATHEMATICAL TASK DESIGN FOR PRE-SERVICE TEACHERS: A PRAXEOLOGICAL ANALYSIS

Giorgia Nieddu^{1,2}, Francesco Beccuti^{2,3}, Maria Cristina Carrisi², Cristina Aliberti⁴, Maria Polo², Berardo Branella⁴, Ornella Robutti⁴

¹Università di Macerata; g.nieddu@unimc.it

²Università di Cagliari; mariacri.carrisi@unica.it ; mpolo@unica.it

³Università di Trieste; francesco.beccuti@units.it

⁴Università di Torino; cristina.aliberti@edu.unito.it ; ornella.robutti@unito.it

We investigate the use of a popular chatbot interface as a resource for designing mathematical problem solving activities by 57 pre-service mathematics teachers enrolled in two quasi-parallel teaching qualification courses in Italy. Employing the Meta-Didactical Transposition framework, we study the participants' initial praxeologies concerning the chatbots which we partially compare with those related to more traditional resources such as textbooks. Our findings show varied levels of engagement in interacting with the chatbot, highlighting tendencies toward reliance on chatbot outputs with limited critical re-elaboration. In doing this, we also discuss the difficulties and opportunities of the employment of chatbots as a task design resource also with reference to some of their key features such as their opacity and the dynamicity of their underlying technology.

Keywords: Mathematics education, Artificial intelligence, Meta-Didactical Transposition, Teacher training

INTRODUCTION

Research in mathematics education has evidenced both the supportive role and limitations of so-called generative artificial intelligence (AI) in instructional design. In particular, significant challenges have emerged as related to the educational use of AI chatbots, including concerns about pedagogical autonomy and a generalized risk of over-reliance on the chatbot (e.g., Sawyer, 2024). Segal and Biton (2025) showed how chatbots can aid pre-service teachers in formulating clearly worded mathematical problems and predicting students' difficulties. Buchholtz and Huget (2024) found that prompting the chatbot to pose questions to pre-service teachers, rather than providing direct ideas, fostered better decision-making and heightened awareness of key lesson planning variables. On the other hand, Sawyer (2024) observed that pre-service teachers using chatbots to develop teaching resources over-relied on the chatbots' mathematical accuracy and overall prioritized superficial aspects such as visual appeal. Noster and colleagues (2024) investigated whether pre-service teachers took into consideration the general unpredictability of chatbots' outputs and their proneness to mistake. They found that these aspects were frequently overlooked, thus suggesting paying explicit attention to this in teacher training. Together, these studies highlight a complex and nuanced landscape: chatbots hold potential as supportive tools in mathematics education, yet their effective integration depends on addressing challenges related to pedagogical autonomy, transparency, and critical engagement. With the aim of contributing to this growing body of research, the present study offers a preliminary analysis of the way in which two cohorts of Italian pre-service teachers engage in mathematical task design while employing, among other resources, an AI chatbot [1].

THEORETICAL FRAMEWORK

We adopt the Meta-Didactical Transposition framework (as formulated by Cusi et al., 2022), which was conceived to describe how teachers' and researchers' meta-didactical praxeologies evolve in various contexts (e.g., teacher-training programs). According to Chevallard (2019), a praxeology for a certain type of tasks (pertaining to any domain, although Chevallard's focus is primarily on mathematical and didactical tasks) is a quadruplet which comprises the said type of tasks, a technique that solves them, a technology (the explanatory discourse that justifies the technique), and a theory (the underlying principles and conceptual structures that give the discourse its validity and connect it to the broader field). Collectively, the type of task and technique are called the *praxis* component of a praxeology, while the technology and theory are called the *logos* component. Correspondingly, a meta-didactical praxeology, again as defined by Cusi and colleagues (2022), is a praxeology related more specifically to the activities of the teachers out of the class, in an institutional professional development course with researchers: it again includes a type of tasks (e.g., designing a classroom laboratorial activity); a technique (e.g., consulting institutional references, asking the chatbot for ideas); a technology and a theory (discourses which justify the techniques employed at a practical as well as at a more theoretical level; in particular a theory encompasses deeper underlying motivations, which may include teachers' beliefs, their general epistemological stances, and their broader attitudes and ways of reasoning about teaching and learning in general or with specific reference to mathematics). These authors argue that a component of a praxeology emerges at the macro level as a result of micro-level interactions among various agents—both human and non-human—involved in mathematical activity, following de Freitas and Sinclair's conceptualization of agency (2014). Cusi and colleagues complement the latter with the classification of agents by Prodromou and colleagues (2018): methodological agents (common teaching practices); disciplinary agents (theoretical or procedural foundations influencing choices, linked to task analysis and mathematical concepts); institutional agents (national curriculum, training programmes etc); technological agents (use of software, hardware, paper-and-pencil environments; not only the tools but the way of using them); motivational agents (teachers and researchers' experiences and beliefs shaping actions and decisions).

