International Electronic Journal of Mathematics Education
2026, 21(4), em0910
e-ISSN: 1306-3030
https://www.iejme.com

MODESTUM

Research Article

OPEN ACCESS

Prompting, evaluating, and adapting AI-generated mathematical tasks: Secondary preservice teachers' curricular decision-making with ChatGPT

Bima Sapkota [1*], Lili Zhou [2]

[1] Department of Teaching and Learning, University of Kentucky, Lexington, KY, 40508, UNITED STATES
[2] Department of Curriculum & Instruction, California State University, Los Angeles, CA, 90032, UNITED STATES
*Corresponding Author: bsapkota@uky.edu

Citation: Sapkota, B., & Zhou, L. (2026). Prompting, evaluating, and adapting AI-generated mathematical tasks: Secondary preservice teachers' curricular decision-making with ChatGPT. *International Electronic Journal of Mathematics Education, 21*(4), em0910. https://doi.org/10.29333/iejme/19970

ARTICLE INFO

Received: 08 Jul. 2026

Accepted: 23 Sep. 2026

ABSTRACT

Despite emerging research on Artificial Intelligence in mathematics education, studies on how secondary preservice mathematics teachers make curricular decisions when using Artificial Intelligence tools (e.g., ChatGPT) to generate mathematical tasks are underdeveloped. Building on prior research that shows how teachers consider factors related to students' needs, valid mathematics disciplinary ideas, contextual and institutional factors, and their own characteristics, we unpack how these factors interacted when secondary preservice mathematics teachers reflected on their prompts used to generate mathematical tasks using ChatGPT and their evaluation of ChatGPT-generated tasks. We also examined the potentials and challenges teachers identified in using ChatGPT for generating tasks. Preservice mathematics teachers highlighted the need for adjusting the difficulty level of ChatGPT-generated tasks, expressing concerns that the tasks at times did not (a) appropriately address students' comprehension, (b) align with given standards/goals, and (c) present mathematics ideas coherently. They also shared formulating right prompts to generate appropriately challenging tasks was more time-consuming than writing their own tasks. By presenting examples of the factors that secondary preservice mathematics teachers considered when evaluating ChatGPT-generated mathematical tasks, we suggest strategies to help teacher and educators use Artificial Intelligence tools more efficiently in their curricular decision-making. We discuss how teachers' engagement with Artificial Intelligence in curating curricular resources repositions teachers' curricular-decision making.

Keywords: ChatGPT, curricular decisions, preservice mathematics teachers, prompt engineering

INTRODUCTION

Recent research has investigated opportunities, challenges, and ethical considerations introduced by Artificial Intelligence (AI) in K-12 contexts and higher education (e.g., Farrokhnia et al., 2024). For example, researchers have explored: teachers' and students' perceptions of AI as a teaching and learning tool (e.g., Chounta et al., 2022), ethical and practical challenges of using AI (e.g., Chiu et al., 2023; Shneiderman, 2020), and strategies to enhance human-centered AI (e.g., Herrmann & Pfeiffer, 2021). More recently, research has focused on the potential and limitations of ChatGPT for teaching and learning specific disciplines, such as how ChatGPT could support generating mathematics curricular resources (e.g., Aqazade et al., 2025; Sapkota & Bondurant, 2024) and solutions to math word problems (e.g., Shakarian et al., 2023). Within this research, the importance of using a series of right prompts to generate relevant and appropriate curricular resources (we define this process *prompt engineering* and use throughout the manuscript) has received greater attention (e.g., Alneyadi & Wardat, 2023) and point to the need for critically evaluating AI-generated products as these models can be unreliable in generating high-quality mathematics curricular resources (e.g., Berryhill et al., 2024; Sapkota & Bondurant, 2024)

A recent literature review conducted by Pepin and colleagues (2025) shows research in mathematics education and AI has focused on the use of AI tools such as ChatGPT for problem solving, lesson planning, assessment and task design, personalized learning, and teachers' and students' attitudes toward using AI. Within these areas, empirical studies have begun to examine how elementary and secondary preservice mathematics teachers (PMTs) engage with and evaluate AI-generated curricular resources (e.g., Aga et al., 2024; Gurl et al., 2025; Sawyer, 2024). These studies suggest PMTs primarily attend to pedagogical aspects of AI-generated curricular resources. Yet, research on teachers' selection and use of curricular resources generally suggests such decisions involve a lot of complexity. In particular, teachers' curricular decision making involves attending to multiple factors,

Copyright © 2026 by Author/s and Licensed by Modestum Open DOO, Serbia. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

including students' needs, institutional rules and regulations, valid mathematical disciplinary practices (Herbst & Chazan, 2011; Sapkota et al., 2026a), and teachers' own beliefs (Arbaugh et al., 2006). For example, Henningsen and Stein (1997) found teachers modified cognitively demanding tasks into less cognitively demanding ones to better address students' needs. Similarly, Webel et al. (2023) found an elementary teacher prioritized students' needs over ambitious teaching practices during instruction.

Thus, while existing research provides insight into the features that teachers attend to when evaluating AI-generated curricular resources, less is known about how PMTs consider this complexity around selecting and evaluating AI-generated tasks. Specifically, which of the factors (students, mathematics, pedagogy, and contexts) are more evident when PMTs reflect on the use of AI to generate mathematical tasks is still being explored. Less is known about the affordances and challenges PMTs perceive as they reflect on the process of using AI for mathematics task generation, including prompt generation as well as evaluating and revising AI-generated outputs.

To build on this research, we investigated how 23 PMTs enrolled in a secondary mathematics technology-focused course at a U.S. university considered these factors as they engaged in and reflected on the prompt-engineering and associated task-generation processes using ChatGPT. We looked at PMTs' written reflections on their prompt construction and their evaluation of ChatGPT-generated tasks collected in a technology-focused mathematics education course to investigate this question.

The PMTs in our study had completed substantial mathematics coursework as part of their preparation for mathematics teaching certification. In addition, as part of the technology-focused course in which the data was collected, PMTs were explicitly introduced to the use of AI for generating curricular resources and were asked to critically examine the outputs produced by AI. This context allowed us to examine how PMTs' mathematical and pedagogical preparation, together with explicit attention to AI use, may shape how they evaluate and make decisions about AI-generated mathematical tasks. Such an examination extends prior research suggesting some PMTs do not always critically evaluate AI-generated outputs (Mahlaba, 2026) and provides insight into the ways PMTs may consider multiple dimensions of curricular decision making, both critically and productively, when using AI to generate mathematical tasks. Even though our focus in this paper is not to offer a shift in the theory related to teachers' curricular decision-making, unpacking these dimensions will be an initial stage in examining whether PMTs' considerations of these factors have shifted, if at all, in the context of AI for the generation of curricular resources such as mathematical tasks.

We therefore asked:

1. What factors did PMTs consider when they reflected on their prompt engineering for task-generation?
2. What potential and challenges did the secondary PMTs identify in the use of ChatGPT for generating Mathematical Tasks?

Investigating these questions allowed us to highlight how, if at all, teachers' curricular decision-making shifts in relation to prompt engineering and curricular decisions in the context of AI. Given many U.S. teachers report using AI, understanding how PMTs reflect on the use of AI for curricular resource generation has implications for teacher preparation.

REVIEW OF THE RELEVANT LITERATURE: AI AS A RESOURCE TO GENERATE CURRICULAR RESOURCES

The integration of AI (e.g., ChatGPT) into mathematics teaching and learning has shown potentials for generating curricular resources. In this study, curricular resources include textbooks, mathematical tasks, lesson plans, and any other instructional materials teachers use for the purpose of mathematics instruction. When curating and implementing curricular resources, teachers engage in processes of negotiation and meaning making "among different actors in society and the education system, such as politicians, national education authorities, teachers, etc." (Wahlström, 2023, p. 260–261). Thus, teachers' curricular curation and implementation are influenced by contextual factors, such as the emergence of online curricular resources (Sapkota et al., 2024) and, potentially, the emergence of AI (Walton Family Foundation & Gallup, 2025).

