CHATGPT'S PEDAGOGICAL APPROACH IN MATHEMATICS

Laurie Rubel, Shimrit Goidel

University of Haifa

We investigate ChatGPT-4's pedagogical approach by studying mathematics lessons it generates, with respect to school context descriptors. We generated 45 eighth-grade probability lesson plans across three conditions: a baseline prompt, a "Successful School" context, and a "Struggling School" context. Content analysis of the lesson plans and 368 teacher-action verbs suggest that the LLM tends towards direct instruction and shifts its responses to school descriptors. "Successful School" prompts generated lessons with additional mathematics concepts and advanced technology integration, whereas "Struggling School" prompts produced lessons with more monitoring of students' progress, reinforcement activities, and attention to socioemotional support.

INTRODUCTION

Large Language Models (LLMs) have become ubiquitous, and teachers are utilizing LLMs to assist with various aspects of their work, including lesson planning (Pepin et al., 2025). Yet it is often overlooked how these systems are produced by human ideologies and values (Akgun & Greenhow, 2021). This study explores the pedagogical approach of ChatGPT's lesson plans for school mathematics, with an interest in the implicit messages about knowledge, authority, and social relationships that are transmitted through curriculum. We investigate the extent to which school context descriptors guide the LLM to differentiate its pedagogical approach.

Particularly relevant to the current study is Etgar et al.'s (2024) study of financial advice generated by LLMs. They examined how embedded cues in the prompts produce different kinds of financial advice, with recommendations shifting significantly based on subtle gender cues embedded in the prompts. Even when financial circumstances were otherwise identical, the LLM offered more cautious, simplified financial guidance to profiles implicitly signalling women (e.g., nursery school teacher) than it gave to profiles that implicitly signalled men (e.g., construction worker). In the current study, we made use of this strategy, whereby the prompts did not include explicit demographic markers. The results suggest that the LLM detected these implicit signals in the context descriptions and adjusted its output in specific ways.

THEORETICAL FRAMEWORK

Pedagogical approach is Remillard and Kim's (2020) term for a curriculum's messaging about how students are to learn: how they are to engage with mathematics, with one another, and with the teacher and how teachers are to facilitate that learning.

2026. In H. Krzywacki, M. Hähkiöniemi, E. Haataja, M.S. Hannula & P. Hästö (Eds.). *Proceedings of the 49th Conference of the International Group for the Psychology of Mathematics Education: Research Reports*, pp. 1251-1258. PME.

Like Remillard and Kim, we make use of Munter et al.'s (2015) continuum, of direct instruction and dialogic instruction at the two poles. Both models aim to develop procedural and conceptual understanding, regard practice as integral, value mathematically rigorous tasks, and conceptualize the monitoring of student reasoning as fundamental to teaching, with essential differences as follows.

Direct instruction is a teacher-centred model with theoretical foundations in behaviourism and cognitivism. This approach posits that students learn by observing clear teacher-led demonstrations, practicing problems with increasing levels of difficulty, and having any errors readily corrected. Mathematical authority resides primarily with the teacher, disciplinary rules, and the textbook. Group work is optional and if used, is typically used as a configuration for additional practice.

Dialogic instruction, in contrast, frames mathematics learning as participation in disciplinary discourse. Its theoretical foundations are in constructivism and sociocultural theories of learning, particularly the Vygotskian view of knowledge as socially mediated and developed through language. From a dialogic perspective, learning occurs as students persevere through solving unfamiliar problems collaboratively, without teacher specific directiveness about exactly how to solve the problem. Key teacher moves include eliciting student thinking, pressing students for explanations, connecting student contributions to disciplinary ideas, and sequencing student work for public evaluation. Feedback is aimed at advancing student intellectual authority to judge the correctness of their own and others' reasoning.