Employing the aforementioned framework, we aim at investigating the following questions:

RQ1: Which are these pre-service teachers' initial task design *praxeis* related to the use of chatbots in mathematical task design?

RQ2: Which are these pre-service teachers' initial *logoi* related to the use of chatbots in mathematical task design and education in general?

We understand these praxeologies as "initial" both in view of the fact that the participants had not received extensive training in activity design and in view of the fact that these chatbots are still relatively recent tools (i.e. tools with which presumably the participants have had no experience or a very limited experience at the time when the experiment was carried out, see below).

In addition to the framework outlined above, we defined an additional *a priori* classification to describe participants' use of the chatbot and to better analyze their *praxeis*. This new classification focuses on three key dimensions: the level of complexity in prompt formulation; the degree of autonomous decision-making (as reflected in the nature and specificity of the commands given to the chatbot); and the extent to which the chatbot's outputs were integrated into the final task design. With regard to task design itself, our assignment to the participants was based on the concept of *laboratorio di matematica* (mathematics laboratory), understood as a set of methodological guidelines aimed at constructing meanings for mathematical objects (Anichini et al., 2004). In this context, students are actively encouraged to socially

engage with and solve problems, are introduced to the cultural tools of mathematics to study facts and phenomena, and are supported in reflecting on the involved mathematical meanings (Anichini et al., 2004). Finally notice that we adopt an agnostic perspective regarding the functioning of chatbots, treating them essentially as black boxes (cf. Wiener, 1961, p. ix), i.e., without making assumptions about their underlying (or emerging) human-like cognitive abilities [2]. This assumption and the consequent assumed opacity of the tool is not only coherent with the behaviouristic theoretical stance of Turing (1950) but also fundamentally reflects the participants' general knowledge on the chatbot, as we will see.

METHODOLOGY

The participants of the present study are 57 pre-service teachers enrolled in two teacher training courses for upper secondary education in two cities: 25 participants attended in Torino, capital of the north-western region of Piedmont, and 32 participants in Cagliari, capital of the region of Sardinia. These are part of two parallel one-year training programs which are organized at a local level but aligned with the same national ministerial framework. Completion of the program is required for accessing the national competition for permanent teaching position in state schools [3]. Access to the program is conditional on holding a two-year master's degree with a sufficient amount of mathematical coursework. Notice that these degrees often do not include compulsory courses in education nor in mathematics education. Notice also that the Italian system allows—and frequently sees—periods of (precarious and/or substitute) teaching prior to obtaining formal certification or completing any kind of professional training. While this is the case for many of the participants, the present is the first formal teacher training experience for most participants. The course began in Torino in June 2024, in Cagliari in September 2024. The third and fifth authors were teacher educators in the course in Cagliari, the seventh author in Torino.