Even though teachers' use of ChatGPT for mathematics instruction is lower than in other subject areas, the overall use of ChatGPT is increasing (Langreo, 2025). One reason for the increasing use of AI is its nonjudgmental and always-available nature, which makes it a low-stakes option for generating mathematical tasks (Walkington & Bainbridge, 2025). Researchers and educators have positioned AI as a thought partner for both in-service and preservice teachers, which offers greater flexibility in generating mathematical tasks (e.g., Pepin et al., 2025; Sapkota & Bondurant, 2024). In this context, what the use of ChatGPT-generated curricular resources, particularly mathematical tasks, entails, and what influences teachers' decisions to work with ChatGPT-generated tasks, are questions worth investigating.

A consistent finding across studies is that generative AI can serve as a productive brainstorming partner, assisting teachers in creating a variety of curricular resources, including mathematical tasks (Berryhill et al., 2024; Sapkota & Bondurant, 2024). For example, teachers reported AI-generated prompts and activities were helpful in incorporating real-world scenarios, using multiple representations, and offering opportunities for peer interaction (Walkington & Bainbridge, 2025). Some studies also showed engaging with AI encouraged teachers to reflect more deeply on how tasks might be adapted for diverse learners and classroom settings (Biton & Segal, 2025; Kim et al., 2024). For many PMTs, interacting with ChatGPT to generate curricular resources and solutions for mathematical tasks has provided a space for refining problem-posing skills, deepening content understanding, and engaging in pedagogical reasoning (Biton & Segal, 2025; Kim et al., 2024). An AI-based chatbot designed to simulate authentic classroom interactions helped PMTs practice and refine their responsive teaching and questioning skills (Lee & Yeo, 2022). Teachers in their study reported ChatGPT was helpful in tailoring lessons to address student needs, enabling them to make content more relevant and personalized.

However, these potentials of AI are often contingent on teachers' ability to critically assess and revise AI's suggestions (Aga et al., 2024; Kim et al., 2024). ChatGPT-generated outputs may involve mathematical inaccuracies, developmentally inappropriate content, and outputs that are overly generic (Aga et al., 2024; Sapkota & Bondurant, 2024; Walkington & Bainbridge, 2025; Wardat et al., 2023). As a result, AI-generated tasks cannot be ready-to use resources as they often require substantial refinement and contextualization prior to implementation (Biton & Segal, 2025). Sapkota and Bondurant (2024) found some ChatGPT-generated tasks were vague and presented incorrect information. Sawyer and Aga (2024) also found ChatGPT was not a reliable source for task generation as it generated inaccurate and developmentally inappropriate problems. These findings underscore the importance of teachers' understanding of prompt engineering techniques and learning how to identify which elements of AI-generated tasks need modification in order to assess accuracy, appropriateness, and efficiency of AI-generated materials (Aga et al., 2024; Biton & Segal, 2025; Kim et al., 2024; Walkington & Bainbridge, 2025).

In many cases, ChatGPT only responded effectively after several rounds of deliberate and specific prompting (which we call prompt engineering). For example, a preservice teacher in Biton and Segal's (2025) study engaged in a series of prompt revisions to improve task clarity, increase cognitive demand, link problems to real-world contexts, and anticipate potential student misunderstandings. Walkington and Bainbridge (2025) noted modifying ChatGPT-generated outputs required significant preparation time to ensure instructional quality and correctness.

This paradox wherein AI simultaneously reduces and increases teachers' workload highlights that PMTs must learn how to craft effective prompts and critically assess AI-generated responses. Situating in this paradox, our research presents examples of the factors that PMTs consider when evaluating ChatGPT-generated mathematical tasks and suggests strategies to help them use AI tools more efficiently in their curricular decision-making processes.

FRAMEWORK: TEACHERS' CURRICULAR DECISION-MAKING

Drawing on existing perspectives and frameworks of teachers' curricular decision-making, we examine how factors influencing curricular decision-making become visible as PMTs generate mathematical tasks using AI and evaluate AI-generated tasks. Teachers make complex curricular decisions throughout the planning and implementation process, including decisions about which curricular resources to use or adapt and how to enact those resources (e.g., Griffin, 1990). These curricular decision-making processes occur across different phases of engagement with written, intended, and enacted curricular resources (Remillard & Heck, 2014) and are influenced by various contextual factors. Teachers' considerations of different stakeholders and contextual factors during decision-making processes have been conceptualized in various ways (e.g., Dingman et al., 2021; Herbst & Chazan, 2011). One such framework is the theory of practical rationality, which describes how teachers' curricular decisions may draw on various, and sometimes competing, considerations related to students, society, disciplinary practices of mathematics, and institutional contexts (Chazan et al., 2016; Herbst & Chazan, 2011). Herbst and Chazan (2011) conceptualized these considerations as individual, institutional, disciplinary, and interpersonal obligations.

According to individual obligations, teachers draw on students' needs, backgrounds, experiences, and engagement. The institutional obligation reflects the requirements and constraints imposed by schools, districts, and states. The disciplinary obligation foregrounds the mathematics represented in curricular resources. The interpersonal obligation concerns how curricular resources enable teachers to support diverse students within a shared classroom environment. Teachers consider whether resources allow flexibility in students' pace and schedules, support the use of varied strategies and resources, enable teachers to negotiate different levels of support, and create opportunities for collaboration and peer-to-peer scaffolding. **Table 1** provides more elaboration of each obligation.

In addition to these practical constraints, teacher characteristics, including teachers' beliefs, experiences, and views of appropriate learning and instruction also influence teachers' curricular selection and implementation. For example, Breyfogle et al. (2010) describe how teachers' beliefs about learning paths for students may influence the resources they trust and use. Arbaugh et al. (2006) found teachers' beliefs about students' ability to learn and teachers' ability to create appropriate learning opportunities influenced how they implemented curricular resources. Wilkins (2008) found teachers' beliefs and attitudes toward inquiry-based curricula influenced their instructional practices. Thus, teacher characteristics provide another perspective for understanding what teachers prioritize when selecting, adapting, or generating mathematical tasks. In our study, we did not specifically ask PMTs to describe their beliefs about AI. Rather, we examined whether and how their reflections revealed their beliefs, experiences, and judgments about students, mathematical tasks, student learning opportunities, and effective mathematics instruction.

Even though these frameworks about teacher curricular decisions do not specifically take into considerations what happens during PMTs' interactions with AI and describe the complexities and various factors involved in teachers' curricular decision-making more generally, they provide us insights on different factors teachers may consider when working with curriculum and this study looks at how PMTs would weigh in those factors. Across these perspectives, teachers' exploration and use of curricular resources is also context-specific (Dietiker et al., 2018; Remillard, 2013). For example, teachers increasingly curated materials from online resources, particularly during the COVID-19 pandemic when many relied on freely available online resources (Sapkota et al., 2026b). More recently, teachers have begun turning to large language model (LLM)-based platforms to curate curricular materials and have reported AI can be useful for personalizing content (Holmes, 2022; Sharma, 2025). The use of AI-generated materials has also expanded rapidly; almost 60% of teachers reported using AI for various purposes, including supporting student engagement, problem solving, critical thinking, and independent thinking (Walton Family Foundation & Gallup, 2025). This