The direct and dialogic instructional models and tensions between them are entwined with political concerns regarding equity, power, and authority. On one hand, proponents of direct instruction cite its effectiveness particularly for students who have fallen behind, to shore essential, foundational skills. Critics, however, contend that a focus on basic skills and procedures is part of a two-tiered educational system, wherein marginalized youth are taught to follow rules, but privileged youth are given opportunities to develop critical reasoning and argumentation skills. Although dialogic instruction is often championed as the more equitable approach, with its promise to democratize intellectual authority and encourage broad participation, several studies reveal the potential for significant gaps between this promise and classroom reality (e.g., Reinholz et al., 2022). Most importantly, our understanding of dialogic instruction's equity potential remains limited because of how it is implemented far less frequently than direct instruction (O'Connor et al., 2017). A sparser implementation means that the field lacks research knowledge about the equity implications of dialogic instruction.

In sum, the debates about mathematics curriculum and pedagogy constitute part of the training data on which LLMs are built. For one, the limited implementation of dialogic instruction suggests that the data available to train AI systems likely predominantly reflect direct instruction. Furthermore, when LLMs are prompted with contextual cues about different school settings, they may reproduce certain patterns embedded in their

training data. These considerations led us to formulate the following research questions: What is the pedagogical approach of LLM-generated mathematics lesson plans, as determined by how they define the roles for the teacher? How, if at all, does this pedagogical approach shift when prompts include contextual cues specifying different school settings?

RELATED RESEARCH

LLMs are emerging as helpful for mathematics teachers to support lesson planning, potentially as a brainstorming partner or in support of lesson personalization. Early studies of teachers' engagement with LLMs around lesson planning document a range of strategic uses: seeking or refining ideas, clarifying concepts, developing outlines, or getting feedback (Broutin, 2024; Gurl et al., 2024; Yanar & Ergene, 2025). Teachers can strategically leverage LLMs to play the role of teaching assistants who pose targeted reflective questions for the teacher (e.g., Buchholtz & Huget, 2024), to simulate a learner of a certain age and his or her engagement with the lesson's tasks (e.g., Naresh et al., 2024), or to improve specific lesson components (e.g., Biton & Segal, 2024). A particular affordance is how LLMs can generate variations of mathematical problems or activities (e.g., Celik et al., 2022) and personalize contexts for story problems (e.g., Walkington, 2025).

Alongside these affordances, studies also document problems and challenges with LLMs for mathematics education (Şimşek, 2025). Under baseline prompting conditions, LLM-generated mathematical tasks tend to be procedural, with low cognitive demand (Sapkota & Bondurant, 2025). Cameron and Mesiti (2024) found that these lessons follow a boilerplate structure consisting of an introduction of key procedures or concepts, demonstration of key skills with examples, and then additional problems for students to practice, like the direct instruction model, with limited support for dialogic mathematical practices.

METHODS

We selected ChatGPT-4 because of its prominence in current discussions regarding GenAI in education. We used temporary chat sessions with the memory feature disabled for each prompt. First, we designed a baseline default prompt that positioned the AI as an educational professional and specified key parameters:

> You are a professional in the field of mathematics education, possessing extensive teaching experience and expertise. You have the ability to integrate educational theory with practice, create instructive and actionable lesson plans to promote effective student learning. Please write a mathematics teaching unit about probability for Grade 8 learners, on grade level. The unit should include five lessons. Provide highly detailed lesson plans, including an explanation about how long every part should take and be handled by the teacher and the expected role of the students. The classroom has a computer and a projector.

We developed two additional prompts. The "Successful School" condition repeated the default prompt with this sentence at the end: "The students attend a successful school

where state math exam scores are high, and all the graduates continue to four-year colleges." The “Struggling School” condition repeated the default prompt but culminated with: "The students attend a struggling school where state math exam scores are low, and most graduates do not continue to four-year colleges." We ran each prompt three times, always in independent sessions, to yield nine distinct unit plans and resulted in a dataset of n = 45 individual lesson plans.