An experiment was conducted in both courses after the first three classes, ensuring participants had basic knowledge of lesson planning. The participants first completed a questionnaire [4] aimed at collecting preliminary data: previous teaching experience (*Yes/No*); type of degree(s) held; prior use of chatbots (*Yes/No*); current classroom use (*Yes/No*); these were followed by three open-ended questions: “Why? Provide a justification for your previous answer”; “What are your thoughts on the use of ChatGPT (and similar tools) for teaching mathematics?”; “Which issues do you think using ChatGPT presents?”. They then attended a short presentation introducing generative AI and Chatbots, and were divided into small groups (3-5 participants for each group, 15 groups in total) and were asked to design an upper secondary school problem solving mathematics laboratory activity using one of three categories of resources: 4 groups were instructed to choose one among a selection of the most widely adopted textbooks in Italy; 5 groups were asked to use any resource they deemed proper to use which they could find by searching on the world wide web; 6 groups were instructed to use ChatGPT [5]. After completing the activity (approximately one hour was dedicated to this), the groups were asked “Do you think that your resource was useful and appropriate for the task? How? Why?”. The data for this study hence are all the individual questionnaires, the group-designed activities, the final group answers, and the chatbot conversations (when applicable).

Thus, to address the research questions, in the following section, we study the data with particular reference to the questionnaire responses and the materials produced by the groups, focusing on those obtained with the help of the chatbot and employing those obtained with other resources to make comparisons. The aim of the analysis is to explore the groups' initial praxeologies in relation to the task they were assigned (activity design with the use of chatbots). In consistence with this aim, we deliberately chose not to provide technical instructions on prompt engineering, so as to not influence the participants and to keep their approaches as close as possible to be considered as initial.

As to RQ1, we first classified the final activities proposed by the groups, evaluating whether these could be considered as part of a mathematics laboratory. Second, we analysed the chatbot conversations according to the classification of agents by Prodromou and colleagues (2018) discussed in the previous section. Finally we classified the entirety of the material (the participants' final proposals, their interaction with the chatbot, together with their final answers) by means of the following the *a priori* categories described in the previous section. As to RQ2, we searched again the materials for sentences or expressions reflecting the participants' *logos* concerning the use of chatbots for task design, and more generally for mathematics education. We selected from the final answers produced by the groups that used the chatbot those phrases containing references to participants' view on the chatbot usefulness, output characteristics, and how chatbots should be used for mathematical task design. We then analysed the questionnaire responses to gain a broader understanding of all participants' concerns and opinions regarding the use of chatbots as educational resources.

ANALYSIS

A first reading of the activities produced by the participants revealed that very few of them proposed a problem solving laboratorial activity as understood by Anichini and colleagues (2004), favouring procedural tasks instead. Indeed, almost all the groups who used books or the web simply copied or slightly modified a portion of text found there. Similarly, those who used the chatbot generally handed in the last output of their interaction with minor changes. Reading the conversations, the chatbot appears to have influenced the participants and, in some cases, to have shifted their attention towards task design variables, indeed acting both as a methodological and a disciplinary agent. It acted as a methodological agent, for instance by suggesting specific collaborative scenarios for the problem, classroom discussions, or more traditional lessons in many cases, though the proposed strategies were often perceived by teachers (as declared in their answers) as simplistic or already familiar. It acted as a disciplinary agent, for example, by conducting *a priori* analyses of the mathematical content, or prompting teachers to formalize their lesson planning with circumscribed learning objectives and prerequisites.

Several patterns emerged regarding the ways in which the 6 aforementioned groups were assigned the use of the chatbot (subsequently called group A, B, C, D, E, and F) engaged with it. Groups A, B, and C typed short and general commands (e.g., "Design a problem solving activity aimed at introducing trigonometry"), mostly displaying low engagement with the tool and asking few follow-up questions. As said above, their final lesson plans mostly reproduced the chatbot's last output (only group A added a brief introduction). Group D also used short prompts but asked the chatbot to elaborate on its own output by focusing on the problem's scenario and what should be done practically in the classroom (e.g., "I need to set up an activity that begins with questions like "how does the signal get to our house?""") and delegated to the chatbot the specific properties of the topic to be included in the activity (e.g., the decision to treat the parabola as a locus of points rather than a real function). Again, their final report coincided with the last output of the chatbot. Group E employed more refined prompting strategies, including hypothetical classroom context and details on the mathematical topics (e.g., "Think like a math teacher. [...] parabola in the Cartesian plane as an interpretation of sign study", "The class is a merged group of two first-year classes: one only works with the discriminant [...]"), but still handed in the chatbot's final output with no change. Group F posed brief and minimally elaborated commands (e.g., "Design a laboratorial activity to understand exponential growth based on the spread of infections at the beginning of the COVID-19 pandemic"). Yet, they submitted a practical laboratorial activity with a clearly defined classroom scenario and a basic outline of its phases, even if not highly detailed. In this case, only the practical part of the activity proposed by the chatbot was adopted by the teachers almost unchanged. However, they also added further details which do not appear in the