Table 1. Aspects of curricular decisions

Aspect	Relevant conceptualization in prior literature	How we conceptualize the aspect in this study
Students	Individual and interpersonal obligations within the theory of practical rationality emphasize teachers' consideration of students' needs, experiences, engagement, and relationships with others in the classroom (Herbst & Chazan, 2011).	PMTs' considerations of students' prior knowledge, backgrounds, experiences, needs, engagement, and access to curricular resources, including the appropriateness, accessibility, level of support, opportunities for collaboration, and flexibility of AI-generated tasks for diverse learners.
Discipline	The disciplinary obligation within practical rationality emphasizes teachers' responsibility to represent and support valid mathematical knowledge and practices (Chazan et al., 2016; Herbst & Chazan, 2011).	PMTs' considerations of the mathematical content and practices represented in AI-generated tasks, including mathematical accuracy and coherence, appropriateness of content, connections to prior learning, opportunities for review and practice, higher-order thinking, problem solving, multiple representations, and real-world connections.
Instructional contexts	The institutional obligation within practical rationality recognizes that teachers' decisions are shaped by institutional requirements and constraints (Herbst & Chazan, 2011).	PMTs' considerations of standards alignment, instructional goals, grade-level appropriateness, available time, curriculum and assessment requirements, and the fit of AI-generated tasks within a particular lesson or unit.
Teacher characteristics	Research on curricular decision-making also highlights the role of teacher characteristics, including teachers' beliefs, experiences, and views of appropriate learning and instruction (e.g., Breyfogle et al., 2010).	PMTs' considerations of their pedagogical beliefs, experiences, ideologies, and judgments about effective mathematics teaching and appropriate curricular resources, including preferences for particular approaches or representations, judgments of task quality, willingness to adapt AI-generated tasks, and perceptions of AI's usefulness and limitations.

emerging curricular context raises questions about how teachers' existing considerations in curricular decision-making and their teacher characteristics play out when AI becomes involved in generating and evaluating curricular resources.

Informed by these perspectives and grounded in the specific context of PMTs using AI for curricular purposes, four aspects of PMTs' curricular decision-making were identified: (1) students, (2) the discipline, (3) instructional contexts, and (4) teacher characteristics. The first three aspects draw on considerations identified across the existing literature, including the theory of practical rationality, while teacher characteristics draws on research highlighting the role of teachers' beliefs, experiences, and views of learning and instruction. these four aspects were used as a lens for examining how PMTs' curricular considerations become visible as they generate mathematical tasks using AI and evaluate the characteristics of AI-generated tasks. **Table 1** summarizes how each aspect was identified from the literature and operationalized in the study. The study examines PMTs' curricular decisions as they reflect on the process of generating mathematical tasks and their associated prompt-engineering techniques.

Identifying these four aspects is not intended to propose a new theory or framework or to suggest that teachers' curricular decision-making can be reduced to four distinct categories. Rather, these aspects were used as an analytic lens to connect this study to existing research on curricular decision-making and to examine the extent to which considerations identified in prior literature are evident when PMTs evaluate AI-generated mathematical tasks. These aspects allowed us to examine whether and how these considerations appear in an emerging curricular context in which AI is involved in generating and evaluating curricular resources. This examination is much needed in the context of rapidly developing curricular landscape created by the use of AI (Karataş et al., 2024) because research on how teachers make curricular decisions when interacting with AI is still being explored. Examining how existing considerations play out in this context can help us understand whether current conceptualizations of curricular decision-making adequately capture teachers' interactions with AI or whether additional dimensions may be needed to account for their evolving curricular practices. In the following subsection, we synthesized literature to discuss how some of these factors may have appeared in the literature related to AI.

Teachers' Curricular Decision-Making in the Context of AI

Gurl et al. (2025) studied secondary PMTs' use of ChatGPT for lesson planning. They found the PMTs critically examined that ChatGPT lacked authentic student work and generated teacher-centered lesson plans. However, they did not point out the wrong mathematical problems; instead they referred those solutions as alternative approaches to problems. This shows PMTs paid attention to some pedagogical approaches but did not attend to mathematical accuracy. This led us to investigate what aspects PMTs pay attention to when they are not given with any framework and are only asked to evaluate the mathematical tasks generated by ChatGPT.

Aga and colleagues (2024) found secondary PMTs critically reflected on ChatGPT-generated lesson plans and shared that lesson plans were vague and generic. The authors suggested the PMTs considered themselves capable of analyzing ChatGPT-generated lesson plans. Sawyer (2024) shared that elementary preservice teachers were not critical enough to evaluate the quality of ChatGPT-generated mathematics curricular resources, pointing to the need for learning systematic prompt-engineering techniques. While these studies provided valuable information about how preservice teachers may evaluate curricular resources generated by AI, they may point to secondary teachers feeling more confident in critically examining AI-generated curricular resources. However, which factors PMTs consider most and how those factors relate to their attention to students, mathematics content, and other contextual aspects remains underdeveloped. Our research addresses this by unpacking the various factors PMTs considered when evaluating ChatGPT-generated prompts, as well as reflecting on opportunities and challenges related to the use of AI.

We note that the extent to which teachers can independently select, adapt, and implement curricular resources varies across educational systems (Organisation for Economic Co-operation and Development [OECD], 2024). In some countries, teachers work within highly prescribed curricula and may have limited freedom and autonomy to modify mandated textbooks and/or other instructional materials. Thus, assumptions about teachers' curricular autonomy cannot be generalized across contexts and our intent is not to claim teachers in every country use AI to generate mathematical tasks. Our study was conducted in the United States, where teachers' curricular decision-making is shaped by state standards, district and school requirements, and mandated curricular resources, but teachers have flexibility to supplement or adapt those resources (Sapkota et al., 2026b) such as teachers may have opportunities to select or design mathematical tasks, provided that these tasks are consistent with given standards and other mandates. This context is relevant to our study because AI tools such as ChatGPT may be used to generate or adapt mathematical tasks that teachers can potentially incorporate into instruction. We therefore interpret PMTs' evaluations of ChatGPT-generated tasks in relation to the U.S. educational context and do not assume that the factors they attend to would be the same in educational systems with more tightly prescribed curricula.

METHODS

Context and Participants

We framed this study as a single case-study design (Yin, 2009). Twenty-three secondary PMTs enrolled in a semester long technology-focused mathematics education course in the mathematics department of a SouthEast U.S. university participated in this study. The half of the semester of the course, we introduced specific non-adaptive technology such as Desmos to teach mathematics and about the half of the course was focused on AI. When focusing on the tasks with AI, we introduced tasks related to proportionality. The activities related to this study are given in **Table 2**.

Table 2. Taks and assignments

Tasks	Types of Assignment	Description of Assignment
PMTs analyzed five ChatGPT-generated tasks on proportionality by identifying strengths, weaknesses, and ways to improve those tasks	Group work, in-class	PMTs' written evaluations of the characteristics of mathematical tasks Note: at this point, PMTs did not know the prompt used. The instructor generated tasks using prompts. They were not asked to describe prompt modifications. Instruction included a mini-lesson on proportionality and a reading on task characteristics (Sapkota, 2022).
PMTs generated and revised two proportionality tasks based on provided learning and performance goals. They evaluated and compared original and revised tasks.	Group work, in-class	Two versions of PMT-generated tasks and associated evaluations.
PMTs analyzed five proportionality tasks generated by ChatGPT and created two alternative prompts to improve task quality.	Individual work in remote setting	PMTs' evaluations of tasks and their alternative prompts.
PMTs completed a survey on their use of ChatGPT, including how they modified tasks, challenges encountered, and strategies to optimize ChatGPT for task generation.	Individual work in remote setting	Survey responses on ChatGPT use, challenges, and plans for future integration in practice. This is the data set used for this study

Note: The data used for this study and the associated activity from which the data came from is highlighted in the Grey Row.

All the PMTs were considered as one case. For the first research question, we examined all the factors that the PMTs collectively considered when evaluating ChatGPT-generated mathematical tasks and their prompts. For the second research question, we considered all the PMTs' descriptions of the affordnances and limitations of using ChatGPT to generate tasks. During the semester of data collection, all PMTs were enrolled in several mathematics content and methods courses, and some of them were doing student teaching, while others were in the semester prior to their student teaching. Six PMTs said they had never used AI prior to this class; the rest of them shared they had some experience using AI prior to this class for a variety of work, including generating lesson plan drafts, using it as a resource for homework prior to this course.