We first conducted an iterative, comparative reading of the units to identify emergent themes. Next, we focused on the pedagogical actions indicated for the teacher by identifying all verbs that specified what the teacher was directed to do, which produced 368 verbs (119 in the Default condition, 108 in the “Successful School” condition, and 141 in the “Struggling School” condition). We began with initial open coding on this set of and then iteratively refined these emergent codes into final categories, which we structured according to our main research goal of locating the pedagogy on a continuum between direct instruction and dialogic instruction.

The five categories of teacher actions are: 1) Show or Explain, 2) Discuss or Elicit, 3) Monitor, Support, or Assess, 4) Manage Logistics, and 5) Sociodimensional Aspects of Classroom. The second author independently coded a representative subset of the verbs (n = 176, approximately 48% of the set of teacher verbs). We achieved an initial inter-coder agreement of 86%. We then resolved discrepancies through collaborative discussion and applied the final coding scheme to the complete dataset.

RESULTS

The basic mathematical trajectory was roughly similar across all the three sets of lessons. Both the "Default" and "Struggling" unit plans began with basic probability, moved to differentiating between theoretical and experimental probability, and concluded with compound events and simulations. However, two of the three units generated by the ”Successful School” prompt included additional mathematics topics that did not appear in the other units: Addition Rules for non-mutually exclusive events as part of the attention to compound events, formal set notation to represent the compound event formulas, and large sample size, digital simulations to support the understanding of experimental and theoretical probabilities.

An additional divergence is in terms of how the lessons indicated uses of technology. In the "Default" as well as the "Struggling School" plans, technology was relegated primarily to a demonstration role; the teacher was mostly directed to use the projector to "display definitions" or "show examples," or create review games using presentation technology sites. In contrast, the "Successful School" plans included directions for uses of technology for computation and data investigation, suggesting advanced platforms like GeoGebra and PhET simulations. There was no mention in the lesson plans of ways of integrating GenAI technologies into the learning of eighth-grade probability.

The Default lessons contained 31 distinct verbs for teacher actions, the “Successful School” lessons 39, and the “Struggling School” lessons 45. Sixteen of these verbs

were common to all three prompt conditions, which we view as a baseline of teacher actions, shown at the top of Figure 1. Some of these common verbs can be seen as aligned with direct instruction and others can be seen as aligned with dialogic instruction.

Figure 1. Common and Unique Verbs for Teaching Actions

Verbs Common to All Three Prompt Categories:		
Ask, circulate, define, demonstrate, discuss, display, elicit, facilitate, monitor, pose, provide, recap, review, show, use, summarize		
Unique to Default Prompt: Calculate, formalize, have them, practice, project	Unique to “Successful school” Prompt	Unique to “Struggling School” Prompt
	Administer, assist, challenge, highlight, illustrate, transition, troubleshoot	Address, assess, build (a supportive atmosphere), connect, celebrate, encourage, invite, reinforce, spark, wrap-up

However, looking at the verbs that were unique to each prompt condition suggests that the introduction of school context labels seems to have shifted how the LLM set forth the teacher’s role. As shown in Figure 1, the "Struggling School” condition generated a unique cluster of verbs centered on socioemotional support, which we elaborate with frequency analysis below. Examples that appeared in these lessons include; "build a supportive atmosphere," “spark curiosity,” "celebrate learning progress," and "provide encouragement."

Next, we conducted a frequency analysis of verbs according to these categories: 1) Show or Explain, 2) Discuss or Elicit, 3) Monitor, Support, or Assess, 4) Manage Logistics, and 5) Sociodimensional Aspects of Classroom. We display these results in Table 1. The Show or Explain category was the most frequent across all prompt conditions. The difference in proportion between Show or Explain, hallmarks of direct instruction, and Discuss or Elicit, hallmarks of dialogic instruction, was widest in the Default prompt lessons, but in all cases were at least 1.5 times more. We conducted a chi-square test of independence to confirm a statistically significant association between the prompt type and the verb categories [χ^2 (8, N = 368) = 28.68, $p < .001$].