chatbots' outputs. These especially concern their activity's final aim which focused on the students' formal mastery of the topic, thus suggesting that (at least a minimal) process of collaborative re-elaboration of the outputs.

In summary, groups A, B, and C showed low prompt complexity and stuck to the chatbot's suggestions. Group D displayed medium prompt complexity, made some decisions about the lesson scenario during the interaction, but still showed high dependence on the chatbot. Group E engaged with the chatbot with complex prompts, displayed high decision-making but ultimately handed in the chatbot's output. Group F used very short prompts, showed a medium degree of autonomy in decision-making, and medium to low dependence on the chatbot's output, reworking its suggestions independently—either by significantly modifying them or by adding missing details. As to these groups' responses to the final questions, all expressed a favourable stance toward the chatbot, generally acknowledging its potential usefulness. Group B additionally noted that not all ideas and content generated by the chatbot during their interaction were useful (e.g., "some activities were limited to drawing triangles and measuring [...]") and, in view of this, cared to specify that chatbots should be regarded as support and hence not substitute teachers. Group C appreciated the "creative ideas" provided by the chatbot, which they felt could (and need to) be adapted to suit teachers' needs. Group D emphasized the necessity of "guiding" the chatbot to avoid falling into "the repetition predefined structures" and stated that they had considered it as "an expert illustrating its knowledge following our viewpoint". Groups A and E focused on how the chatbot could "facilitate" and "speed up" lesson planning, reducing teachers' workload.

If we read the responses to the individual questionnaires, we find that almost all participants raised concerns on the use of chatbots in education (e.g., "The answer within reach, the solution to a problem readily available, without students engaging in processes or formulating conjectures"). Some participants focused on students' misbehaviour or tendency to avoid thinking autonomously (e.g., "the excessive and inappropriate use of the tool without any educational purpose by the students"). On the other hand, some participants emphasized the teacher's responsibility, sometimes framing it in the first-person plural, meaning "we as teachers", (using expressions as "we should teach"), or showing awareness of the need for careful planning (e.g., "We must teach them to master it rather than be subjected to it"; "to teach kids how to use this resource at its best and with a critical spirit"; "Useful if used [...] after careful activity planning").

CONCLUSION

In summary, the chatbot appeared to play a dominant role in the activity design, as the participants relied almost exclusively on it, with minimal use of other resources (except some who consulted the national guidelines for teaching). This was expected as these participants were explicitly instructed to use the chatbot in their process (and the instructions mentioned no other tool). As seen, in almost all cases the chatbot directed the participants' attention to important elements of activity design (e.g., learning objectives, task analysis). Overall, almost all the groups exhibited a strong tendency to employ the chatbot as an oracle of sort (cf. Zurru, 2024), perhaps expressing an initial *praxis* related to the use of chatbots in mathematical task design. Interestingly, we observed a similar tendency in other groups as well, particularly among those who used textbooks. What we observed as to the groups that used the chatbot aligns with findings from previous literature (e.g., Noster et al., 2024; Sawyer, 2024), which highlight pre-service teachers' general overreliance on chatbots as well as their generally limited consideration of factors such as chatbots' opacity and proneness to errors. In line with these authors, we stress the need to address these issues explicitly within teacher training programs.