Data and Analytic Methods

PMTs' responses were collected using Google Forms and exported to an Excel spreadsheet, which was used to organize and manage the coding process. We used both deductive and inductive coding methods (Saldaña, 2021) and analyzed the data in two rounds. In the first round, the first author first deductively examined PMTs' responses for evidence of the four aspects identified in the literature: *Discipline*, *Students*, *Curricular Contexts*, and *Teacher Characteristics*. The concepts of these four aspects were used from the literature but the exact descriptors in **Table 3** were developed inductively from PMTs' responses. Thus, although the literature informed the identification of the four broad factors, the specific descriptors within each factor were generated from participants' responses inductively.

Table 3. Factors PMTs' considered

Factors PMTs' Considered	No of PMTs	How Many times has it appeared
Appropriately challenging	23	41
Alignment with goals and standards	19	33
Grade-level appropriate	17	19
Multiple representations	13	14
Variation in the nature of the problem	12	13
Students' prior knowledge	13	13
Contextual and connected with real-life applications	9	11
Accurate mathematical concepts, procedures, and representations	9	10
Use of visuals and illustrations	7	8
Clear and detailed problem-solving instructions	7	7
Multiple approaches to solve problems	6	6
Logical progressions of tasks/questions	5	5
Encouraging critical thinking	3	5
Available time in classroom	2	3

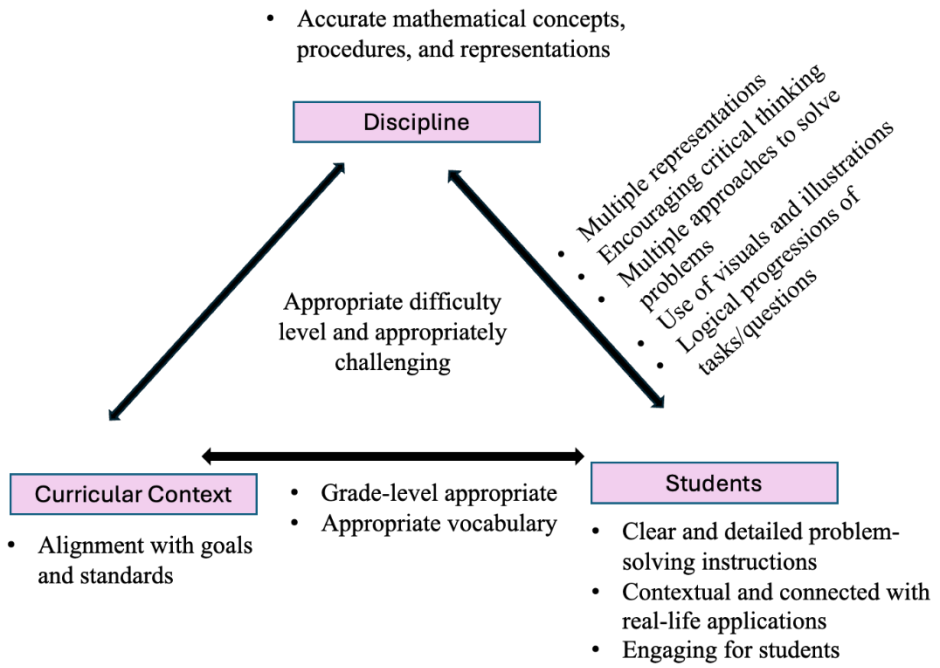


Figure 1.PMTs' considerations of mathematics, students, and curricular contexts during assessment of ChatGPT-generated tasks and associated prompts (Source: Authors' own elaboration)

In the second round, the second author reviewed the first author's coding and emerging descriptors, focusing on the identification and categorization of factors and descriptors. The authors discussed differences in interpretation and reached consensus on the coding and categorization. When the authors identified differences, they returned to the participants' responses and refined the descriptors as needed. The first author then revisited the data to ensure that the revised framework was applied consistently across responses. The coding process illustrates how PMTs' responses informed the development of the descriptors and their placement within the broader factors represented in **Figure 1**. For example, in the quote below, Mr. Jn's reference to "the learning objectives" and ensuring that a task "aligns with the specific math concept I'm teaching" was initially coded as *Aligning with goals and standards* and categorized under *Curricular Context*. His reference to "students' prior knowledge" was coded as *Students' prior knowledge* and categorized under *Students*. Similarly, his emphasis on "real-world examples" and varying problem types was coded as *Contextual and connected with real-life applications* and *Variation in the nature of problems/tasks*, respectively. These codes contributed to the *Students* factor in **Figure 1**.

> To optimize ChatGPT for generating mathematical tasks, I would consider factors like the learning objectives, student readiness, and the context of the task. First, I'd ensure the task aligns with the specific math concept I'm teaching, like proportionality or linear equations. Next, I'd adjust the difficulty level based on the students' prior knowledge and gradually increase complexity as they master concepts. I'd also incorporate real-world examples to make the tasks more engaging. Finally, I'd modify the prompts to include a mix of problem types (multiple choice, open-ended) and varying contexts to keep students interested and challenged.

Other PMT responses provided evidence for other descriptors shown in **Table 3** and **Figure 1**. For example, when PMTs emphasized that ChatGPT-generated tasks should provide opportunities for students to use different representations or approaches to solve a problem, these responses were inductively coded as *Multiple representations* and *Multiple approaches* to solve problems. These descriptors were then categorized under *Discipline* because PMTs were attending to the mathematical characteristics and opportunities for mathematical reasoning embedded in the task. Similarly, PMTs' attention to whether a task involved accurate mathematical concepts, procedures, and representations contributed to the descriptor *Accurate mathematical concepts*, *procedures*, and *representations* within *Discipline*.

The analysis also shows some descriptors belonged to more than one factor. For instance, *Appropriate difficulty level* and *appropriately challenging* was associated: with *Students* because PMTs considered students' readiness and prior knowledge, with *Discipline* because the level of mathematical complexity mattered, and with *Curricular Context* because PMTs considered the grade level and intended learning goals. Thus, this descriptor helped to establish connections between multiple factors that show the interaction among mathematical content, students, and curricular considerations shown in **Figure 1**. We did not identify evidence of *Teacher Characteristics* in the data; therefore, this factor is not represented in our findings and in **Figure 1**.

We used a fully inductive analysis to answer the second research question. We examined PMTs' responses for evidence of the strengths and weaknesses they identified and inductively developed themes that represented the range of PMTs' perspectives. The first author conducted the initial inductive analysis and then shared the emerging themes with the second author. The two authors discussed the coding and resolved discrepancies.

FINDINGS

We organize the findings into two sections to answer our two research questions. In the first section, we address the factors PMTs considered when evaluating and modifying prompts and ChatGPT-generated mathematical tasks. In the second subsection, we share the potential and challenges PMTs identified in using ChatGPT as a resource for generating mathematical tasks.

Factors Considered When Evaluating Prompts and ChatGPT-generated Tasks

PMTs considered factors related to the following three aspects of curricular decision-making when evaluating ChatGPT-generated mathematical tasks and modifying their prompts to generate those tasks (i.e., prompt engineering): (1) Mathematics as a discipline, (2) Student needs, interests, and knowledge, and (3) Curricular context, such as grade-level and standards. In **Table 3**, we present the frequencies of each factor, and in **Figure 1**, we categorize these factors into three aspects to show the interactions among the three aspects of curricular reasoning.

In **Figure 1**, the three boxes show three categories with the factors that solely belonged to each dimension above or below the boxes. For example, the accuracy of mathematical solutions is PMTs' consideration of discipline when evaluating ChatGPT-generated tasks and prompts. The arrows between each box show that PMTs considered two or more factors simultaneously when evaluating their prompts and resulting tasks. The text in the middle of all the boxes show that PMTs' considerations of all three dimensions simultaneously. In the subsections below, we explain each of these factors with examples from PMTs' explanations, and explain how PMTs considered mathematics, contexts, and students, including the relationship among these factors.

We present one example to show how PMTs' consideration of difficulty-level revealed their attention to *Discipline*, *Students*, and *Context* as seen in Ms. Al's quote:

> The prompt was modified to introduce complexity and a wider range of problem-solving strategies. The initial prompt focused on one proportionality relationship and could be answered with a simple calculation. The revised prompt included real-world scenarios like comparing unit prices, applying discounts, and price increases. These modifications help engage students into a deeper understanding while making sure that the task aligns with the performance goals for 7th-grade mathematics. If I had to make another revised prompt, I would make sure to ask for visual representations such as tables and graphs.