One main difference is the emergence of the socioemotional support verbs in the “Struggling Schools” lessons. Eleven, or 8%, of the teacher verbs in the “Struggling School” were in this category, in comparison with zero instances in the default conduction and only one in the “Successful School” condition. A second point of contrast in how both school context prompts seemed to direct the model more towards

assessment: these verbs occurred far more often in those two sets (18% in each) of lessons than in the default lessons (5%).

Table 1. Frequency of Teacher Verbs By Prompt Condition

Verb Category	Default	“Successful School”	“Struggling School”
Show or Explain	69 (58%)	51 (47%)	61 (43%)
Discuss or Elicit	40 (34%)	31 (29%)	39 (28%)
Manage Logistics	4 (3%)	6 (6%)	5 (4%)
Monitor, Support or Assess	6 (5%)	19 (18%)	25 (18%)
Socioemotional Aspects	0 (0%)	1 (1%)	11 (8%)
Total	119 (100%)	108 (100%)	141(100)

DISCUSSION

This study examined the pedagogical approaches of ChatGPT-4 generated mathematics lesson plans and explored the effects of introducing contextual cues about school settings. The teacher verbs in the generated lessons leaned towards the direct instruction model, as characterized by teacher-centered transmission of knowledge, explicit modeling, and structured practice sequences. This finding corroborates and extends Cameron and Mesiti's (2024) findings.

We also explored how pedagogical approaches shift when prompts include different school context cues. These results show a pattern of differentiation that could be suggestive of historical patterns of educational stratification. Even though all three prompts indicated that students are on-grade level, two of the three "Successful School" units included advanced mathematics topics absent from other prompt conditions. Furthermore, in response to the "Struggling School" descriptor, ChatGPT-4 generated lessons with more monitoring and assessment and a nearly exclusive presence of socioemotional learning verbs. This differentiation could suggest that the LLM is reflecting a false tension in mathematics education that perhaps exists implicitly in the training data, between mathematical rigor and socioemotional care.

A primary limitation being its relatively small dataset (45 lessons), only one mathematics topic, a single LLM, and around one base prompt. A second limitation is that we treated ChatGPT's responses as static, one-off outputs, whereas in practice, teachers could refine outputs through iterative prompting. Thus, our findings capture the model’s initial pedagogical approach rather than the potentially more nuanced plans that could emerge through a teacher’s sustained interaction with it.

CONCLUSIONS

Future research should build on this study's preliminary findings to explore lesson planning of GenAI across lesson topics, grade levels, and AI models. LLM outputs are determined by prompting strategies, so supporting the capacity of teachers to develop prompts embedded with explicit pedagogical directions seems. Given that this study suggests that contextual labels might guide the LLM towards different outputs, teachers need strategies to recognize and counter these and other potential biases in their use of GenAI. Finally, future research could employ think-aloud protocols or observational studies to better understand how the field can support teachers in learning to identify and modify potentially biased outputs through sustained, extended prompting that guides the LLM in terms of both pedagogical and equity considerations.

References

Akgun, S., Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. AI Ethics, 2, 431–440. https://doi.org/10.1007/s43681-021-00096-7

Biton, Y., & Segal, R. (2025). Learning to craft and critically evaluate prompts: The role of generative AI (ChatGPT) in enhancing pre-service mathematics teachers' TPACK and problem-posing skills. *International Journal of Education in Mathematics, Science, and Technology*, *13*(1), 202–223. https://doi.org/10.46328/ijemst.4654

Broutin, M. S. T. (2024). Exploring mathematics teacher candidates' instrumentation process of generative artificial intelligence for developing lesson plans. *Yükseköğretim Dergisi, 14*(1), 165–176. https://doi.org/10.53478/yuksekogretim.1347061

Buchholtz, N., & Huget, J. (2024). ChatGPT as a reflection tool to promote the lesson planning competencies of pre-service teachers. In E. Faggiano, A. Clark-Wilson, M. Tabach, & H.-G. Weigand (Eds.), *Proceedings of the 17th ERME Topic Conference MEDA 4* (pp. 129–136). University of Bari Aldo Moro.