The groups (respectively of the resource employed) generally did not articulate clear justifications for the design choices made. On the contrary, in some cases, the chatbot proposed its own reasoning, broader

didactic approaches (and implicit biases) which the participants simply accepted (i.e., they again used the tool as an oracle, as said). To address this issue, an approach like the one proposed by Buchholtz and Huget (2024) mentioned in the introduction may offer a solution, as explicitly asking the chatbot to support the pre-service teachers in their design processes, rather than simply providing solutions, was proven by the authors to be effective in fostering pre-service teachers' reflection on their choices. Notice that, in our case, while many of the chatbots' design suggestions were proper, the teaching practices and the justifications provided often lacked depth and theoretical grounding.

The statements by the participants in the questionnaire concerning, for instance, the risk of overreliance on the chatbot and on the role of teachers in guiding students towards an appropriate use of chatbots, indicate that they had already engaged in independent reflections regarding the future role of chatbots in education. In some groups, as seen, a shared initial *logos* appears to be based on a shared discourse on the responsibility of teachers as a professional category, an agreed need for ad hoc training, and a general preoccupation for careful planning.

Generally speaking, also in view of the extreme dynamicity of the development of chatbot technology, we may consider these chatbots as fluctuating agents, sometimes reinforcing, sometimes conflicting with pre-existing praxeologies. In particular, a further deeper study of pre-service teachers' *logoi* in connection to the use of chatbots as educational resources appears to be pressing. This will be relevant not only to investigate the impact of these tools in education, but also because the ways in which these resources are used can make visible pre-service teachers' underlying views about mathematics, didactic methods, as well as the teaching paradigms which are institutionally prevalent where they operate (cf. Beccuti & Robutti, 2022). This is particularly significant in the case of chatbots not only because their underlying technology appears to be still in full development but also because their functioning appears to this day to be uniquely opaque both to pre-service teachers and researchers alike.

NOTES

1. We employed the chatbot ChatGPT (version available between June - September 2024), a conversational interface based on the Large Language Model Generative Pretrained Transformer owned by the private company OpenAI Inc., based in the United States of America. Our choice of this specific tool was motivated by the intention to exploit the most famous and popular chatbot with a user-friendly interface, to summon participants' opinions, beliefs and impressions about Large Language Models in general. Indeed, the specific chatbot is not the main focus: as they are usually owned by private companies and change rapidly, we stress the importance of researching their main features rather than specific models. ChatGPT, in this sense, only serves as a representant of the broad category of chatbots.

2. A more definite alternative would be, for instance, the so-called stochastic parrot metaphor (cf. Bender et al., 2021), which would tend to frame chatbots environment, however refined, as systems that generate responses based solely on learned patterns from data, without genuine understanding. This and similar perspectives have been critiqued, and several legitimate alternative viewpoints on the cognitive capabilities of machines can be also applied in the context of contemporary chatbot developments (cf., e.g., Turing, 1950; Chalmers, 2019). As a more promising metaphor for educational researchers, that of the extended mind (cf., e.g., Murphy Paul, 2021), in turn tends to underline the collaborative aspects of human-machine interaction, by emphasizing how chatbots and other technology based on contemporary machine learning can serve as an extension of human cognition (cf. also Kersting, 2024). While this metaphor appears promising, further theoretical research would be needed to effectively integrate it within our Meta-Didactical framework and to employ it in praxeological analysis (particularly in view of the fact that a comprehensive understanding of these models against the central categories of mathematics education is still in its infancy). In particular, further theoretical elaboration would be necessary to explore how basic educational normative assumptions could influence (or be influenced) by human-machine interactions.

3. The Italian system comprehends three distinct modalities for teaching mathematics in upper secondary schools. Thus, even if they attended the same course, some trainees were enrolled in the program to qualify as teachers of Mathematics, Mathematics and Physics, or Mathematics in technical institutes, corresponding to Italian “Classi di concorso” A026, A027 and A047. The courses were organized by the institutions CEDIAF (University of Cagliari) and CIFIS (University of Torino).
4. This and the other materials collected—including the excerpts by participants presented in the following section—were originally in Italian and have been translated into English by the authors for the purposes of this paper.
5. The suggested topics for the activity were quadratic, exponential, or trigonometric functions.