In this example, Ms. Al shared maintaining the difficulty level of tasks were related to what *Mathematical* practices (i.e., procedures- "simple calculation" vs contextual tasks- "comparing unit prices, applying discounts, and price increases") a task includes. She also shared other aspects such as maintaining appropriate level of difficulty level or challenge is also needed for developing students' "deeper understanding" (related to *Students*) given the "7th-grade context" (related to curricular contexts). Thus, collectively we categorized *maintaining appropriate difficulty level* as all three aspects (i.e., *Mathematics disciplines*, *Students*, and *Contexts*.)

Maintaining an appropriate level of challenges in tasks

The most frequent factor that all the PMTs highlighted consistently was the importance of maintaining an appropriate level of difficulty or cognitive challenge for students (the text between all three boxes in **Figure 1**). As seen in **Table 3**, all 23 PMTs shared this factor 41 times. They shared ChatGPT-generated mathematical tasks were often either too easy or too difficult for students; therefore, the tasks typically required adjustment, or their prompts needed to be revised. The PMTs attended to *discipline*, *students*, and *curricular context* when describing why maintaining appropriate level of difficulty was important in ChatGPT-generated tasks and prompts. As such, the PMTs considered which mathematics concepts and reasoning were presented in the tasks, the grade level being considered, and how their students might comprehend or engage with the tasks. As seen in the quote

by Ms. Al above, her main motivation for revising the prompt was to introduce an appropriate level of difficulty for her students. Her explanation reflected attention to *mathematics* (multiple representations), *curricular context* (7th-grade standards), and *student needs* (engagement with the concept of proportionality).

In another example, Mr. El addressed all three aspects: the emphasis on the equation $y = kx$ as a good way to represent proportionality, its appropriateness for students, and alignment with instructional goals:

> I found challenging to ensure [mathematical tasks] were both challenging and appropriate for all students. The main modification [to prompt] I made was including the learning and performance goals that students needed to meet. I also directed ChatGPT to emphasize the equation $y = kx$ and to generate questions that would help students develop a solid understanding of proportionality.

Overall, when describing the appropriate level of challenges for students, the PMTs considered three aspects in their curricular decision-making, including what mathematical content was represented in tasks, whether the task would appropriately challenge students, and how it fits within the curricular contexts (e.g., grade level).

Alignment with goals and standards and appropriate for specific grade levels

The second highest category that appeared in PMTs' evaluation of their prompts and associated ChatGPT-generated mathematical tasks was their description of curricular contexts such as how learning and performance goals, as well as standards, are represented in tasks, and if those aspects were not represented in those tasks. This factor was mentioned by 19 PMTs, a total of 33 times (**Table 3**). We identified this consideration as PMTs' attention to the curricular context. As an example, Ms. Jy shared,

> It was important to ask for the AI to generate something with specific standards and goals. Once you established those goals, the AI would generate richer tasks that would allow for students to reinforce their understandings of the topic.

In another example, Ms. Ja shared that she decided to revise her prompt because the task she generated was not appropriate for the students of the grade level she intended for. She shared:

> When I think about the task I gave the AI before revision, I would say the quality of the task was very low. It was the opposite when I gave AI a better prompt. The [first] task wasn't exactly appropriate for students of 7th grade. So, I came up with my other prompt that was a bit more specific and I would say that the quality was better than the first one. My revised prompt was, 'Generate five mathematical tasks to teach the concept of proportionality for seventh-grade students.'

In this example, Ms. Ja shared that because her original prompt was not specific enough to generate appropriate tasks for seventh-grade students, she modified it. Similarly, Ms. Dy also shared she "modified [her] prompts by specifying the grade level." Mr. An noted similar examples:

> To get better results, I learned to be more specific with my prompts. Instead of just asking for 'math problems,' I would ask for 'algebra problems for 9th grade involving solving linear equations.' This helped ensure that the problems were more relevant to my students' needs.

The PMT shared that specifying grade-level in the prompt will also help him get tasks that are relevant to his students.

Ensuring multiple mathematical representations, diverse problem-types, and multiple approaches to solve problems

Many PMTs (13 out of 23) shared one criterion for determining the quality of ChatGPT-generated tasks was whether a task provided opportunity for students to use and understand multiple mathematical representations. Additionally, the PMTs shared that they valued the tasks that allow students to use various problem-solving strategies. The PMTs shared having multiple representations of tasks and allowing students to use multiple approaches are good disciplinary practices (*Mathematics*), and also those tasks with multiple mathematical representations allow for better *Student* engagement and develop in-depth understanding of the concept. The PMTs recognized ChatGPT sometimes did a good job creating tasks that offered students with opportunities to use multiple representations, which they described as supporting better student understanding. They also stated the use of multiple representations as a strong disciplinary practice because mathematical concepts such as proportionality can often be better communicated this way. Mr. Jn shared content around why he would seek for tasks that allow students to engage with multiple representations,

> [The task] could be expanded to include additional strategies, such as…asking students to represent the relationship in a table or graph. This would encourage a deeper understanding of proportionality and offer students a more well-rounded approach to solving such problems.

Ms. Jy also shared she specifically included the prompts that allowed the students use multiple representations: "The first prompt I gave it was, 'Generate five mathematical tasks to teach the concept of proportionality for seventh-grade students, please be sure to allow students to explain their work and use different forms of representation." Pointing out the challenges of using ChatGPT-generated tasks, Mr. Er shared the importance of tasks with multiple representations. He shared, "Another challenge is creating tasks that involve diagrams or visual representations, as ChatGPT currently relies on textual descriptions for these."

In addition to multiple mathematical representations, the PMTs also valued multiple problem-solving approaches in ChatGPT-generated tasks and shared they would consider changing their prompts to generate tasks that potentially allow students to use

and explore multiple problem-solving strategies. PMTs noted offering multiple approaches to solve a problem is not only good disciplinary practice but also gives students more opportunities to learn mathematics effectively. Jn shared:

> To improve the prompt, I modified it to encourage a more comprehensive exploration of proportionality. I included multiple strategies for solving the problem. Instead of just asking for the amount of flour needed to make 36 muffins, I could expand the prompt to ask students to solve the problem using both a unit rate and a scale factor, as well as represent the relationship in a table or graph. A revised prompt might look like this: 'A recipe calls for 3 cups of flour to make 12 muffins. How many cups of flour are needed to make 36 muffins? Solve the problem using both a unit rate and a scale factor. Then, represent the proportional relationship in a table or graph.' This modification encourages students to apply different methods, deepening their understanding of proportionality and offering opportunities to visualize the relationships.

Mr. Jn revised his prompt to ensure the task allowed students to use multiple approaches, such as unit rate and scale factor, to understand proportionality. He believed this would support a more "comprehensive exploration" of the concept, which he saw as a strong way to represent mathematics as a discipline and to help students build a deeper understanding. In the same quote, Jn emphasized the use of multiple representations, like tables and graphs, which shows he was considering both the mathematics and students' needs.

PMTs also appreciated ChatGPT for generating various forms of tasks, acknowledging that it allows to present mathematics in various ways. Ms. Ma shared "ChatGPT created various forms of task and approach to these tasks." Similarly, Mr. Ax shared that he valued the tasks that ChatGPT created because "it [could] quickly provide multiple versions of a question, which [was] beneficial for creating practice sheets or quizzes." Ms. Al also "like[ed] that [ChatGPT] can create a variety of question types."

Considering both *Mathematics* and *Students*, PMTs also assessed the sequencing of ChatGPT-generated tasks and revised their prompts accordingly. As an example, Ms. Na shared: "The sequence of questions in the [ChatGPT-generated tasks] sometimes felt disjointed, making it harder to follow the logical progression. This lack of clarity could create further confusion and hinder students' overall comprehension."