Cameron, S., & Mesiti, C. (2024). What kind of mathematics teacher is ChatGPT? Identifying the pedagogical practices referenced by generative AI tools when preparing lesson plans. In J. Višňovská, E. Ross, & S. Getenet (Eds.), Surfing the waves of mathematics education: Proceedings of the 46th annual conference of the Mathematics Education Research Group of Australasia (pp. 135–142). MERGA.

Diversity in Mathematics Education Center for Learning and Teaching. (2007). Culture, race, power and mathematics education. In F. K. Lester (Ed.), Second handbook of research on mathematics teaching and learning (pp. 405–433). Information Age Publishing.

Etgar, R., Guttmann, R., Taub-Tabib, H., & Shmueli, E. (2024). *Implicit bias in LLMs: Bias in financial advice based on implied gender* (SSRN Scholarly Paper No. 4930361). https://doi.org/10.2139/ssrn.4930361

Gurl, T., Markinson, M., & Artzt, A. F. (2025). Using ChatGPT as a lesson planning assistant in secondary mathematics teacher education. *Digital Experiences in Mathematics Education*. https://doi.org/10.1007/s40751-024-00162-9

Munter, C., Stein, M. K., & Smith, M. S. (2015). Dialogic and direct instruction: Two distinct models of mathematics instruction and the debate(s) surrounding them. *Teachers College Record*, *117*(11), 1–32. https://doi.org/10.1177/016146811511701102

Naresh, N., Yilmaz, Y., & Galanti, T. (2024). Leveraging the potential of AI as a partner in teaching. *AMTE Connections* https://amte.net/connections/2024/05/connections-thematic-articles-artificial-intelligence-mathematics-teacher

O'Connor, C., Michaels, S., Chapin, S., Harbaugh, A.G. (2017). The silent and the vocal: Participation and learning in whole-class discussion. *Learning and Instruction, 48*, 5-13. https://doi.org/10.1016/j.learninstruc.2016.11.003

Pepin, B., Buchholtz, N. & Salinas-Hernández, U. (2025). A scoping survey of ChatGPT in mathematics education. *Digital Experiences in Mathematics Education*, 11, 9–41. https://doi.org/10.1007/s40751-025-00172-1

Reinholz, D., Johnson, E., Andrews-Larson, C., Stone-Johnstone, A., Smith, J., Mullins, B., Fortune, N., Keene, K., & Shah, N. (2022). When active learning is inequitable: Women's participation predicts gender inequities in mathematical performance. *Journal for Research in Mathematics Education*, *53*(3), 204–226. https://doi.org/10.5951/jresematheduc-2020-0143

Remillard, J. T. & Kim, O. K. (2020). *Elementary mathematics curriculum materials: Designs for student learning and teacher enactment*. Springer. https://doi.org/10.1007/978-3-030-38588-0_4

Sapkota, B. & Bondurant, L. (2024). Assessing concepts, procedures, and cognitive demand of ChatGPT-generated mathematical tasks. *International Journal of Technology in Education* (IJTE), 7(2), 218-238. https://doi.org/10.46328/ijte.677

Şimşek, N. (2025). Integration of ChatGPT in mathematical story-focused 5E lesson planning: Teachers and pre-service teachers' interactions with ChatGPT. *Education and Information Technologies*, *30*, 11391–11462. https://doi.org/10.1007/s10639-024-13258-x

Walkington, C. (2025). The implications of generative artificial intelligence for mathematics education. *School Science and Mathematics*, 1–10. https://doi.org/10.1111/ssm.18356

Yanar, B., & Ergene, Ö. (2025). Preservice teachers' views on using ChatGPT for mathematics lesson planning: Opportunities and threats. *Hacettepe Journal of Education*, *40*(1), 199–219. https://doi.org/10.16896/HUJE.2024127226