REFERENCES

- Anichini, G., Arzarello, F., Ciarrapico, L., & Robutti, O. (Eds.) (2004). *Matematica 2003. La matematica per il cittadino: Attività didattiche e prove di verifica per un nuovo curriculum di Matematica. Ciclo Secondario*. Matteoni Stampatore.
- Beccuti, F. & Robutti O. (2022). Teaching mathematics in today's society: Didactic paradigms, narratives and citizenship. *For the Learning of Mathematics*, 42(2), 29–34
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 Association for Computing Machinery Conference on Fairness, Accountability, and Transparency* (pp. 610–623).
- Buchholtz, N., & Huget, J. (2024). ChatGPT as a reflection tool to promote the lesson planning competencies of pre-service teachers. *Proceedings of the 17th ERME Topic Conference (MEDA 4)*, pp. 129–136). Università di Bari Aldo Moro.
- Chalmers, D. (2019). The singularity and the mind-body problem. *Philosophy & Technology*, 32(4), 595–614.
- Chevallard, Y. (2019). Introducing the anthropological theory of the didactic: An attempt at a principled approach. *Hiroshima Journal of Mathematics Education*, 12, 71–114. <https://doi.org/10.24529/hjme.1205>
- Cusi, A., Robutti, O., Panero, M., Taranto, E., Aldon, G. (2022). Meta-Didactical Transposition.2: The Evolution of a Framework to Analyse Teachers’ Collaborative Work with Researchers in Technological Settings. In: Clark-Wilson, A., Robutti, O., Sinclair, N. (eds) *The Mathematics Teacher in the Digital Era. Mathematics Education in the Digital Era*, vol 16. Springer, Cham. https://doi.org/10.1007/978-3-031-05254-5_14
- De Freitas, E., & Sinclair, N. (2014). *Mathematics and the body: Material entanglements in the classroom*. Cambridge University Press.
- Kersting, M. (2024). *Metaphors in Motion: Navigating the Embodied Landscape of Science Education in the Age of Artificial Intelligence*. Abstract from 1st Specialist Seminar on Embodied Education. Bozen-Bolzano, Italy.
- Murphy Paul, A. (2021). *The extended mind*. Houghton Mifflin Harcourt
- Noster, N., Gerber, S., & Siller, H.-S. (2024). Tasks incorporating the use of ChatGPT in mathematics education—Designed by pre-service teachers. *Proceedings of the 17th ERME Topic Conference (MEDA 4)*, pp. 303–310). Bari, Italy.
- OpenAI, et al. (2023). GPT-4 Technical Report. <https://arxiv.org/abs/2303.08774>.
- Prodromou, T., Robutti, O., & Panero, M. (2018). Making sense out of the emerging complexity inherent in professional development. *Mathematics Education Research Journal*, 30(4), 445–473. <https://doi.org/10.1007/s13394-017-0229-z>

- Sawyer, A. G. (2024). Artificial Intelligence Chatbot as a Mathematics Curriculum Developer: Discovering Preservice Teachers' Overconfidence in ChatGPT. *International Journal on Responsibility*, 7(1). <https://doi.org/10.62365/2576-0955.1106>
- Segal, R., & Biton, Y. (2025). Mathematical problem-posing according to defined educational needs utilizing generative AI as a lever to enhance pre-service mathematics teachers' TPACK. *Proceedings of the Fourteenth Congress of the European Society for Research in Mathematics Education (CERME14)*. Free University of Bozen-Bolzano
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59, 433–460.
- Wiener, N. (1961). *Cybernetics: Or the control and communication in the animal and the machine*. MIT Press.
- Zurru, A. L. (2024). Reverting to the oracle. Anthropological, ethical, and pedagogical questions about artificial intelligence. *Journal of Inclusive Methodology and Technology in Learning and Teaching*, 4(1), Article 1. <https://doi.org/10.32043/jimtl.v4i1.158>