Aligning with students' context and prior knowledge to ensure better student engagement

The PMTs considered factors that were specifically related to *Students*, such as student engagement, clear directions, and the inclusion of contextual tasks when assessing the quality of ChatGPT-generated mathematical tasks and considered their prompt revisions. These elements often came up together. They explained that whether a task includes clear directions can influence how well students engage with it, and often expressed their concerns related to ChatGPT-generated tasks lacking clear task-solving directions and guidance. For example, Ms. Ya shared:

> First, I ensure that the wording is clear and precise, so students don't get confused by ambiguous phrasing. Sometimes, I notice that the problem lacks real-world context, so I modify my prompt to include that, such as *write a word problem involving linear equations related to shopping or travel.* By tweaking my prompts this way, I ensure that the generated tasks are clear, engaging, and appropriate for my students.

Ms. Na also expressed a concern about unclear phrasing and how it might impact student engagement with the task that she generated with ChatGPT: "I found it challenging to visualize some of the graph and table content, which made it difficult to interpret what was being asked. If I struggled to understand the instructions, I imagine my students would face similar difficulties."

PMTs also considered students' cultural backgrounds such as the language and their access to other resources. Ms. Ya shared, "I focus on differentiation as one factor. I keep changing prompts to generate multiple versions of a task at different levels of difficulty or scaffolds. This could support diverse learners, including English Language Learners." PMTs also positioned students as explorer as seen in Ms. El shared, "when modifying my prompts the factors that I mostly considered were placing the students as explorers." Ms. Saa also shared that "I see how its [task] allows the students to explore the task [to find] if I will have to make changes to it." In explaining how task allows students to communicate their ideas, Mr. Jn shared that he modifies prompts so … students have ample opportunity for students to explain their thinking.

Ensuring accurate representation, visuals, illustrations, and logical progressions

PMTs often modified their prompts and ChatGPT-generated tasks to improve how mathematics is represented in those tasks. They noted ChatGPT sometimes produced incorrect or incomplete mathematical reasoning. A PMT, Mr. An, shared his focus on "ensuring the accuracy and precision of the problems and solutions provided [as] it sometimes struggles with interpreting prompts and [produces] wrong or incomplete tasks." Another PMT, Mr. Ja shared, "The problem doesn't explicitly require the calculation of dividing by a fraction (3 pizzas /5 friends), which is the core concept of the problem."

The PMTs also shared that they considered a logical and meaningful progression of questions and the presentation of problems when evaluating ChatGPT-generated tasks. When evaluating this factor, PMTs often considered mathematics (e.g., a particular type of sequencing and also students (if problems are not presented in a good sequential order, students do not understand). As an example, Ms. Na shared that "the sequence of questions [in ChatGPT-generated tasks) sometimes felt disjointed, making it harder to follow the logical progression. This lack of clarity could create further confusion and hinder students' overall comprehension." In this example, the PMTs considered a particular sequencing of questions that was not logical and this was harder to follow for students.

Opportunity for students to engage with big mathematical ideas

When evaluating ChatGPT-generated mathematical tasks and their prompts, the PMTs shared that they considered the extent to which mathematical tasks align with the concepts that students have learned prior to and upcoming topics, highlighting both curricular context such as which topic is placed in the prior grades and upcoming grades as well as mathematics (i.e., teaching big mathematical ideas. Ms. Le shared, "I would see how it [ChatGPT-generated task] relates to previous or upcoming topics to make sure students would understand connections among various concepts, such as geometry and equations, both have proportionality." Ms. Le shared that previous and upcoming topics as curricular contexts and seeing proportionality as a big mathematical idea as a good representation of mathematics when assessing ChatGPT-generated tasks and considering their prompt revisions. Even though this was only from one (4%) PMT (**Table 3**), it was worth to acknowledge all the PMTs' responses.

AI to Generate Mathematical Tasks: Time Saver Versus Time Consuming

In this section, we answer the second research question related to the opportunities and challenges PMTs saw in using ChatGPT as a tool to generate curricular resources such as mathematical tasks. While all the PMTs recognized the potential of ChatGPT as a valuable tool for generating mathematical tasks, they also identified challenges in coming up with right prompts as seen in Ms. Ja's comments:

> When generating mathematical tasks with ChatGPT, I found it can be a bit tedious. The tediousness comes from having to be specific when asking to generate a task. If you're too vague, the task that would be generated would be very base-level and basic.

PMTs identified opportunities in terms of efficiency, variety, and instructional utility whereas challenges were related to demanding efforts to come up with right prompts, not having visuals, and mathematics usefulness not being accurately represented as explained in the prior section. Ms. Da shared, "Instructors have to be specific with their prompts to generate useful tasks, which takes practice and can be time-consuming for novice teachers." Ms. Li also shared,

> The main challenge I encountered when generating mathematical tasks using ChatGPT was difficulty wording my prompts in order to get a 'good' task. Additionally, I found it challenging to get ChatGPT to generate what I had in mind. For instance, after noticing some flaws and modifying my prompt ChatGPT continued to generate a very similar task with not much improvement. Finally, although not much of a challenge, I found having to look for the flaws in the generated mathematical tasks to be time consuming.

Some PMTs described it as a dependable resource when prompts were iteratively refined. Others appreciated its capacity to serve as a thought partner during the initial stages of task development. Ms. Te explained, "ChatGPT is a useful tool. It gives a good enough coverage and application of what students are expected to do given the learning objectives, though not perfect (to my standards)."

Many PMTs appreciated ChatGPT could be used as a tool to save the demanding time they need for planning because they could have something to start off and they could get a variety of tasks (e.g., word problems, problems with different difficulty levels). Ms. Sa noted,

> It is simpler to put your goals and expectations and get task in a fast and advance way. That way, teachers can have extra time to focus on how to deliver the instructions and how to conduct the class.

Likewise, Mr. Rn emphasized how AI could support teachers in managing their workload:

> I found it useful for ChatGPT to generate mathematical task because it makes it easier to assign exercises or homework. If you think of it as if being a teacher, it saves lots of time since you have thousands of tasks at your fingertips.

Another opportunity PMTs identified was the ability of ChatGPT to generate tasks across a range of difficulty levels, which they saw as valuable for differentiated instruction. Mr. Ja noted, "Cognitive differentiation is the last factor I would take account in because a mathematical task should target different levels of learners and have diversity." The tool's ability to generate sample student responses, including common misconceptions, also supported teachers' planning for instruction as Ms. Le stated, "It is helps us anticipate possible answers from students." Ms. Saa also had similar thoughts. She shared, "ChatGPT can generate examples of student responses so that I can prepare for what will be asked during class."

PMTs also articulated several challenges in using ChatGPT effectively. They recognized that crafting prompts that yield mathematically correct and appropriately challenging tasks is the biggest challenge. Ms. Sa captured this challenge, noting,

> I also find [it] pretty hard to get an interesting task since ChatGPT would give you something very professional which can be something hard for some students and if you correct yourself and ask for something for high school students it would give you something a little bit too dumb like how many apples.

The PMTs also discussed the tool's inability to generate visual representations or incorporate diagrams, which are often essential in mathematical instruction. Mr. Er remarked, "Another challenge is creating tasks that involve diagrams or visual representations, as ChatGPT currently relies on textual descriptions for these." Ms. Mna echoed this concern: "One main challenge I found with ChatGPT is for it to present visuals of the mathematical tasks. For example, graphing was an option which ChatGPT could not perform. It wouldn't give me another option or websites to [go to] when I could graph these tasks."

DISCUSSION AND IMPLICATIONS

Our findings highlight how engagement with AI surfaces the complexity of PMTs' curricular decision-making as they considered a wide range of factors around correct mathematical representations, addressing students' needs, and following institutional and contextual constraints. We discuss this complexity in the sections below.

Curricular Decision-Making in AI-Generated Mathematical Tasks

Our findings challenge the concerns that AI could be replacing human intellectual work. The participating PMTs were involved in navigating the complex interactions among multiple stakeholders when assessing AI-generated tasks. To put the context, teachers' curricular decision-making considerations became visible as PMTs interacted with ChatGPT and adjusted their prompts in response to unsatisfactory outputs, which is important step to assess given that teachers' decision-making may sometimes are elusive (Siuty et al., 2022). The participating PMTs' consideration of difficulty-level of ChatGPT-generated mathematical tasks involved how accurately mathematics is represented as a discipline, students' potential difficulty levels and potential student engagement, and contextual factors such as which grade-level they had to implement their mathematical tasks. This again suggests AI-mediated task generation did not replace PMPs' professional judgment; instead, it made that judgment more explicit.

Based on the findings suggesting multiple factors PMTs considered (**Table 3**), our study provides concrete ways and specific aspects that teachers and educators should attend to when assessing ChatGPT-generated mathematical tasks. In particular, as identified by our PMTs, one consideration is whether the tasks generated by ChatGPT are appropriately aligned with students' prior knowledge. Thus, the diagram shown in **Figure 1** could serve as a useful tool for teachers and educators when assessing the quality of mathematics tasks. This tool may help them feel more prepared to evaluate task quality, rather than becoming confused or overwhelmed, as some studies have suggested (Langreo, 2025).

Aligning with prior research (e.g., Alneyadi & Wardat, 2023; Sapkota & Bondurant, 2024), our research reinforces the need to critically examine ChatGPT-generated curricular resources. The secondary PMTs in our study were aware of what prior research has established how ChatGPT-generated outputs may involve mathematical inaccuracies, developmentally inappropriate content, and outputs that are overly generic (Aga et al., 2024; Sapkota & Bondurant, 2024; Walkington & Bainbridge, 2025; Wardat et al., 2023). As PMTs shared that ChatGPT-generated resources can be mathematically inaccurate, vague, logically disorganized, and misaligned with instructional goals and other contextual factors, teachers and educators need to critically assess these various aspects.

Reconfiguring Instructional Planning Through AI Use

The findings also point to a reconfiguration, not a reduction of teachers' work when using AI tools. Although ChatGPT was perceived as offering efficiency, its effective use depended on PMTs' ability to use appropriate prompts and critically evaluate and refine outputs. Lee and Yeo (2022) suggested (and our findings similarly highlight) assessing the difficulty levels of ChatGPT-generated mathematical tasks is critical and directly related to why educators may use ChatGPT to differentiate tasks for students. However, our participants had dilemmas regarding whether ChatGPT was a helpful tool for generating mathematical tasks—particularly the difficulty of developing prompts that yielded mathematically accurate tasks. This raises questions about the extent to which AI is useful for instructional purposes. These concerns may reflect participants' limited opportunities for systematic preparation in how to critically and strategically use AI tools, which is different from the shortcomings of AI itself.

In contrast to some prior studies (e.g., Gurl et al., 2025; Sawyer, 2024), our findings suggest a different pattern in PMTs' evaluations of AI-generated tasks. While Gurl et al. (2025) reported that PMTs were primarily critical of pedagogical features such as teacher-centered versus student-centered instruction and paid limited attention to mathematical accuracy, PMTs in our study closely attended to logical sequencing, mathematical representations, and the correctness of mathematical content. The PMTs in our study were mathematics majors who had completed many content courses; therefore, they may have felt confident and authoritative in judging the quality of ChatGPT-generated mathematical tasks. Future research may consider closely examining how teacher preparation shape teachers' evaluations of AI-generated tasks.

Potential Reason for the Absence of Teacher Characteristics

Although prior studies, such as Breyfogle et al. (2020), have noted that teacher characteristics influences curricular decision-making, our study did not reveal how PMTs' characteristics, such as their beliefs about effective mathematics instruction, shaped their decisions. There are two possible reasons for this finding. First, our reflection questions (see Methods section) did not explicitly invite PMTs to reflect on their beliefs about effective mathematics instructional practices or how those beliefs informed their decisions. Although some factors they considered may have been related to their beliefs, our methodology did not allow us to determine this connection. For example, a PMT's decision to modify a task to provide greater opportunities for problem solving could reflect a belief that problem solving is an effective approach to mathematics instruction. However, because we did not ask PMTs to explain their beliefs about pedagogical practices, we cannot conclude that such decisions were explicitly driven by teacher characteristics. Second, the course emphasized more on examining the characteristics of mathematical tasks and less on considering how PMTs might implement or adapt those tasks based on their belief about effective mathematics instruction. This may have guided their reflection to be more focused on task features rather than on how their instructional perspectives influenced their decisions to accept, modify, or reject them.

Future research could address this limitation by explicitly prompting PMTs to consider how their instructional perspectives inform their use of AI-generated curricular resources. For example, prompts could ask PMTs to explain how they would modify a task to support instructional practices such as problem solving, mathematical discourse, group work, or individual inquiry. Such

prompts could provide greater insight into how teacher characteristics shape PMTs' decisions when using AI to generate and revise mathematical tasks.

LIMITATIONS

Our findings should be interpreted in the context of mathematics domain in which PMTs interacted with AI. Because the participating PMTs only used the concept of proportionality, the findings may look different when PMTs interact with AI in other mathematics content areas. For example, in geometry, PMTs may highlight more on spatial and deductive reasoning. Thus, PMTs reflections on those tasks may look different; for example, in terms of how PMTs attend disciplinary rigor and integrity may look different for other domains. Future research should examine PMTs' interactions with AI across a wider range of mathematics topics to investigate how these aspects shape curricular decision-making in different disciplinary contexts.

We note that our findings provide insight into how participants reflected on their decisions about generating, revising, and evaluating AI-generated mathematical tasks. Thus, their reflections are different from their moment-to-moment reasoning during their interactions with ChatGPT and may not capture all of the decisions they made during the process. Future research could address this limitation by examining PMTs' successive prompts and interviewing them as they revise AI-generated outputs. Comparing these sources could provide a more nuanced understanding of how PMTs' prompts evolve and how they respond to AI-generated outputs.

Author contributions: BS: conceptualization, formal analysis, investigation, methodology, project administration, supervision, visualization, writing – original draft, writing – review & editing; **LZ:** conceptualization, investigation, methodology, writing – original draft, writing – review & editing. Both authors have agreed with the results and conclusions.

Funding: The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Ethical statement: The authors stated that research ethics approval was obtained by the Institutional Review Board of the University of Texas Rio Grande Valley on 1/31/2025 with IRB number #IRB-25-0024 and by the Institutional Board of the University of Kentucky on 6/8/2026 with IRB number #118686.

AI statement: The authors stated that generative AI or AI-based tools were not used in the research, analysis, or preparation of this manuscript.

Declaration of interest: No conflict of interest is declared by the authors.

Data sharing statement: Data supporting the findings and conclusions are available upon request from the corresponding author.

REFERENCES

Aga, Z. G., Sawyer, A. G., & Wolfe, M. (2024). More confidence more critical: Investigating secondary mathematics preservice teachers' use of an artificial intelligence chatbot as a curriculum development tool. *American Journal of Creative Education, 7*(2), 63–78. https://doi.org/10.55284/ajce.v7i2.1278

Alneyadi, S., & Wardat, Y. (2023). ChatGPT: Revolutionizing student achievement in the electronic magnetism unit for eleventh-grade students in Emirates schools. *Contemporary Educational Technology, 15*(4), Article ep448. https://doi.org/10.30935/cedtech/13417

Aqazade, M., Mauntel, M., & Atabas, S. (2025). Empowering mathematics teacher educators: Exploring artificial intelligence-driven mathematical tasks. *School Science and Mathematics*. https://doi.org/10.1111/ssm.18339

Arbaugh, F., Lannin, J., Jones, D. L., & Park-Rogers, M. (2006). Examining instructional practices in Core-Plus lessons: Implications for professional development. *Journal of Mathematics Teacher Education, 9*(6), 517–550. https://doi.org/10.1007/s10857-006-9019-3

Berryhill, A., Chandler, L., Bondurant, L., & Sapkota, B. (2024). Using ChatGPT as a thought-partner in lesson planning. *Association of Mathematics Teacher Educators (AMTE) Connections, 33*(3). https://www.amte.net/connections

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

Breyfogle, M. L., Roth McDuffie, A., & Wohlhuter, K. A. (2010). Developing curricular reasoning for grades pre-K–12 mathematics instruction. In B. J. Reys, R. E. Reys, & R. Rubenstein (Eds.), *Mathematics curriculum: Issues, trends, and future directions* (pp. 307–320). National Council of Teachers of Mathematics.

Chazan, D., Herbst, P., & Clark, L. (2016). Research on the teaching of mathematics: A call to theorize the role of society and schooling in mathematics instruction. In D. H. Gitomer & C. A. Bell (Eds.), *Handbook of research on teaching* (5th ed., pp. 1039–1097). American Educational Research Association.

Chiu, T. K. F., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 4*, Article 100118. https://doi.org/10.1016/j.caeai.2022.100118

Chounta, I.-A., Bardone, E., Raudsep, A., & Pedaste, M. (2022). Exploring teachers' perceptions of artificial intelligence as a tool to support their practice in Estonian K–12 education. *International Journal of Artificial Intelligence in Education, 32*(3), 725–755. https://doi.org/10.1007/s40593-021-00243-5

Dingman, S., Teuscher, D., Olson, T. A., & Kasmer, L. A. (2021). Conceptualizing curricular reasoning: A framework for examining mathematics teachers' curricular decisions. *Investigations in Mathematics Learning, 13*(4), 267–286. https://doi.org/10.1080/19477503.2021.1981742

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

Griffin, G. A. (1990). Curriculum decision making for teacher education. *Theory Into Practice, 29*(1), 36–41. https://doi.org/10.1080/00405849009543428

Gurl, T. J., Markinson, M. P., & Artzt, A. F. (2025). Using ChatGPT as a lesson planning assistant with preservice secondary mathematics teachers. *Digital Experiences in Mathematics Education, 11*(1), 114–139. https://doi.org/10.1007/s40751-024-00162-9

Henningsen, M., & Stein, M. K. (1997). Mathematical tasks and student cognition: Classroom-based factors that support and inhibit high-level mathematical thinking and reasoning. *Journal for Research in Mathematics Education, 28*(5), 524–549. https://doi.org/10.5951/jresematheduc.28.5.0524

Herbst, P., & Chazan, D. (2011). Research on practical rationality: Studying the justification of actions in mathematics teaching. *The Mathematics Enthusiast, 8*(3), 405–462. https://doi.org/10.54870/1551-3440.1225

Karataş, F., Eriçok, B., & Tanrikulu, L. (2024). Reshaping curriculum adaptation in the age of artificial intelligence: Mapping teachers' AI-driven curriculum adaptation patterns. *British Educational Research Journal, 51*(1), 154–180. https://doi.org/10.1002/berj.4068

Kim, Y. R., Park, M. S., & Joung, E. (2024). Exploring the integration of artificial intelligence in math education: Preservice teachers' experiences and reflections on problem-posing activities with ChatGPT. *School Science and Mathematics*. https://doi.org/10.1111/ssm.18330

Langreo, L. (2025). Why some math teachers don't want professional development on AI. *Education Week*. https://www.edweek.org/technology/why-some-math-teachers-dont-want-professional-development-on-ai/2025/04

Lee, D., & Yeo, S. (2022). Developing an AI-based chatbot for practicing responsive teaching in mathematics. *Computers & Education, 191*, Article 104646. https://doi.org/10.1016/j.compedu.2022.104646

Mahlaba, S. C. (2026). How preservice teachers evaluated the accuracy of ChatGPT-generated solutions of curriculum-based mathematical word problems in a self-directed context. *Frontiers in Education, 11*, Article 1783496. https://doi.org/10.3389/feduc.2026.1783496

Organisation for Economic Co-operation and Development [OECD]. (2024). *Curriculum flexibility and autonomy: Promoting a thriving learning environment*. OECD Publishing. https://doi.org/10.1787/eccbbac2-en

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

Remillard, J. T., & Heck, D. J. (2014). Conceptualizing the curriculum enactment process in mathematics education. *ZDM–Mathematics Education, 46*(5), 705–718. https://doi.org/10.1007/s11858-014-0600-4

Saldaña, J. (2021). *The coding manual for qualitative researchers* (4th ed.). Sage.

Sapkota, B. (2022). Preservice teachers' conceptualizations of mathematical tasks. *The Mathematics Educator, 30*(2), 3-32. https://files.eric.ed.gov/fulltext/EJ1339807.pdf

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

Sapkota, B., Fulwider, D., & Newton, J. (2024). *Exploring Elementary Teachers' Post-COVID Mathematics Curricular Use: Focus on Curricular Reasoning.* In K. Kosko, J. Caniglia, S. Courtney, & M. Zonlfaghari (Eds.), *Proceedings of the forty-six annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 101-105). Kent State University. https://doi.org/10.51272/pmena.46.2024

Sapkota, B., Jiang, N., Hendrickson, L., & Wood, M. (2026a). Elementary teachers' professional obligations reflected in mathematics curricular complexity. *Educational Studies in Mathematics*. https://doi.org/10.1007/s10649-026-10512-6

Sapkota, B., Olson, A. M., & Baniahmadi, M. (2026b). Elementary mathematics curriculum: State policy, COVID-19, and teacher control. Eurasia Journal of Mathematics, Science and Technology Education, 22(9), Article em2913. https://doi.org/10.29333/ejmste/19278

Sawyer, A. G. (2024). Artificial intelligence chatbot as a mathematics curriculum developer: Discovering preservice teachers' overconfidence in ChatGPT. *International Journal on Responsibility, 7*(1), Article 1.

Sawyer, A. G., & Aga, Z. G. (2024). Counterexamples to demonstrate artificial intelligence chatbot's lack of knowledge in the mathematics education classrooms. *Connections*. Association of Mathematics Teacher Educators.

Shakarian, P., Koyyalamudi, A., Ngu, N., & Mareedu, L. (2023). An independent evaluation of ChatGPT on mathematical word problems (MWP). In A. Martin, K. Hinkelmann, H.-G. Fill, A. Gerber, D. Lenat, R. Stolle, & F. van Harmelen (Eds.), *Proceedings of the AAAI 2023 Spring Symposium on Challenges Requiring the Combination of Machine Learning and Knowledge Engineering (AAAI-MAKE 2023)*.

Sharma, S., Mittal, P., Kumar, M., & Bhardwaj, V. (2025). The role of large language models in personalized learning: A systematic review of educational impact. *Discover Sustainability, 6*, Article 243. https://doi.org/10.1007/s43621-025-01094-z

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction, 36*(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118

Wahlström, N. (2023). Curriculum policy, teaching, and knowledge orientations. *Teachers and Teaching, 29*(3), 259–275. https://doi.org/10.1080/13540602.2022.2159364

Walkington, C., & Bainbridge, K. (2025). Mathematics teachers' use of generative AI to create active learning experiences. *Social Innovations Journal, 30*(2).

Walton Family Foundation, & Gallup. (2025). *Teaching for tomorrow: Unlocking six weeks a year with AI*. https://static.waltonfamilyfoundation.org/df/fb/eba12807470a9402d7433cc47dba/teaching-for-tomorrow-unlocking-six-weeks-a-year-with-ai-report.pdf

Wardat, Y., Tashtoush, M. A., AlAli, R., & Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *Eurasia Journal of Mathematics, Science and Technology Education, 19*(7), Article em2286. https://doi.org/10.29333/ejmste/13272

Webel, C., Conner, K. A., & Sheffel, C. (2023). Explaining differences in one teacher's instruction across multiple tracked fifth-grade classes. *School Science and Mathematics, 122*(1), 54–69. https://doi.org/10.1111/ssm.12504

Yin, R. K. (2009). *Case study research: Design and methods* (5th ed.). Sage.